IDENTIFICATION, EXPRESSION, AND CHARACTERISATION OF GEOSMIN SYNTHASES IN CYANOBACTERIA

By

Steven Giglio

Bachelor of Applied Science in Medical Laboratory Science

A thesis submitted to the University of South Australia for the degree of Doctor of Philosophy in Biomedical Sciences

October 2009

Per Mama è Papa, Nonna è Nonno

TABLE OF CONTENTS

CHAPTER 1 GENERAL INTRODUCTION 1

1.1 Public water supplies 2
1.2 Cyanobacteria 2
1.2.1 History and geographic distribution 2
1.2.2 Basic structural biology of cyanobacteria 4
1.2.3 Taxonomy of cyanobacteria 6
1.2.4 Significance of cyanobacteria and blooms 7
1.2.4.1 General overview 7
1.2.4.2 Toxic cyanobacteria 8
1.2.4.3 Taste and odours in waters 9
1.3 Control of cyanobacteria 11
1.3.1 Engineering methods 11
1.3.2 Treating blooming waters 13
1.4 Economic impacts of cyanobacteria 14
1.5 Overview of molecular methods for detecting environmental organisms 16
1.5.1 Polymerase chain reaction 16
1.5.2 Alternative DNA amplification methods 17
1.5.3 Real-time PCR 18
1.5.3.1 Real-time PCR overview 18
1.5.3.2 Detection of real-time PCR products 19
1.5.3.2.1 Overview 19
1.5.3.2.2 Intercalating dyes - Overview 22
1.5.3.2.3 Intercalating dyes – Melting curve analysis 23
1.5.4 Future technologies 24
1.5.5 Mobilisation of molecular technologies to the field 25
1.6 Geosmin 26
1.6.1 Background 26
1.6.2 Human sensory perception of geosmin 27
1.6.3 Cellular distribution of geosmin within cyanobacteria 28
1.7 Biochemistry and Genetics of geosmin production - OVERVIEW 28
1.7.1 Biochemical pathway of geosmin production 28
1.7.2 Genetics of geosmin production 31
1.8 Finding the geosmin gene in cyanobacteria 32
1.9 Aims and objectives 34
CHAPTER 2 GENERAL MATERIALS AND METHODS 35

2.1 General Materials 36
2.1.1 Algal culture media 36
2.1.1.1 ASM medium 36
2.1.1.2 BG-11 medium 37
2.1.1.3 Nostoc 819 medium 38
2.1.2 Reagents and chemicals 38

2.1.3 Antibiotics 38
2.1.4 Microbial Media 38
2.1.4.1 Luria-Bertani (LB) agar 38
2.1.4.2 LB broth 39
2.1.4.3 Starch-casein agar 39
2.1.5 Oligonucleotides 39
2.1.6 SDS-PAGE materials 39
2.1.6.1 Coomassie Blue Stain 39
2.1.6.2 2x SDS-PAGE sample buffer 39
2.1.6.3 SDS-PAGE de-staining solution 40
2.1.6.4 Pre-cast SDS-PAGE gels 40
2.1.7 Enzymes 40
2.2 General Methods 40
2.2.1 DNA extraction 40
2.2.2 Plasmid extraction 40
2.2.3 Restriction enzyme digestion 40
2.2.4 Agarose gel electrophoresis 41
2.2.5 Purification of PCR products 41
2.2.6 *In-vitro* cloning 41
2.2.7 DNA sequencing 42
2.2.8 Geosmin analysis by GC-MS 42
2.3 Sequence Data analysis software 42
CHAPTER 3 ISOLATION AND CHARACTERISATION OF *NPUN02003620* (ZP_00109187) FROM *NOSTOC PUNCTIFORME* ATCC 29133 (PCC 73102) 43
3.1 Introduction 44
3.2 Aims 47
3.3 Specific materials and methods 47
3.3.1 Isolation of *npun02003620* 47
3.3.1.1 Culture and extraction of *Nostoc punctiforme* ATCC29133 DNA 47
3.3.1.2 Amplification of *npun02003620* and SCO6073 47
3.3.2 Transformation of PCR products into pET-21d(+) protein expression plasmid 48
3.3.2.1 Preparation of PCR products for ligation 48
3.3.2.2 Preparation of pET-21d(+) plasmid for ligation 48
3.3.2.3 Ligation of SCO6073 and *npun02003620* PCR products with pET-21d(+) vector 50
3.3.2.4 Transformation of ligated PCR product and pET-21d(+) into XL1-Blue 50
3.3.2.5 Screening of successful transformants using G-PCR1 50
3.3.2.6 Transformation of pET-21d(+) plasmid into BL21(DE3)pLysS *E. coli* 51
3.3.2.7 Sequencing primers 52
3.3.3 Expression of *npun02003620* and SCO6073 52
3.3.3.1 Induction of protein production 52
3.3.3.2 Separation of expressed soluble and insoluble protein fractions 52
3.3.3.3 SDS-PAGE 53
3.3.3.4 Incubation of *npun02003620* protein with farnesyl diphosphate (FPP) 53
3.4 Results and Discussion 54
3.4.1 Introduction of *npun02003620* into BL21(DE3)pLysS protein expression system 54

3.4.2 Preliminary expression screening of SCO6073 and *npun02003620* protein.....58

3.4.3 Sequencing of SG7.....60

3.4.4 Successful *npun02003620* expression.....61

3.4.5 Biochemical activity of *npun02003620*.....64

3.5 Summary.....67

CHAPTER 4 ISOLATION AND CHARACTERISATION OF NPUNMOD, A PREDICTED GEOSMIN SYNTHASE FROM *NOSTOC PUNCTIFORME* ATCC 29133 (PCC 73102).....69

4.1 Introduction.....70

4.2 Aims.....71

4.3 Specific materials and methods.....71

4.3.1 Isolation of NPUNMOD.....71

4.3.2 Transformation of NPUNMOD PCR product into pET-21d(+).....72

4.3.3 Screening of NPUNMOD inserts.....72

4.3.4 NPUNMOD sequencing primers.....73

4.3.5 Expression of NPUNMOD protein product.....73

4.3.6 Site-directed mutagenesis to remove His_6-tag from NPUNMOD.....73

4.3.7 Co-expression of NPUNMOD with GroES/GroEL and purification of active protein 74

4.3.8 Incubation of NPUNMOD with FPP.....74

4.4 Results and Discussion.....74

4.4.1 Expression of NPUNMOD.....74

4.4.2 Sequencing of NPM2_58.....76

4.4.3 Purification of NPUNMOD and incubation with FPP.....76

4.5 Summary.....80

CHAPTER 5 EXPRESSION OF GEOSMIN SYNTHASE IN *ANABAENA CIRCINALIS*.....82

5.1 Introduction.....83

5.1.1 Geosmin synthesis and photosynthesis.....83

5.1.2 The comparative C_T method ($2^{-\Delta\Delta CT}$) for the examination of mRNA synthesis.....87

5.2 Aims.....89

5.3 Specific Materials and Methods.....89

5.3.1 Test organism and cell counting.....89

5.3.2 Culture and sampling conditions.....89

5.3.3 Geosmin analyses.....90

5.3.4 RNA extraction and quantification.....90

5.3.5 PCR primers.....91

5.3.6 Reverse transcription.....91

5.3.7 PCR conditions.....91

5.3.7.1 Primer validations 91
5.3.7.2 Reverse Transcriptase - PCR 92
5.3.7.3 Summary of sample testing regime for geosmin synthase gene expression studies described in this Chapter 93
5.4 Results and discussion 93
5.4.1 Verification of 173-PCR for geosmin synthase 93
5.4.2 Fulfilling the requirements of PCR primer efficiency for use in comparative C_T analysis of geosmin synthase and 16s RNA 94
5.4.3 Growth of ANA318 and measurement of intracellular geosmin – Vessel 1, continuous light 96
5.4.4 Growth of ANA318 and geosmin concentrations – Vessel 2, darkness induced on day 9 100
5.4.5 Gene expression profile of geosmin synthase in ANA318 102
5.4.5.1 Vessel 1 – Continuous light 102
5.4.5.2 Vessel 2 – examination of geosmin synthase expression in response to absence of light 105
5.5 Summary 110
CHAPTER 6 DEMONSTRATION OF GEOSMIN SYNTHASES IN ENVIRONMENTAL CYANOBACTERIA 113
6.1 Introduction 114
6.2 Aims 117
6.3 Specific materials and methods 118
6.3.1 DNA extraction 118
6.3.2 PCRs 118
6.3.2.1 G-PCR1 118
6.3.2.2 G-PCR2 118
6.3.3 Extraction of DNA following agarose gel electrophoresis 118
6.3.4 Generation of PCR product for G-PCR2 quantification 118
6.3.5 Collection of water samples from the River Murray, South Australia 119
6.3.6 Reference Strains 119
6.4 Results and Discussion 121
6.4.1 Screening of reference strains for geosmin synthase 121
6.4.2 G-PCR2 - real-time PCR assay for geosmin synthase 124
6.4.3 Monitoring geosmin synthase levels in the River Murray, South Australia 127
6.4.3.1 Geosmin levels in sampling locations 127
6.4.3.2 Generation of a geosmin synthase standard curve 129
6.4.3.3 Analysis of River Murray samples 129
6.4.3.3.1 Melting curve analysis 129
6.4.3.3.2 Quantification of geosmin synthase 130
6.5 Summary 137
CHAPTER 7 SUMMARY AND DISCUSSION 140
7.1 Summary 141
7.2 Discussion and Future Works 143

BIBLIOGRAPHY 150

APPENDIX 1 DNA and Amino Acid Sequences and Alignments 164

LIST OF FIGURES

Figure 1-1. *Anabaena spiroides* (A) and *Anabaena circinalis* (B). Vegetative cells are granular in nature, while akinetes (AK) are larger and more ovoid, and heterocysts (H) are similar in size to vegetative cells, but more refractile.....5

Figure 1-2. Schematic of photosynthesis.....5

Figure 1-3. Morphological differences between cyanobacteria. *Anabaena planktonica* (A), *Cylindrospermopsis raciborskii* (B), *Phormidium* sp (C), and *Spirulina* sp (D). Photos courtesy of Peter Baker (Australian Water Quality Centre). 7

Figure 1-4. A toxic *Microcystis aeruginosa* bloom in the Torrens River, Adelaide.....9

Figure 1-5. An *Anabaena circinalis* (geosmin producing) bloom in Myponga reservoir. 11

Figure 1-6. Stratification and temperature of a typical reservoir storage system. 12

Figure 1-7. Dosing an Australian reservoir with copper sulphate. 13

Figure 1-8. Paddle boats closed for business due to a toxic cyanobacterial bloom. 15

Figure 1-9. Simple schematic of the polymerase chain reaction 17

Figure 1-10. Generation of a standard curve using real-time PCR. Each line and point represent a different amount of initial DNA concentration. 19

Figure 1-11. Real-time PCR chemistries (reproduced from Monis and Giglio, 2006).Schematic of commonly used real-time PCR chemistries. 21

Figure 1-12. Melting curve patterns for A) *Naegleria fowleri*, b) *Naegleria lovaniensis*, and C) *Naegleria australiensis* (Modified from Rasmussen *et al*, 2007). 24

Figure 1-13. Mobile PCR set up (arrows pointing to PCR machine set up). 26

Figure 1-14. The isoprenoid pathways (from Ajikumar *et al*., 2008).....30

Figure 3-1. Geosmin synthase (GS)-catalysed cyclization of farnesyl diphosphate (FPP) to geosmin, germacradienol, germacrene D, and 8,10-dimethyl-1-octalin. Germacradienol and germacrene D are formed by the N-terminal domain of the bi-functional protein, and geosmin is generated from germacradienol via the octalin by the C-terminal domain (adapted from Jiang *et al*, 2007). 45

Figure 3-2. Restriction map and cloning insertion sites of pET-21d(+) expression vector (reproduced from pET manual, Novagen). 49

Figure 3-3. Amplified *npun02003620* (M: 500 bp marker; 1-3: replicate *Nostoc punctiforme* DNA). 54

Figure 3-4. Screening of XL1-Blue and BL21(DE3) transformants with G-PCR1. Isolates SG4-8 labelled. M, 500 bp marker; C, control DNA. 56

Figure 3-5. *NcoI* and *XhoI* double digested plasmids from SG14, 15, 23, 24, and 7. M, 500 bp marker. SG14 and 15 are BL21(DE3) transformants using plasmid from SG7, SG23 and 24 are BL21(DE3) transformants using plasmid from SG8 (SG8 not pictured). 58

Figure 3-6. SCO6073 transformant expression profiles for isolates SG23, 24, and 25. NI, non induced fraction; IS, induced soluble fraction; II, induced inclusion body fraction. Molecular marker, left lane. 59

Figure 3-7. Transformed *npun02003620* expression profiles for isolates SG13, 14, and 15. NI, non induced fraction; IS, induced soluble fraction; II, induced inclusion body fraction. Molecular marker, left lane; SCO23 II (induced inclusion body fraction) positive control. 60

Figure 3-8. Demonstration of NJ2 double digest pattern. M, 500 bp marker.....61

Figure 3-9. Additional "T" nucleotide at position 1847 in the deposited sequence (boxed in black within the red box, top figure), and resultant corrected sequence and translated product (bottom figure).....62

Figure 3-10. Multiple alignment of experimentally derived NJ2 and *npun02003620* demonstrating an additional "T" nucleotide boxed in yellow. Full sequences available in Appendix 1-1.....63

Figure 3-11. Multiple alignment of amino acid sequences of SCO6073 from *S. coelicolor*, SAV2163 from *S. avermitilis*, ZP_00109187 from *Nostoc punctiforme*, and experimentally derived NJ2. The relieved frame shift from the additional "T" nucleotide in protein product ZP_00109187 unveiled a previously masked NSE triad that was apparent in experimental NJ2. Full sequences available in Appendix 1-2.....63

Figure 3-12. SDS-PAGE of NJ2 expression. 1, pooled non-induced fractions; 2-5, replicate induced soluble fractions; 6-9, replicate induced inclusion body fractions; M, Molecular weight marker.....64

Figure 3-13. GC-MS trace for the incubation of NJ2 with FPP, with major products of germacradienol (retention time 11.29 min) and germacrene D (retention time 10.34 min).....65

Figure 5-1. Relationship of the isoprenoid pathways and geosmin and chlorophyll formation (modified from Qin *et al*., 2007).....85

Figure5-2. Microscopic image of ANA318 (x400 DIC). Photo courtesy of Zong Lai.....89

Figure 5-3. Summary of sample testing regime.....93

Figure 5-4. Electropherogram summary of PCR product on Bioanalyser software. Peaks at 43 and 113s are marker peaks, while the peak at 63.76s is the specific 173 PCR amplicon (arrowed, virtual gel on right hand side) 94

Figure 5-5. Analysis and suitability of primer pairs for relative expression experiments - ΔCT of 173 V 16S PCRs; $MgCl_2$, 2.5 mM 96

Figure 5-6. Analysis and suitability of primer pairs for relative expression experiments - ΔCT of 173 V 16S PCRs; $MgCl_2$, 2.0 mM 96

Figure 5-7. Vessel 1, Growth of ANA318 97

Figure 5-8. Intracellular geosmin concentration of replicate experiments for vessel 1. 98

Figure 5-9. Vessel 1, ratio of intracellular : extracellular geosmin in ANA318 99

Figure 5-10. Vessel 2, growth and intracellular geosmin concentration for the dark induced culture at day 9 (red bar indicates the start of dark conditions). 101

Figure5-11. Graphical representation of fold changes for vessel 1 over time. 104

Figure5-12. Graphical representation of fold changes for vessel 2 over time. 107

Figure5-13. Vessel 2, ratio of intracellular : extracellular geosmin in ANA318 (red bar indicates the start of dark conditions). 107

Figure 5-14. Regulation of carotenoids production in *Synechocystis*. The greater the width of the arrow, the greater the light dependence increase (reproduced from Schafer *et al*., 2006). 111

Figure 6-1. Sampling locations, state level (boxed). 120

Figure 6-2. Sampling locations in detail. 120

Figure 6-3. Gel electrophoresis demonstrating the presence of putative geosmin synthase homologs in cyanobacteria. (M; 500 bp DNA ladder; 1, *Nostoc punctiforme* ATCC 29133 (geosmin producer); 2, *Nostoc* sp SWC701 (geosmin producer); 3, *Pseudanabaena limnetica* (geosmin non-producer); 4, *Phormidium calcicola* (geosmin and MIB producer); 5, *Anabaena laxa* (geosmin producer). 121

Figure 7-1. Schematic for the proposed inclusion of geosmin synthase PCR in monitoring of waters in T&O events.....147

LIST OF TABLES

Table 3-1. Sequencing primers used for *npun02003620.* 52

Table 3-2. Sigma protein standard subunit molecular weight markers. 53

Table 3-3. Laboratory strains generated from pET-21d(+) transformations 57

Table 3-4. Sequence comparison *of S. coelicolor* SCO6073 geosmin synthases with bacterial orthologs (reproduced from Jiang *et al*., 2007, supplementary information Table 3). ...67

Table 4-1. NPUNMOD sequencing primers 73

Table 5-1. PCR primers used for relative expression determinations. 91

Table 5-2. Primer validation data for 173A and 16S PCRs. 95

Table 5-3. Intracellular geosmin centrations, vessel 1 98

Table 5-4. Fold changes of vessel 1 over time. 103

Table 6-1. Protein BLAST of *Phormidium calcicola* sequence. 124

Table 6-2. Comparison of G-PCR2 with GCMS geosmin analysis on additional cyanobacterial isolates. 125

GLOSSARY OF ABBREVIATIONS AND ACRONYMS

aa	Amino acid
ADWG	Australian Drinking Water Guidelines
APHA	American Public Health Association
ATP	Adenosine triphosphate
AWQC	Australian Water Quality Centre
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CCD	Charge-coupled device
chl *a*	Chlorophyll a
CPD	Critical Population Density
C_T	Threshold cycle
CYN	Cylindrospermopsin
DMAPP	Dimethylallyl diphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
DXP	1-deoxy-D-xylulose-5-phosphate
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
FAM	6-carboxyfluorescein
FGPP	Farnesyl geranyl diphosphate
FPA	Flavour Profile Analysis
FPP	Farnesyl diphosphate
FRET	Fluorescence Resonance Energy Transfer
FU	Fluorescent units
Ga	Giga annums
GC-MS	Gas Chromatography – Mass Spectroscopy
GGPP	Geranylgeranyl diphosphate

GOI	Gene of Interest
GPP	Geranyl diphosphate
GS	Geosmin synthase
HKG	House Keeping Gene
HPLC	High performance liquid chromatography
iFRET	Induced Fluorescence Resonance Energy Transfer
IPP	Isopentenyl diphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobases
KDa	Kilo-Daltons
LB	Luria-Bertani
LBAKC	Luria-Bertani with Ampicillin, Kanamycin, and Chloramphenicol
LC-MS	Liquid Chromatography – Mass Spectroscopy
LCR	Ligase Chain Reaction
LLA	Linear Linked Amplification
M	Molar
MEP	2-C-methyl-D-erythritol 4-phosphate
MGB	Minor Groove Binding
MIB	Methylisoborneol
Milli-Q	Millipore UltraPure Water
Min	Minute
mRNA	Messenger RNA
MV	Mevalonate
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate-oxidase
NHMRC	National Health and Medical Research Council
nm	Nanometer
NOM	Natural Organic Matter
nt	Nucleotide
PAC	Powdered Activated Carbon

PCR	Polymerase Chain Reaction
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal RNA
s	Second
SDA	Strand Displacement Amplification
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SPME	Solid phase microextraction
Taq	*Thermus aquaticus*
TBE	Tris/Borate/EDTA
TET	6-carboxy-2',4,7,7'-tetrachlorofluorescein
V	Volts
w/v	Weight per volume
X-Gal	5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside

PUBLICATIONS AND AWARDS ARISING FROM THIS STUDY

GIGLIO, S., JIANG, J., SAINT, C. P., CANE, D. E. & MONIS, P. T. (2008) Isolation and characterization of the gene associated with geosmin production in cyanobacteria. Environ Sci Technol, 42, 8027-32.

GIGLIO, S., SAINT, C.P. & MONIS, P.T. (2009) Expression of geosmin synthase in *Anabaena circinalis. Manuscript in preparation*.

PATENT

CANE, D.E., GIGLIO, S., JIANG, J., SAINT, C.P., MONIS, P.T. (2008). Assays and methods of detecting two VOCs, Geosmin and MIB. U.S. Patent application No. 61/192,309.

AWARDS

2008 – International Travel Grant, University of South Australia.

2009 – Australian Water Association Post Graduate Award (SA Branch).

EXECUTIVE SUMMARY

The presence of algal metabolites in source and finished waters is a common occurrence worldwide. One such metabolite, geosmin, is responsible for the earthy/musty taste and odour in affected waters, and the detection of this compound by customers suggests that the water may be unsafe to drink, seriously affecting consumer confidence, and undermining water utility performance. Current practice to deal with issues surrounding an increase in geosmin concentrations in waters is often employed after the problem is realised, but in order to take a more proactive approach, a fundamental understanding of the mechanisms involved in geosmin production is required.

In this study, a model geosmin-producing cyanobacterium, *Nostoc punctiforme* ATCC 29133, was used to isolate the gene responsible for geosmin production. One such candidate gene, *npun02003620* encoding for the putative ZP_00109187 protein, was isolated and shown not to produce geosmin, but it produced the geosmin precursor intermediates germacradienol and germacrene D when incubated with farnesyl diphosphate (FPP). A single, but vitally important, sequencing error was subsequently discovered in the deposited sequence of *npun02003620*. Correction of this sequence resulted in a frame shift of the coding region, and unveiled an additional 100 amino acids. This new protein, dubbed NPUNMOD, when incubated with FPP produced germacradienol, germacrene D, and geosmin as major products.

Examination of the gene expression of geosmin synthase in *Anabaena circinalis* suggested that the expression is co-ordinate with cell growth and likely to be constitutive in nature under conditions of high light exposure. Moreover, the activity of the geosmin synthase enzyme, or the translation of mRNA geosmin synthase transcripts, appeared to be active for two days post incubation in complete darkness, and post stationary phase, when grown in batch cultures.

Using a novel and rapid real-time PCR assay, geosmin synthase homologs were detected in several genera of geosmin-producing cyanobacteria, including *Anabaena* sp, *Phormidium* sp, *Geitlerinema* sp, and *Nostoc* sp. In addition to the detection of cyanobacteria capable of producing geosmin, the real-time PCR assay was also able to differentiate between key geosmin-producing species using DNA melting curve analysis of amplicons. The utility of this

novel tool was successfully demonstrated using field samples from the River Murray during a period where low-moderate geosmin concentrations were detected.

This study provides a pivotal point from which to propel future work into the detection, control, and investigation of taste and odour episodes. With water being one of our most valuable commodities, the works described herein will prove to be timely, and aid in our understanding and management of taste and odour events.

DECLARATION

I declare that:

this thesis presents work carried out by myself and does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; to the best of my knowledge it does not contain any materials previously published or written by another person except where due reference is made in the text; and all substantive contributions by others to the work presented, including jointly authored publications, is clearly acknowledged.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying

_____ Date: _____

Steven Giglio

ACKNOWLEDGEMENTS

First and foremost, I wish to thank my Principal Supervisor Dr Paul Monis for the opportunity to undertake this journey under his auspices. Paul's wisdom, patience, and genuine thirst for knowledge have given me the opportunity to develop my skills that will hold me in good stead for many years to come. A sincere thank you is also extended to Prof. Christopher (Ozzy) Saint, who, along with Paul, has been nothing but supportive with every aspect of my study, career, and aspirations while at SA Water. This post graduate study would not have been possible if it was not for the financial and developmental support from SA Water.

A long arm of continual thanks is extended to Prof. David Cane, Dr. Jiaoyang Jiang, and members of the Cane lab for their patience in dealing with me and interest in my project. A special thanks to my Brown buddies Wayne (The Destroyer) Chou and Tony (True Dat) Comeau for entertaining me (to say the least!) while in Rhode Island. My time spent at Brown University was more than memorable, extremely rewarding both professionally and personally, and was only possible through the provision of an International Travel Grant awarded by the University of South Australia.

My sanity was often saved by the entertaining people that I work with. A special thanks to Daniel Hoefel, who gave me more than 20 min while critiquing my thesis (urrgh!). Special thanks also to **DR** Brendon King, Alex Keegan, an always animated Bec Campbell, Bret Robinson, Stella Fanok, and all past members of the R&D team. A thank you also to Peter Hobson, Peter Baker, Mike Burch, and the algal research team for their time and assistance.

A "wow" to my confidantes Jim, Adam, and Gus, who have been with me through thick and thin. The PhD pain was certainly eased by the endless fishing adventures, trips, poker games, fine dining, and general hoo-hah and rubbish we get up to. A special thanks also to Annaliese whose assistance and skills in főrmattinG of tH-is DoCuMeNt are cle@rly evidĚnt.

Last but not least I would like to thank my family and friends for their support through this sometimes painful saga. The deepest thank you possible goes to my girl, Sonia, whose smile has soothed my savage soul on many an occasion. Sonia, you have been an absolute breath of fresh air when I needed it most, thanks for all your patience, care, support, and encouragement. Your gentle nature and zest for life has been a godsend.

Chapter 1

GENERAL INTRODUCTION

More than just blind ambition

More than just simple greed

More than just a finish line

Must feed this burning need

Marathon, Rush

1.1 PUBLIC WATER SUPPLIES

The settlement and expansion of civilised communities is usually centred on the ability to trade commodities, the ability to successfully culture produce for self sustenance, and the presence of fresh water. A reliable water supply has underpinned most successful settlements and supported modern day business infrastructure requirements, and as such is perhaps our most valuable commodity.

The quality of the water supplied to consumers is scrutinised on several levels. Consumers demand the water to be both free of pathogens and aesthetically pleasing. To ensure the quality of water supplied to consumers in Australia, the Australian Drinking Water Guidelines (ADWG) was developed to set safe and operable parameters to ensure the consistency and quality of water supplied to households and industry (ADWG, 2004). These guidelines set strict limits for chemical and microbiological parameters to which the performance of water utilities are ultimately measured, and the safety of the water is ultimately assessed.

Whilst a relationship exists between the presence of pathogens or chemicals in water, and the risk of human pathologies, (Donovan *et al*., 2008, Howard *et al*., 2006, van Lieverloo *et al*., 2007, Westrell *et al*., 2004, Dewettinck *et al*., 2001, Eisenberg *et al*., 1996), the determination of what is aesthetically pleasing is subject to perception and interpretations based on individual sensitivities to various compounds. One of the most common aesthetic complaints is related to perceived earthy or musty taste and/or odours that are more often than not cyanobacterial in origin. Consumers correlate an earthy/musty taste with improperly treated water, and often assume that the water may be unsafe to drink. The reality is that the compounds mainly responsible for these taste and odours, geosmin (earthy taste and odour) and 2-methylisoborneol (MIB, musty taste and odour), have no known adverse health effects. Nonetheless, water utilities treat taste and odour events as a serious operational problem, and as such are the focus of much attention for most water providers.

1.2 CYANOBACTERIA

1.2.1 History and geographic distribution

Cyanobacteria, or “blue-green algae” as they are often referred to, are amongst the most primitive forms of life on earth. Debate continues in relation to the exact timing for the

emergence of these organisms, and their implied origin for the development of the photosynthetic machinery subsequently passed on to more complex life forms. Cyanobacteria contain a blue/green pigment called phycocyanin (hence the name blue-green algae), an essential compound involved in photosynthesis. They also contain chlorophyll a, which is the same photosynthetic pigment found in plants; the presence of these similar photosynthetic sub units in plants are thought to have evolved from a symbiotic relationship between ancestral plant cells and cyanobacteria several billion years ago (Raven, 2002).

The evolution of cyanobacteria and their position in the history of the earth is important as they have long been considered the first organisms to be photosynthetic, giving rise to the machinery necessary for the fixation of carbon, and for oxygenation of the earth which resulted in complex changes in flora and fauna. Current consensus is that cyanobacteria existed at least 2.4 Ga (Giga annums, billions of years ago) with controversy surrounding speculative 3.5 Ga claims by Schopf and Packer, who discovered cyanobacterial structures in early archean fossil remains (Schopf and Packer, 1987). The claims of Schopf and Packer were challenged by Brasier and colleagues (Brasier *et al*., 2002) and counter-challenged again by Schopf and co-workers (Schopf *et al*., 2002). These claims and counter claims were based largely on geochemical studies and Raman spectroscopy, but an evaluation of the phylogeny of photosynthetic organisms has again challenged this timing and places the earliest origin of cyanobacteria to be 2.2-2.4 Ga (Hedges *et al*., 2001). Evidence has emerged, however, that there may have been photosynthetic organisms around 3.5 Ga that most likely used hydrogen as the primary electron donor (Tice and Lowe, 2006), thus raising speculation as to the exact source of the modern day photosynthetic apparatus.

Cyanobacteria are ubiquitous, having evolved to survive in a variety of ecological niches. Cyanobacteria are primarily autotrophic, photosynthetic, prokaryotic organisms that sequester atmospheric carbon and release oxygen using photosystems I and II (Castenholz and Waterbury, 1989). Some cyanobacteria have a remarkable ability to tolerate up to 30% w/v NaCl in hyper-saline lakes (Brock, 1976), where it is believed the salt concentration inhibits photosystems I and II and directs energy to anaerobic respiration (Allakhverdiev and Murata, 2008, Allakhverdiev *et al*., 2005). Cyanobacteria survive extreme temperatures in hot springs (Mohamed, 2008, Roeselers *et al*., 2007), and Arctic and Antarctic lakes (Fernandez-Valiente *et al*., 2001, Kastovska *et al*., 2005, Waleron *et al*., 2007), as well as

inhabiting complex microbial communities in crude oils found in Japan and Saudi Arabia (Al-Thukair *et al*., 2007, Tanaka *et al*., 2008). Additionally, through structures known as akinetes (discussed in 1.2.2), cyanobacteria have the ability to remain dormant until conditions are suitable for akinete germination, giving rise to viable cyanobacterial cells that can thus populate water bodies when the conditions are ideal (Mohamed, 2007).

1.2.2 Basic structural biology of cyanobacteria

Cyanobacteria are prokaryotes, which by the simplest definition, indicates a nuclear membrane is absent (compared to eukaryotes that have a distinct nuclear membrane). Depending on the species, cyanobacteria can be unicellular, form filaments of chains, or form a clump of cells in mucilage, and also exhibit gliding motility (Rippka *et al*., 1979). Many cyanobacteria contain gas vesicles that are an important buoyancy regulatory mechanism for movement through the water column; moving upwards to source sunlight when available, and downwards to source nutrients leaching from the sediment (Rabouille *et al*., 2003, Walsby, 1972, Walsby *et al*., 2006). Cyanobacteria are oxygenic photo-autotrophs, i.e. they use CO_2 as their carbon source and light as an energy source. Some cyanobacteria have specialised cell types knows as heterocysts that have a thickened cell wall (impermeable to oxygen) and are the site of nitrogen fixation. Some cyanobacteria have akinete forms as part of their morphology under various conditions (Figure 1-1). These are very thick-walled organelles that are resistant to desiccation and heat. Akinetes often reside in sediments and are “reactivated” or germinated when conditions are suited for repopulation of this species (Adams and Carr, 1981). Cellular inclusions such as polyphosphate and cyanophycin granules are important energy sources, storing vital phosphorous and nitrogen/carbon respectively (Berg *et al*., 2000).

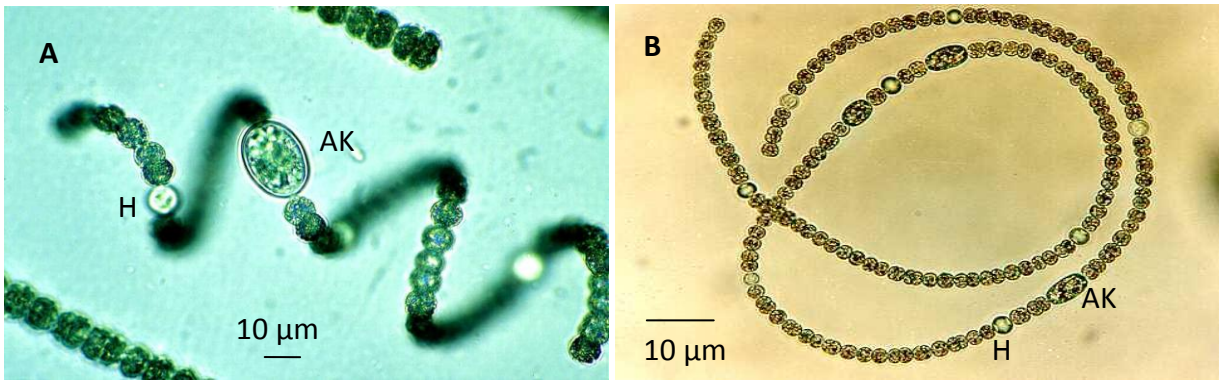


Figure 1-1. *Anabaena spiroides* (A) and *Anabaena circinalis* (B). Vegetative cells are granular in nature, while akinetes (AK) are larger and more ovoid, and heterocysts (H) are similar in size to vegetative cells, but more refractile.

Cyanobacteria use the same photosynthetic pathway as eukaryotic cells, such as higher plants, where the "C3" or "Calvin" cycle is utilised to fix carbon into available sugars for cell energy, with the subsequent release of oxygen. Light acts on chlorophyll a in the thylakoid membrane resulting in the release of oxygen and NADP from photosystems I and II respectively. NADP and subsequently generated ATP are then input into the Calvin cycle for the generation of simple carbohydrates while consuming CO_2 (Stryer, 1981).

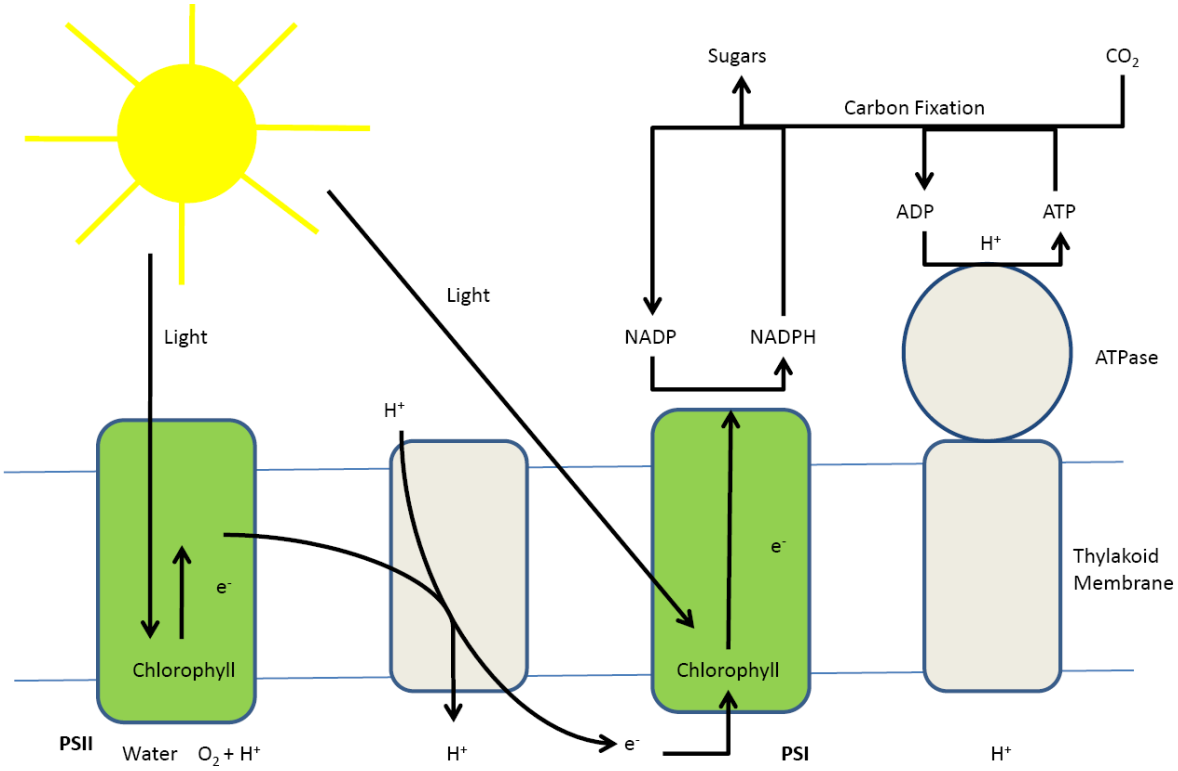


Figure 1-2. Schematic of photosynthesis

Cyanobacteria, in addition to chlorophyll a, utilise phycobiliproteins that are arranged in phycobilisomes as part of the photosystem II. Examples of these are phycoerythrin (absorption maximum at 550 nm), phycocyanin (absorption maximum at 620 nm), and allophycocyanin (absorption maximum at 650 nm), each with different absorption maxima to maximise available light utilisation under available conditions.

1.2.3 Taxonomy of cyanobacteria

Most cyanobacterial taxonomists refer to classification according to Rippka *et al.* (Rippka *et al*., 1979), where cyanobacteria are divided into five sections based on morphology and reproductive traits. The taxonomy of cyanobacteria is rapidly changing and other morphological taxonomic systems also exist (Stanier and Cohen-Bazire, 1977); readers are directed to these references for complete information and classical criteria used in cyanobacterial taxonomy. In this molecular era, however, the historical classification schemes are now being continually challenged, and cyanobacteria are being reclassified by fusing genetic data and morphological data to make the taxonomy more stringent. The most up to date classification scheme encompassing both morphological and genetic data is available at the taxonomy browser - http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?name=cyanobacteria, and is best used in conjunction with Bergey's Systematic Manual of Bacteriology (Castenholz, 2001)

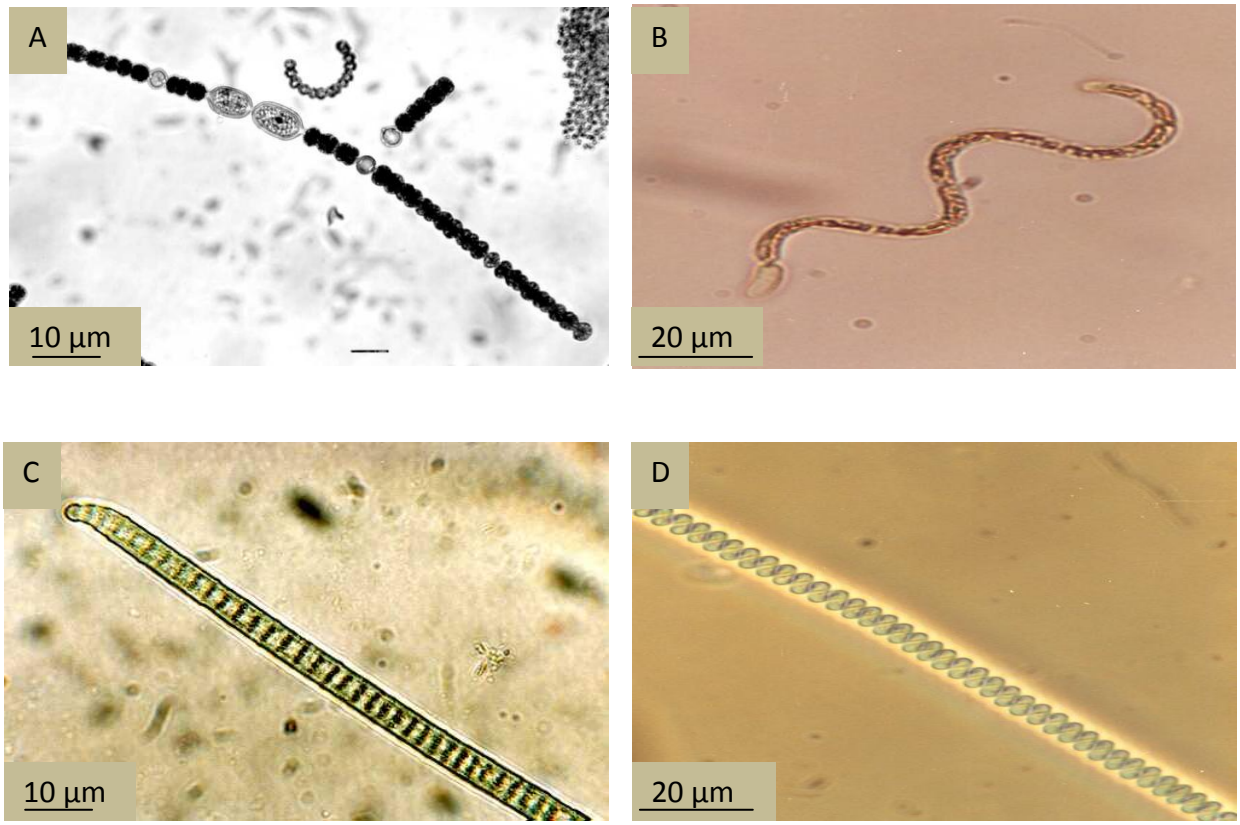


Figure 1-3. Morphological differences between cyanobacteria. *Anabaena planktonica* (A), *Cylindrospermopsis raciborskii* (B), *Phormidium* sp. (C), and *Spirulina* sp. (D). Photos courtesy of Peter Baker (Australian Water Quality Centre).

1.2.4 Significance of cyanobacteria and blooms

1.2.4.1 General overview

The fundamental importance of cyanobacteria should not be overlooked. They are the mediators of carbon and nitrogen fixation in exchange for the release of vital oxygen. Cyanobacteria provide a significant amount of the earth's total oxygen content and are vital to the survival of the planet. Some cyanobacteria co-exist with many plants, particularly legumes, where they are responsible for fixation of atmospheric nitrogen to ammonia or nitrate which can subsequently be absorbed through their roots. This has been exploited in the production of rice (*Oryza sativa*) where the floating *Azolla* fern is distributed in rice fields. This fern has a symbiotic relationship with *Nostoc* sp. and *Anabaena* sp. (Nilsson *et al*., 2005, Boison *et al*., 2006, Prasanna *et al*., 2002, Irisarri *et al*., 2001) which effectively makes the *Azolla* fern a floating fertiliser pod, leading to significantly larger and more robust crops. *Spirulina* cultures have been utilised by Aztecs and Orientals for many years, and recently the use of *Spirulina* products as a health supplement has gained momentum amongst those looking for "natural"

sources of carotenoids, protein, and probiotic factors (Kay, 1991). The virtuosity of cyanobacteria is further demonstrated in recent findings that suggests cyanobacteria were involved in the bio-mineralisation of gold in the world's largest gold deposit, the Witwatersrand QPC deposit, in South Africa (Reith *et al*., 2007).

Cyanobacteria also produce several bioactive secondary metabolites, currently the focus of much research, especially in the fields of drug discovery for anti-cancer therapies (Tan, 2007) and antimicrobials (Ramaswamy *et al*., 2007). These metabolites show promise, but are still in the early stages of development and will need rigorous evaluations before being released to the market. Additionally, as world crude oil stocks continue to decline as demand increases, alternative energy sources are becoming increasingly attractive, and several cyanobacterial species are currently under investigation for possible exploitation and provision of alternative fuels (Haag, 2007).

1.2.4.2 Toxic cyanobacteria

Some of these bioactive metabolites produced by cyanobacteria are also undesirable. Several species of cyanobacteria produce potent cyclic peptide hepatotoxins such as microcystin, cylindrospermopsin (CYN), and nodularin; and neurotoxins such as anatoxin and saxitoxin. Cyanotoxin bloom forming organisms are a frustrating occurrence. Not only do they present potentially fatal scenarios to humans, livestock, and aquatic life (Blyth, 1980, Thomas *et al*., 1998), but the management of these types of blooms are also challenging both economically and politically. In 2005, the release of a report from the international hydrological programme CYANONET "A Global Network for Cyanobacterial Bloom and Toxin Risk Management -Initial Situation Assessment and Recommendations" (Codd *et al*., 2005) reported that water utilities worldwide frequently experience problems with cyanotoxin bloom forming organisms. The reported occurrence of these cyanotoxins is not localised to a single country and climate, and microcystin and cylindrospermopsin were reported to be the most prevalent cyanotoxins detected. This is perhaps due to the increased interest in these toxins, or the availability of specialised testing equipment and established protocols in which these countries examine blooms for these toxins routinely (for an extensive summary of cyanotoxins and distribution, readers are directed to Codd *et al.*, 2005).



Figure 1-4. A toxic *Microcystis aeruginosa* bloom in the Torrens River, Adelaide.

1.2.4.3 Taste and odours in waters

Several of the tastes and odours that can be present in waters can be attributed to bacteria, algae, or cyanobacteria. Sulphate reducing bacteria are often responsible for the "rotten egg" hydrogen sulphide tastes in ground waters; the algal group including chryseophytes, diatoms, and dinoflagellates, are often responsible for the detection of fish and septic like tastes and odours which are attributed to specific aldehydes produced by, for example, *Uroglema* and *Ceredium* and *Perenumdium* species respectively (Rashash *et al*., 1996). Several other species of algae also produce "hay" tastes and odours (e.g. *Chlamadomonas peterfii*), and several reports of cyanobacteria producing grassy, cucumber, or rotting fish smells have been reported by the Australian Water Quality Centre (AWQC) laboratory (personal communication, Peter Hobson). The most commonly reported taste and odours, however, are those of an "earthy" or "musty" taste and odour that is related to the presence of geosmin and/or MIB.

Geosmin and MIB-producing actinomycetes are normally present in low numbers in waters and sediments and relatively few episodes have (questionably) been linked to the presence of these organisms. The high abundance of actinomycetes in a Canadian river (Jensen *et al*., 1994) and associated geosmin levels, and the presence of *Streptomyces* in a geosmin-tainted stream

supplying a trout farm (Klausen *et al*., 2005) linked the presence of geosmin to the presence of these organisms purely, it seems, by association. In these reports the authors attribute the taste and odour event to these organisms because they demonstrated an abundance of these organisms, and also demonstrated they produce geosmin, but failed to investigate any benthic sources, which are common sources of taste and odour-producing organisms. Additionally, the enumeration of actinomycetes may have been the result of detecting spore bodies, and not actively growing cells, and may falsely suggest an active population of actinomycetes. The role of the actinomycetes, therefore, remains speculative, particularly with the absence of further significant reported episodes. Certainly, an investigation into actinomycetes in Australian waters failed to draw any definitive conclusions relating to their contribution to the detection of taste and odour compounds in waters (unpublished data, Peter Hobson, Steven Giglio *et al*).

Several other secondary metabolites produced by cyanobacteria are also undesirable. Human detection and sensitivity to some of these compounds are very low, often at part per trillion or part per billion levels. These compounds do not have any known biological function, but their detection is undesirable from an aesthetic perspective, and consumers relate the detection of these compounds to a failed disinfection regime and that the water may be unsafe to drink. Considerable effort from utilities worldwide is spent on managing taste and odour cyanobacterial blooms; the economic impacts are considerable, and can be politically troublesome, particularly as the security and safety of water supplies is politically sensitive.



Figure 1-5. An *Anabaena circinalis* (geosmin producing) bloom in Myponga reservoir.

1.3 CONTROL OF CYANOBACTERIA

1.3.1 Engineering methods

The design and engineering of catchments and associated infrastructure is a crucial factor in controlling taste and odour events. Physical control of nutrients, stratification events, and off take levels are the first lines of defence. Eutrophication (an increase in nutrients and decrease in available oxygen) of fresh waters is often the result of an increase in phosphorous and/or nitrogen. The control of nitrogen as a critical factor seems a debateable topic, as often the investigations into eutrophic events are short term (i.e. less than 3 years). However, a 37-year investigation into Lake 227 (a deliberate experimental lake) suggests that nitrogen levels are not as critical as phosphorous, and indeed low nitrogen levels will favour the nitrogen-fixing bacteria (including taste and odour producers)(Schindler *et al*., 2008).

Stratification of a water column is a term used to describe the separation of water into 3 sections; the upper layer, epilimnion; the central layer; thermocline or metalimnion, and the

bottom hypolimnion layer (Figure 1-6). The epilimnion is close to the surface and typically the warmest layer, and is subject to mixing by wind and addition of nutrients from surrounding areas, and therefore provides an ideal environment for cyanobacteria to grow. The hypolimnion is a region of cold dense water that typically has a low level of oxygen, thus promoting the release of inorganics, such as nitrogen and phosphorous, from the sediments. The metalimnion acts as a barrier between the epilimnion and hypolimnion, causing an equilibrium between anoxic conditions in the hypolimnion and the epilimnion. The removal of these stratification layers will disrupt this equilibrium and discontinue favourable conditions for cyanobacterial growth and sediment leakage of inorganics. The use of destratification devices, such as bubble plume aerators and mechanical mixers, remains a viable option for controlling eutrophication in reservoirs, and these are widely used as a preventive measure in cyanobacterial risk management frameworks (Burch *et al*., 2000).

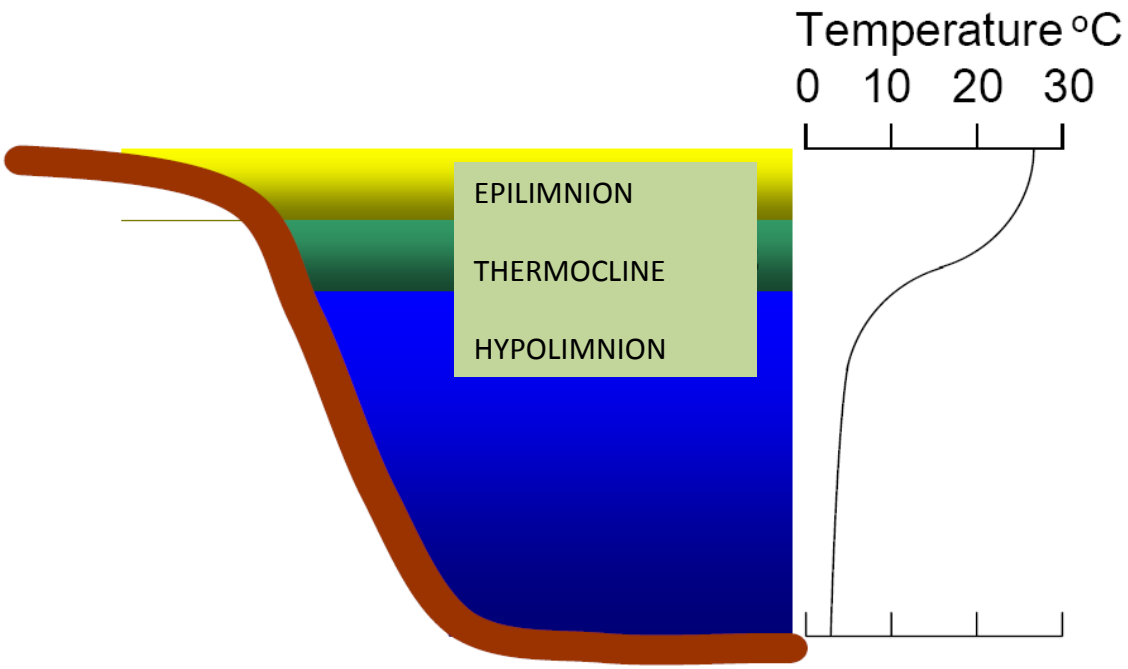


Figure 1-6. Stratification and temperature of a typical reservoir storage system.

Off-take levels of source water can also be used to lessen the risk of contaminated water entering the treatment plant or catchment area. Unfortunately, this is a massive engineering exercise and is specific to each catchment region, and requires detailed investigations often post a series of dramatic (or politically troubling!) taste and odour or toxic events.

1.3.2 Treating blooming waters

Control of blooming cyanobacteria is limited primarily to the use of copper sulphate dosing or other algicides, such as potassium permanganate (Fitzgerald, 1966) or copper citrate (Fitzgerald and Faust, 1963).The distribution of copper sulphate when dosed into a water body is not homogenous (Sklenar and Horne, 1999), nor is the distribution of cyanobacteria. The resultant situation is that not all cells have sufficient contact with copper sulphate, thereby decreasing the efficacy of the dosing program. Additionally, copper-tolerant species have been reported (Izaguirre, 1992) and marked differences in susceptibility have been reported for cyanobacterial species (McKnight, 1981). The use of copper sulphate, whilst a semi-effective measure for controlling cyanobacterial blooms in the short term, may exacerbate taste and odour problems because lysis of the cells also results in release of geosmin and places further pressure on downstream treatment plants. Additionally, if the suspected (or unsuspected) culprit is a benthic cyanobacterium, algicide dosing on the surface will have minimal impact unless physical disruption techniques are employed to release the benthic organisms.



Figure 1-7. Dosing an Australian reservoir with copper sulphate.

1.4 ECONOMIC IMPACTS OF CYANOBACTERIA

Economic impacts of cyanobacterial events can be significant. Cyanobacterial blooms, whether toxin or taste and odour-related, have immediate and prolonged financial impacts. An immediate reaction is to increase monitoring of waters for toxins and/or tastes and odour compounds. The costs for these analyses can range from AUD$150-1000, and turnaround times can vary. The problem must then be dealt with directly, either through the dosing of algicides into reservoirs, or through the use of powdered activated carbon (PAC) for water treatment. As an example, SA Water spends in excess of $1 million per annum on the dosing and removal of copper sulphate in various systems. Another option to control these blooms is to flush water from upstream (at a cost), but this practice is now rarely utilised in Australia due to extreme drought conditions that started in early 2000. If the water is severely compromised, and tainted water has reached the consumer, then bottled spring water may need to be distributed to affected areas, at a significant cost to water utilities.

Unsightly blooms also negatively impact on hospitality and tourism, particularly where water bodies are used for recreational purposes such as canoeing, paddle boating (Figure 1-8), skiing, and swimming. Walker and Greer (Walker and Greer, 1992) examined the economic impact in three separate bloom episodes in New South Wales, and estimated a loss of $1.2 – 6.7 million tourism dollars. Several episodes, either toxic or taste and odour-related, do not have detailed investigations into economic impacts, and the total cost of cyanobacterial blooms is more than likely severely under-reported and under-estimated.



Figure 1-8. Paddle boats closed for business due to a toxic cyanobacterial bloom.

Aside from the immediate impacts described above, several long term economic impacts arise from bloom episodes. These costs are primarily focused on preventative measures to avoid repeat episodes, and provide a mechanism to initiate management frameworks for mitigation of future occurrences. These may include the installation of covers on small storage tanks, and efforts to control eutrophication of lakes and rivers by decreasing nutrient levels from wastewater treatment plants and urban storm water run offs. Management efforts to control future blooms have been compiled into a Water Quality Management Framework (ADWG, 2004).

Early detection of the occurrence, or possible occurrence, of taste and odour episodes is the key to successful management. Detection of an organism in large numbers presumed to a be a taste and odour producer seems to be the only indicator employed by utilities at present (Izaguirre and Taylor, 2004). A more rigorous examination of the organism(s) involved has been suggested (Persson, 1983) that relies on a) concurrence of the odour and the organism in the field, b) isolation of the organism and proof of odour production in the laboratory, and c) chemical identification and sensory characterization of the odour compounds. Obviously, if the latter were employed, several taste and odour episodes would run their course before mitigating protocols were deployed. A rapid and accurate test to determine if organisms are

taste and odour producers without the need for expensive, and sometimes slow, laboratory tests would therefore be an ideal tool to use in the challenges of taste and odour management. The use of GC-MS, however, for the detection of geosmin concentrations in waters will continue to remain an important measurement, particularly as this value is important for calculating the amount of PAC or GAC to add in the water treatment plant.

1.5 OVERVIEW OF MOLECULAR METHODS FOR DETECTING ENVIRONMENTAL ORGANISMS

1.5.1 Polymerase chain reaction

The emergence of the polymerase chain reaction (PCR) in 1987 (Mullis and Faloona, 1987) was perhaps the second most influential discovery in molecular biology, apart from Watson and Crick's pioneering work on the structure of DNA. PCR is an *in vitro* enzymatic reaction used to make multiple copies of a target DNA sequence by the use of sequence specific primers (short oligonucleotides) and thermal cycling. PCR has paved the way for a revolution in the scientific world from the discovery of disease, to criminal identification in courts of law.

Multiple copies of DNA are achieved by the denaturing the DNA strands by heat (94 °C), then allowing it cool for primer annealing (50-60 °C), followed by allowing the extension of the separate strands of DNA by incorporation of nucleotides catalysed by *Taq* DNA polymerase at 72 °C (Figure 1-9). By this process, theoretically, 2^n copies of target DNA will be produced, where n is the number of cycles of heating, cooling, and extension.

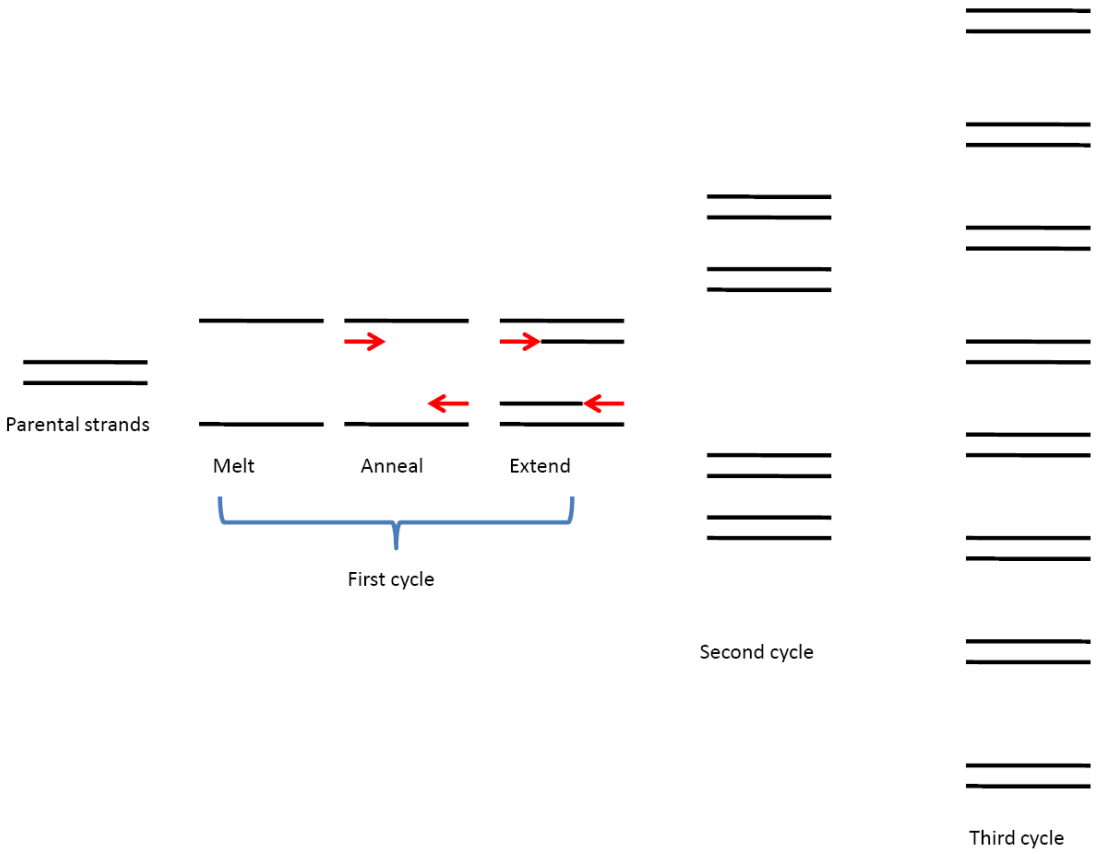


Figure 1-9. Simple schematic of the polymerase chain reaction

1.5.2 Alternative DNA amplification methods

Other DNA based amplification techniques have been developed. These include the ligase chain reaction (LCR), strand displacement amplification (SDA), and linked linear amplification (LLA), but applications of these technologies has been limited, and primarily focussed on clinically important pathogens or detection of mutations linked to human genetic disorders. There has been limited applications in other areas, and alternative amplification technologies are thoroughly reviewed elsewhere (Monis and Giglio, 2006).

1.5.3 Real-time PCR

1.5.3.1 Real-time PCR overview

Perhaps the greatest revolution since the advent of PCR has been the introduction of real-time PCR. Real-time PCR is the monitoring of specific PCR products at the end of every cycle using a fluorescent molecule (i.e. in "real-time"), rather than the detection of an end product (usually by agarose gel electrophoresis) in conventional PCR and other PCR-based methodologies. From the relatively humble yet pioneering beginning of Higuchi's experiments, (Higuchi *et al.*, 1993) who monitored amplification changes using a mounted fluorescent CCD camera and carcinogenic ethidium bromide as a DNA binding dye, several chemistries for detecting PCR product formation have now been developed, without the use of known carcinogens. Real-time PCR is also significantly quicker than conventional PCR, offering results in as little as 30 minutes compared to up to 5 hours when using conventional PCR protocols.

Real-time PCR has significant advantages over conventional PCR. Aside from being a completely closed tube system and therefore decreasing the risk of airborne PCR product contamination, it offers the ability to quantify DNA over a wide dynamic range of concentrations. Real-time PCR exploits the relationship between the points at which exponential amplification is detected, and the starting DNA concentration. This linear correlation is established by generating a standard curve over a large magnitude of orders, typically 10^2-10^8, and expressing the amount of starting DNA as a function of a "threshold cycle" (or C_T) in the form of a regression equation (Figure 1-10). The enumeration of a test sample can then be calculated from its measured C_T and relative position on the regression equation.

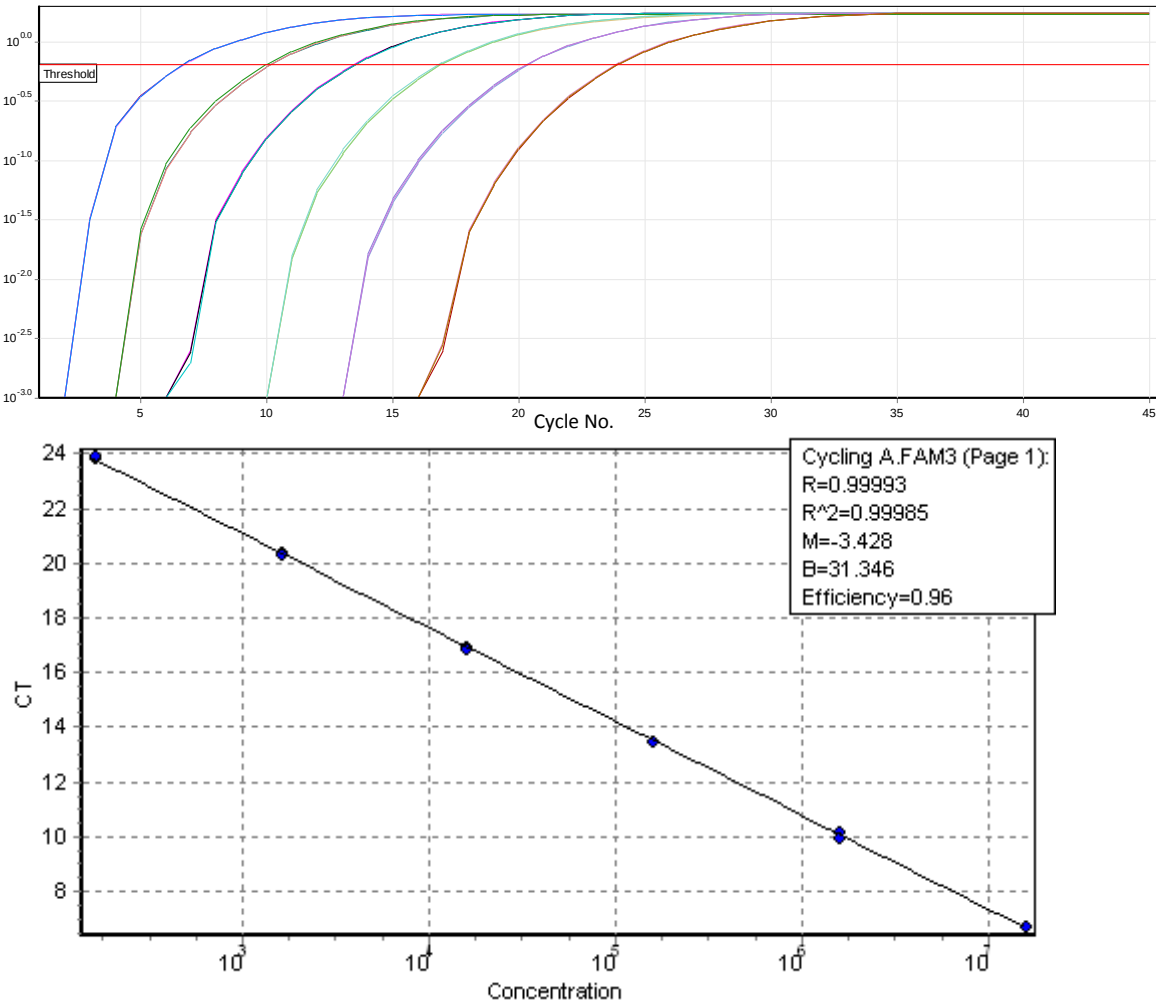


Figure 1-10. Generation of a standard curve using real-time PCR. Each line and point represent a different amount of initial DNA concentration.

1.5.3.2 Detection of real-time PCR products

1.5.3.2.1 Overview

Real-time PCR utilises several chemistries for detection, species identification, and quantification of target organisms. Additionally, a myriad of testing platforms have flooded the market, and differ primarily in the way the reaction chamber is heated and cooled, and the type of optics available for detection of specific fluorophores. Often the choice of testing platform and detection chemistry is driven by the need of the end user and the information required. In essence, all of the platforms operate on the same detection principle, which is the detection of emitted light from an excited fluorophore. The optics of the testing platforms is set up in such a way to detect, for example 6-carboxyfluorescein (FAM) or SYBR Green I, which emits light at 510 nm when excited with light at 470 nm. Other reporter molecules such as JOE, HEX, and TET require excitation at 530 nm, with emission at 550 nm. Other molecules such as ROX and Cy5

require excitation at 585 nm and 625 nm respectively, with emissions detected at 610 nm and 660 nm. The complexity of the detectable fluorophores one chooses to use is totally dependent on the platform in use, and the availability of the optics to excite and detect these molecules. Readers are directed to a complete review of detection chemistries for further information (Monis and Giglio, 2006). Applications of this technology to environmental waters and nuisance organisms are frequently described in the literature, and examples of applications range from the detection of Norovirus (Ngazoa *et al*., 2008) and *Helicobacter pylori* (Nayak and Rose, 2007), to *Naegleria fowleri* (Behets *et al*., 2007), *Bacteroides* sp. (Layton *et al*., 2006), and toxic cyanobacteria including *Cylindrospermopsin* sp., *Microcystis* sp., and *Nodularia* sp. (Rasmussen et al., 2008, Rantala et al., 2006, Koskenniemi et al., 2007).

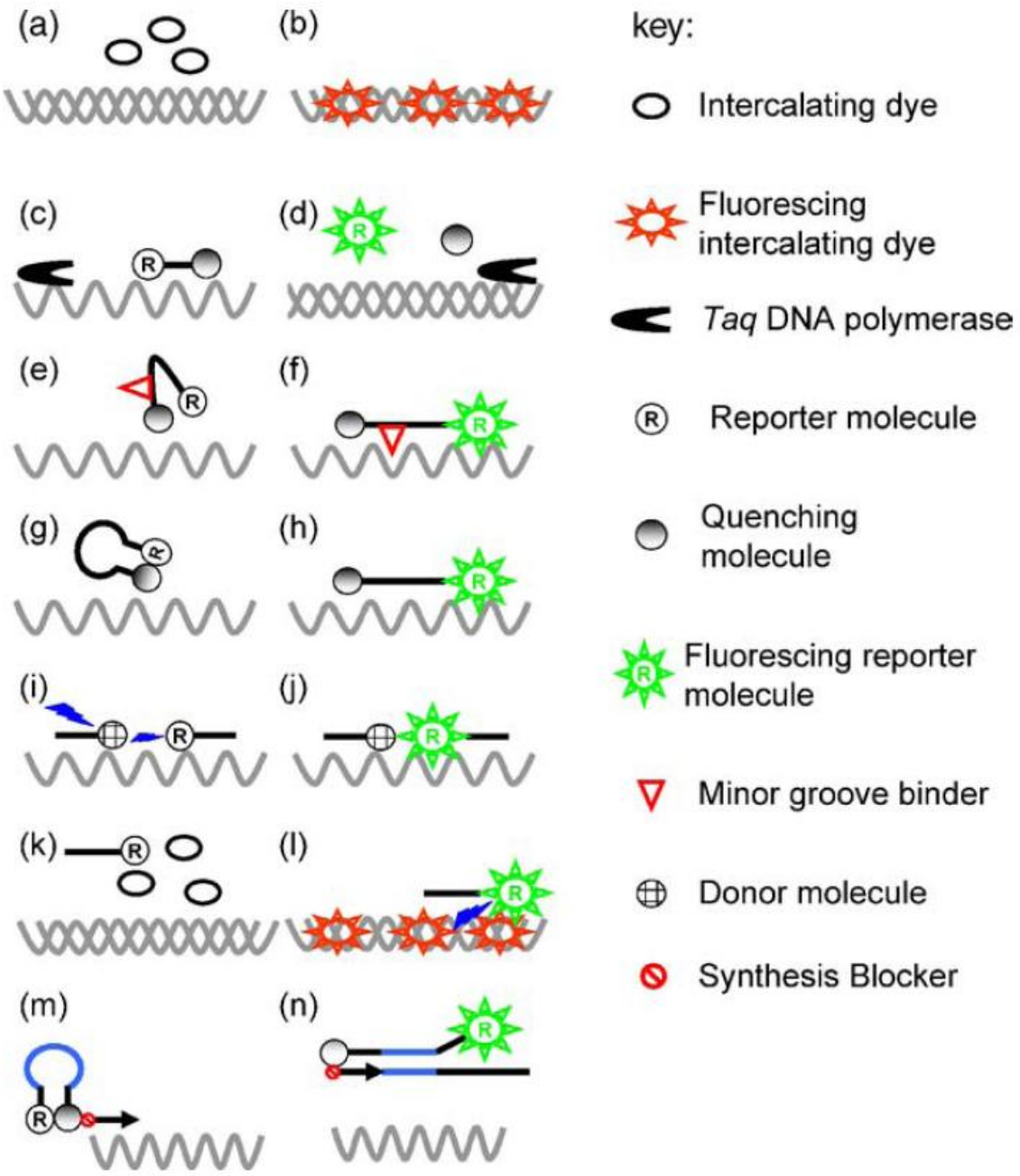


Figure 1-11. Real-time PCR chemistries (reproduced from Monis and Giglio, 2006).Schematic of commonly used real-time PCR chemistries.

In Figure 1-11, a summary of commonly used real-time PCR detection chemistries is presented. Unbound intercalating dye is not fluorescent (a) but will produce a quantum yield increase in fluorescence upon binding to double-stranded DNA (b). Taqman probes cannot fluoresce when intact due to the proximity of the reporter and quencher molecule (c), but will produce fluorescent signal following hydrolysis by Taq polymerase and release of the reporter molecule (d). The secondary structure of MGB Eclipse probes causes the quencher and reporter molecules to be in close proximity so that the reporter will not fluoresce (e), but once bound to

target DNA the probe is stabilised by the minor groove binder and separates the quencher and reporter sufficiently to allow fluorescence (f). Molecular beacons hold the quencher and reporter in close proximity via a stem loop structure, preventing fluorescence when not bound (g) but allowing fluorescence once bound to the target DNA (h). FRET probes anneal to target sequences and energy transfer from the donor to reporter molecule (i) results in increased fluorescence of the reporter molecule (j). Intercalating dyes will not fluoresce unless bound to double stranded DNA and so cannot act as energy donors when in the unbound state for iFRET reactions (k). Upon binding of the intercalator and iFRET probe, energy transfer can occur, resulting in production of fluorescence (l). Scorpion probes are held in a stem-loop structure that prevents probe fluorescence (m), but once incorporated into an amplicon the structure is opened by binding of the loop to complementary sequence within the amplicon (n).

1.5.3.2.2 Intercalating dyes - Overview

Perhaps the most common detection chemistry used in real-time PCR is intercalating dyes. Free in solution, intercalating dyes exhibit low fluorescence, but once bound to double stranded DNA and excited with the appropriate wavelength, there is a substantial increase in fluorescence. Several intercalating dyes have been described including BEBO, YOYO, LC Green, SYTO9, Evagreen and SYBR Green I (Mao *et al*., 2007), all of which are becoming increasingly popular in diagnostic settings. The advantage of intercalating dyes is that the amplicon identity can also be confirmed by melt curve analysis, rather than just an increase in fluorescent signal as is the case with other detection chemistries. The cost of intercalating dyes is also much less than other detection chemistries, and optimisation of PCRs is often easier.

1.5.3.2.3 Intercalating dyes – Melting curve analysis

Melt curve analysis is conducted post thermal amplification and cycling, and involves raising the temperature incrementally from, typically, 70 °C to 95 °C, and measuring the dissociation of the intercalating dye from the PCR generated amplicon. As the temperature increases, and depending on the base pair composition of the sequence, the double stranded DNA continues to dissociate slowly until a specific time that the amplicon completely dissociates, typically over a 2-3 °C range. This rate of dissociation, measured as a rate of change of fluorescence over time (dF/dT), is plotted against temperature and a "melting pattern" is observed, typically with a maximum dF/dT at a given temperature. This is known as the melting temperature or *T*m, and is characteristic of a specific PCR amplicon.

Selected regions of amplified target DNA may also exhibit multiple melting domains. This is usually determined by the absolute sequence and relative juxta-positioning of sequence strings. It is conceivable therefore, that the design of a PCR can be done in such a way to use primers to conserved regions and exploit possible internal variations in melting domains to resolve differences between genera or species. Up until recently, this was done by empirical trial and error. Recent work by Rasmussen *et al.* (Rasmussen *et al*., 2007) discusses the use of bioinformatic resources, such as POLAND and MELTSIM to predict the melting pattern of amplicons. Indeed, the authors describe the application of both software in predicting complex melting patterns for *Cryptosporidium*, *Naegleria* sp. (Figure 1-12), and *Cylindrospermopsis* and *Aphanizomenon*, which were confirmed experimentally on real-time PCR instruments. Robinson *et al.* also describes an extensive validation of a complex melting curve assay for *Naegleria* spp (Robinson *et al*., 2006), a test that is now routinely used to screen *Naegleria* isolates rather than isozyme analysis in that laboratory.

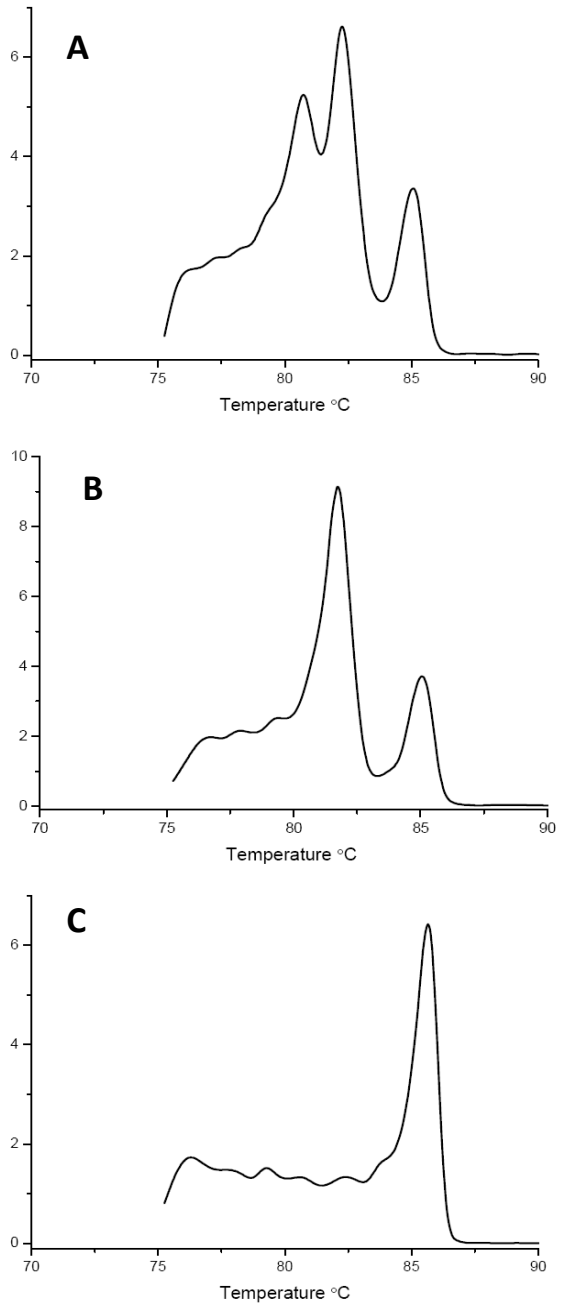


Figure 1-12. Melting curve patterns for A) *Naegleria fowleri*, b) *Naegleria lovaniensis*, and C) *Naegleria australiensis* (Modified from Rasmussen *et al*, 2007).

1.5.4 Future technologies

Several exciting technologies for the detection of nuisance organisms are readily becoming available, although their efficacy is yet to be demonstrated. Moreover, the costs associated with these technologies are currently prohibitive for market uptake. Several lab-on-a-chip biosensor platforms for the detection of target DNA hybridisation have been developed. These platforms differ in the way the target DNA hybridisation event is detected; for example the

Biacore system uses surface plasmon resonance, while the AMBRI system uses ion-channel switches to detect hybridisation events. Both systems look promising, but are yet to be demonstrated in a practical setting. Nanotechnology and the use of sub-microlitre volumes for analyses are in vogue, and much effort to mature this technology is being resourced. No commercial systems have emerged, but the technology and applications look exciting.

1.5.5 Mobilisation of molecular technologies to the field

Although exciting technologies are emerging as described above, currently the application of real-time PCR is mostly within clinical diagnostic laboratories, and is fast becoming an almost "routine" testing protocol. Several clinical diagnostic laboratories utilise this technology on a daily basis, and it is also slowly gaining momentum for use in environmental applications.

In a study by Rasmussen *et al.*, a portable real-time PCR machine was taken on site (Figure 1-13) for the testing of *Cylindrospermopsis* (Rasmussen *et al*., 2008). In this study, the authors firstly detail the specificity of the assay, and secondly deploy this technology to North Pine Dam in Queensland, Australia, to demonstrate the spread and toxicity within the reservoir. This study provided "real-time" data within hours of the sample being collected and the results are discussed later in Chapter 6.

As a result of the 9/11 terrorist attacks in the USA and the anthrax mail attack a few weeks later, The Biological Aerosol Sentry and Information System of the office of Homeland Security in the USA rapidly set up a task force to deal with biological warfare agents such as anthrax, ricin, smallpox and botulinum toxin, as well as biological pathogens, including *Listeria, E. coli, and Salmonella* (http://www.globalsecurity.org/security/library/news/2003/03/sec-030326-afps04.htm). The instrument used in these rapid response teams is the same instrument as used by Rasmussen *et al*. (2008), and thus demonstrates that this type of mobile technology is indeed practical and lends itself to numerous applications that require immediate and important answers.

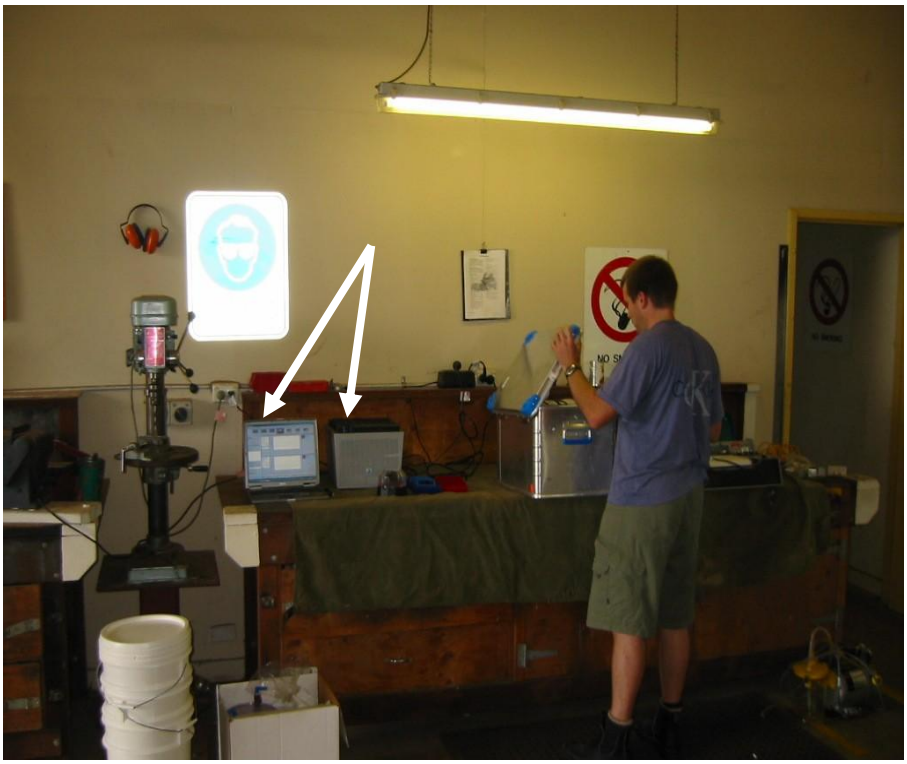


Figure 1-13. Mobile PCR set up (arrows pointing to PCR machine set up).

1.6 GEOSMIN

1.6.1 Background

Geosmin ("ge" from the Greek meaning earth, and "osme" meaning odour) was first identified by Gerber and Lechevalier (Gerber and Lechevalier, 1965) after investigating the earthy odour produced from actinomycetes. Further investigations characterised the structure and it was first proposed that geosmin was likely a degraded sesquiterpene, based on the observed incorporation of labelled acetate in cultures of *Streptomyces antibioticus* (Bentley and Meganathan, 1981).

Geosmin is found in a variety of micro-organisms including actinomycetes (Gerber, 1979), fungi (Schnurer *et al*., 1999), cyanobacteria (Watson, 2003), as well as in symbiotic bacteria (Hayes *et al*., 1991). The distribution of geosmin producing organisms in the environment is large, and the taste and odour it imparts is both desirable and undesirable. For example, the presence of actinomycetes and cyanobacteria in aquaculture and potable water catchment areas is troublesome, and presents significant challenges to the water and aquaculture industry (Sugiura *et al*., 1998, Gagne *et al*., 1999, Van der Ploeg *et al*., 1992, Van der Ploeg and Boyd,

1991, Lloyd and Grimm, 1999, Izaguirre *et al*., 1982, Izaguirre *et al*., 1999, van Breemen *et al*., 1992, Smith *et al*., 2002). Conversely, geosmin is also responsible for the characteristic flavour in beets (Lu *et al*., 2003), and the desirable tastes in some cheeses, nuts, and some vegetables (Frisvad *et al*., 1997).

1.6.2 Human sensory perception of geosmin

Detection of geosmin in potable waters by consumers can occur at levels as low as 10 ng/L (Cook *et al*., 2001), but this is dependent on consumer taste and odour sensitivity. In order to assess this, water utilities routinely subject several source and product waters to some form of qualitative assessment by a panel of trained examiners. Three methods for the assessment and perception of taste and odours; the flavour threshold method, the flavour rating scale, and the flavour profile analysis (FPA) have been previously described (APHA, 1998). Typically, a panel of 4-5 examiners would assess the taste and odour of waters and rate the taste and/or odour on a linear scale. Recently, a simplified profile analysis was developed (Taylor *et al*., 2006) that decreased the number of standards used and eased the interpretive scale, and lends itself to a more streamlined approach to increase turnaround times for these analyses.

Consumers perceive the taste and odour of geosmin to be unsafe, even though it complies with health guidelines (McGuire, 1995), and whilst no enforceable standard limit has been set in Australian drinking waters, action is taken for removal of these compounds when levels reach a defined operational limit. Taste and odour compounds are not easily removed by conventional water treatment techniques (such as flocculation, sedimentation, filtration, and chlorination) and therefore removal of these compounds requires additional treatment on the water before distribution. Differences in operational acceptance limits occur between water authorities, and range from 20 ng/L (Burlingame *et al*., 1986) to 5 ng/L as is set in South Australian waters (Peter Baker, personal communication). These action levels are set for water entering the treatment plant, and when exceeded, remedial action such as the addition of activated carbon is required. The success of these treatments is dependent on the physiochemical composition of the water and the quality of PAC used in dosing (Ho *et al*., 2002, Mouchet and Bonnelye, 1998, Bruce *et al*., 2002, Cook *et al*., 2001). High levels of natural organic matter (NOM) compete with absorptive sites on the PAC (Lalezary *et al*., 1986) therefore decreasing efficiency, and using ozonation has also resulted in the emergence of fruity flavours in treated water (Baker *et al*., 1973). The ability of these technologies, however, may be compromised when dealing with

extremely high levels of geosmin where systems cannot remove geosmin to below an operational threshold (Sklenar and Horne, 1999). Alternative technologies are being investigated, and recent work has suggested that biological filtration using aged sand filters may represent a suitable alternative to remove these undesired compounds (Ho *et al*., 2007), and is currently the focus of much research by these investigators.

1.6.3 Cellular distribution of geosmin within cyanobacteria

In cyanobacteria, relatively little work has been performed investigating the mechanisms of geosmin production, storage, and release. The investigations into these phenomena all but stalled since 1991 when Wu and co-workers demonstrated that geosmin can be found dissolved in the aqueous cytosol, or bound to intracellular or membrane proteins, and that geosmin production may be related to nitrogen assimilation (Wu and Juttner, 1988, Wu *et al*., 1991).

1.7 BIOCHEMISTRY AND GENETICS OF GEOSMIN PRODUCTION - OVERVIEW

1.7.1 Biochemical pathway of geosmin production

The biochemical origin and mechanisms involved in the formation of geosmin are fundamental to propel our understanding of the factors influencing geosmin biosynthesis by organisms in the environment. Early work using radio-labelled acetate demonstrated that the derivation of geosmin is likely to be the result of a sesquiterpene precursor, produced via the mevalonate (MV) isoprenoid pathway (Bentley and Meganathan, 1981). Isoprenoids are a large and diverse group of naturally occurring compounds with varied biological roles, including hormone response and photosynthetic pigment production (chlorophyll and carotenoids), and are also industrially useful as antibiotics and anti-cancer drugs (Ajikumar *et al*., 2008).

Up until the 1950s, the MV pathway was considered the sole avenue for isoprenoid synthesis, but another isoprenoid pathway, the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway (also known as 1-deoxy-D-xylulose-5-phosphate (DXP) pathway), was also discovered (Flesch and Rohmer, 1988). The occurrence of one or both of these pathways within an organism is varied and is especially well summarised by Lange *et al.* (Lange *et al*., 2000) and associated supplementary material. In brief (and in general), the MV pathway is used in plants and animals for the production of major sterol products such as cholesterol, and usually occurs in the cytosol. The MEP pathway is the dominant pathway for isoprenoid synthesis in eubacteria, and in phototrophic organisms it is the MEP pathway involved in the formation of photosynthetic pigments such as chlorophyll and carotenoids. Utilisation of both pathways, however, in

Streptomyces aeriouvifer has been demonstrated (Seto *et al*., 1996). In those organisms with both pathways present (typically higher plants), the MVA pathway is predominantly used for sterol synthesis in the cytosol, while the MEP pathway is used to generate carotenoids, chlorophyll and other terpenoid products within plastid-bearing organelles.

For cyanobacteria specifically, the MEP pathway has been demonstrated in *Synechocystis* (Proteau, 1998, Disch *et al*., 1998) and it seems likely that, given the cyanobacterial origin of photosynthetic apparatus in eukaryotes, the MEP pathway existed in cyanobacteria before being subsequently incorporated within eukaryotes. It is likely that, although there is a common cyanobacterial ancestor for MEP genes, a divergence occurred, at which time a definite and distinct lineage of MEP genes began to emerge in eukaryotes as a result of horizontal gene transfer (Boucher and Doolittle, 2000, Lange *et al*., 2000). However recent compounding, and somewhat confusing, evidence has emerged that challenges the presence of the MEP pathway within *Synechocystis* (Ershov *et al*., 2002), and it seems likely that isoprenoid biosynthesis in photosynthetic organisms may be far more complicated than initially suspected.

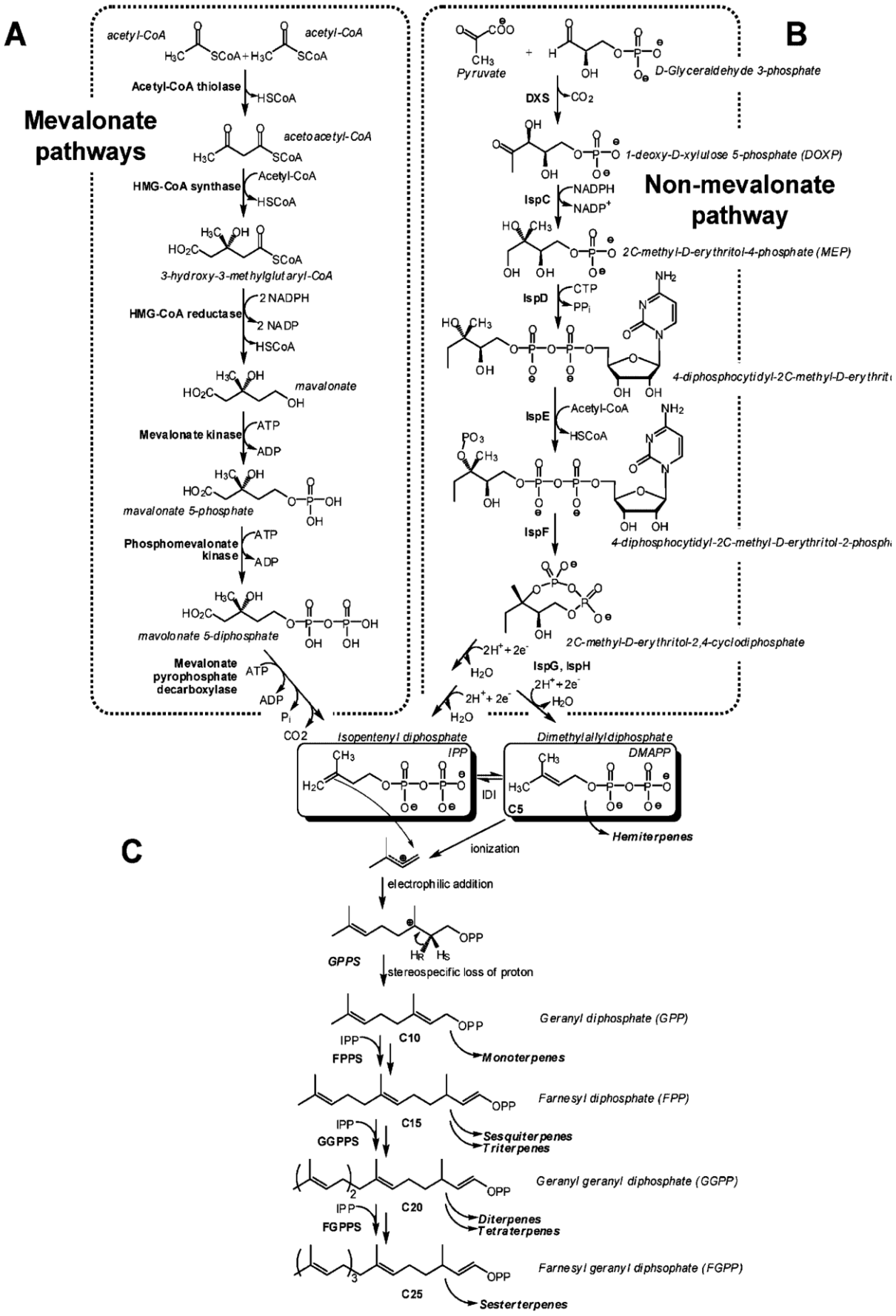


Figure 1-14. The isoprenoid pathways (from Ajikumar *et al*., 2008).

Both the MV and MEP pathways produce a common universal precursor for isoprenoid production, which is isopentenyl diphosphate (IPP). IPP is isomerised to dimethylallyl diphosphate (DMAPP) in those organisms that contain the MV pathway, and formation of some DMAPP occurs in those organisms that have the MEP pathway. IPP and DMAPP are the basic building blocks, where IPP is added to DMAPP to generate the terpenoid precursors geranyl diphosphate (GPP, C10), farnesyl diphosphate (FPP, C15), geranylgeranyl diphosphate (GGPP, C20), and farnesyl geranyl diphosphate (FGPP, C25). These precursors are further converted to produce monoterpenes (from the C10 GPP precursor, mono- indicating a single IPP unit incorporated), sesquiterpenes (from the C15 FPP precursor), diterpenes (from the C20 GGPP precursor), and sesterterpenes (from the C25 FGPP) (Figure 1-14, from Ajikumar *et al*., 2008).

The initial experiments by Bentley and Meganathan (1981) hypothesised that MIB was derived from a C10 precursor and geosmin from the FPP C15 precursor, although little work was done after these experiments to confirm this hypothesis directly. Since 2003, however, much work has been performed on elucidating the biochemical pathway for geosmin production. With the genome of several organisms now fully sequenced and annotated, researchers have been able to assign novel functions to hypothetical proteins and solve long standing biosynthetic mysteries. One such example is the demonstration of the geosmin synthase enzyme in *Streptomyces coelicolor* A3(2) and *Streptomyces avermitilis* (Cane *et al*., 2006, Cane and Watt, 2003).

1.7.2 Genetics of geosmin production

Independently pioneering work by Cane and Watt (Cane and Watt, 2003) and Gust *et al.* (Gust *et al*., 2003) led to the discovery of the function of the SCO6073 gene in *Streptomyces coelicolor* A3(2). SCO6073 encodes a 726 aa protein that contains two similar halves, and through knockout mutants and enzymatic functional analyses, both research groups concluded that SCO6073 was responsible for the formation of an essential intermediate, germacradienol, which was an absolutely required step towards the formation of geosmin.

In these early works, it was postulated that the N-terminal domain was responsible for the formation of germacradienol, and that the similar C-terminal domain was catalytically silent. The experimental data proved that the N-terminal domain catalysed the conversion of farnesyl diphosphate (FPP) to germacradienol, and although these works led to the discovery of an

essential intermediate, the remaining steps in the formation of geosmin from germacradienol still remained unanswered.

With an essential intermediate conclusively involved in the formation of geosmin, speculation grew as to the remainder of the steps involved. Cane and Watt reported a possible pathway for geosmin formation in their initial report (Cane and Watt, 2003), and works on the synthesis of volatiles (including geosmin) on Myxobacteria and liverworts suggested alternative pathways for geosmin production from the essential intermediate germacradienol (Dickschat *et al*., 2005, Schulz *et al*., 2004), but no firm or substantiated claims could be supported.

In 2006 and 2007, rapid advancement on the virtuosity of the SCO6073 enzyme was made. Jiang *et al.* irrefutably demonstrated that the formation of geosmin from FPP was *solely* by the SCO6073 enzyme (Jiang *et al*., 2006), and did not involve additional enzymes as first thought. The failure to detect geosmin in the original experiments by Cane and Watt (Cane and Watt, 2003) was due to the low amount of enzyme used in the functional assays and subsequent inability to detect geosmin. Additionally, supporting work on another actinomycete, *Streptomyces avermitilis*, demonstrated that the highly similar SAV2163 enzyme also possessed similar geosmin synthase functionality (Cane *et al*., 2006). Finally, the SCO6073 enzyme was fully characterized as a unique bi-functional terpene synthase, where the N-terminal domain catalysed the conversion of FPP to germacradienol, and the C-terminal domain, previously thought to be catalytically silent, converted germacradienol finally to geosmin (Jiang and Cane, 2008, Jiang *et al*., 2007), thus dispelling any arguments that additional enzymes were involved in geosmin synthesis in *Streptomyces* sp. Extensive mechanistic details of this remarkable reaction are described in Chapters 3 and 4.

1.8 FINDING THE GEOSMIN GENE IN CYANOBACTERIA

At the core of all the exciting technologies described above, whether it be nanotechnology or real-time PCR, is knowledge of the DNA sequence required for detection. Often the target gene of an assay also has a specific biological function assigned; in the case of toxic cyanobacteria, the *pks* and *ps* genes are part of the cylindrospermopsin (CYN) gene cluster that is reported to be associated with the mechanistic assembly of cylindrospermopsin toxin, and thus detection of either of these genes is associated with toxicity (Fergusson and Saint, 2003, Rasmussen *et al*., 2008, Wilson *et al*., 2000). This makes the *pks* and *ps* gene sequences ideal candidates for an

assay, as the detection of these genes in a cyanobacterium leaves little doubt that the cyanobacterium will produce CYN. Similarly the *mcy*E gene for microcystin production and *nda*F for nodularin production have been successfully exploited in real-time PCR assays for the detection and monitoring of these toxin-producing cyanobacteria (Rantala et al., 2006, Koskenniemi et al., 2007).

Knowing the DNA sequence of the origin of the nuisance or toxic compound is imperative to propel future technologies. The micro-fluidic devices and lab-on-a-chip systems show great promise but still require the detection of a specific DNA:ligand interaction. Elucidation of the gene and specifically determining its functionality may also allow the generation of antibodies to the subsequent protein product, which in turn may also lead to new technologies or methodologies being developed for detection.

Similarly, the determination of what gene is responsible for geosmin production in cyanobacteria would be a valuable target for use in PCR assays to directly detect geosmin producing organisms in waters. The underlying mechanism of geosmin production in cyanobacteria, at the time of commencing this dissertation, was yet to be determined. Investigating candidate gene(s) for geosmin production, and thus linking it to a specific metabolic pathway, will provide the scientific community with the foundation to start our understanding of how geosmin is produced in a cyanobacterium. With this information, the mechanisms by which geosmin production is regulated (if in fact there is regulation) or actively transported may be realised, and thus broaden our conceptual thinking on possible ways to control cyanobacteria and associated noxious blooms. At the very least, the use of the geosmin gene as a target for real time PCR assay will provide water utilities with another management tool to a) predict what preventative measures may be necessary in the short term; b) track the emergence of taste and odour producers in the longer term; and c) anticipate what trends may occur in the future.

1.9 AIMS AND OBJECTIVES

The primary objective of the work presented in this dissertation was to examine the model cyanobacterium *Nostoc punctiforme* for gene(s) involved in geosmin production. With this knowledge in hand, an additional aim was to investigate the regulation of geosmin synthase in an environmental isolate of *Anabaena circinalis*, which is a common problem for SA Water. Furthermore, these works also aimed to determine the prevalence of these genes in environmentally isolated cyanobacteria, with a view towards developing a rapid molecular based tool for use.

Chapter 2

GENERAL MATERIALS AND METHODS

Ultimately they will become empty, mindless spectres…..

The Necromancer, Rush

2.1 GENERAL MATERIALS

2.1.1 Algal culture media

2.1.1.1 ASM medium

Compound	Final concentration (mg/L)
$NaNO_3$	170
K_2HPO_4	17.4
Na_2HPO_4	14.2
$MgCl_2.6H_2O$	40.62
$MgSO_4.7H_2O$	49.33
$CaCl_2.2H_2O$	29.4
$FeCl_3.6H_2O$	1.08
H_3BO_3	2.47
$MnCl_2.4H_2O$	1.37
$ZnCl_2$	0.44
Na_2EDTA	6.64
$CaSO_4$	0.022
$CuCl_2.2H_20$	0.00013

ASM media was autoclaved at 121 °C for 15 min. When cooled, pH was adjusted to 7.6.

2.1.1.2 BG-11 medium

Compound	Amount
$NaNO_3$	1.5 g
K_2HPO_4	0.04 g
$MgSO_4 \cdot 7H_2O$	0.075 g
$CaCl_2 \cdot 2H_2O$	0.036 g
Citric acid	0.006 g
Ferric ammonium citrate	0.006 g
EDTA (disodium salt)	0.001 g
Na_2CO_3	0.02 g
Trace metal mix A5	1.0 mL
Distilled water	1.0 L

Trace Metal Mix A5	Amount
H_3BO_3	2.86 g
$MnCl_2 \cdot 4H_2O$	1.81 g
$ZnSO_4 \cdot 7H_2O$	0.222 g
$NaMoO_4 \cdot 2H_2O$	0.39 g
$CuSO_4 \cdot 5H_2O$	0.079 g
$Co(NO_3)_2 \cdot 6H_2O$	0.049
Distilled water	1.0 L

BG-11 media was autoclaved at 121 °C for 15 min. When cooled, pH was adjusted to 7.1.

2.1.1.3 Nostoc 819 medium

Compound	Amount
K_2HPO_4	0.04 g
$MgSO_4 \cdot 7H_2O$	0.075 g
$CaCl_2 \cdot 2H_2O$	0.036 g
Citric acid	0.006 g
Ferric ammonium citrate	0.006 g
EDTA (disodium salt)	0.001 g
Na_2CO_3	0.02 g
Trace metal mix A5	1.0 mL
Distilled water	1.0 L

Nostoc 819 media was autoclaved at 121 °C for 15 min. When cooled, pH was adjusted to 7.1.

2.1.2 Reagents and chemicals

All reagent and chemicals used were of the highest molecular or analytical grade, unless otherwise specified.

2.1.3 Antibiotics

Ampicillin, chloramphenicol, and kanamycin were purchased from Sigma-Aldrich. Ampicillin and kanamycin were prepared in Milli-Q water (Millipore Ultra Pure Water), and chloramphenicol was prepared in methanol. All antibiotic preparations were filter sterilized through 0.2 µm Acrodisc filters (Pall Life Sciences) prior to use.

2.1.4 Microbial Media

2.1.4.1 Luria-Bertani (LB) agar

10 g tryptone, 5 g of yeast extract, 10 g of NaCl, and 15 g bacteriological agar #1 (Oxoid) were added to 1 L of Milli-Q water and pH adjusted to 7.5. The mixture was autoclaved at 121 °C for 15 min, cooled to 50 °C and poured into 90 mm petri dishes with or without the addition of antibiotics as required.

2.1.4.2 LB broth

As in 2.1.4.1, without the bacteriological agar.

2.1.4.3 Starch-casein agar

10 g soluble starch, 0.3 g casein, 2 g KNO_3, 2 g K_2HPO_4, 0.05 g $MgSO_4.7H_2O$, 0.02 g $CaCO_3$, 0.01 g $FeSO_4.7H_2O$, 0.08 g cyclohexamide, 15 g bacteriological agar #1, and 2 g NaCl were added to 1 L Milli-Q water and autoclaved for 121 °C for 15 min, cooled to 50 °C and poured into 90 mm petri dishes and used as required.

2.1.5 Oligonucleotides

Synthetic oligonucleotides were purchased from Geneworks, Adelaide, Australia, and resuspended in sterile nuclease free water (Promega, USA) to 100 µM (as a stock) and stored at -20 °C prior to use.

2.1.6 SDS-PAGE materials

2.1.6.1 Coomassie Blue Stain

1.2 g Brilliant Blue R, 800 mL methanol, 200 mL acetic acid, and 1 L water were combined and mixed thoroughly. The solution was stored at room temperature, and wrapped in foil until used.

2.1.6.2 2x SDS-PAGE sample buffer

0.6 g Tris-base, 2 mL of 0.25 M EDTA, 50 mL glycerol, 1 mL of 1% bromophenol blue, 1.25 mL of 10% sodium azide, and 10 g sodium dodecyl sulphate (SDS) were mixed to a final volume of 200 mL using Milli-Q water, and pH adjusted to 8.0. After the pH was adjusted, a further 50 mL of Milli-Q water was added.

4.5 mL of above sample buffer was combined with 0.5 mL of 2-mercaptoethanol prior to use. An equal volume of sample buffer (with 2-mercaptoethanol) and sample to be tested were combined and heated at 95 °C for 5 min before loaded into SDS-PAGE wells.

2.1.6.3 SDS-PAGE de-staining solution

As described in 2.1.6.1, minus the addition of Brilliant Blue R.

2.1.6.4 Pre-cast SDS-PAGE gels

Pre-cast 10% SDS-PAGE gels were purchased from BioRad (product code 161-1101) and used and stored as per manufacturer's instructions.

2.1.7 Enzymes

The following DNA polymerases were purchased from Invitrogen: Platinum *Taq* DNA polymerase, and Elongase *Taq* DNA polymerase.

*Nco*I and *Xho*I restriction endonucleases were obtained from Promega and used with appropriately supplied buffers.

T4 DNA ligase was obtained from Promega.

2.2 GENERAL METHODS

2.2.1 DNA extraction

Unless otherwise stated, genomic DNA was extracted from cultures and water samples using the QIAGEN DNA minikit using the tissue protocol procedure, except that pure cyanobacterial cultures were digested with proteinase-K overnight, instead of 1-3 hours.

2.2.2 Plasmid extraction

After overnight incubation in relative LB broths, transformed *E. coli* XL1-Blue and BL21(DE3)pLysS plasmid DNA was extracted using the QIAGEN QiaPrep Spin Mini Prep kit according to manufacturer's instructions.

2.2.3 Restriction enzyme digestion

Approximately 1 µg of DNA of interest was digested with either *Nco*I and/or *Xho*I as required in the relevant chapters. DNA was digested for 3 hours at 37 °C with the restriction enzyme(s) according to manufacturer's protocol, utilising restriction enzyme-specific buffers along with

bovine serum albumin in a final reaction volume of approximately 20 µl. Endonuclease activity was terminated by incubation of the mixture at 65 °C for 30 min.

2.2.4 Agarose gel electrophoresis

PCR product or restriction digest products were combined in equal volumes with 2X loading buffer (Geneworks, Australia). Visualisation of PCR products and restriction digests were performed using 1% (w/v) agarose gel with 10 µl of SYBR safe DNA gel stain incorporated (Invitrogen). DNA Products combined with loading buffer were loaded into the agarose gel in an electrophoresis tank containing TBE buffer (0.1% (w/v) Tris, 0.93% (w/v) EDTA, 0.055% (w/v) boric acid) and a current of 80-100V applied for approximately 1 hour. Upon completion, the gel was removed from the electrophoresis tank and DNA bands visualized using a dark reader illuminator, and images were captured using a digital camera. PCR and digest product sizes were estimated by comparing the relative mobility with DNA ladders included in the electrophoresis.

2.2.5 Purification of PCR products

When required for sequencing or cloning, PCR products were purified from PCR reactions shown to produce a single band of interest. PCR reactions were purified using Montage PCR centrifugal filter devices as per manufacturer's instructions.

2.2.6 *In-vitro* cloning

PCR products were cloned using the pCR 2.1® TOPO ® vector kit (Invitrogen) and supplied chemically competent cells as per manufacturer's instructions.

2.2.7 DNA sequencing

DNA amplicons of interest were sequenced using Big Dye III terminator chemistry by the Australian Genome Research Facility, Adelaide, Australia.

2.2.8 Geosmin analysis by GC-MS

Samples for geosmin analysis (50 mL total volume) were combined with pre-baked sodium sulphate (anhydrous, 10-60 mesh) in a sterile extraction vial. Internal standards (Ultrafine Chemicals, UK) were injected below the sample surface and incubated in a 45 °C water bath for 20-30 minutes. Samples were concentrated using a solid phase microextraction (SPME) polydimethylsiloxanedivinylbenzene (PDMS-DVB) syringe fibre (Supelco, Australia) and analysed on a Hewlett Packard 5890 Series II Gas Chromatograph (using a temperature of 50-280 °C, 20 °C per minute) with Hewlett Packard 5971 Series Mass Selective Detector (Agilent Technologies, Australia) against quantified labelled internal standards.

2.3 SEQUENCE DATA ANALYSIS SOFTWARE

DNA and protein sequence multiple alignments were generated using the CLUSTAL algorithm (Thompson *et al*., 1997).

Analysis of DNA sequencing data was performed using Lasergene v 7.0 software, incorporating Seqman for analysis of sequencing files, EditSeq for manipulations of sequence data, and SeqBuilder for multiple sequence tool analyses.

Primers for sequencing and PCR were designed and evaluated using Primer Premier v 5.0, (Premier Biosoft), and Kodon (Applied Maths).

Chapter 3

ISOLATION AND CHARACTERISATION OF *npun02003620* (ZP_00109187) FROM *Nostoc punctiforme* ATCC 29133 (PCC 73102)

Driven in...Driven to the edge

Driven out...On the thin end of the wedge

Driven off...By things I've never seen

Driven on...By the road to somewhere I've never been

Driven - Rush

3.1 INTRODUCTION

Within the last several years, the mechanism of geosmin biosynthesis in *Streptomyces* and myxobacteria has been investigated in detail (Dickschat *et al*., 2005, Dickschat *et al*., 2007, Schulz *et al*., 2004). Despite earlier confusion regarding how many enzymes were involved, and indeed if the mechanism of geosmin production was different in different organisms, it is now known that in *Streptomyces coelicolor* A3(2) a single 726-amino acid protein, encoded by the 2181-bp SCO6073 gene (*cyc2*) (Cane and Watt, 2003, Gust *et al*., 2003), catalyses the Mg^{2+}-dependent cyclization of farnesyl diphosphate (FPP), the universal acyclic precursor of all sesquiterpenes, to a mixture of geosmin and the sesquiterpene alcohol germacradienol, accompanied by smaller amounts of the bicyclic hydrocarbons germacrene D and 8,10-dimethyl-1-octalin (Jiang and Cane, 2008, Jiang *et al*., 2006, Jiang *et al*., 2007, Nawrath *et al*., 2008) (Figure 3-1). The *S. coelicolor* geosmin synthase is in fact a bi-functional enzyme in which both the N-terminal and C-terminal halves show a high degree of sequence similarity to the well-studied 336-amino acid sesquiterpene synthase, pentalene synthase (Jiang *et al*., 2007). Experiments with individually expressed recombinant proteins corresponding to the N-terminal and C-terminal domains have shown that the N-terminal half of the geosmin synthase catalyses the cyclization of FPP to a 85:15 mixture of germacradienol and germacrene D, accompanied by traces of the octalin, while the C-terminal domain catalyses the Mg^{2+}-dependent cyclization-fragmentation of germacradienol to geosmin, with release of the 2-propanol side chain as acetone (Jiang and Cane, 2008, Jiang *et al*., 2007). Site-directed mutagenesis experiments have confirmed that the N- and C-terminal domains each harbor catalytically independent active sites (Jiang *et al*., 2007)

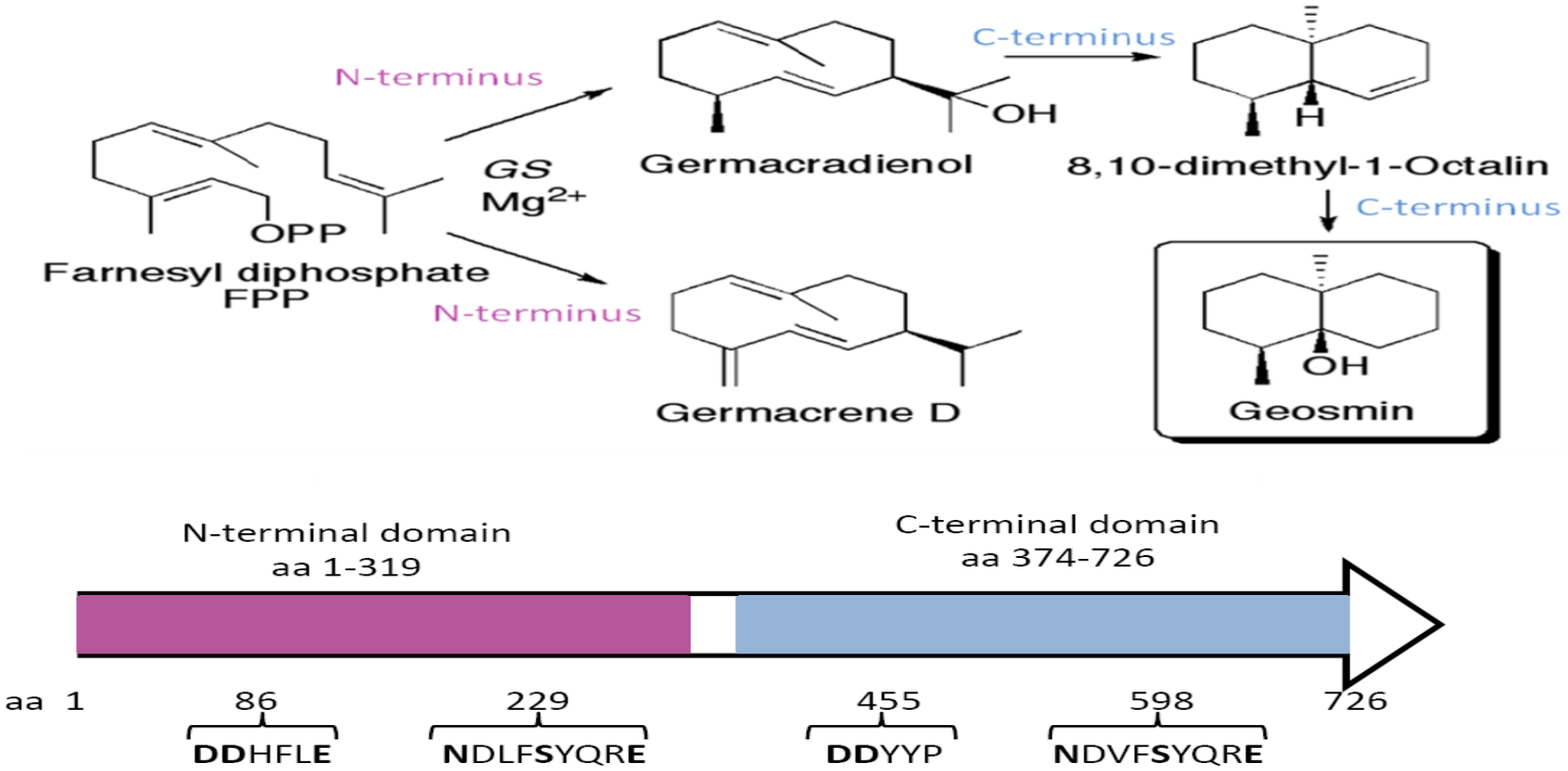


Figure 3-1. Geosmin synthase (GS)-catalysed cyclization of farnesyl diphosphate (FPP) to geosmin, germacradienol, germacrene D, and 8,10-dimethyl-1-octalin. Germacradienol and germacrene D are formed by the N-terminal domain of the bi-functional protein, and geosmin is generated from germacradienol via the octalin by the C-terminal domain (adapted from Jiang *et al*, 2007).

It has also been shown that the closely related SAV2163 protein of *Streptomyces avermitilis* (78% sequence identity and 85% similarity to SCO6073) catalyses the same biochemical reaction (Cane *et al*., 2006), while more than a dozen known or presumed geosmin synthases deduced from a variety of *Streptomyces*, *Frankia, Saccharopoly*sp*ora*, and myxobacterial genome sequences share correspondingly high levels of sequence conservation over all 730-740 amino acids (45-75% identity, 57-85% similarity). In all these proteins, the N-terminal region contains two strictly conserved motifs, a **DD**HFL**E** sequence, typically 80-100 amino acids from the N-terminus, and a **N**D(L/I)F**S**Y(Q/E)R**E** motif approximately 140-amino acids downstream, corresponding to the universally conserved aspartate-rich **DD**XX**D** motif and NSE triad (**N**/**D**)DXX(**S**/**T**)XXX**E,** respectively, that are found in all sesquiterpene synthases and that are known to be involved in the binding of the essential cofactor Mg^{2+} (Christianson, 2006, Komatsu *et al*., 2008). Similarly the C-terminal half of geosmin synthase has a canonical variant of the aspartate rich motif, **DD**YYP, as well as a downstream **N**D(V/I/L)F**S**YQK**E** variant of the NSE triad.

The *npun02003620* gene in the reported genome sequence of the cyanobacterium *Nostoc punctiforme* ATCC 29133 (PCC 73102) encodes a hypothetical protein ZP_00109187 with 55% amino acid sequence similarity to the N-terminal region of SCO6073, including the presence of the universally conserved **DD**HFL**E** and **N**DLF**S**YQR**E** motifs characteristic of this class of enzyme. The predicted protein, however, consists of only 630 amino acids, approximately 100 amino acids shorter than the SCO6073 protein or any other known or predicted geosmin synthase. Equally importantly, the C-terminal half of ZP_00109187 also is predicted to harbor a typical **DD**YFP motif, while the essential NSE triad is apparently absent.

At the commencement of this dissertation, only the N-terminal domain was suspected of being part of a suite of enzymes responsible for geosmin production in *Streptomyces coelicolor*, although this was corrected three years later. The initial hypothesis of only a N-terminal functional enzyme was the result of experimental data from two independently working groups (Cane and Watt, 2003, Gust *et al*., 2003). Ironically, the true activity of the SCO6073 was not fully realised in earlier experiments because the amount of protein initially used was insufficient to detect any geosmin produced (Jiang *et al*., 2007), and secondly, the result of a C-terminal knockout strain from Gust and co-workers was contaminated with a wild-type geosmin

producing strain (personal communication, Gregory L. Challis to David Cane). Given that the reported *npun02003620* sequence had good N-terminal similarity (and hence suspected in being involved in the committed step from FPP towards geosmin via the essential intermediate germacradienol), this gene was investigated further here to determine its biochemical function in cyanobacteria, and moreover, if it could in fact *solely* support the production of geosmin in cyanobacteria, in contrast to the reported activity in *Streptomyces coelicolor* at that time, and in the absence of any other putative geosmin synthase sequence similarity within the deposited *Nostoc* genome sequence.

3.2 AIMS

The aims of the research presented in this chapter were to

- Isolate *npun02003620* from *Nostoc punctiforme,*
- Introduce *npun02003620* into a suitable host protein expression vector, and
- Isolate, purify, and perform functional assays to determine enzymatic activity of the expressed protein product.

3.3 SPECIFIC MATERIALS AND METHODS

3.3.1 Isolation of *npun02003620*

3.3.1.1 Culture and extraction of *Nostoc punctiforme* ATCC29133 DNA

Nostoc punctiforme was cultured in medium 819 and incubated at 20 °C with 20 µmol m^{-2} s^{-1} irradiated in 12 hour cycles of light and dark. DNA was extracted (as described in 2.2.1) when the culture was approximately 3 weeks old. An aliquot of culture was also subcultured onto starch-casein agar to determine if there were any contaminating (and possibly geosmin producing) actinomycete organisms. The subculture was free of such contaminants.

3.3.1.2 Amplification of *npun02003620* and SCO6073

It is with great thanks that SC9B1.20 cosmid (isolate containing the SCO6073 geosmin synthase from *S. coelicolor*, from John Innes Centre,U.K.) was provided as control material for much of this work. SCO6073 was amplified using the primers rmwstartb and rmwhaltb as described previously (Gust *et al*., 2003), and the DNA sequence corresponding to the reported full length *npun02003620* was amplified using the primers npunstart1 (5′-ATTTTATCCATGGTTATGCAACCCTTTGAACTGCCAGAA-3′) and npunhalt1 (5′-

TAATAACTCGAGTTATGGATTTCGCCCTCG-3'), with *NcoI* and *XhoI* restriction sites incorporated (underlined). Fragments were PCR amplified using proof-reading Elongase *Taq* polymerase (Invitrogen) as described by the manufacturer, using 35 cycles consisting of 94 °C for 30 s, 55 °C for 30 s, and 68 °C for 120 s, with a final hold at 4 °C until required. PCR products were screened by agarose gel electrophoresis to check for expected size and to confirm purity. Expected sizes for SCO6073 and *npun02003620* were approximately 2.2 kb and 1.9 kb respectively.

3.3.2 Transformation of PCR products into pET-21d(+) protein expression plasmid

3.3.2.1 Preparation of PCR products for ligation

PCR products from 3.3.1.2 were double digested using *NcoI* and *XhoI* as directed by the manufacturer (Promega). Briefly, 1 µg of PCR product was incubated with 10 U each of *NcoI* and *XhoI* together in the corresponding restriction enzyme buffer solution for 2.5 hours at 37 °C. Restriction enzymes activity was inactivated by incubation at 65 °C for 20 minutes. The resultant digested PCR products were purified using the Montage PCR purification kit as outlined in 2.2.5.

3.3.2.2 Preparation of pET-21d(+) plasmid for ligation

Similarly, the pET-21d(+) vector (Novagen) was double digested as in 3.3.2.1, except 20 U each of *NcoI* and *XhoI* were used, and 3 µg of vector was used.

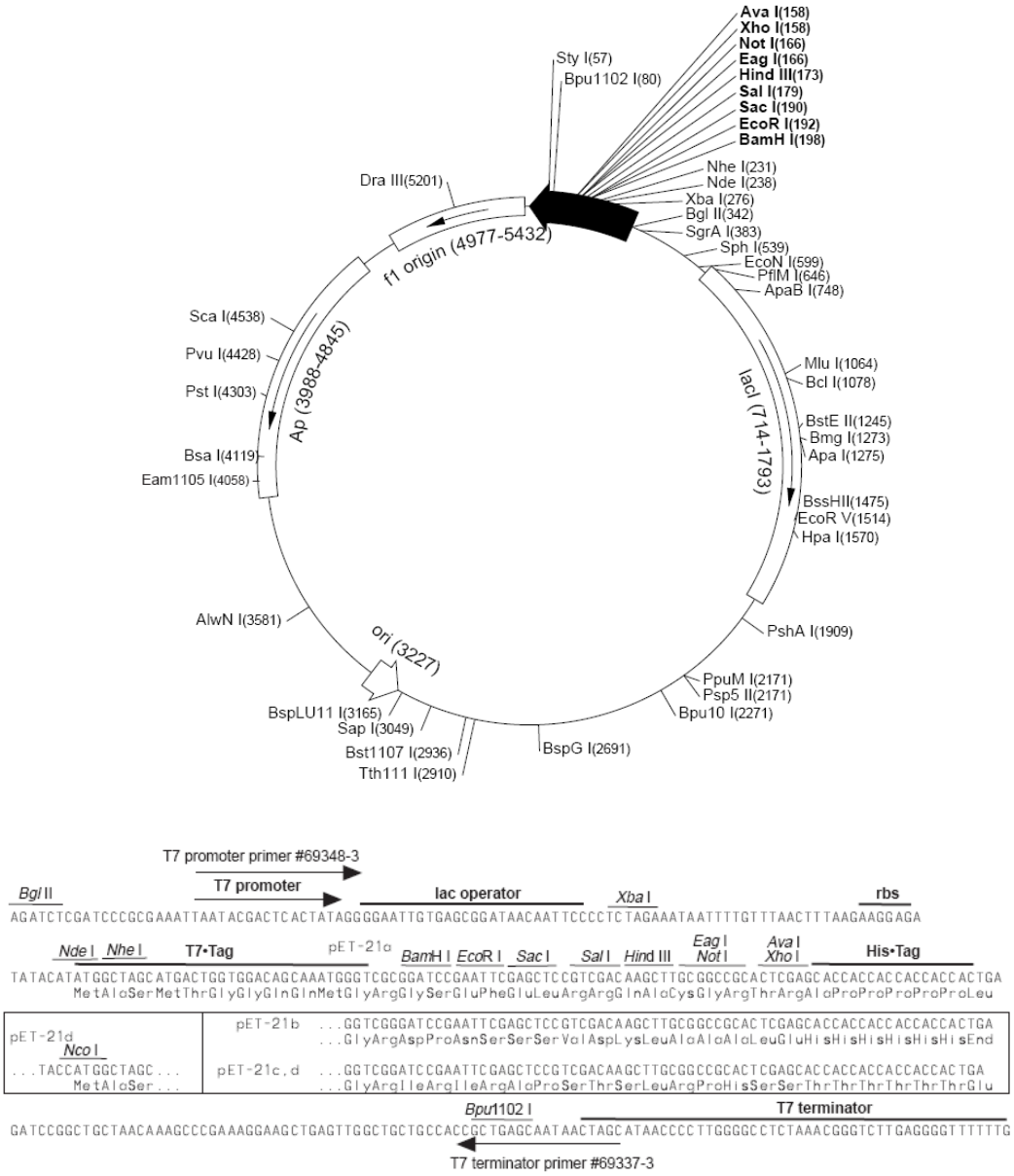


Figure 3-2. Restriction map and cloning insertion sites of pET-21d(+) expression vector (reproduced from pET manual, Novagen).

3.3.2.3 Ligation of SCO6073 and *npun02003620* PCR products with pET-21d(+) vector

Ligation of the vector with the PCR products from 3.3.1.2 was optimized using a 1:3 molar ratio of vector : insert. 100 ng of plasmid was required for the ligation and the amount of PCR product required to satisfy the molar ratio was calculated as below:

ng of PCR product insert required = (100 ng vector/size of vector) x 3,

where pET-21d(+) vector size was approximately 5 kb.

The required amount of plasmid and PCR insert were combined with 1.5 U of T4 DNA ligase and appropriate buffers (Promega) and the ligation reaction performed overnight at 4 °C.

3.3.2.4 Transformation of ligated PCR product and pET-21d(+) into XL1-Blue

The ligated product from 3.3.2.3 was transformed into chemically competent *E. coli* XL1- Blue cells (Stratagene) according to manufacturer's instructions. This step was performed to generate multiple copies of the ligated product (pET-21d(+) plasmid and PCR product) for confirmation of inserts and for the later introduction into a protein expression host *E. coli*. Briefly, 50 ng of ligated product was introduced into XL1-Blue by heat-pulsing the plasmid and pre-conditioned XL1-Blue cells together for 45 seconds, followed by chilling on ice for 2 minutes. 0.9 mL of pre-heated LB broth was added to the mixture and incubated at 37 °C, shaken at 250 rpm, for 1 hour. Aliquots were then subcultured onto LB agar+ 50 µg/mL ampicillin and incubated overnight at 37 °C. The resultant colonies were subcultured into LB broth + 50 µg/mL ampicillin, and incubated overnight at 37 °C. The plasmid was extracted using the QIAGEN Plasmid Extraction Kit (see 2.2.2), and screened for the presence of the desired insert by G-PCR1 (see 3.3.2.5) and subsequently digested with *NcoI* and *XhoI* to further confirm the presence of the full length PCR product along with the presence of pET-21d(+).

3.3.2.5 Screening of successful transformants using G-PCR1

Because the pET vector system was not amenable to the traditional "blue/white" screening utilising isopropyl β-D-1-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-Gal) for selection of successful clones, the screening of candidate clones for a) successful insertion of desired PCR product, and b) successful orientation of the inserted

PCR product, required the inserted sequence to be confirmed by PCR detection, restriction enzyme digest and DNA sequencing.

Screening of successful inserts was performed using a PCR designed to flank both the DDHFLE and NSE triad of the N-terminal half of the SCO6073 gene, which had good similarity to N-terminal functional motifs in *npun02003620*. This PCR, named G-PCR1, consisted of 2.5 mM $MgCl_2$, 1 X PCR buffer, 200 µM dNTP, 0.3 µM each of forward primer 250F (5′-TTCTTCGACGAYCACTTCC-3′) and reverse primer 971R (5′-CCCTYGTTCATGTARCGGC-3′), 5% dimethyl sulfoxide, 1 U of Platinum *Taq* DNA polymerase (Invitrogen), and 2 µL of extracted plasmid DNA. Reactions were run on a Perkin-Elmer 2400 thermal cycler with an initial denaturation step of 95 °C for 5 min, followed by 55 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 120 s, with a final extension step of 10 min at 72 °C. Samples were then run on a 1 % agarose gel with 10 µL of SYBR Safe (Invitrogen) for 45 min at 80 V and bands were visualized using a Dark reader transilluminator (Clare Chemical Research, Dolores, CO). The expected band size for this product was 743 bp.

The plasmid DNA was also double digested with *NcoI* and *XhoI* as described earlier and visualised using agarose gel electrophoresis. The expected digested products were approximately 5 kb of pET-21d(+) and an accompanying band of 2.2 kb for SCO6073, and similarly 1.9 kb for *npun02003620*.

3.3.2.6 Transformation of pET-21d(+) plasmid into BL21(DE3)pLysS *E. coli*

Successfully transformed plasmids from 3.3.2.4 were introduced into chemically competent *E. coli* BL21(DE3)pLysS cells (Stratagene) according to manufacturer's instructions. Briefly, 50 ng of ligated product was introduced into BL21(DE3)pLysS by heat pulsing the plasmid and pre-conditioned BL21(DE3)pLysS cells for 45 seconds, followed by chilling on ice for 2 minutes. 0.9 mL of pre-heated LB broth was added to the mixture and incubated at 37 °C, shaken at 250 rpm, for 1 hour. Aliquots were then subcultured onto LB agar + 50 µg/mL ampicillin + 50 µg chloramphenicol and incubated overnight at 37 °C. A selection of resultant colonies were subsequently subcultured into LB broth + 50 µg/mL ampicillin and 50 µg/mL chloramphenicol, and incubated overnight at 37 °C. Plasmid DNA was extracted and screened using G-PCR1 as described in 3.3.2.5.

3.3.2.7 Sequencing primers

Table 3-1. Sequencing primers used for *npun02003620.*

Primer name	Sequence (5' – 3')
250F	TTCTTCGACGAYCACTTCC
971R	CCCTYGTTCATGTARCGGC
261AR	GTTGATATTGGCGAGTT
173AF	ATGTGAGTACCCAAGAGG
288AF	AACGACCTGTTCTCCTA
173AR	CTGCCAATCCTGAAGTCCTTTT
npunstart1	ATTTTATCCATGGTTATGCAACCCTTTGAACTGCCAGAA
npunhalt1	TAATAACTCGAGTTATGGATTTCGCCCTCG
NA1F	TGACGAAACGCAACAGC

3.3.3 Expression of *npun02003620* and SCO6073

3.3.3.1 Induction of protein production

Protein induction in transformed BL21(DE3)pLysS was performed according to the manufacturer's instructions. Briefly, successful BL21(DE3)pLysS transformants from 3.3.2.6 were cultured in LB broth + 50 µg/mL ampicillin + 50 µg/mL chloramphenicol overnight at 37 °C. The next morning, 50 µL of overnight culture was subcultured into 1 mL of LB broth, with no antibiotics, and shaken at 250 rpm at 37° C for 2 hours. To induce protein production, IPTG was added to a final concentration of 1 mM, and shaken 250 rpm at 37 °C for a further 2 hours.

3.3.3.2 Separation of expressed soluble and insoluble protein fractions

The cells from 3.3.3.1 were harvested by centrifugation at 4,000 g for 10 minutes. Cells were resuspended in BugBuster Protein extraction reagent (Novagen) and incubated at room temperature for 20 minutes with slow shaking. The solution was centrifuged at 16,000 g for 20 minutes at 4 °C. The supernatant was removed and kept aside for determination of protein content present in the soluble fraction. The following steps were performed using reagents from His-bind purification kit (Novagen). The remaining insoluble inclusion body pellet was resuspended in 1 mL of 1X Binding buffer containing 6 M urea and incubated on ice to

resolubilise the inclusion bodies. Insoluble material was removed by centrifugation at 16,000 g for 30 minutes at 4 °C. The supernatant contained resolubilised inclusion bodies.

3.3.3.3 SDS-PAGE

20 µL of protein extracts (soluble and insoluble inclusion bodies) were separately combined with 20 µL of 2X SDS sample buffer and incubated at 95 °C for 5 minutes. 25-30 µL of each sample was loaded into Ready Gel Tris-HCl Gel, 10% resolving gel, 4% stacking gels (Biorad) and electrophoresed in a Mini-Protean Gel System (Biorad) at 150 V for 1 hour. Protein bands were visualised by staining with Coomassie Blue and de-staining according to standard protocols (Sambrook *et al*., 1989). The relative mobility and size of the experimental protein were compared to a reference protein standard (Sigma M2789) run in parallel (Table 3-2).

Table 3-2. Sigma protein standard subunit molecular weight markers.

Protein	Subunit Molecular Weight
α-lactalbumin, Bovine Milk	14,200
Tryspin inhibitor, Soybean	20,100
Carbonic anhydrase, Bovine erythrocyte	29,000
Ovalbumin, Chicken egg	45,000
Albumin, Bovine serum	66,000
Phospohorylase b, Rabbit muscle	97,400
β-galactosidase, E. coli	116,000
Myosin, Rabbit muscle	205,000

3.3.3.4 Incubation of *npun02003620* protein with farnesyl diphosphate (FPP)

These analyses were initially performed by Jiaoyang Jiang, Brown University, Rhode Island, USA, and reproduced by S. Giglio during an 8 week study visit at Brown University. For GC-MS analysis, protein from the soluble fraction was incubated with 500 nmol of FPP in 5.0 mL of 50 mM Tris-HCl, pH 8.2, containing 20% glycerol, 4 mM $MgCl_2$, overlaid with 3 mL of HPLC-grade pentane in 18 × 150 mm glass tubes fitted with a rubber septum for 6 h at 30 °C. The reaction mixture was extracted with 3 × 3 mL pentane/CH_2Cl_2 (5:1), and the combined organic layers

were dried by passage through a Pasteur pipette containing 1 g of $MgSO_4$ before concentration under reduced pressure at 0 °C to 200 µL. A 1 µL portion of the concentrated extract was analysed by GC-MS (30 mm× 0.25 mm HP5MS capillary column, using a temperature program of 50–280 °C, 20 °C per min). Each compound was identified by direct comparison with authentic standards or with the spectra of the corresponding reference compounds in the MassFinder 3.0 Library (http://www.massfinder.com).

3.4 RESULTS AND DISCUSSION

3.4.1 Introduction of *npun02003620* into BL21(DE3)pLysS protein expression system

The gene encoding the hypothetical protein ZP_00109187 encoded by *npun02003620* was successfully amplified using npunstart1 and npunhalt1 primers, producing a band of approximately 1.9 kb (Figure 3-3). Using rmwstart1 and rmwhalt1 primers, the SCO6073 gene was similarly amplified.

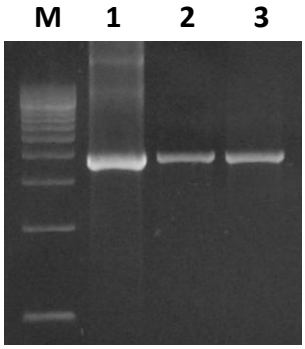


Figure 3-3. Amplified *npun02003620* (M: 500 bp marker; 1-3: replicate *Nostoc punctiforme* DNA).

SCO6073 and *npun02003620* genes were introduced into pET-21d(+) after both the vector plasmid and PCR fragments were double digested with *XhoI* and *NcoI* and subsequently ligated with T4 DNA ligase. The ligated products were introduced into XL1-Blue *E. coli* cells in order to generate a large amount of plasmid for future transformation into BL21(DE3)pLysS, and characterisation of the plasmid with the inserted sequence.

After overnight incubation on LB agar + ampicillin, four colonies were evident for the SCO6073 transformation, and thirty colonies for the *npun02003620* transformations. All colonies from both the SCO6073 and *npun02003620* agar plates were then subcultured into LB broth + ampicillin and the plasmid extracted after overnight incubation.

To test for the successful insertion and orientation of *npun02003620* and SCO6073, G-PCR1 was used to screen extracted plasmid from transformed XL1-Blue cells (Figure 3-4). The result of the screening process was that out of the four chosen SCO6073 transformants, only one produced the desired PCR fragment when screened with G-PCR1. Similarly, five from the thirty screened *npun02003620* transformants produced the desired PCR band of expected size. These isolates were stored at -80 °C and named SG4, 5, 6, 7, and 8 (Table 3-3).

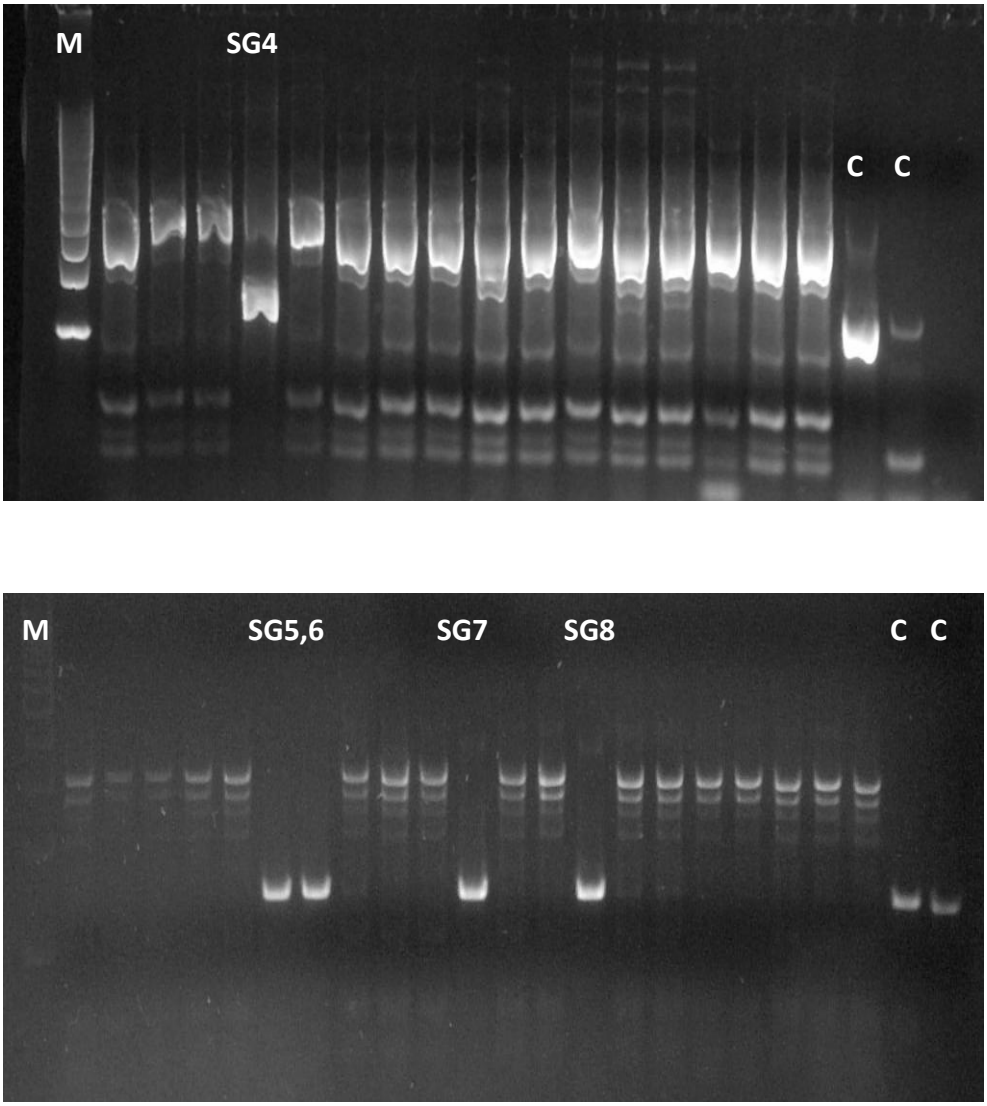


Figure 3-4. Screening of XL1-Blue and BL21(DE3) transformants with G-PCR1. Isolates SG4-8 labelled. M, 500 bp marker; C, control DNA.

Table 3-3. Laboratory strains generated from pET-21d(+) transformations

SG#	Lab ID	Description
4	nos7	XL1 BLUE, pet21d *npun02003620* insert
5	nos6	XL1 BLUE, pet21d *npun02003620* insert
6	nos14	XL1 BLUE, pet21d *npun02003620* insert
7	nos11	XL1 BLUE, pet21d *npun02003620* insert
8	sco2	XL1 BLUE, pet21d sco6073 insert
9	nost1	BL21(DE3) pLysS *npun02003620* insert, protein expression ready
10	nost2	BL21(DE3) pLysS *npun02003620* insert, protein expression ready
11	nost3	BL21(DE3) pLysS *npun02003620*insert, protein expression ready
12	nost4	BL21(DE3) pLysS *npun02003620* insert, protein expression ready
13	nost5	BL21(DE3) pLysS *npun02003620* insert, protein expression ready
14	nost6	BL21(DE3) pLysS *npun02003620* insert, protein expression ready
15	nost7	BL21(DE3) pLysS *npun02003620*insert, protein expression ready
16	nost8	BL21(DE3) pLysS *npun02003620* insert, protein expression ready
17	nost9	BL21(DE3) pLysS *npun02003620* insert, protein expression ready
18	nost10	BL21(DE3) pLysS *npun02003620* insert, protein expression ready
19	sco1	BL21(DE3) pLysS sco6073 insert, protein expression ready
20	sco2	BL21(DE3) pLysS sco6073 insert, protein expression ready
21	sco3	BL21(DE3) pLysS sco6073 insert, protein expression ready
22	sco4	BL21(DE3) pLysS sco6073 insert, protein expression ready
23	sco5	BL21(DE3) pLysS sco6073 insert, protein expression ready
24	sco6	BL21(DE3) pLysS sco6073 insert, protein expression ready
25	sco7	BL21(DE3) pLysS sco6073 insert, protein expression ready
26	sco8	BL21(DE3) pLysS sco6073 insert, protein expression ready
27	sco9	BL21(DE3) pLysS sco6073 insert, protein expression ready
28	sco10	BL21(DE3) pLysS sco6073 insert, protein expression ready

Plasmid DNA from SG7 (with *npun02003620* insert) and SG8 (with SCO6073 insert) were selected to transform BL21(DE3)pLysS cells. Using the protocol detailed in 3.3.2.6, there was a high level of successful transformation into the BL21(DE3)pLysS host, with >200 colonies observed per agar plate for both SG7 and SG8 isolates. Ten SG7-BL21(DE3)pLysS- and ten SG5-BL21(DE3)pLysS- transformed colonies were randomly chosen from the agar plates, confirmed for the presence of G-PCR1 PCR fragment (data not shown), and subsequently stored at -80 °C as isolates SG9-18 and SG19-28 (Table 3-3) respectively until required.

Plasmid DNA from representative isolates SG7, 14, 15, 23, and 24 was extracted, double digested with *NcoI* and *XhoI* and visualised by agarose gel electrophoresis. Figure 3-5 shows that the expected insert size for *npun02003620* of approximately 1.9 kb (confirmed by sequencing, see 3.4.3) was apparent in isolates SG14, 15, and 7, as was the expected insert size of 2.2 kb SCO6073 for isolates SG23 and 24.

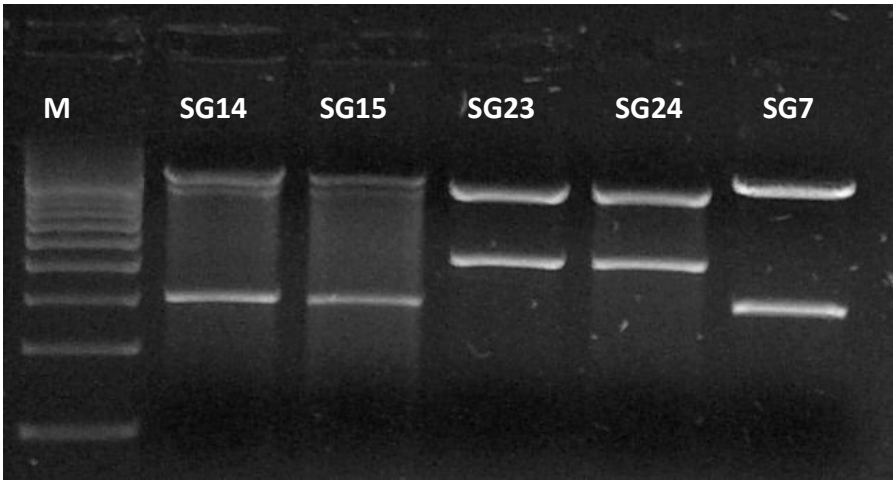


Figure 3-5. *NcoI* and *XhoI* double digested plasmids from SG14, 15, 23, 24, and 7. M, 500 bp marker. SG14 and 15 are BL21(DE3) transformants using plasmid from SG7, SG23 and 24 are BL21(DE3) transformants using plasmid from SG8 (SG8 not pictured).

3.4.2 Preliminary expression screening of SCO6073 and *npun02003620* protein

Sequencing of plasmid DNA from SG7 (the original plasmid DNA from XL1-Blue used to transform the BL21(DE3)pLysS cells) occurred concurrently with the following expression studies.

Several attempts to induce the *npun02003620* and SCO6073 proteins were initially unsuccessful because it became apparent that the quality of IPTG used was suspect. However, once new IPTG was sourced, SCO6073 protein was successfully expressed and extracted from inclusion bodies as demonstrated by the SDS page band of expected size of approximately 85 KDa in isolates SG23, 24, and 25 (Figure 3-6). The expressed protein was absent in the non-induced control, highly expressed in the insoluble fraction from induced cells containing the inclusion bodies, and absent (or below detection limit) in the soluble fraction from the same cells. This is consistent with the original findings of Cane and Watt, who demonstrated that the SCO6073 protein is >95% expressed within inclusion bodies (Cane and Watt, 2003). Isolates SG13, 14, and 15, however, failed to express *Nostoc* protein as indicated by the absence of a protein band on the SDS PAGE at an expected molecular weight of 71KDa. (Figure 3-7).

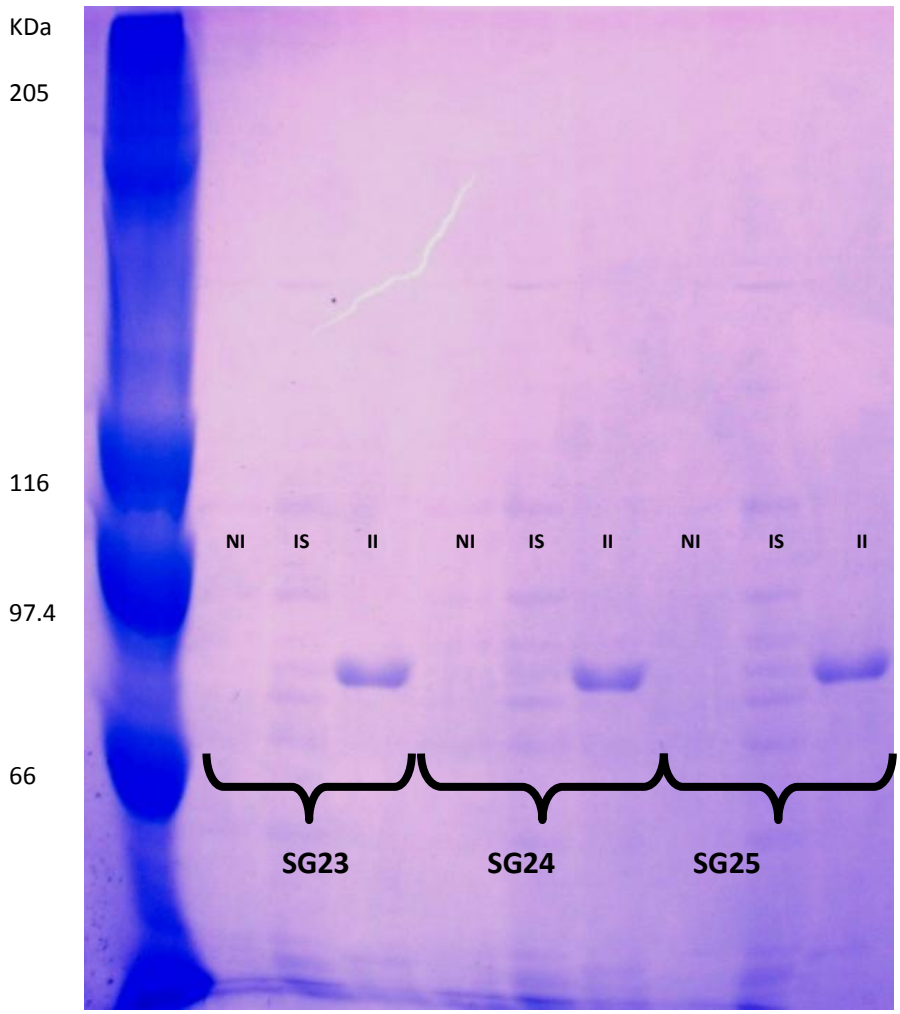


Figure 3-6. SCO6073 transformant expression profiles for isolates SG23, 24, and 25. NI, non induced fraction; IS, induced soluble fraction; II, induced inclusion body fraction. Molecular marker, left lane.

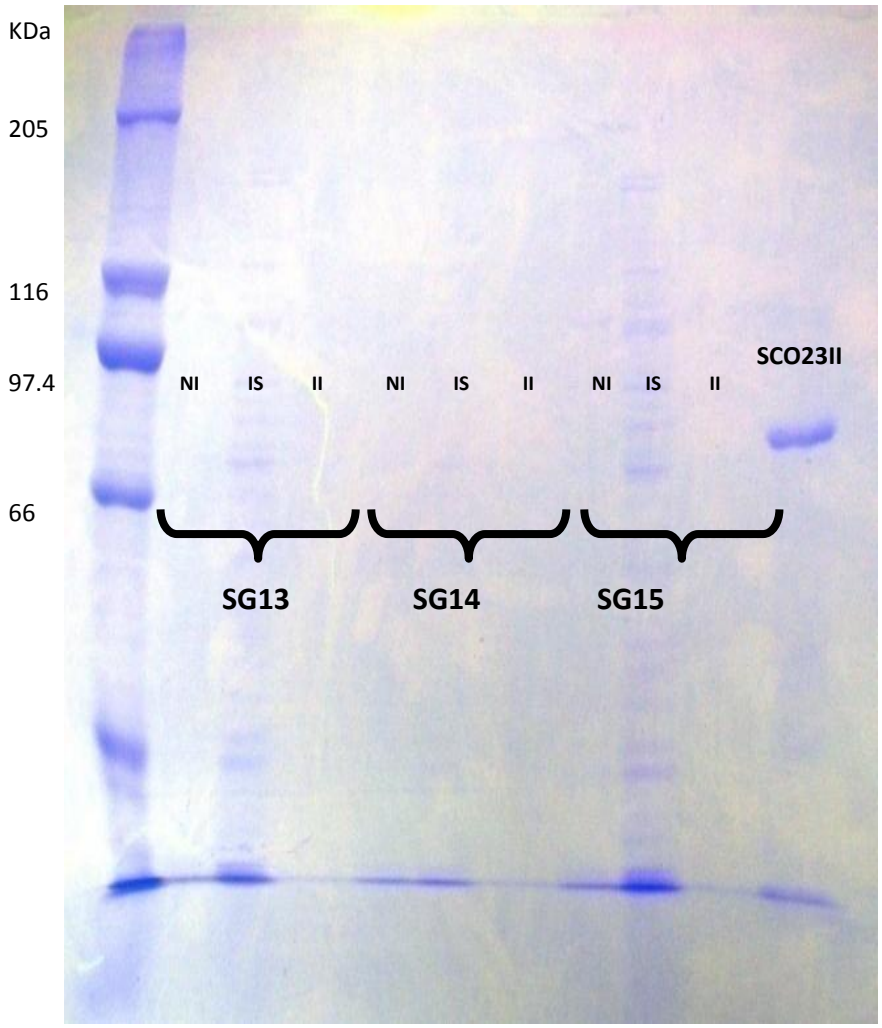


Figure 3-7. Transformed *npun02003620* expression profiles for isolates SG13, 14, and 15. NI, non induced fraction; IS, induced soluble fraction; II, induced inclusion body fraction. Molecular marker, left lane; SCO23 II (induced inclusion body fraction) positive control.

3.4.3 Sequencing of SG7

Sequencing primers described in 3.3.2.7 were used to sequence the plasmid used for the *npun02003620* transformations, as isolated from SG7. Sequencing of the insert and flanking regions revealed that although the *npun02003620* sequence had successfully been introduced and ligated into the pET-21d(+) plasmid (as demonstrated by the positive G-PCR1 screens on isolates SG4-7 and SG9-18, and expected double digest products, and approximately 1.9 kb of sequence data), the ribosome binding site (GAAGGA) which should be 21nt upstream of the start codon, and the T7 promoter region (TAATACGACTCACTATA) which should be 104nt upstream of the start codon, were both absent. This was a highly unusual phenomenon, and

may have resulted from experimental processes, quality of the original pET21d(+), or some other recombination event of unknown aetiology. Both of these sites are crucial to the initiation of transcription and subsequent translation of protein from the BL21(DE3)pLysS cells, and the absence of these sites explains the lack of protein expression in isolates SG13, 14, and 15, even after several optimisation conditions were trialled (modification of induction time, initial OD_{600} inoculation levels, and temperature; data not shown, nor detailed in methods). Given the difficulties with obtaining a successful *npun02003620* transformed isolate, the entire process was re-initiated in order to generate a successful clone.

3.4.4 Successful *npun02003620* expression

As detailed previously, *npun02003620* was re-amplified from *Nostoc punctiforme*, digested, introduced into pET-21d(+), transformed into XL1-Blue and subsequently transformed into BL21(DE3)pLysS with high efficiency. G-PCR1 screening confirmed the presence of a 743 bp fragment, and double digests with *NcoI* and *XhoI* confirmed the presence of a 1.9 kb insert in isolate NJ2, along with another unexplainable recombination event in a similarly transformed isolate NJ1 (Figure 3-8).

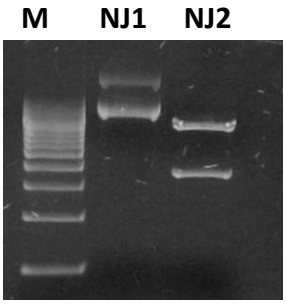
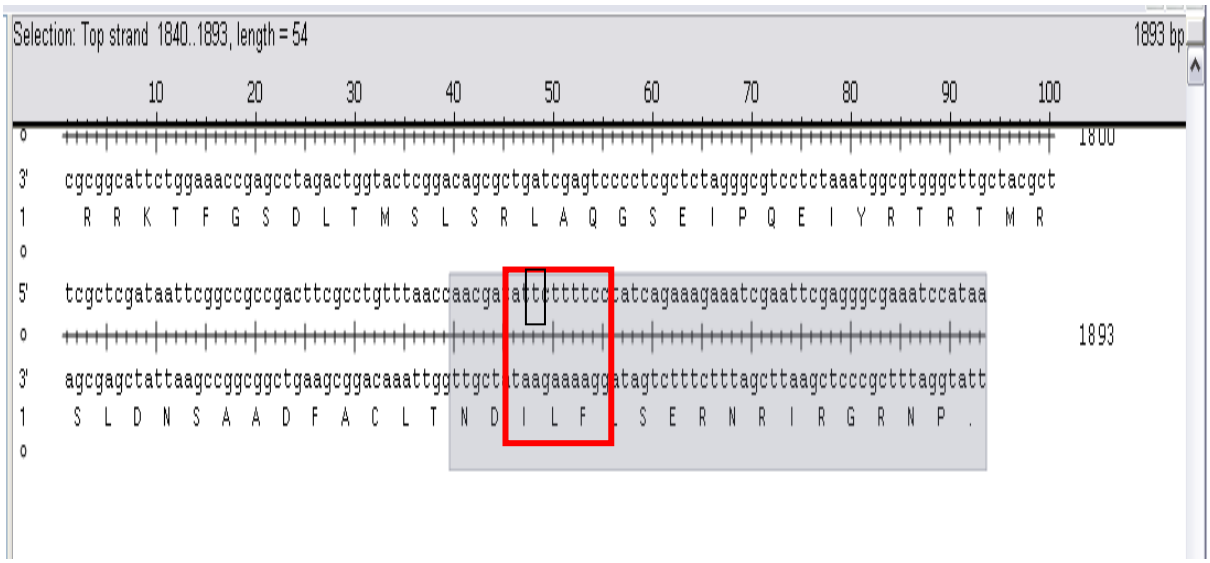


Figure 3-8. Demonstration of NJ2 double digest pattern. M, 500 bp marker.

Sequencing of NJ2 confirmed the correct insert and presence of both the T7 promoter region and the ribosomal binding site. Apart from the last 7 nucleotides that failed to sequence successfully, the experimental NJ2 DNA sequence was almost identical to the *npun02003620* sequence published on the GenBank database. At position 1847 of the sequence (corresponding to position 33677 in the whole genome sequence), an additional “T” nucleotide was apparent in the published sequence when compared to the experimentally derived NJ2 sequence (GenBank accession number FJ010202). When the NJ2 DNA sequence was translated, the corresponding amino acid position 616 (isoleucine) sequence and surrounding sequences were changed from NDILFLSER in the deposited ZP_00109187 sequence, to NDIFSYQKE in the

experimentally derived NJ2. When this apparent change was compared to the SCO6073 and SAV2163 proteins, the NDIFSYQKE was similar to corresponding sequence positions, and a previously unidentified NSE triad was revealed at the extremity of the deposited ZP_00109187 sequence (Figures 3-9, 3-10, 3-11). Full sequences are available in Appendices 1-1 and 1-2.



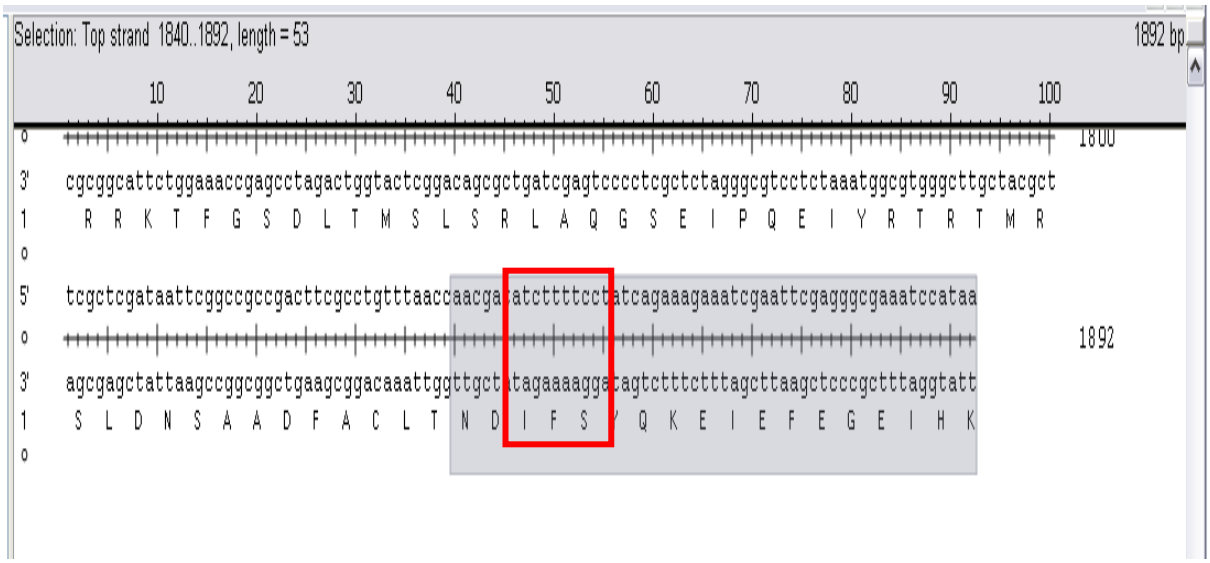


Figure 3-9. Additional "T" nucleotide at position 1847 in the deposited sequence (boxed in black within the red box, top figure), and resultant corrected sequence and translated product (bottom figure).

```
NJ2                 CCGATAGATTATGTTGAGATGCGCCGTAAGACCTTTGGCTCGGATCTGACCATGAGCCTG 1740
NPUN02003620        CCGATAGATTATGTTGAGATGCGCCGTAAGACCTTTGGCTCGGATCTGACCATGAGCCTG 1740
                    ************************************************************

NJ2                 TCGCGACTAGCTCAGGGGAGCGAGATCCCGCAGGAGATTTACCGCACCCGAACGATGCGA 1800
NPUN02003620        TCGCGACTAGCTCAGGGGAGCGAGATCCCGCAGGAGATTTACCGCACCCGAACGATGCGA 1800
                    ************************************************************

NJ2                 TCGCTCGATAATTCGGCCGCCGACTTCGCCTGTTTAACCAACGATAT-CTTTTCCTATCA 1859
NPUN02003620        TCGCTCGATAATTCGGCCGCCGACTTCGCCTGTTTAACCAACGATATTCTTTTCCTATCA 1860
                    ********************************************** ************

NJ2                 GAAAGAAATCGAATTCGAGGGCGATA------- 1885
NPUN02003620        GAAAGAAATCGAATTCGAGGGCGAAATCCATAA 1893
                    ************************* *
```

Figure 3-10. Multiple alignment of experimentally derived NJ2 and *npun02003620* demonstrating an additional "T" nucleotide boxed in yellow. Full sequences available in Appendix 1-1.

```
SCO6073             VDDMTEAWLWELSNQIQNRVPDPVDYLEMRRATFGSDLTLGLCRAGHGPAVPPEVYRSGP 582
SAV2163             VDVMTESWVWELSNQLQHRIPDPIDYLEMRRATFGADLTLSLCRVGHGPKVPPEIYRSGP 581
ZP_00109187         IQDMTDSWVWELANQIQNRIPDPIDYVEMRRKTFGSDLTMSLSRLAQGSEIPQEIYRTRT 598
NJ2                 IQDMTDSWVWELANQIQNRIPDPIDYVEMRRKTFGSDLTMSLSRLAQGSEIPQEIYRTRT 598
                    :: **::*:***:**:*:*:***:**:**** ***:***:.*.* .:*. :* *:**: .

SCO6073             VRSLENAAIDYACLLNDVFSYQKEIEYEGEIHNAVLVVQNFFGVDYPAALGVVQDLMNQR 642
SAV2163             VRSLENAAVDYGMLINDVFSYQKEIEYEGEVHNAILVVQNFFGCDYPTALGVINDLMTQR 641
ZP_00109187         MRSLDNSAADFACLTNDILFLSERNRIRGRNP----------------------------- 630
NJ2                 MRSLDNSAADFACLTNDIFSYQKEIEFEGI------------------------------ 628
                    :***:*:* *:. * **::  .:. . .*
```

Figure 3-11. Multiple alignment of amino acid sequences of SCO6073 from *S. coelicolor*, SAV2163 from *S. avermitilis*, ZP_00109187 from *Nostoc punctiforme*, and experimentally derived NJ2. The relieved frame shift from the additional "T" nucleotide in protein product ZP_00109187 unveiled a previously masked NSE triad that was apparent in experimental NJ2. Full sequences available in Appendix 1-2.

Induction and extraction of the NJ2 protein was performed as previously described. The expected molecular weight of ZP_00109187 (and NJ2) was approximately 71 KDa. SDS-PAGE demonstrated the presence of a protein band in the inclusion body fraction approximately of this size (Figure 3-12).

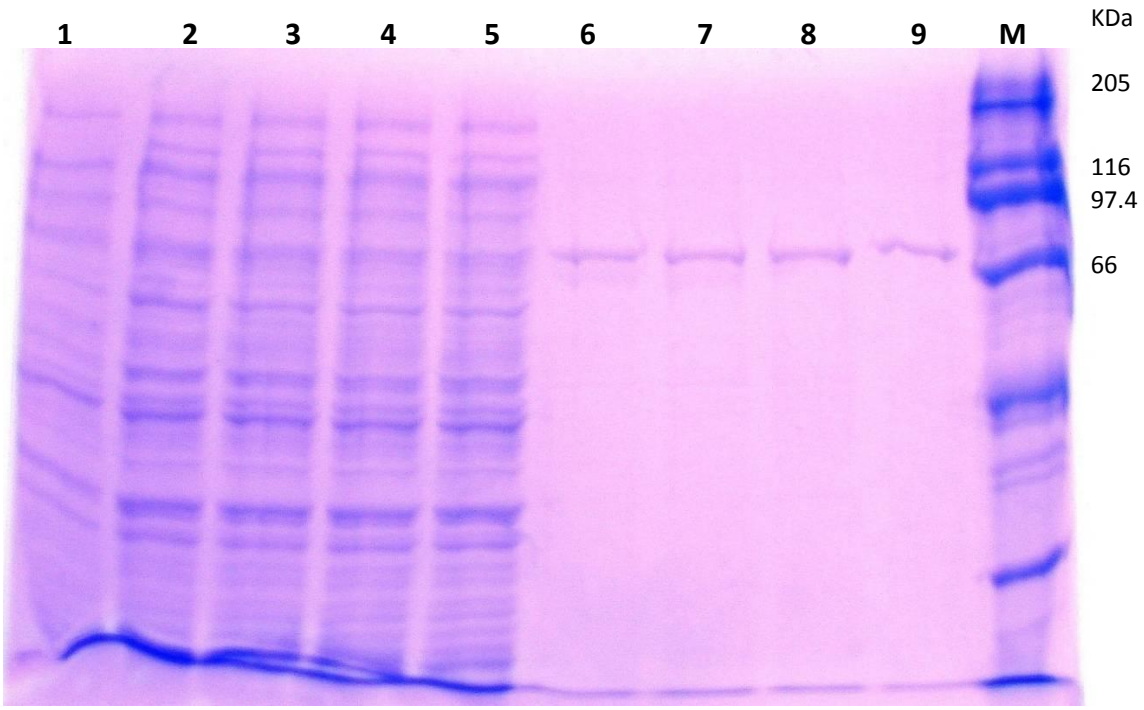


Figure 3-12. SDS-PAGE of NJ2 expression. 1, pooled non-induced fractions; 2-5, replicate induced soluble fractions; 6-9, replicate induced inclusion body fractions; M, Molecular weight marker.

3.4.5 Biochemical activity of *npun02003620*

NJ2 protein from the soluble fraction was incubated with FPP as described in 3.3.3.4. Incubation of recombinant NJ2 protein with FPP yielded a mixture of germacradienol (94%, retention time 11.29 min) and germacrene D (6%, retention time 10.34 min) accompanied by a trace of octalin, but without any detectable geosmin according to standard GC-MS analysis (Figure 3-13).

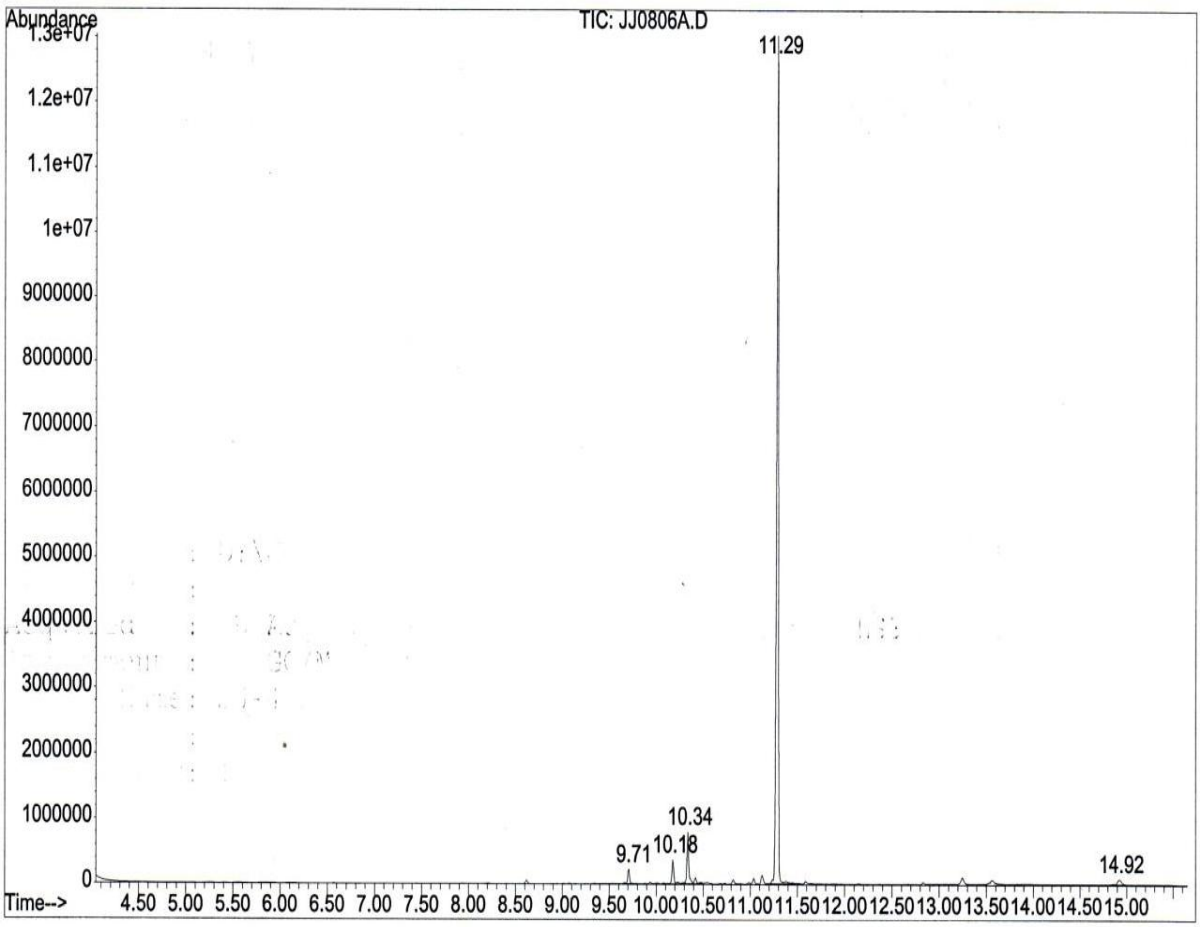


Figure 3-13. GC-MS trace for the incubation of NJ2 with FPP, with major products of germacradienol (retention time 11.29 min) and germacrene D (retention time 10.34 min).

The reported 1893-bp *npun02003620* gene of *Nostoc punctiforme* ATCC 29133 (PCC 73102), annotated as encoding a hypothetical protein (ZP_00109187), has 55% amino acid sequence similarity to the N-terminal region of both known geosmin synthases, *S. coelicolor* A3(2) SCO6073 and *S. avermitilis* SAV2163, thereby suggesting the likely biochemical function of this cyanobacterial gene may be similar to the N-terminal mediated reactions of the reported *Streptomyces* geosmin synthase. The reported open reading frame is, however, almost 300 bp shorter than either the SCO6073 or SAV2163 gene or any of the homologous actinomycete and myxobacterial geosmin synthase genes. The predicted ZP_00109187 protein also lacks a NSE triad in the C-terminal domain that is found in all other geosmin synthases as well as in all other known terpene synthases. To assess the biochemical function of the reported *npun02003620* open reading frame, recombinant ZP_00109187 protein was generated based on the start and stop codons predicted by the deposited sequence, resulting in a 71 KDa protein of the expected size as determined by SDS-PAGE, as demonstrated by NJ2. Incubation of recombinant NJ2

expressed protein with FPP yielded a mixture of germacradienol (94%) and germacrene D (6%) accompanied by a trace of the octalin, but without any detectable geosmin according to the standard GC-MS analysis. The generation of germacradienol indicated that the N-terminal half of the NJ2 expressed protein had properly folded and had the expected germacradienol/germacrene D synthase activity, consistent with the demonstrated properties of the homologous N-terminal domain of *S. coelicolor* SCO6073 geosmin synthase (Cane and Watt, 2003, Gust *et al*., 2003). On the other hand, the absence of geosmin production was consistent with the apparent truncation of the C-terminal domain and the absence of a functional active site for geosmin formation. Indeed the biochemical properties of the recombinant *Nostoc* NJ2 protein were similar to those of the previously reported truncated mutant derived from the N-terminal half of *S. coelicolor* geosmin synthase, as well as variants of full-length SCO6073 protein carrying mutations in the essential C-terminal DDYYP or NSE triad regions, all of which could convert FPP to germacradienol and germacrene D but were completely defective in supporting geosmin formation (Jiang *et al*., 2007).

Comparison of the experimentally determined DNA sequence of the 3'-region of the PCR-amplified NJ2 construct with the published *npun02003620* sequence revealed that the sequence recorded in the *Nostoc punctiforme* genome database contained an extra "T" nucleotide at position 33677, at the amino acid isoleucine at position 616, inserting a stop codon after amino acid 630 of the predicted protein ZP_00109187. Excluding this additional "T" nucleotide removed the implied frame shift and eliminated the predicted stop codon. The corrected open reading frame encoded for an additional 126 amino acids, corresponding to a predicted protein length of 756 amino acids (85 kDa) (See Appendix 1-3). Most importantly, the full-length C-terminal domain of the corrected hypothetical protein sequence, dubbed NPUNMOD, now had the universally conserved DDYFP and NDI/VFSYQKE motifs that are found in the *Streptomyces* geosmin synthases SCO6073 and SAV2163 as well as all other actinomycete and myxobacterial orthologs (Table 3-4). Although the PCR-amplified DNA sequence encoding recombinant NJ2 did retain the C-terminal NSE triad just before the C-terminus, the premature truncation presumably prevented proper assembly of the active site of the C-terminal half of the suspected geosmin synthase in *Nostoc punctiforme,* and hence may explain the failure to produce geosmin when incubated with FPP.

Table 3-4. Sequence comparison *of S. coelicolor* SCO6073 geosmin synthases with bacterial orthologs (reproduced from Jiang *et al*., 2007, supplementary information Table 3).

Organism	Uniprot ID	aa	Identity (%)	Similarity (%)
Streptomyces scabies	[a]	738	78	85
Streptomyces avermitilis	Q82L49	725	76	85
Streptomyces peucetius ATCC 27952	sterp13	732	64	74
Frankia sp. strain Cc13	Q2J565	751	60	72
Saccharapolyspora erythraea NRL 2338	A4FE18	759	58	70
Frankia alni ACN14a	Q0RBQ4	758	58	70
Frankia sp. EANIpec	Q3WJX6	750	59	72
Myxococcus xanthus strain DK1622	Q1CYZ7	755	57	72
Saccharapolyspora erythraea NRL 2338	A4FJE8	763	56	69
Stigmatella auranticaca DW4/3-1	Q09A24	704	55	67
Saccharapolyspora erythraea NRL 2338	A4FGS3	45	57	57

[a] *S. scabies* chromosome , nt 2284449-2282248; Sanger Centre, http://www.sanger.ac.uk/Projects/S_scabies/.

3.5 SUMMARY

At the commencement of the work outlined in this dissertation, a molecular understanding of geosmin synthases was only starting to become realised. Initially, simultaneous work by independent research groups, concluded that SCO6073 was involved in geosmin production (Cane and Watt, 2003, Gust et al., 2003). Again, both research groups concluded that only the N-terminal half appeared to be essential for geosmin production as the C-terminal knockout generated by Gust and co-workers produced geosmin, and both the full length and N-terminal expressed proteins produced germacradienol when incubated with FPP. Combined, the results suggested that germacradienol was an essential intermediate in geosmin production, and the final conversion of germacradienol to geosmin may have required additional enzymes. Conflicting possible biochemical pathways for the conversion of germacradienol to geosmin were then postulated in different organisms, raising the possibility that perhaps the mechanism for the production of geosmin in cyanobacteria may also differ. In light of this, a search of the

GenBank database revealed that a putative gene encoding a hypothetical protein in *Nostoc punctiforme* had good similarity to the N-terminal domain, and as such was an ideal initial candidate to investigate for geosmin production in cyanobacteria.

Some three years later, however (while much of the work outlined in this chapter was conducted), the initial assumption that only the N-terminal half of SCO6073 was required for geosmin production was corrected by the demonstration that the full length enzyme had geosmin synthase ability (Jiang *et al*., 2006). Late in 2007, Jiang and co-workers demonstrated in SCO6073 that the N-terminal domain was responsible for the conversion of FPP to germacradienol, and the C-terminal half was responsible for the conversion of germacradienol to geosmin (Jiang *et al*., 2007). Recombinant ZP_00109187 protein from *Nostoc punctiforme* failed to produce geosmin, and was likely due to absence of the full length C-terminal domain. This was supported by the experimental data produced when NJ2 was incubated with FPP with germacradienol as the major product, similar to the C-terminal mutants produced by Jiang.

Of major importance was the additional "T" nucleotide in the deposited *npun02003620* sequence compared to the experimentally derived sequence from this work. This extra nucleotide created a crucial frame shift that prematurely truncated the ZP_00109187 protein. This frame shift masked the vital NSE triad in the C-terminal domain of the deposited sequence, a domain that has now been identified to have crucial enzymatic activity in bacterial geosmin synthases. When this frame shift was relieved, the resultant protein (dubbed NPUNMOD) was longer, harboured the characteristic NSE motif in the C-terminal domain, and potentially represented a fully functional geosmin synthase similar to those found in *Streptomyces coelicolor* and *Streptomyces avermitilis*. The investigation into this predicted NPUNMOD protein is the focus of Chapter 4.

Chapter 4

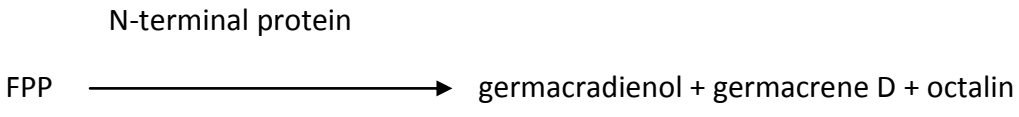
ISOLATION AND CHARACTERISATION OF NPUNMOD, A PREDICTED GEOSMIN SYNTHASE FROM *Nostoc punctiforme* ATCC 29133 (PCC 73102)

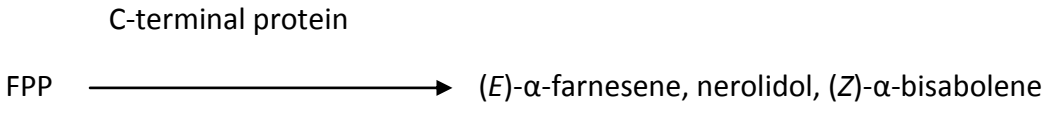
Like a million little crossroads
Through the back streets of youth
Each time we turn a new corner
A tiny moment of truth

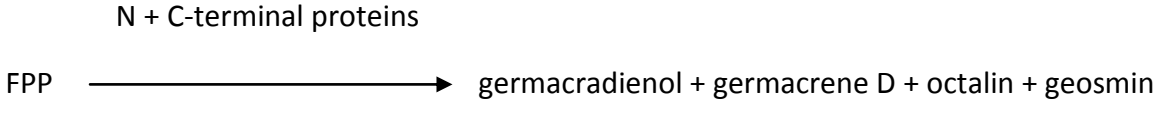
Ghost of a chance, Rush

4.1 INTRODUCTION

Recent work from Jiang *et al.* uncovered some remarkable properties of geosmin synthase from *Streptomyces coelicolor* (Jiang *et al*., 2007). Jiang and co-workers firstly produced N-terminal and C-terminal constructs of SCO6073, and expressed and purified the relative halves. When only the N-terminal half of the protein was incubated with FPP, germacradienol, germacrene D, and small quantities of octalin were produced. When the separately generated N-terminal and C-terminal protein products were combined and incubated with FPP, germacradienol, germacrene D, octalin and geosmin were formed, with a concomitant decrease in germacradienol when compared to the amounts generated whilst using N-terminal protein only. Incubation of FPP with only the C-terminal half of the protein resulted in the production of minor amounts of acyclic solvolysis products (*E*)-α-farnesene, and nerolidol, along with a trace of (*Z*)-α-bisabolene.

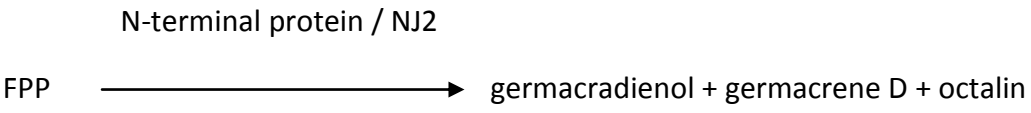






Secondly, mutants were generated to alter key catalytic sites; to the DDHFLE and NSE motifs in the N-terminal domain, and DDYYP and NSE motifs in the C-terminal domain. These mutants were incubated as above with FPP and the results demonstrated that perturbation to any of these active sites resulted in the abolishment of expected products, and ultimately geosmin formation. Together, these results clearly demonstrated the key catalytic activities of both the N-terminal and C-terminal domains (in particular the NSE motifs) of geosmin synthase in *Streptomyces coelicolor*, and as such was the first reported occurrence of a bi-functional terpene synthase (see Figure 3-1 for more detail).

The results generated by Jiang *et al.* (2007) supported the findings generated herein with the NJ2 protein reported in Chapter 3. Incubation of NJ2 with FPP produced germacradienol, germacrene D, and octalin; similarly produced by the N-terminal construct of the SCO6073 protein in *Streptomyces coelicolor*.



The discovery of an additional "T" nucleotide in the deposited *npun02003620* sequence (outlined in Chapter 3) was especially important. As Jiang and co-workers demonstrated, the NSE motif in the C-terminal domain was absolutely essential to the activity of the full length enzyme to produce geosmin (Jiang et al., 2007). When the deposited *npun02003620* sequence was corrected, the previously masked NSE motif became apparent, and the length of the predicted protein product increased from 630 aa to 754 aa. This predicted protein, dubbed NPUNMOD, was now similar to other geosmin synthases in length, key motifs, and presumed geosmin synthase activity. It was hypothesised that this, previously undiscovered, full length predicted protein (NPUNMOD) would be sufficient for the production of geosmin in cyanobacteria, and the investigations into the activity of NPUNMOD are the focus of this chapter.

4.2 AIMS

The aims of the research presented in this chapter were to

- Isolate, characterise, and express NPUNMOD from *Nostoc punctiforme,* and
- Isolate, purify, and perform functional assays using expressed NPUNMOD protein product to determine enzymatic activity.

4.3 SPECIFIC MATERIALS AND METHODS

4.3.1 Isolation of NPUNMOD

Based on the sequence data derived from Chapter 3,the NPUNMOD gene was amplified from *Nostoc punctiforme* as previously in 3.3.1.2, except that the new antisense primer was npunhalt4 (5'-TAATAACTCGAGTAATTGACCGAGTAATGAC-3'), which still carried the *XhoI* restriction site (underlined), and allowed the generation of a His_6-tag.

4.3.2 Transformation of NPUNMOD PCR product into pET-21d(+)

NPUNMOD PCR product was inserted into pET-21d(+), transformed into XL1 Blue and subsequently into BL21(DE2)pLysS as described in 3.3.2, but with one modification; both NPUNMOD and pET- 21d(+) were treated with 0.14 U of Calf Alkaline Phosphatase (Promega) prior to the initial ligation in accordance with manufacturer's instructions.

4.3.3 Screening of NPUNMOD inserts

In addition to the G-PCR1 described in 3.3.2.5, a T7 promoter screening PCR was developed and applied as described in 3.3.2.5, except the primers T7 promoter NEW (5'-TAATACGACTCACTATAGGG-3') and 288AR (5'-GCTCGATCTCATGTGCC-3') replaced 250F and 971R. This PCR was designed to confirm the presence of a T7 promoter region in transformants, in light of the unusual observations of T7 deficient transformants described in Chapter 3.

4.3.4 NPUNMOD sequencing primers

Table 4-1. NPUNMOD sequencing primers

Primer name	Sequence (5′ – 3′)
NPUNMOD1448	TGGCTGGTGCGAAAGTATTCA
NPUNMOD1513	AGCGAGGTTCCGACCAATC
NPUNMOD173	ACGACTACGCCTTGCTTTGC
NPUNMOD1766	TCCCGCAGGAGATTTACCG
NPUNMOD1988	AACTCATCGTCGCCACCG
NPUNMOD2096	GCGGCGTCCTCAAGTGG
NPUNMOD509	TCGCCAATATCAACCAGGATC
NPUNMOD696	ACGCAACGACCTCTTCTCCTA
NPUNMOD956	CGGGCGGTCACGAGTGG
NOSTOC130R	CTTCCTCTTTGAGCCAAGTATCCC
T7 promoter NEW	TAATACGACTCACTATAGGG
NPM56R	GCTGACCATACAAACCCACAAG
NPM136R	TTCCTCATAGAGCCGTCTAATGG
288AR	GCTCGATCTCATGTGCC

4.3.5 Expression of NPUNMOD protein product

The ability of the NPUNMOD construct to express protein was demonstrated as described previously in 3.3.3.1, 3.3.3.2, and 3.3.3.3.

4.3.6 Site-directed mutagenesis to remove His_6-tag from NPUNMOD

The work in this section was carried out by Jiaoyang Jiang, Brown University, Rhode Island, USA. QuikChange Site-Directed Mutagenesis Kit (Stratagene, U.S.) was used to introduce a stop codon (and hence stop the generation of His_6-tag) on the initial NPUNMOD construct as directed. The mutagenic primers were NPM2_58-notag (5′-TTACTCGGTCAA**TGA**TTACTCGAGCAC-3′) and NPM2_58-notaga (5′-GTGCTCGAGTAA**TCA**TTGACCGAGTAA-3′) (stop codon in bold, *XhoI* restriction sites underlined).

4.3.7 Co-expression of NPUNMOD with GroES/GroEL and purification of active protein

The work in this section was initially carried out by Jiaoyang Jiang, Brown University, Rhode Island, USA, and then by S. Giglio during an 8 week study visit at Brown University. The NPM2-58 protein was co-expressed along with the chaperone protein combination GroES/GroEL by co-transformation of the two plasmids, NPM2-58-pJJ32 (ampicillin-resistant) and pGro12 (kanamycin-resistant) into *E. coli* BL21(DE3)pLysS (chloramphenicol resistant), selecting a single colony displaying ampicillin-kanamycin-chloramphenicol multi-resistance. Cells from this colony were used to inoculate LB broth containing 100 µg/mL of ampicillin, 50 µg/mL of kanamycin, and 34 µg/mL of chloramphenicol (LBAKC), and the overnight culture was transferred to 500 mL of LBAKC broth and incubated at 37 °C and 250 rpm for 2 hours. Chaperone expression was induced by addition of arabinose to a final concentration of 4 mg/mL, and the desired protein expression was then induced by the addition of IPTG to a final concentration of 0.4 mM. Incubation was continued for an additional 18 hours at 18 °C, 250 rpm. The cells were harvested by centrifugation and resuspended in 30 mL of lysis buffer (20 mM Tris-HCl, 10% glycerol, 0.1 mM DTT, pH 8.0). The cells were disrupted by sonication and the cell lysate was clarified by centrifugation. The supernatant was purified by 12% ammonium sulphate precipitation followed by purification on a 25 mL *n*-butyl-Sepharose column that had been pre-equilibrated in buffer A (0.5 M ammonium sulphate, 50 mM NaH_2PO_4, 0.1 mM DTT, pH 7.0). After loading of the supernatant onto the column, the resin was washed with 60 mL of buffer A followed by a 180 mL linear gradient from buffer A to buffer B (50 mM NaH_2PO_4, 0.1 mM DTT, pH 7.0). The purified protein eluted in buffer B as a 1:1 mixture with chaperone protein, as determined by SDS-PAGE.

4.3.8 Incubation of NPUNMOD with FPP

Incubation of NPUNMOD with FPP was performed as described previously in 3.3.3.4.

4.4 RESULTS AND DISCUSSION

4.4.1 Expression of NPUNMOD

Re-ligation and transformation irregularities seen in Chapter 3 were not as marked in the NPUNMOD investigation, perhaps due to the use of alkaline phosphatase. The use of the T7 promoter screening PCR combined with G-PCR1 significantly reduced the number of potential expression candidates that needed to be screened, and avoided the inclusion of T7 deficient clones. Several XL1-Blue transformed candidates that were positive for G-PCR1, T7 promoter

PCR, and the correct digest pattern, were selected for further propagation into BL21(DE3)pLysS for final protein expression (NPM2 isolates). It was anticipated that isolates positive for all the above tests would have the NPUNMOD sequence inserted in the correct orientation, with an active T7 promoter region and downstream ribosomal binding site, because it was apparent in the initial investigations of *npun02003620* that the T7 promoter and ribosomal binding sites were absent in some constructs, which directly related to the isolates inability to express the desired protein.

After successfully screening isolates using G-PCR1 PCR, T7 promoter PCR, and restriction enzyme digestion patterns, isolates NPM2_58, NPM2_71, and NPM2_109 were chosen and investigated further for protein production as described previously in 3.3.3.1, 3.3.3.2, and 3.3.3.3. Figure 4-2 demonstrates the excellent expression of hypothetical NPUNMOD protein for isolates NPM2_58, NPM2_71, and NPM2_109 with an approximate size of 85 KDa as expected. Comparison with lane 5, NJ2 protein from Chapter 3, clearly demonstrates the difference in protein size as expected. As with NJ2, the protein was predominantly expressed within inclusion bodies of the host BL21(DE3)pLysS cells.

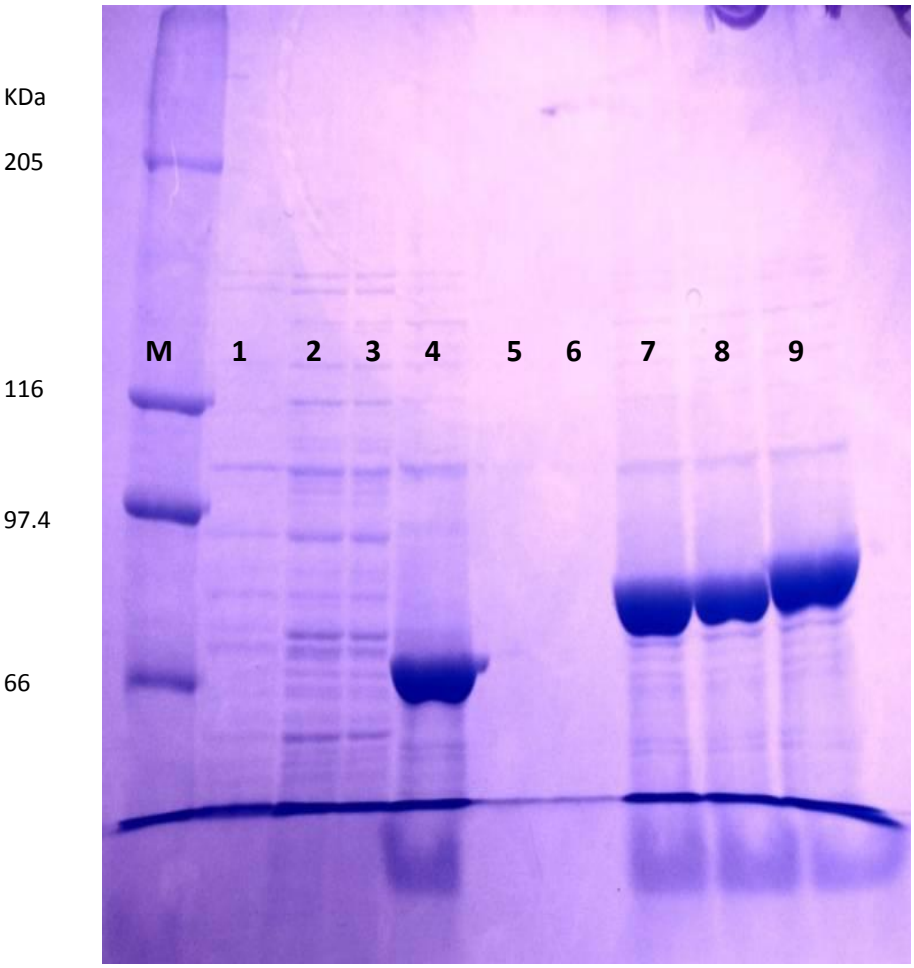


Figure 4-1. SDS-PAGE of hypothetical NPUNMOD protein. 1, Molecular weight marker (KDa); 1, pooled non-induced NPM2 fractions; 2, NJ2 induced soluble fraction; 3, pooled NPM2 induced soluble fractions; 4, NJ2 induced inclusion body fraction; 5-6 T7 deficient isolate - induced soluble fractions; 7, NPM2_58 induced inclusion body fraction; 8, NPM2_71 induced inclusion body fraction; 9, NPM2_109 induced inclusion body fraction.

4.4.2 Sequencing of NPM2_58

The hypothetical NPUNMOD and experimental NPM2_58 had excellent sequence similarity, having only one mismatch at the amino acid level (see Appendices 1-4 and 1-5). The NPM2_58 sequence was deposited on GenBank as a geosmin synthase (after confirmation studies in 4.4.3), and annotated as NPUNMOD (GenBank accession number FJ010203).

4.4.3 Purification of NPUNMOD and incubation with FPP

From this section onwards, experimental NPM2_58 expressed protein is referred to as NPUNMOD in subsequent experimental dealings.

As shown in Figure 4-1, NPUNMOD was highly expressed within inclusion bodies. Despite several refolding conditions, active protein could not be purified as was evident by the formation of precipitous aggregates in several refolding buffers trialled (data not shown). The terminal His_6-tag was initially suspected to be hindering re-folding of packed protein into its native active state, as has been reported previously in the Cane laboratory (personal communication, Jiaoyang Jiang). Therefore, site directed mutagenesis was employed to remove the His_6-tag. However, the resultant protein (void of His_6-tag) could still not be refolded into an active protein. The formation of precipitate in both of these preparations suggested that neither of these proteins were correctly folded, and therefore did not warrant incubation of these probably biologically inactive proteins with FPP.

In order to generate soluble and active protein, the plasmid used for producing NPUNMOD protein (NPM2_58_pJJ32) was co-expressed in BL21(DE3)pLysS with the chaperonin protein combination of GroES/GroEL as described in 4.3.7. Chaperonin proteins aided in the proper folding of the NPUNMOD protein to an active soluble form by multiple rounds of ATP mediated binding and release of target protein to chaperone protein (Horwich *et al*., 2007, Young *et al*., 2004) (Figure 4-2), rather than packaged as inactive protein within inclusion bodies. Soluble active protein was isolated as described in 4.3.7, and subsequently incubated with FPP as previously detailed in 3.3.3.4.

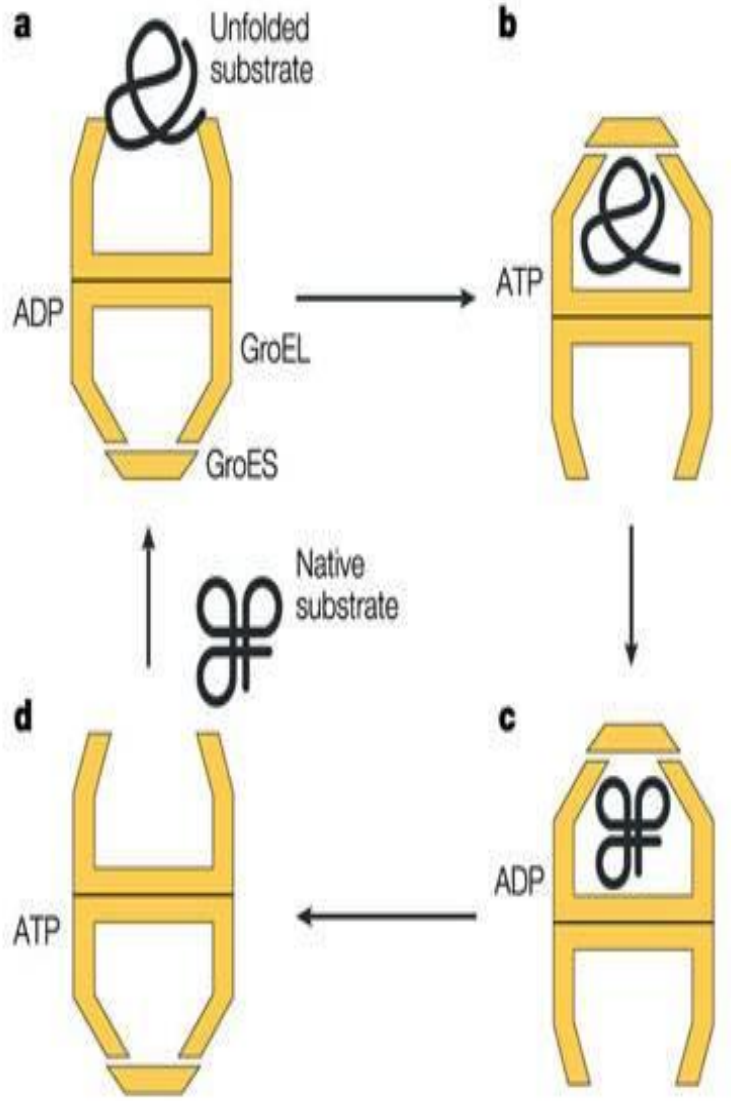


Figure 4-2. Chaperone mediated folding of inactive unfolded enzyme (a) into a native and biologically active form (d). Reproduced from Young *et al.*, 2004.

Recombinant full-length NPUNMOD protein was shown to be a fully functional geosmin synthase by incubation with FPP under standard conditions. GC-MS analysis of the hexane-soluble extract confirmed the formation of a typical product mixture consisting of geosmin (4%), germacradienol (78%), germacrene D (8%), octalin (1%), and (*E*)-nerolidol (9%) (Figure 4-3). The formation of nerolidol is unusual and may have resulted from some degree of improper folding of the recombinant geosmin synthase or by chaperone-mediated solvolysis of FPP.

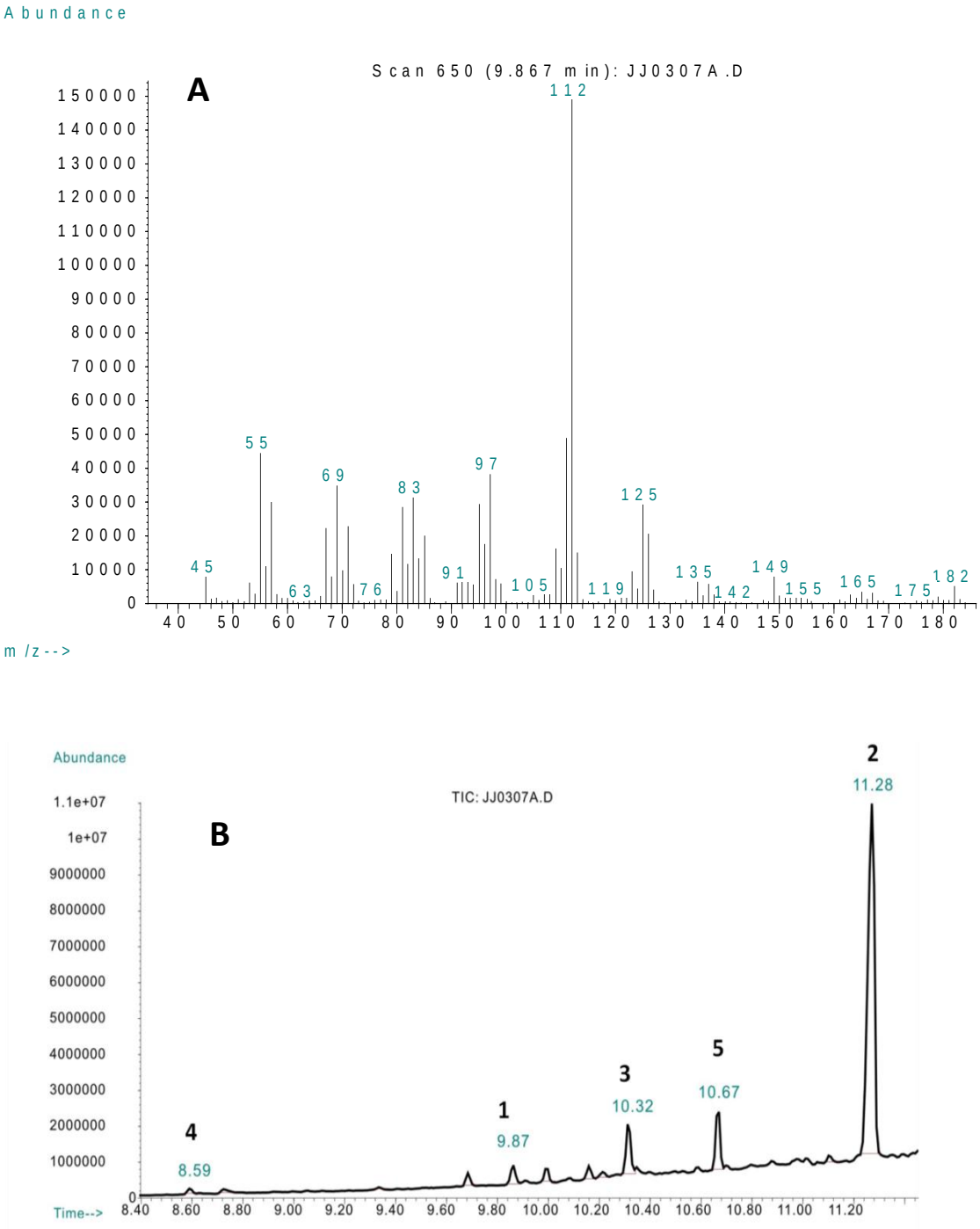


Figure 4-3. Mass spectrum (A) of geosmin produced in the incubation of FPP with recombinant NPUNMOD protein, the geosmin synthase of *Nostoc punctiforme*, and GC-MS trace (B) showing geosmin (1), germacradienol (2), germacrene D (3), octalin (4), and (E)-nerolidol (5).

Studies performed by Agger *et al.*, also discovered the *npun02003620* sequencing error and frame shift to reveal a theoretical geosmin synthase (Agger *et al*., 2008). The authors were also unsuccessful in generating a fully active protein from the inclusion bodies and were only able to demonstrate the formation of germacradienol and germacrene D, presumably due to refolding difficulties as experienced in this work, and were not able to demonstrate geosmin formation. The proper folding of the C-terminal half, and hence geosmin synthase activity, appeared to require extremely specific conditions *in vitro* as described here.

4.5 SUMMARY

Terpenes are naturally occurring products that have reported anti-cancer, anti-inflammatory, and anti-microbial activities (Ajikumar *et al*., 2008). The discovery of geosmin synthase in *S. coelicolor* by the Cane and Gust groups solved a long standing biosynthetic mystery. While these groups were investigating naturally occurring terpenes (and terpene and sesquiterpene synthases), one of several candidate terpene synthases was SCO6073, and the subsequent finding that this enzyme was responsible for production of geosmin was a fortuitous one for the water industry.

At the time of reporting their major discoveries from 2006-2007 (Jiang *et al*., 2007, Jiang *et al*., 2006), the only sequenced geosmin producing cyanobacteria (partially annotated) was *Nostoc punctiforme*, and this contained a putative gene with reasonable similarity to SCO6073, but only within the N-terminal half. Through Chapters 3 and 4, it was shown that the published ZP_00109187 protein was incorrect, and that this enzyme was incapable of supporting geosmin formation, as demonstrated by experimental NJ2. Once the sequence was corrected, NPUNMOD was revealed and the experimental evidence presented in this chapter demonstrated that this protein was a fully functional geosmin synthase.

The road to discovery of geosmin synthases in both *S. coelicolor* and now *Nostoc punctiforme* was a relatively rapid though convoluted one. Despite initial confusion surrounding the apparently catalytically silent C-terminal half by virtue of experimental inconsistencies in initial works (Cane and Watt, 2003, Gust *et al*., 2003), and several proposed mechanisms for the conversion of germacradienol to geosmin in various organisms (Cane and Watt, 2003, Dickschat *et al*., 2005, Schulz *et al*., 2004), it is now known that the uniquely bi-functional geosmin

synthase present in *Streptomyces* sp. is also present in *Nostoc punctiforme*. It is also highly likely, given the high level of conservation seen in prokaryotic organisms (see Table 3-4), that the presence of this unique enzyme will also be present in higher life forms such as plants and fungi that also produce geosmin.

The demonstration of the same biosynthetic pathway for geosmin in *Streptomyces* sp and now cyanobacteria (and most likely myxobacteria) dispels earlier conundrums regarding possible biosynthesis via different mechanisms in different organisms. For cyanobacteria, this finding represents an opportunity to examine the expression of the geosmin synthase gene under defined conditions (explored in Chapter 5). These findings will also facilitate the development of new and rapid techniques to identify geosmin producing cyanobacteria, and add to the taste and odour "toolbox" for use in monitoring and investigations of taste and odour events (explored in Chapter 6).

Chapter 5

EXPRESSION OF GEOSMIN SYNTHASE IN *Anabaena circinalis*

We sometimes catch a window
A glimpse of what's beyond
Was it just imagination
Stringing us along?
More things than are dreamed about
Unseen and unexplained
We suspend our disbelief
And we are entertained

Mystic Rhythms, Rush

5.1 INTRODUCTION

5.1.1 Geosmin synthesis and photosynthesis

At this point in time, the majority of evidence suggests that the MEP isoprenoid pathway is utilised for the production of phytols and carotenoids within the thylakoid membranes of cyanobacteria; the final product of this pathway being IPP and its isomer DMAPP, which are subsequently channelled into a carbon pool directed towards the production of other isoprenoid precursors (See General Introduction for greater detail). One of the intermediate products from the isoprenoid pathway is geranylgeranyl diphosphate (GGPP), which is a precursor for the synthesis of several molecules. One such molecule is phytyl, which is eventually incorporated into the chlorophyll molecule (Qin *et al*., 2007) (Figure 5-1, modified from Qin *et al*., 2007).

Several reports have investigated the production of geosmin under various light and temperature conditions, and have attempted to link geosmin synthesis with these experimental conditions. Naes *et al.* (1985) examined the effect of various light fluencies and subsequent geosmin production in *Oscillatoria brevis* (Kutz) and showed a direct correlation between chlorophyll *a* (chl *a*) and the production of geosmin (Naes *et al*., 1985), i.e. if there was a decrease in chl *a*, there was a decrease in geosmin dry weight under experimental conditions at a constant temperature of 20 °C. These initial findings were later supported by similar studies using *Anabaena circinalis* (Bowmer *et al*., 1992) and benthic and planktonic *Oscillatoria* (Utkilen and Froshaug, 1992), which correlated the production of geosmin with chl *a* and/or carotenoids. All authors hypothesised that the concomitant decrease of geosmin and chl *a*/carotenoids probably linked geosmin to reactions within an isoprenoid pathway.

Investigations using isoprenoid synthesis inhibitors (fluridone and acifluorofen) on a MIB-producing *Pseudanabaena articulata* demonstrated that these inhibitors did not direct any additional carbons from the isoprenoid carbon pool directly into MIB synthesis, thus suggesting that MIB is not an overflow product of this pathway, but more likely that any increase in MIB is the result of overall increased isoprenoid biosynthetic activity (Zimba *et al*., 1999). A similar experiment using geosmin-producing *Oscillatoria brevis*, however, suggested that isoprenoid precursors were directed to geosmin synthesis when norflurazon (an isoprenoid synthesis inhibitor) was present (Naes *et al*., 1989). The effect of this post-phytoene inhibitor for geosmin production in this report differed from the reported effect of similar inhibitors on MIB

production as outlined by Zimba *et al.* (1999), and it is uncertain why this may be so, although direct comparison on the physiological effects of these inhibitors in two different organisms would be speculative in the absence of supporting and reproducible data.

The preferential channelling of carbons into the isoprenoid pool for geosmin synthesis was also supported by Saadoun *et al.* (Saadoun *et al*., 2001), who demonstrated an inversely proportional relationship between chl *a* and geosmin, i.e. a decrease in chl *a* with a corresponding increase in geosmin concentration in *Anabaena circinalis*. However, this was the converse of some of the other experimental data outlined earlier in this section. Temperature also had an apparent effect on geosmin production in *Lyngbya kuetzingii* by possibly inhibiting chl *a* synthesis, where optimum geosmin concentration in this organism correlated with a low chl *a* at 10 °C (Zhang *et al*., 2009). Blevins and co-workers also demonstrated that there may be channelling of isoprenoid precursors in *Anabaena* sp, but this was highly dependent on temperature and light, with maximal geosmin production at 20 °C and 17 µmol m^{-2} s^{-1} light (Blevins *et al*., 1995). Blevins *et al.* also demonstrated that higher light decreased chl *a* demand with greater geosmin synthesis, while increasing temperature stimulated chl *a* production at the expense of geosmin production (in contrast to the *Lyngbya* reported by Zhang *et al.* (2009) that demonstrated maximal geosmin production at lower temperature). All of the aforementioned reports used different organisms and the incubation conditions lacked consistency, and a direct relationship between chl *a* and geosmin synthesis cannot be reliably established from these data alone.

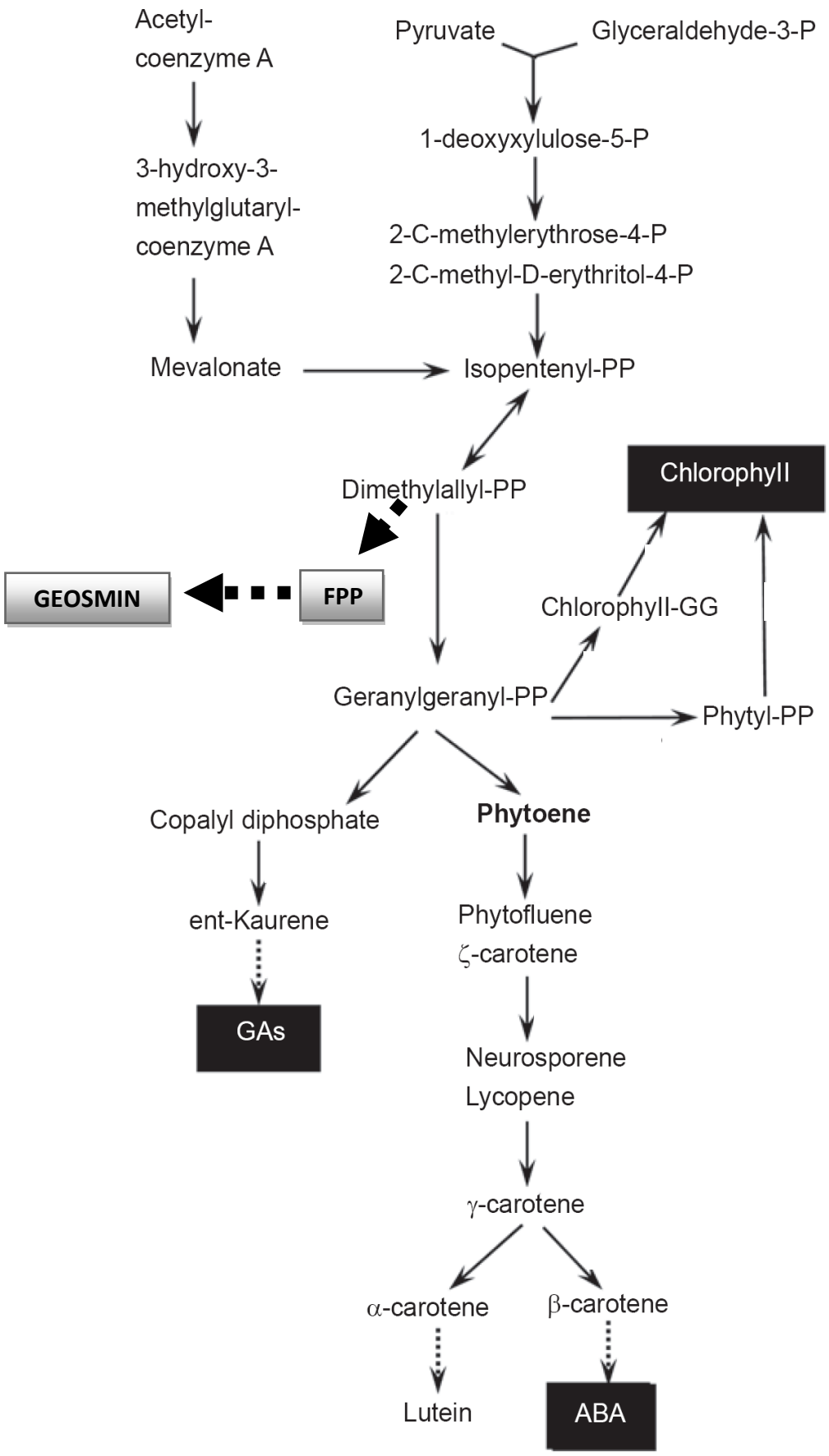


Figure 5-1. Relationship of the isoprenoid pathways and geosmin and chlorophyll formation (modified from Qin *et al*., 2007)

The reports detailing cyanobacterial experiments so far cited in this chapter were inconsistent in experimental approach, had variable results, and lacked a common model organism. It was apparent, however, that there seemed to be an effect of light and/or temperature on the production of secondary metabolites such as geosmin and MIB, but to what extent seemed dependant on the organism studied. Given that each genus of cyanobacteria occupies an environmental niche, each with an optimal temperature and photon fluence range, it would be surprising that the control of secondary metabolite production is solely regulated by only one or both of these factors, and that it would also be consistent between genera. The examination of a benthic *Lyngbya* by Zhang *et al.* (2009) demonstrated that optimal geosmin production was obtained at low temperature and low light, probably representing optimal conditions for the growth and survival of this organism. In comparison, most of the planktonic organisms, such as the *Anabaena* described, favoured moderately higher temperatures with variable responses to light, again probably representing conditions closer to their ideal growth conditions, because these organisms normally occupy the warmer epilimnion and can regulate buoyancy in response to photosynthetic conditions when required.

Indeed, the control and regulation of photosynthetic genes is complex. Hihara *et al.* examined the response of 3,079 genes in *Synechocystis* when cultures were exposed to a marked increase of photon fluence rate (20 µmol to 300 µmol photons m^{-2} sec^{-1}), and determined that over 160 genes were responsive, including both growth-related regulatory components and photosynthetic components (Hihara *et al*., 2003). In addition, chlorophyll and phycocyanin content within *Synechocystis* were drastically reduced within 3 hours, presumably to avoid absorption of excess light energy and accumulation of photo-reactive by-products. Interestingly, in *Synechocystis*, photosynthetic mRNA transcripts were detected in the dark (*psbA* and *psnB* – genes encoding core D1 and D2 polypeptides of photosystem II), indicating that either transcription occurs in the dark, or that the transcripts were relatively stable for up to 48 hours (Smart and McIntosh, 1991). A similar half life for the *psnB* transcript has also been reported in barley chloroplasts (Kim *et al*., 1993a, Kim *et al*., 1993b).

Much work is continuing in examining the regulatory mechanisms of cyanobacteria using *Synechocystis* as a model organism. The fully annotated genome sequence of this organism is available, making genome-wide investigations into responsive and adaptive mechanisms possible, and allowing discernable information to increase our knowledge in this field. Given

the association with photosynthesis and isoprenoid (chlorophyll) synthesis, it is likely that photosynthetic regulation and isoprenoid regulation are intertwined at some level. It is also likely that, as with photosynthesis, regulation of geosmin synthesis is equally as complex.

5.1.2 The comparative C_T method ($2^{-\Delta\Delta CT}$) for the examination of mRNA synthesis

It is only within the last few years, and through the course of the work detailed in this dissertation, that the gene responsible for geosmin synthesis in various organisms has become characterised. Whilst there is little practical interest in the production of geosmin in *Streptomyces* or *Myxobacteria*, there is considerable interest in the expression of geosmin synthase in cyanobacteria because of the complications geosmin causes for water management. It is anticipated that insight into the expression of geosmin synthase may provide valuable and fundamental insights into what conditions are required for geosmin synthesis, with an ultimate view to perhaps one day providing an impetus to develop alternative control strategies if possible.

The advent of real-time PCR has provided a powerful tool to examine the expression of genes in detail. Compared to Northern blotting, which examines mRNA expression on a semi-quantitative level, reverse transcription coupled with real-time PCR is a powerful method to accurately quantify gene expression (Giulietti *et al*., 2001). Quantification of mRNA transcripts can be absolute, i.e. the transcripts are expressed as a whole number by comparison to a standard curve, or mRNA transcripts can be quantified relative to a housekeeping gene that has a constant level of expression under experimental conditions, i.e. the relative or comparative expression (or $2^{-\Delta\Delta CT}$) method. Compared to absolute quantification, relative expression data describes “fold” increases or decreases in the expression of transcripts, and can provide data that is more easily interpreted within the context of experimental design when compared to using absolute numbers.

Relative expression analysis of mRNA transcripts using the comparative expression method requires some fundamental assumptions to be true. Firstly, the expression of the housekeeping gene should be constant under the experimental conditions. One such housekeeping gene (HKG) previously used in *Anabaena* sp PCC7120 is the 16S rRNA gene which is constitutively expressed under various light conditions (Hong *et al*., 2007). The second assumption is that the PCR amplification efficiency of the gene of interest (GOI) and the housekeeping gene are almost

identical, as any differences will bias the differential between the two and subsequent interpretation of the results. The latter assumption needs to be validated for each primer pair used in the PCR by preparing a serial dilution over a 100-fold range and examining the difference in C_T (threshold cycle, the point at which amplifications becomes exponential) between the target gene of interest and reference keeping gene (Livak and Schmittgen, 2001). The ΔC_T at any given time is defined as:

$$\Delta C_T = C_T \text{ GOI} - C_T \text{ HKG}$$

The $\Delta\Delta C_T$, compared to any given time point '*x*' is

$$(C_T \text{ GOI} - C_T \text{ HKG})_{\text{Time } 'x'} - (C_T \text{ GOI} - C_T \text{ HKG})_{\text{Time } 0}$$

where time 0 represents 1 X expression of the GOI normalised to the housekeeping gene (essentially a basal level for experimental conditions).

For a valid combination of primers to be used in relative expression analysis, a plot of ΔC_T versus DNA concentration is necessary. If the slope of the line is zero, i.e. there is no difference in the ΔC_T over a 100-fold concentration range, this implies that the primer combination is highly efficient and suitable for use in further analyses. A 5% tolerance of primer inefficiency is allowed, i.e. the slope of the line of best fit can be +/- 0.05x. The relative amount of target transcript at a given point, relative to time 0, is calculated by

$$2^{-\Delta\Delta CT}.$$

For more detailed information and derivation of the formulas used in this dissertation, readers are directed to an excellent review for the full derivation of the $2^{-\Delta\Delta CT}$ equation and examples for the validation of primer set efficiency (Livak and Schmittgen, 2001).

5.2 AIMS

The aims of the research presented in this chapter were to:

- Develop a comparative C_T method for use in monitoring the expression of geosmin synthase in *Anabaena circinalis*,
- Apply the comparative C_T method to examine the expression of geosmin synthase under continuous illumination with 90 µmol m^{-2} s^{-1} light, and
- Apply the comparative C_T method to examine the expression of geosmin synthase in response to the absence of light.

5.3 SPECIFIC MATERIALS AND METHODS

5.3.1 Test organism and cell counting

Anabaena circinalis AWQC318 (ANA318) was used to investigate geosmin synthase gene expression. This strain was isolated by the AWQC in 1995 from Pejar Dam in Gouldburn, NSW, Australia, during a taste and odour episode. It is an unusual *Anabaena* in that it produces short chains of 2-8 cells in an un-agitated culture, and an average of 2-4 cells in an agitated culture, with many single cells also present. Cells were counted on a Beckman Coulter Cell Lab Quanta SC flow cytometer on FL3 (red fluorescent channel), monitoring cyanobacterial red auto-fluorescence.

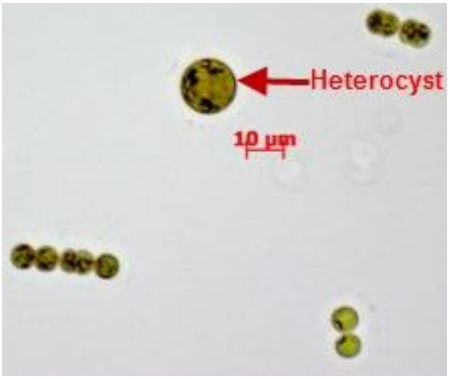


Figure5-2. Microscopic image of ANA318 (x400 DIC). Photo courtesy of Zong Lai.

5.3.2 Culture and sampling conditions

ANA318 was cultured in ASM-1 medium. 200 mL cultures in early log phase were photo-acclimatised in 90 µmol m^{-2} s^{-1} light, 25 °C, for 7 days before being transferred to a 5 L flask containing 2 L of ASM-1 in order to achieve an approximate count of 10,000 cells per mL. Flasks

were inoculated with photo-acclimatised ANA318 and incubated for 16 days at 25 °C (shaking at 100 rpm) at 90 µmol m^{-2} s^{-1} light (vessel 1). This experiment was conducted in duplicate, several months apart, experiment 1 serving as a pilot experiment for overall system integrity and method verification.

To investigate any rapid changes of geosmin synthase expression in response to removing the stimulus of light, induction of a dark state in a separate vessel (vessel 2, run in parallel with the second experiment of vessel 1) was achieved by wrapping the vessel in aluminium foil on day 9, at which point samples were taken every 2 hours for 6 hours to measure any rapid change in geosmin synthase expression. Vessel 2 was then re-incubated as described above for a further 7 days still wrapped in foil.

Samples (50 mL) were collected from all vessels for cell count analysis, RNA extraction, reverse transcription, PCRs, and geosmin analyses. Samples were collected at Days 0 (time of inoculation), 2, 4, 7, 9, 11, 14, and 16.

5.3.3 Geosmin analyses

Geosmin analyses were performed by standard techniques as described in 2.2.8. Total geosmin was analysed from 10 mL of whole culture. Additionally, 10 mL of culture was filtered through glass-fibre filter paper (Whatman WHA1822-150) and the filtrate analysed for extracellular geosmin. 10 mg/mL of $HgCl_2$ was added to collection containers to stop any bacterial degradation of geosmin. Intracellular geosmin was calculated by subtracting the extracellular determination from the total geosmin determination. Single geosmin analyses were performed at each time point, on each sample.

5.3.4 RNA extraction and quantification

For each vessel and at each sampling point, 10 mL of culture underwent centrifugation at 4,000 g for 20 min and RNA was subsequently extracted using Ambion Ribo-Pure Bacteria RNA kit (AMBION AM7020) as per the manufacturer's instructions (including DNA-se treatment). In order to standardise the amount of RNA added to reverse transcription reactions, RNA was quantified using an RNA Pico Kit on the Agilent Bioanalyser (catalogue # 5067-1513). A single RNA extract was performed on each sample at each sampling time.

5.3.5 PCR primers

Table 5-1. PCR primers used for relative expression determinations.

PCR Primer	Sequence (5' – 3')	Reference
173AF	ATGTGAGTACCCAAGAGG	Developed this study
173AR	CTGCCAATCCTGAAGTCCTTT	Developed this study
PF1	TGTAAAACGACGGCCAGTCCAGACTCCTACG	(Shaw *et al*., 1999)
PR1	CGCGTTAGCTACGGCACGGCTCG	(Shaw *et al*., 1999)

5.3.6 Reverse transcription

Reverse transcription was performed using the reverse transcription enzyme SuperScript III (Invitrogen) as described in the manufacturer's instructions, including the incorporation of RNAse OUT (Invitrogen). For the first strand synthesis, primers 173AF and PF1 for geosmin synthase and 16S rRNA respectively were used in separate reactions. Approximately 1 ng of extracted RNA was included in each reaction. Reverse transcription reactions were conducted in triplicate using the RNA extracts obtained as described in 5.3.4.

5.3.7 PCR conditions

5.3.7.1 Primer validations

A requirement for using primers in the comparative C_T method is that the efficiency of the PCR reaction is similar for the GOI (in this case geosmin synthase) and the housekeeping gene. Several candidate PCRs were developed for geosmin synthase, and established PCRs for 16S rRNA and *rpoC* genes were investigated as target housekeeping genes. Several primer combinations and PCR conditions were trialled, but only optimised conditions are described here.

DNA from ANA318 was extracted from a logarithmic culture, quantified, and subsequently serially diluted 1:3 to cover a 100-fold dilution range. 2 µl of each DNA dilution was added to a PCR mixture containing 2.5 mM $MgCl_2$, 0.3 µM each of 173A or 16S forward and reverse primers (separate reactions for 16S and geosmin synthase PCRs), 200 µM dNTPs, 1 U *Taq* DNA polymerase, 1 X reaction buffer, and 2.5 µM SYTO9. Real-time PCR was performed using a Corbett Research Rotorgene 6000HRM (Corbett Research, Australia) under the same conditions

as G-PCR1 (see 3.3.2.5), except that the final extension step was replaced with a DNA melting analysis from 75 to 95 °C, with data being acquired every degree with a 5 s hold at each step. All data were acquired on the "Green" channel, with excitation at 470 nm and emission at 510 nm. All PCR reactions were performed in triplicate.

5.3.7.2 Reverse Transcriptase - PCR

cDNA products of the first-strand reverse transcription (from 5.3.6) were analysed using the PCR protocol described in 5.3.7.1, except that 2 µl of reverse transcription product was incorporated into the reaction instead of diluted DNA.

5.3.7.3 Summary of sample testing regime for geosmin synthase gene expression studies described in this Chapter

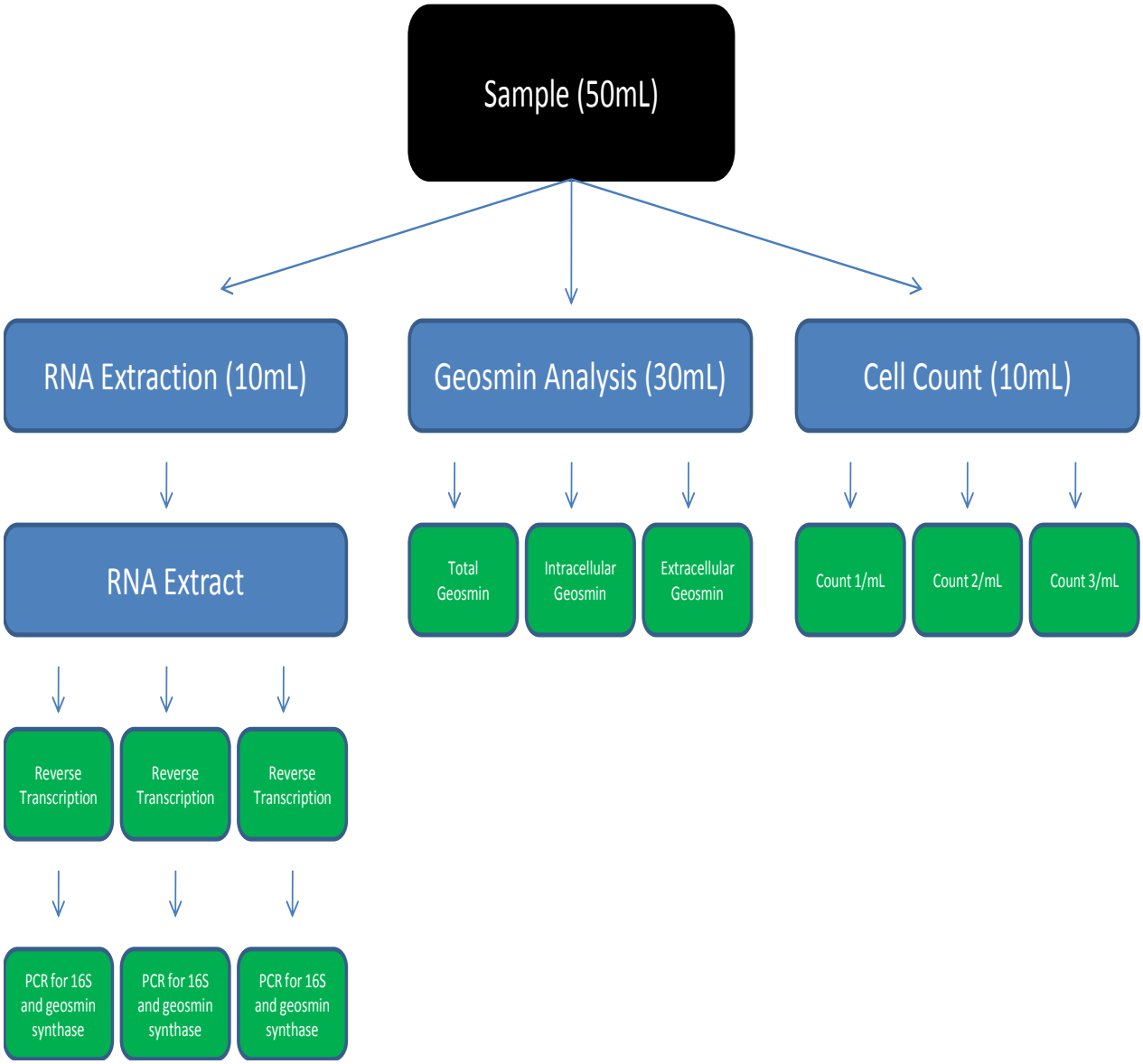


Figure 5-3. Summary of sample testing regime.

5.4 RESULTS AND DISCUSSION

5.4.1 Verification of 173-PCR for geosmin synthase

A primer set was developed to conserved regions within the N-terminal domain of geosmin synthase to amplify a 173 bp fragment. Analysis of the PCR reaction using the Agilent Bioanalyser confirmed the presence of the expected PCR product. Only a single PCR product, from the ANA318 geosmin synthase, was detected on the virtual gel electropherogram (Figure 5-4). This assay, therefore, was suitable for use in determining the expression of geosmin

synthase relative to 16S rRNA under the experimental conditions described. However, sufficient validation on additional strains would be required before it could be used as a potential screening tool for geosmin synthase in other organisms, or for any diagnostic use.

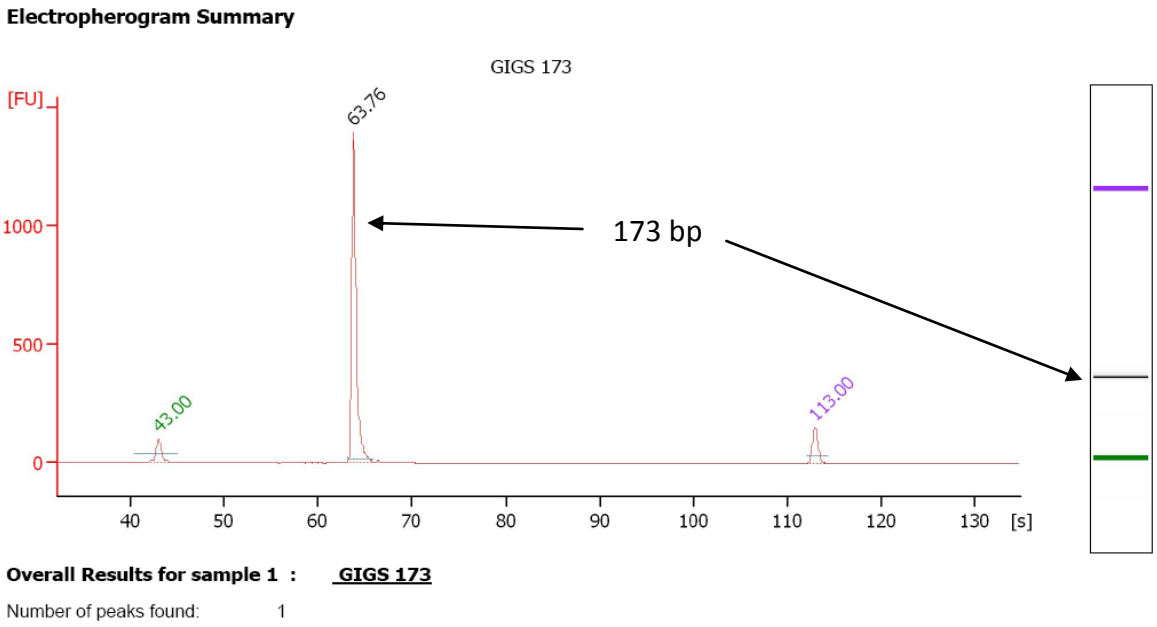


Figure 5-4. Electropherogram summary of PCR product on Bioanalyser software. Peaks at 43 and 113s are marker peaks, while the peak at 63.76s is the specific 173 PCR amplicon (arrowed, virtual gel on right hand side)

5.4.2 Fulfilling the requirements of PCR primer efficiency for use in comparative C_T analysis of geosmin synthase and 16s RNA

Approximately 40 ng/µl of ANA318 DNA was purified and serially diluted 1:3 to approximately 0.49 ng/µl, covering an approximate 100-fold concentration range. Triplicate PCRs for each dilution were performed as described in 5.3.7.1, using 2 µl of diluted stock DNA per reaction, equating to a range of 80 – 0.98 ng per PCR reaction, and the data is presented in Table 5-2.

Table 5-2. Primer validation data for 173A and 16S PCRs.

DNA conc (ng/PCR)	173A C_T	Average 173A C_T	16S C_T	Average 16S C_T	ΔC_T[1]
80.00	19.55	19.57	18.08	18.09	1.48
80.00	19.59		18.08		
80.00	19.56		18.12		
26.66	21.24	21.42	20.37	20.21	1.21
26.66	21.89		20.12		
26.66	21.13		20.13		
8.88	23.10	23.11	22.27	22.10	1.01
8.88	23.07		22.01		
8.88	23.17		22.02		
2.96	25.16	25.12	23.68	23.95	1.17
2.96	25.38		23.99		
2.96	24.81		24.19		
0.98	26.60	26.67	24.98	25.37	1.30
0.98	26.84		25.94		
0.98	26.57		25.19		

[1] $\Delta C_T = C_T173 - C_T16S$ for each given dilution

Plotting ΔC_T versus the concentration of DNA results in the line of best fit: y=-0.0397x+1.3517 (Fig. 5-5). The slope of -0.0397 is within the 5% range of primer efficiency tolerance allowed, and thus fulfils the selection criteria required for use in relative expression analysis (Livak and Schmittgen, 2001). Optimisation of PCRs for use in the comparative C_T method for gene expression is absolutely essential; several annealing temperature and combinations of primer sets were trialled (including *rpoC* as a house keeping gene and a suite of specifically developed geosmin synthase PCRs, data not shown), but no other PCR combinations, other than the one described herein, fulfilled the criteria for amplification efficiency correlation. As an example, even the slightest change in stringency of PCR conditions from a $MgCl_2$ concentration of 2.5 mM to 2.0 mM resulted in a highly incompatible primer combination (Figure 5-6, slope -0.24 or 24% difference in primer efficiency).

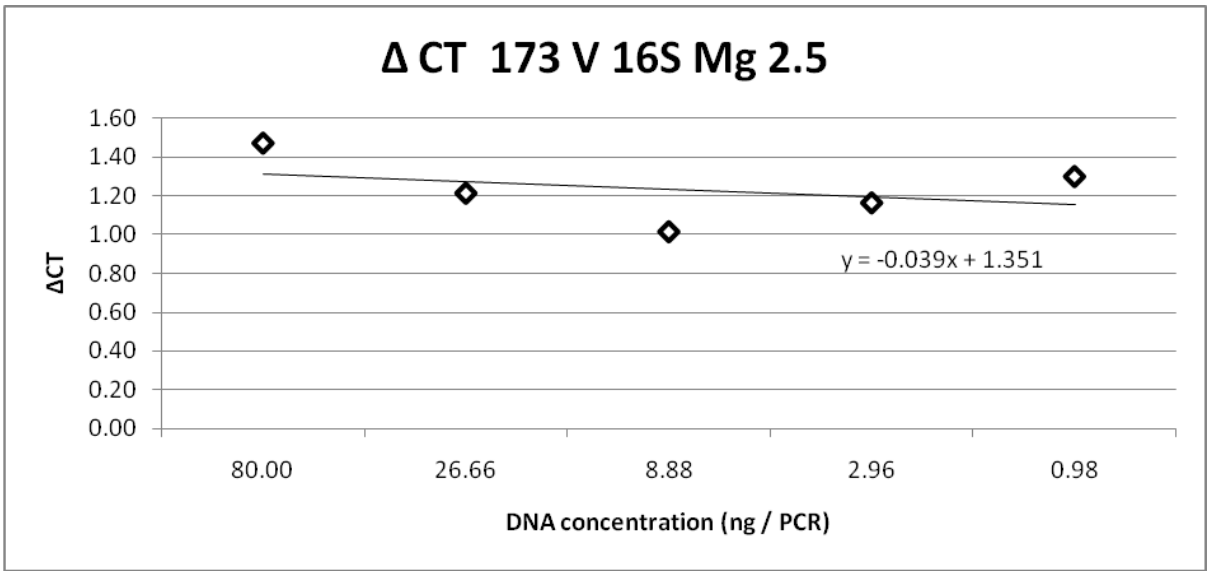


Figure 5-5. Analysis and suitability of primer pairs for relative expression experiments - ΔCT of 173 V 16S PCRs; $MgCl_2$, 2.5 mM

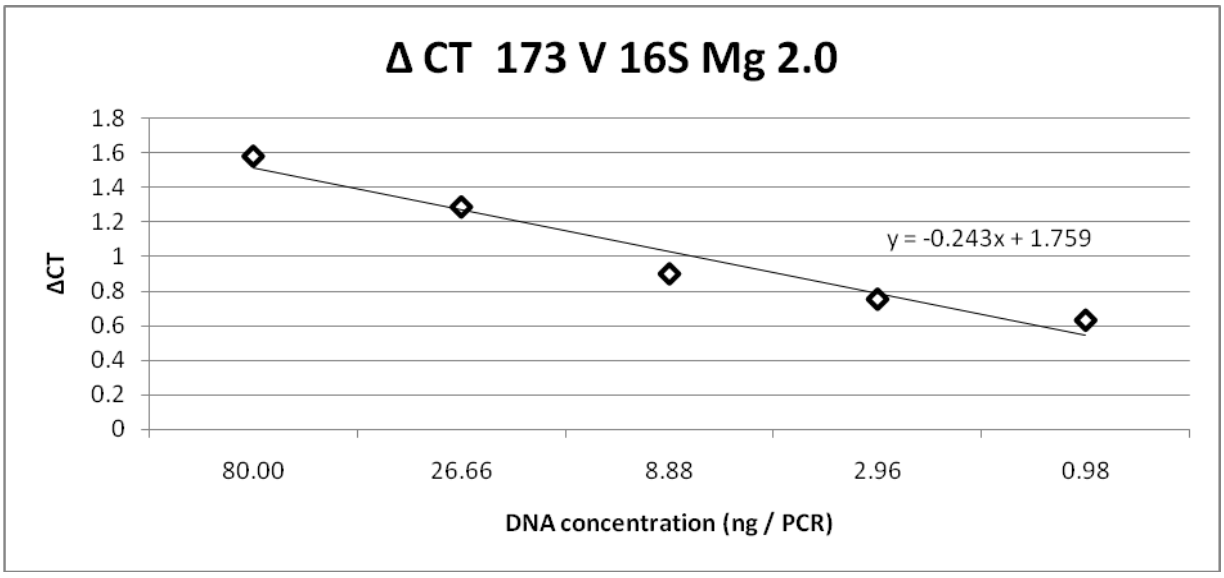


Figure 5-6. Analysis and suitability of primer pairs for relative expression experiments - ΔCT of 173 V 16S PCRs; $MgCl_2$, 2.0 mM

5.4.3 Growth of ANA318 and measurement of intracellular geosmin – Vessel 1, continuous light

The growth of ANA318, expressed as cells per mL over time, is presented in Figure 5-7. The first 4-7 days revealed a decrease in intracellular geosmin (Table 5-3) with a modest increase in cell numbers, followed by a steady increase of intracellular geosmin up until days 11-14, with a

decline at day 16 for both replicate experiments (Figure 5-8) The initial decline in the *calculated* intracellular concentration of geosmin to day 7 can be explained as follows; firstly, the inoculation of geosmin-laden cells from the acclimatised cultures would have had high intracellular geosmin concentrations, and the lysis of some cells upon initial inoculation would have left a population of cells purely with high levels of intracellular geosmin, with some extracellular geosmin in the medium (as indicated by the low intracellular : extracellular geosmin ratio at day 0 (Figure 5-9)). Secondly, over the following 7 days, the calculated intracellular geosmin was nominally diluted amongst increasing cell numbers (newly dividing cells) that would have not had as large a geosmin levels as the originally inoculated cells, therefore decreasing the relative *calculated* amount per cell over time until day 7. As cells grew and cell numbers increased from day 7, cells continued to accumulate geosmin and the overall calculated intracellular geosmin concentration per cell increased, probably representing a stage where cells started to become synchronous in cell growth and *true* intracellular geosmin concentrations became apparent.

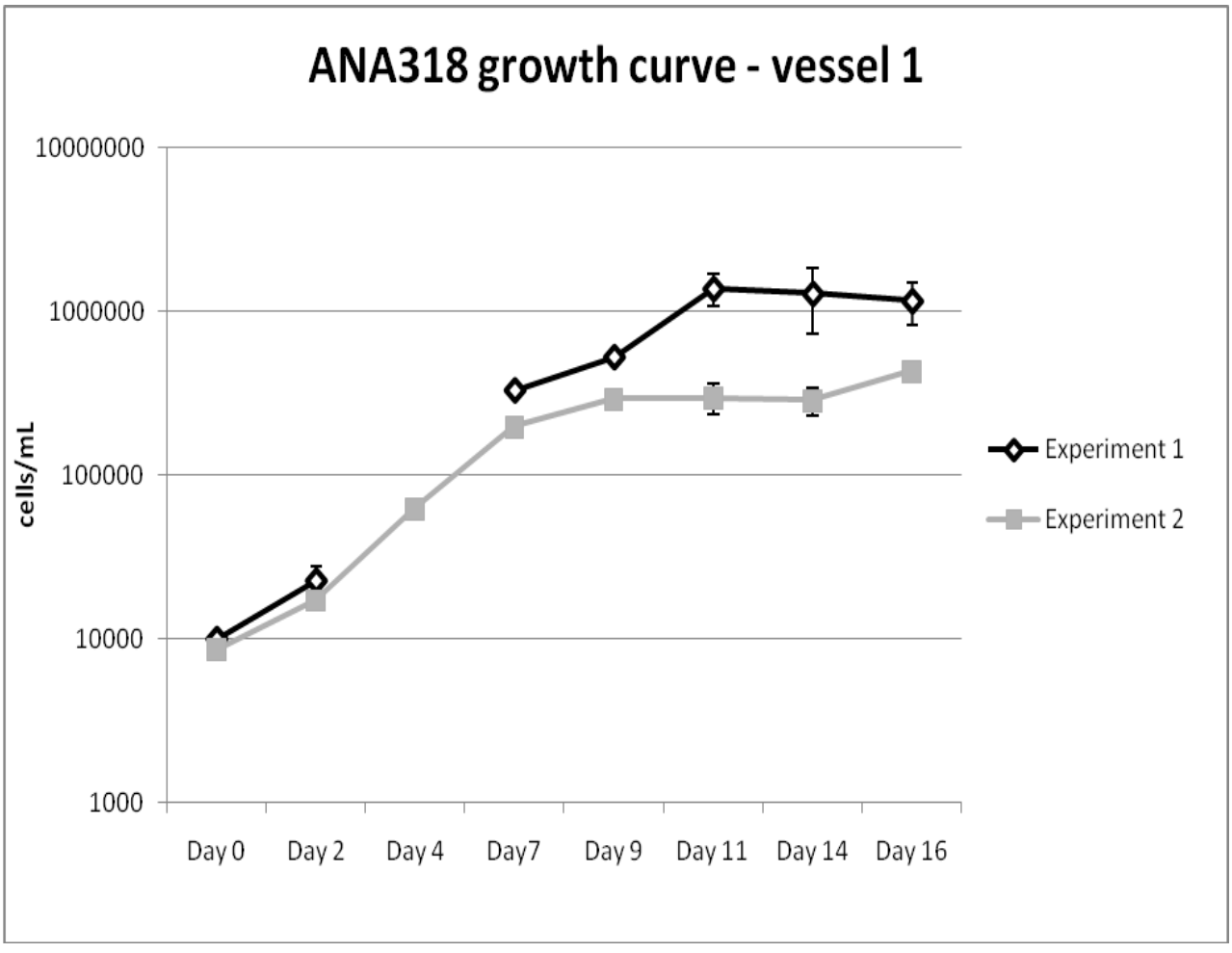


Figure 5-7. Vessel 1, Growth of ANA318

Table 5-3. Intracellular geosmin centrations, vessel 1

Day	Geosmin (fg/cell) Experiment 1	Geosmin (fg/cell) Experiment 2
Day 0	118.9	140.2
Day 2	53.44	106.3
Day 4	-*	68.71
Day 7	15.70	46.80
Day 9	20.40	61.01
Day 11	23.13	80.45
Day 14	19.06	97.68
Day 16	17.58	53.88

* cell count data not obtained

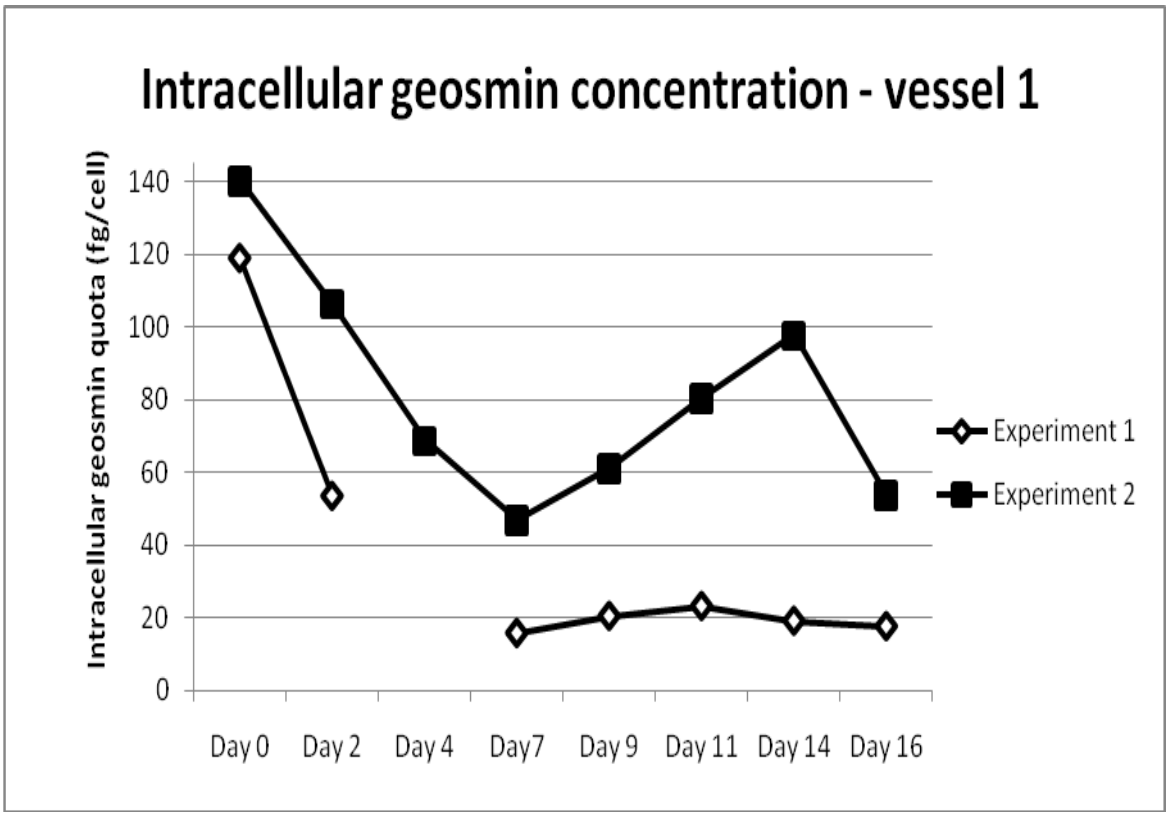


Figure 5-8. Intracellular geosmin concentration of replicate experiments for vessel 1.

Asynchronous growth and DNA replication/cell division has been demonstrated in *Synechococcus* (Binder and Chisholm, 1995), and is likely to be a common phenomenon amongst all cyanobacteria, when examined in both laboratory cultures and natural waters. The amount of intracellular geosmin increased during the incubation period until day 11 in experiment 1 and day 9 in experiment 2, at which point both cultures reached stationary phase. A decline in intracellular geosmin then occurred as the cultures reached late stationary/early decline phase and cell lysis occurred, resulting in an increase in extracellular geosmin in the medium, measured as a decrease in the intracellular : extracellular geosmin ratio (Figure 5-9). The majority of geosmin remained intracellular throughout exponential growth, and that it is probably through cell death that geosmin is liberated, rather than active transport across cell membranes into the medium.

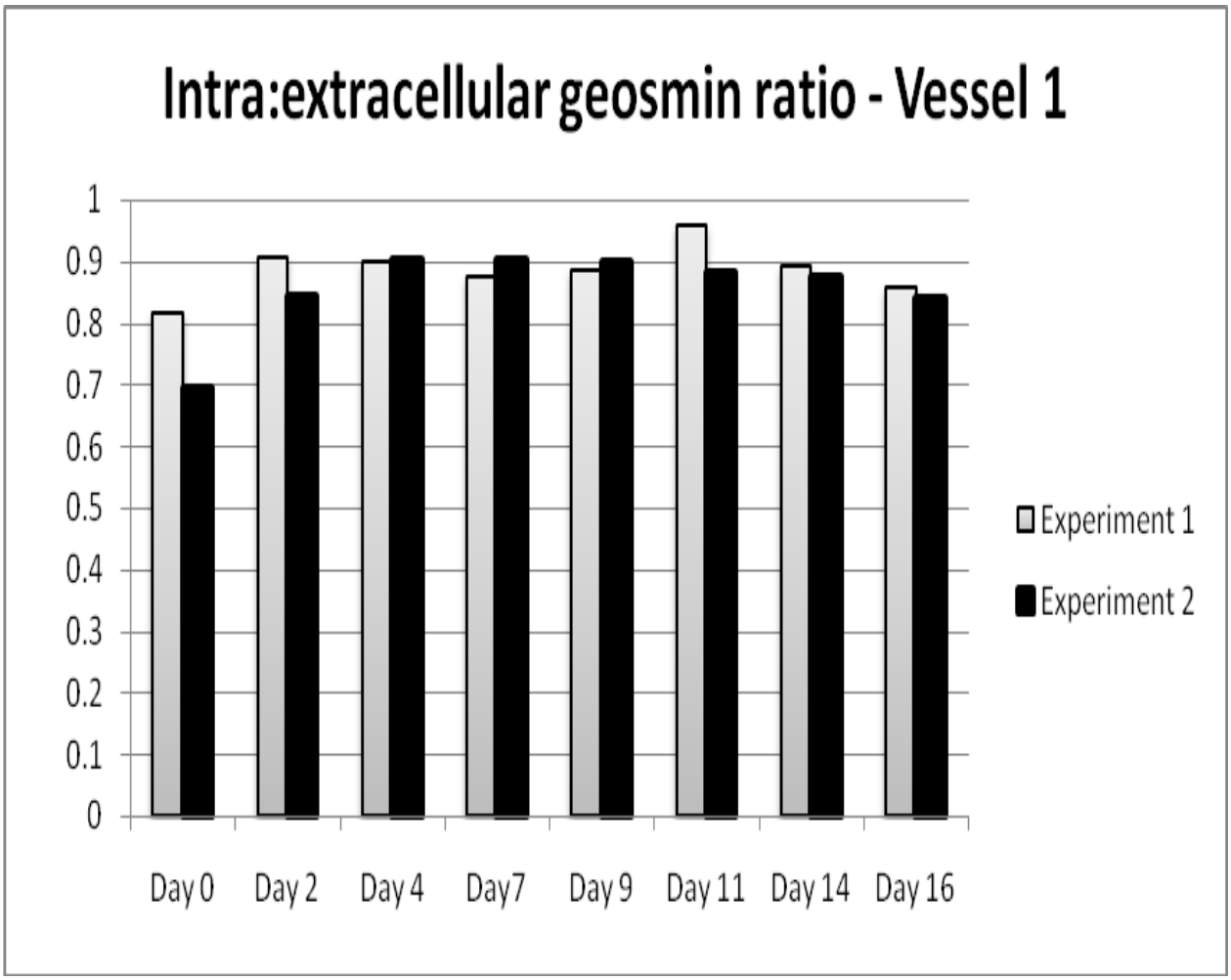


Figure 5-9. Vessel 1, ratio of intracellular : extracellular geosmin in ANA318

The approximate 2-3-fold difference in the absolute intracellular geosmin concentrations between experiment 1 and experiment 2 (presented in Table 5-3 and Figure 5-8) is more than likely a result of lower cell counts being obtained for experiment 2 (by 2 or 3 fold on average, see Figure 5-7), despite similar values obtained for the raw total and intracellular geosmin data in both replicate experiments (data not shown). Several weeks after the completion of the second experiment, critical issues were identified with the flow cytometer that resulted in the under-estimation of cell counts across a variety of assays utilised on this instrument. This critical issue resulted in the eventual decommissioning and replacement of the flow cytometer. It is likely, therefore, that this underestimation of ANA318 cell numbers seen in experiment 2 would have resulted in the over-estimation of intracellular geosmin concentration as reported here. Previous work using ANA318 has also demonstrated the intracellular geosmin concentration to be approximately 6-20 fg/cell (Peter Hobson, personal communication), in line with the values obtained for experiment 1, and in line with the values obtained for experiment 2 if a 2 or 3-fold reduction factor was introduced.

5.4.4 Growth of ANA318 and geosmin concentrations – Vessel 2, darkness induced on day 9

For the dark induced culture (vessel 2, incubated under the same conditions as vessel 1 for 9 days), the same growth characteristics and trend of intracellular geosmin concentration values were observed up until day 9. When darkness was induced at day 9, the intracellular geosmin concentration continued to increase until day 11, with no net increase in cell number, with a subsequent decline of intracellular geosmin concentration to day 16 (Figure 5-10). It is interesting to note that for both vessel 1 and vessel 2, geosmin continued to be synthesised 2 days post maximum cell numbers (at day 9-11 for vessel 1, and day 9 for vessel 2) and this observation is explored later.

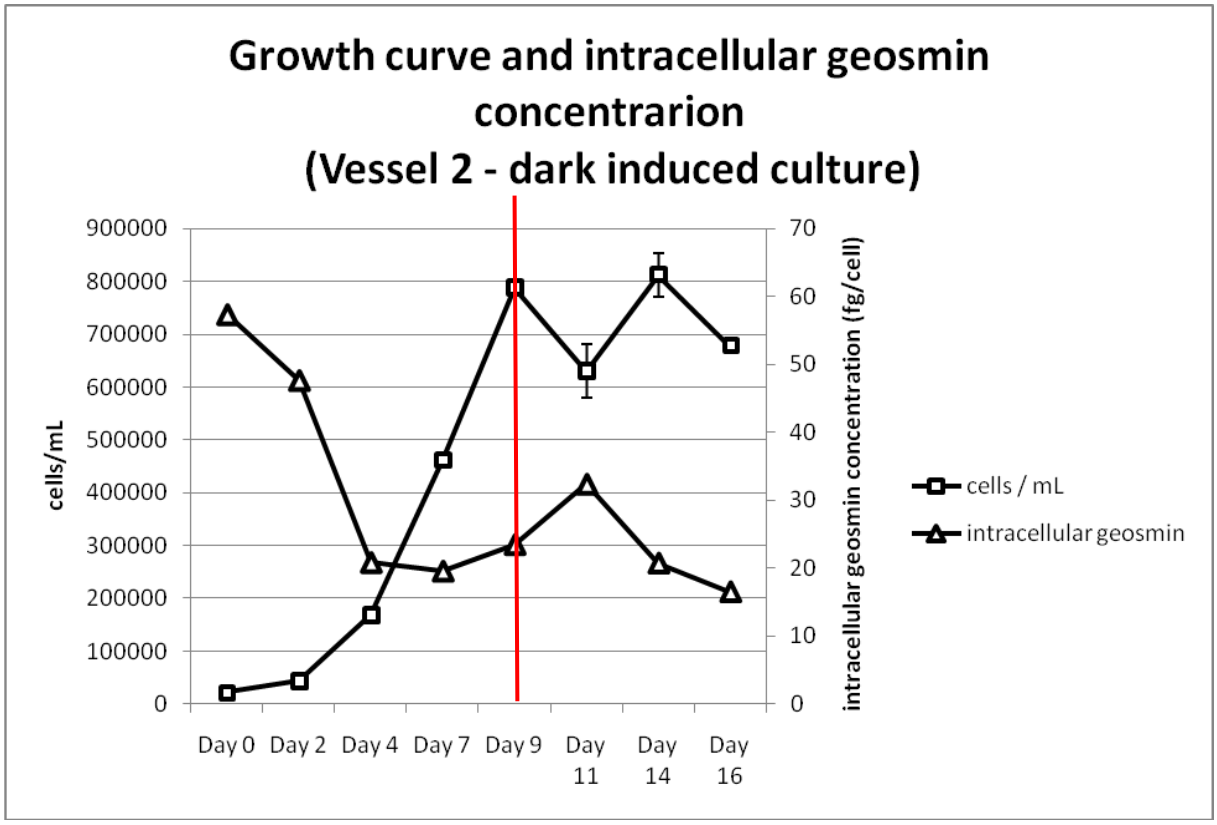


Figure 5-10. Vessel 2, growth and intracellular geosmin concentration for the dark induced culture at day 9 (red bar indicates the start of dark conditions).

5.4.5 Gene expression profile of geosmin synthase in ANA318

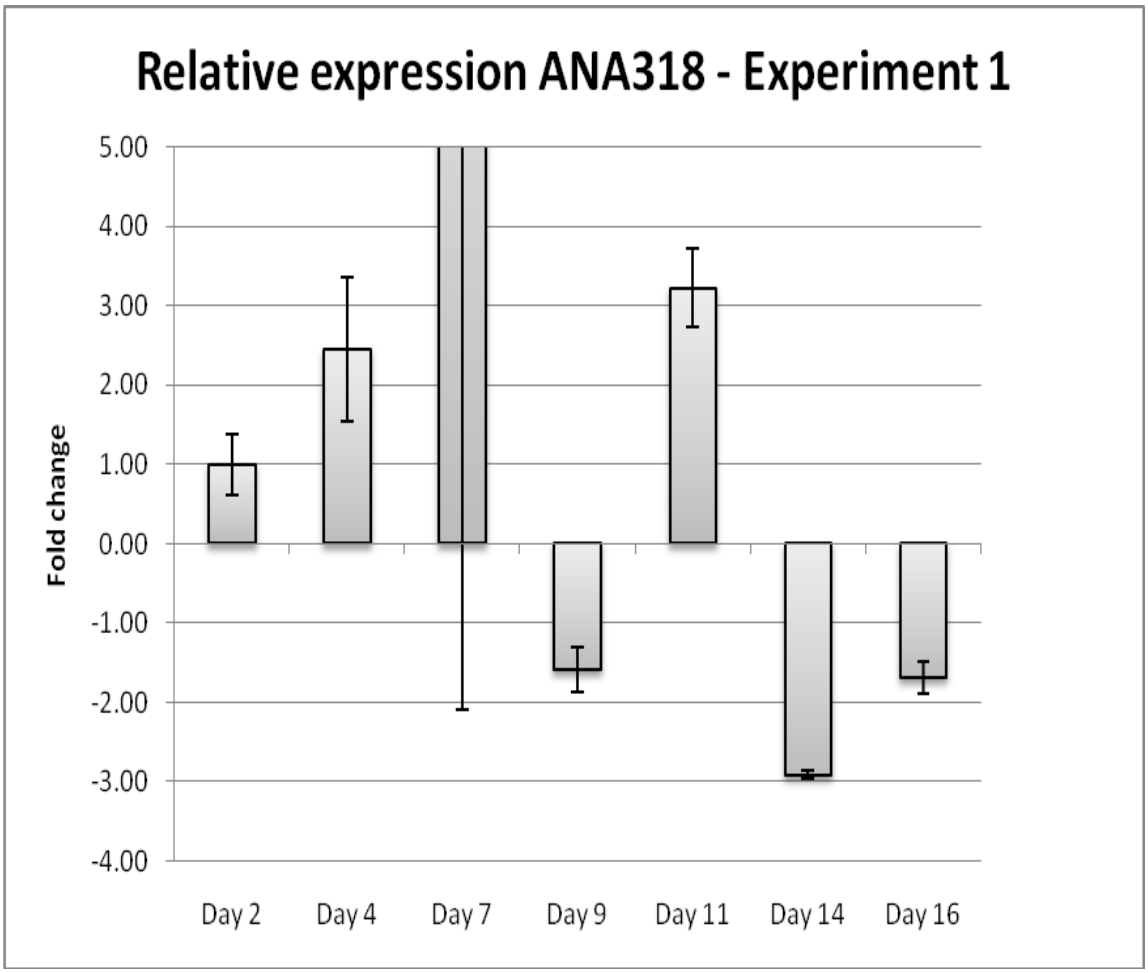
5.4.5.1 Vessel 1 – Continuous light

Relative expression, or fold change of geosmin synthase mRNA levels, is expressed relative to normalised data from Day 2. A significant fold-change in gene expression is defined as one that is at least 2-fold different to the normaliser range (Choe *et al*., 2005), in this case Day 2. A negative fold-change, indicates a decrease or down-regulation of gene activity, whereas a positive fold-change indicates an increase or up-regulation of gene activity.

A limitation of any expression analysis is that analyses fail to factor into account the asynchronous nature of cyanobacterial cultures (or any other cultures); the relative expression data for all data points is calculated from the population of cells extracted at a given comparative time point. It also represents an average value and it is assuming the cells are synchronous in culture and metabolic activity to represent a basal level of expression for the gene of interest. In early batch cultures, it is unlikely that cells in each theoretical phase of growth are completely co-ordinate with that growth phase, i.e. in the initial lag phase period, cells transferred from the acclimatised culture may have lysed or begun to have lysed, some cells may have continued in log-stationary phase from the transferred culture, and some cells may have started to become adjusted to a genuine lag phase. This is certainly supported by examining the intracellular geosmin concentrations (Figures 5-8 and 5-10), where there is a decline in calculated geosmin as a function of cell number, probably due to the inoculation of cells from acclimatised cultures with maximal intracellular geosmin concentration as previously discussed. The *calculated* expression of geosmin synthase at day 0 is, therefore, likely to be compromised and not truly indicative of geosmin synthase basal expression. Given the scenario of events possible at Day 0, representing the day photo-acclimatised cells were inoculated into the medium, the normalising comparative C_T data obtained from this day is unlikely to represent a truly synchronous and stable culture, which may complicate interpretation of relative changes when used as a normaliser. Thus, day 2 was chosen to obtain a better basal level of expression and a normaliser value, as the state of culture at this time would have become more stabilised than the culture at day 0.

Table 5-4. Fold changes of vessel 1 over time.

Day	Experiment 1		Experiment 2	
	$2^{-\Delta\Delta CT}$ (fold changes)	Std. Dev	$2^{-\Delta\Delta CT}$ (fold changes)	Std. Dev
Day 2	1.00	0.38	1.00	0.14
Day 4	2.45	0.91	-1.46	0.14
Day7	10.80	12.88	1.20	0.24
Day 9	-1.58	0.28	-1.00	0.09
Day 11	3.23	0.49	-3.67	0.01
Day 14	-2.91	0.05	-4.21	0.06
Day 16	-1.70	0.20	-6.15	0.05



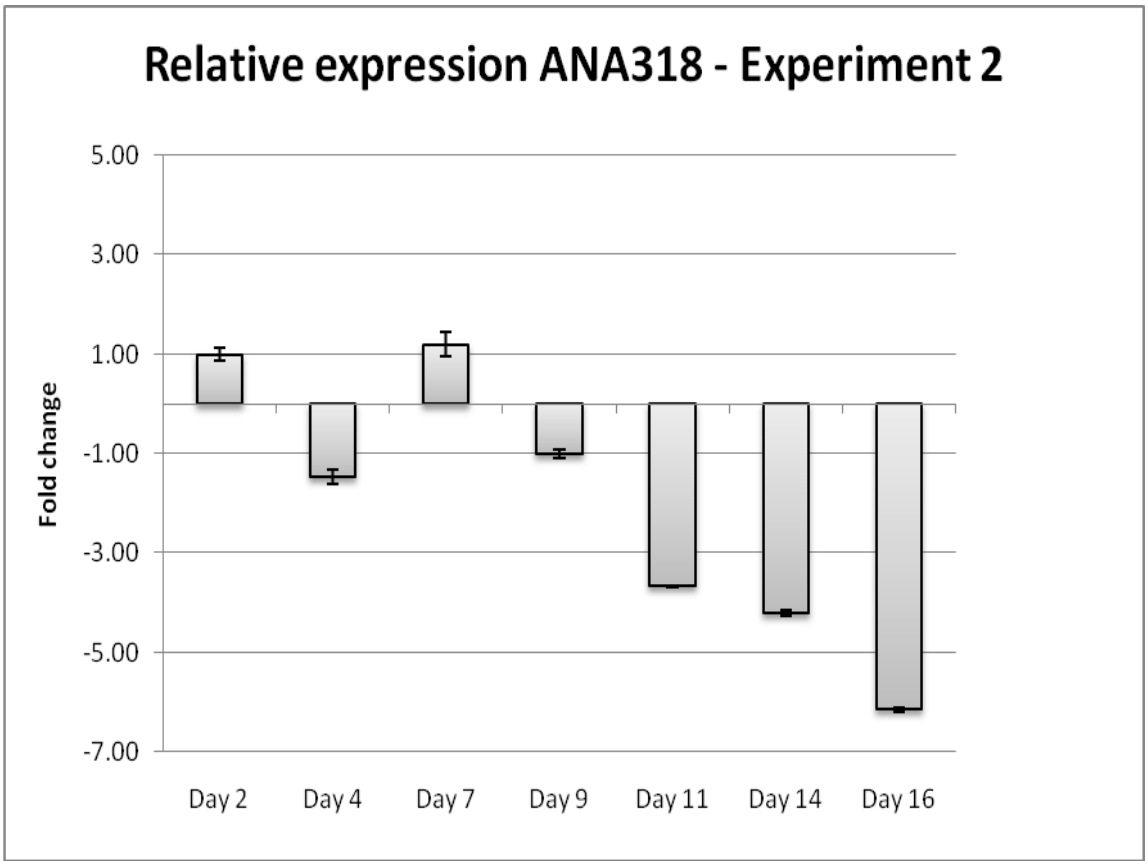


Figure5-11. Graphical representation of fold changes for vessel 1 over time.

Expression of geosmin synthase relative to expression of 16S rRNA is presented in Figure 5-11 and Table 5-4. Relative expression data for vessel 1, experiment 1 showed that a significant fold decrease was observed on days 14 and possibly 16, equivalent to -2.91+/-0.05 and 1.7 +/-0.2 fold changes respectively (the fold changes from days 4 and 7 were not considered significant due to the large standard error observed). Up until day 11, there are no significant changes in fold expression of geosmin synthase (fold differences <2) as cells grow to a maximum number and reach stationary phase. The increase in intracellular geosmin concentrartion from day 7-11 (Figure 5-8) appears to be linear and related to the synchronous growth of ANA318, therefore suggesting that the expression of geosmin synthase may be constitutive and co-ordinate with cell growth and metabolism. In between day 11 and 14, there is a significant decrease in geosmin synthase activity, yet the intracellular geosmin concentration for the same period increased.

A similar pattern was seen with experiment 2, where the culture appeared to reach stationary phase at day 9, with significant fold decreases of geosmin synthase activity on days 11, 14, and 16 (-3.67 +/- 0.01, -4.21 +/-0.06, and -6.15 +/-0.05 respectively). Similarly to experiment 1, the intracellular geosmin concentration increased from day 7 until day 11 (2 days after the start of stationary phase at day 9), with a decrease to day 16 (Figure 5-8). The increase of intracellular geosmin concentration that is concomitant with the apparent significant fold decreases in gene expression from days 9-11 is similar to experiment 1 , and could be due to 1) the geosmin synthase gene continuing to be transcribed and translated for a period of time before a decrease in geosmin synthase gene expression is detected, or 2) the geosmin synthase enzyme itself may be stable for at least a few days and continues to be functional.

5.4.5.2 Vessel 2 – examination of geosmin synthase expression in response to absence of light

Vessel 2 was incubated similarly to vessel 1, except that on day 9 the vessel was wrapped in aluminium foil to induce a period of darkness for the remainder of the incubation period. Day 9 represented a time point where the culture was mid-exponential and likely to be synchronous, and as such was an ideal normalising time point to monitor short and long term changes in geosmin synthase expression relative to that time.

The induction of darkness to the ANA318 culture did not significantly affect the expression of geosmin synthase within a 6 hour period, but there was a significant fold decrease from days 11-16 (Figure 5-12). As seen with vessel 1, the same phenomenon of decreased geosmin synthase expression and increase of intracellular geosmin concentration occurred between day 9 and day 11 for vessel 2, where the intracellular geosmin concentration increased (Figure 5-10), but was associated with a significant fold decrease in geosmin synthase expression at day 11 (Figure 5-12). From days 11-16, the intracellualr geosmin concentration then steadily declined, indicating no net geosmin increase per cell, with variable cell numbers over the corresponding time course. The increase in extracellular geosmin (measured as a decrease in the intracellular : extracellular geosmin ratio, Figure 5-13) indicated that cell death and lysis had also occurred during this period.

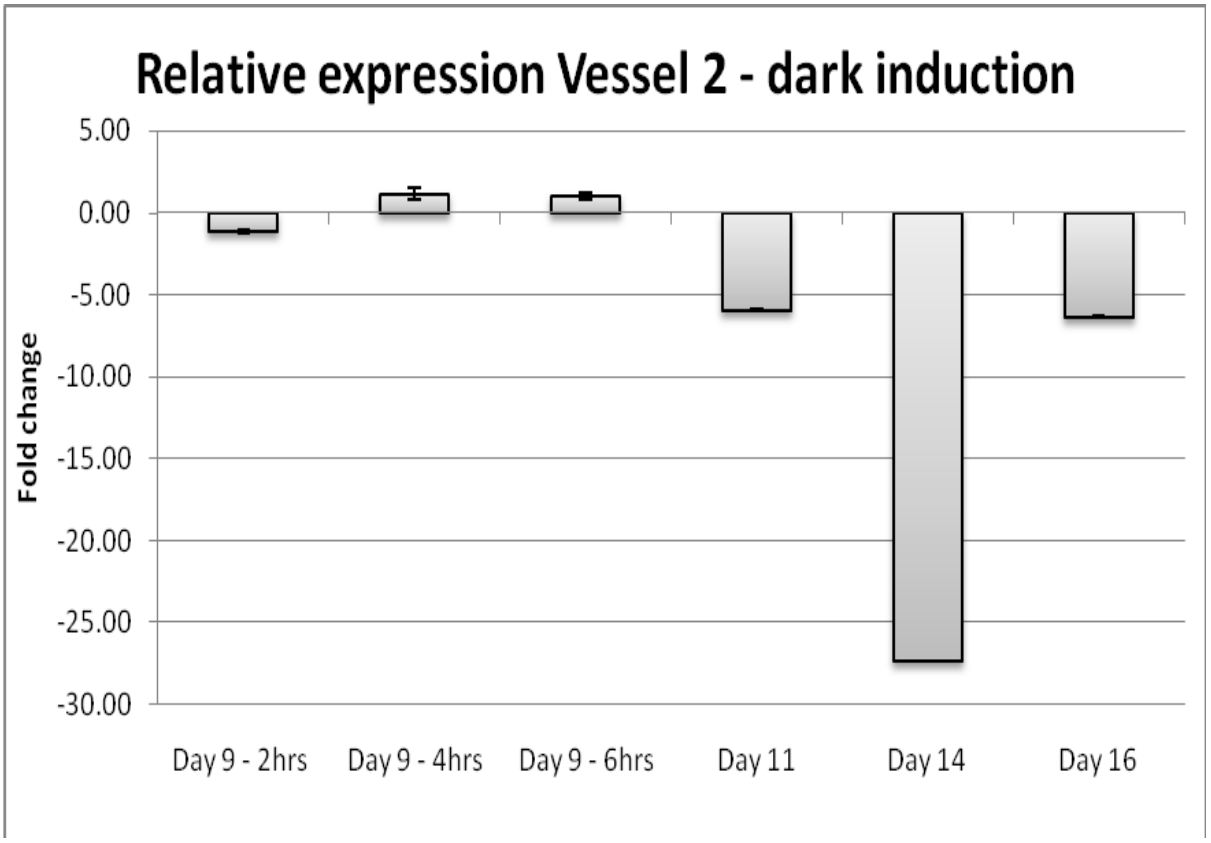


Figure5-12. Graphical representation of fold changes for vessel 2 over time.

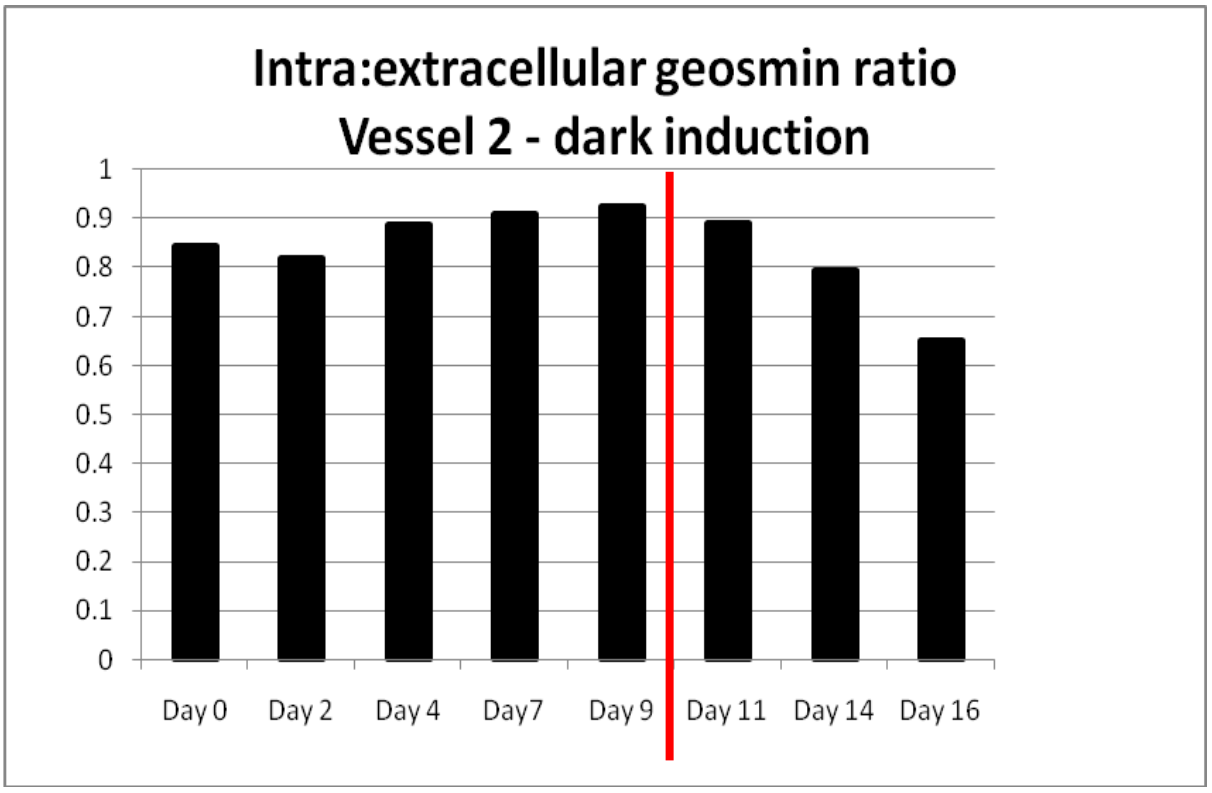


Figure5-13. Vessel 2, ratio of intracellular : extracellular geosmin in ANA318 (red bar indicates the start of dark conditions).

The combined data therefore strengthen the possibility that one of the following scenarios had occurred: that the geosmin synthase gene continued to be transcribed and translated in the dark for a certain period of time before observing a detectable change in gene expression, or that the geosmin synthase enzyme may have been stable for at least a few days and continued to be active.

When examining all the experiments, a similar trend was evident. Firstly, the initial 4 days after inoculation represented a time where the cyanobacterial cultures were acclimatising and the *calculated* intracellular geosmin concentration decreased, where this value is representative of the combined total geosmin from the presence of geosmin-laden cells initially inoculated, and the presence of new cells as cell division commenced. As the cultures became stable from day 7, and most likely synchronous in division, intracellular geosmin concentrations increased at a linear rate until maximum cell numbers occurred (stationary phase in vessel 1, just prior to the induction of a dark state in vessel 2). When cultures reached maximum cell numbers (in both vessel 1 and vessel 2), there was an apparent down regulation in the geosmin synthase gene. For 2-3 days after this event, geosmin continued to be synthesised by the cells, even in the dark, before cell lysis began and geosmin was liberated into the medium. Given that geosmin continues to be synthesised for 2 days in the dark, production is unlikely to be diurnal, suggesting that geosmin production in natural waters is likely to be a continual process so long as the cell is metabolically active and not approaching cell death.

Regulation of gene (and ultimately protein) activity can occur at several points from the DNA coding template, to protein function; levels of control can occur at transcription, translation, and post translation. From the data presented here, it is apparent that there was no immediate regulatory response to the abolishment of light (i.e. no fold changes >2 within 6 hours, Figure 5-12), and that the down regulation of transcription occurred in between 6 hours and 2 days post stationary phase and in response to the abolishment of light, suggesting that changes in geosmin synthase transcription occurred as cells approached the death/decline phase. As cells approach this phase, a cascade of events is likely to occur before cell death, where geosmin synthase transcription decreases prior to cell death.

The control at any of the possible regulatory levels may be species and genus dependant. For example, examining the fatty acid composition in a psychrotolerant *Nostoc* sp. SO-36, showed

that *desA*, *desB*, *desC*, and *desC2* (encoding lipid desaturases) were constitutively expressed at 10 °C, while in mesophilic cyanobacteria there was rapid up-regulation of these genes when these organisms were subjected to lower temperatures (Chintalapati *et al*., 2007). Also, post transcriptional control of *hspA* (encoding a heat shock protein) has been demonstrated in *Synechococcus vulcanus*, where the HspA protein was present at 42 °C but not at 30 °C, yet *hspA* was constitutively expressed at 30°C and 42 °C, suggesting that even though RNA was being transcribed at both temperatures, the protein was not translated at 30 °C (Kojima and Nakamoto, 2005).

The above two examples reinforce evolutionary successes of organisms and control of vital processes, namely conserving metabolic processes for functions when they are actually required. Organisms have evolved strict control mechanisms to conserve burdensome metabolic processes unless absolutely required. In the above example of fatty acid synthesis in mesophilic cyanobacteria, an increase in fatty acid production during conditions where they are not needed is metabolically exhaustive for the cell, and the production of protective fatty acids is only "turned on" when needed, i.e. in colder temperatures. The opposite is true for psychrotolerant cyanobacteria, where it is an absolute necessity for these fatty acids to be continually produced in order to provide tolerance at lower temperatures, thus conserving cell energy required for unnecessary control mechanisms for lipid desaturases.

Little work has been performed to investigate the correlation between geosmin synthesis and geosmin synthase gene expression. The sole report of a culture of *Phormidium* sp. examined the expression of a then suspected geosmin synthase in light and dark conditions, and demonstrated that suspected RNA transcripts were present in cultured organisms at 12 hours incubation in the light, and that they were absent at the end of a 24 hour period of incubation in the dark (Ludwig *et al*., 2007). The study by Ludwig and co-workers, however, did not measure geosmin concentrations of the culture at the time of extracting RNA, and also found an unusual second geosmin synthase separated by a 13bp spacer region in this organism. Unfortunately in the study of Ludwig *et al.* (2007) the cultures were not re-introduced to light for an extended period of time to see if a) the organisms were capable of regeneration, or b) whether the expression of geosmin synthase once again commenced. It is apparent from the work here and by Ludwig and co-workers, there is a down regulation of geosmin synthase

expression by geosmin-producing cyanobacteria when cultures are incubated in dark conditions for 24 hours or greater, but not within 6 hours.

From the data presented in this chapter, geosmin synthesis (and geosmin synthase expression) appears to be co-ordinate with cell growth, lending support to the expression of geosmin synthase being constitutive as a function of cell growth, under the conditions described here. The decrease in geosmin synthase transcription post maximum cell numbers and stationary phase suggests that a decrease of isoprenoid synthesis may occur before a decrease in the transcription of ribosomal units as the process of cell death is initiated.

5.5 SUMMARY

Earlier studies outlined in the introduction of this chapter attempted to link geosmin production to the isoprenoid pathway, and directly with chlorophyll *a*, light, and temperature. Conflicting results emerged between the relationship of chlorophyll and geosmin production, and indeed was further complicated with the introduction of temperature as a variable. It is likely that optimal geosmin synthesis is genus-dependent, and that the production of geosmin may not necessarily be related to chlorophyll synthesis alone.

Several biosynthetic arms are linked to the isoprenoid pathways, and the production of a multitude of products, including secondary metabolites and volatile organic compounds, are the result of a complex and organised series of reactions that are likely to be regulated on several levels. It would be difficult to assume that there is an absolutely direct relationship between chlorophyll and geosmin synthesis alone, without any other contributing drivers. For example, investigations of carotenoid production in high intensity light using *Synechococcus* have demonstrated that phytoene synthase, an enzyme that catalyses GGPP to the C40 phytoene, co-ordinates up regulation of the genes involved to both β-carotene and zeaxanthin, but there is light mediated regulation of even the intermediate genes within this pathway (Figure 5-14, the greater the width of the arrows, the greater the up regulation by light) (Schafer *et al*., 2006). Additionally genome wide analysis of *Synechococcus* has also shown that over 100 genes are regulated in response to changes in light exposure, and that changes in the photosynthetic arm are more globally controlled rather than a site specific event (Hihara *et al*., 2003), and it seems likely that there is a cascade of events involved in the regulation of photosynthetic and isoprenoid elements.

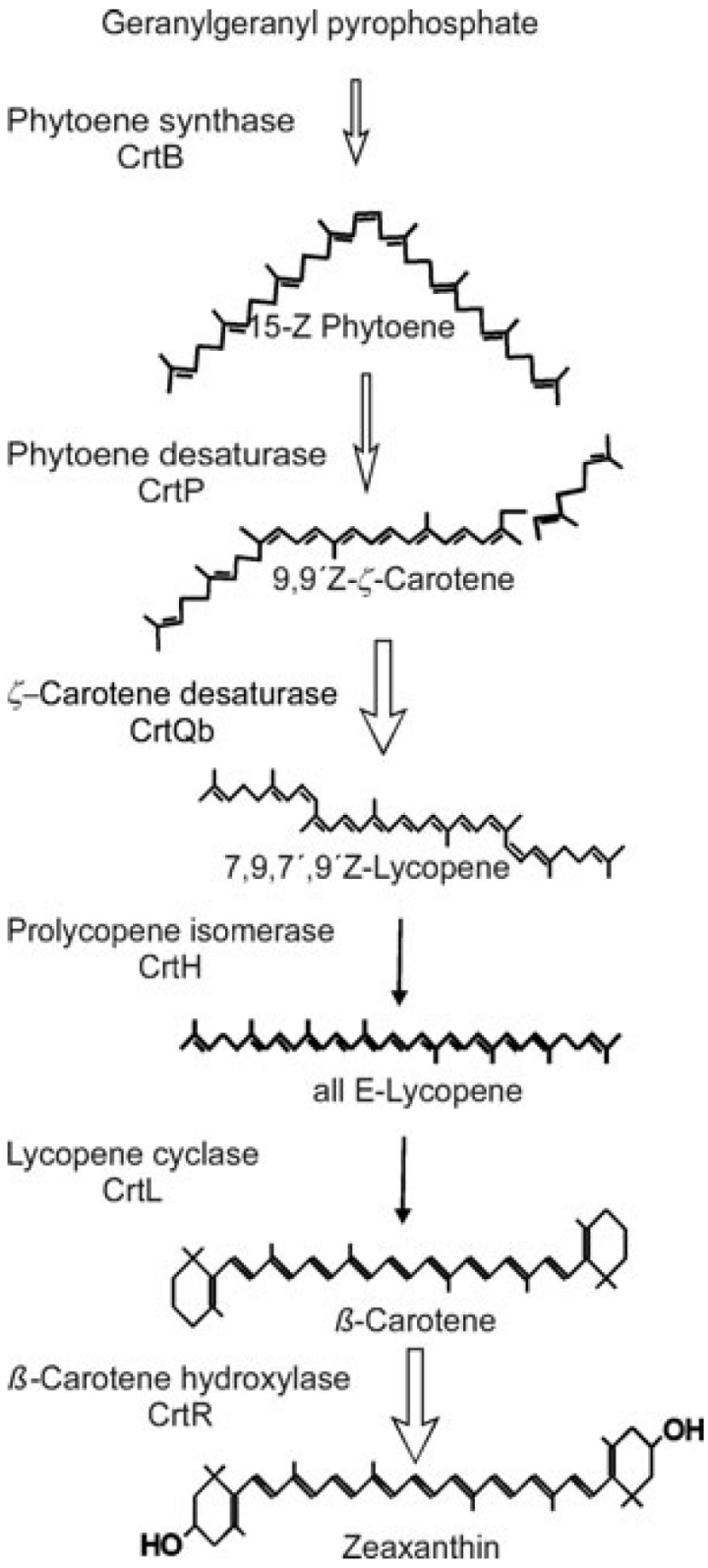


Figure 5-14. Regulation of carotenoids production in *Synechocystis*. The greater the width of the arrow, the greater the light dependence increase (reproduced from Schafer *et al.*, 2006).

The objective of this chapter was to examine geosmin synthase expression as a function of a single stimulus, light. The work here provides evidence that geosmin synthase expression appears to be constitutive, as a function of cell growth. Of particular interest is the observed phenomenon that although there is a down-regulation of geosmin synthase mRNA, geosmin synthesis continues for approximately 2-3 days afterwards. This raises several important questions: 1) is the geosmin synthase mRNA transcript stable for a period of time, or 2) is the geosmin synthase enzyme active for extended period of times? Whilst the majority of transcripts in most organisms are stable to some extent (Lam *et al*., 2001), the specific half life of each mRNA is generally related to its physiological role (Yang *et al*., 2003), and it is unsure why geosmin synthase would have extended stability. Additionally, changes to mRNA stability can occur in response to various stimuli such as hormones (Sahmi *et al*., 2006) and UV-B irradiance (Gowrishankar *et al*., 2005), and further work should be directed to distinguish between transcript and/or protein stability of geosmin synthases to answer these fundamental questions.

It is an exciting time to be a molecular scientist - the genomes of several organisms are becoming fully sequenced and annotated, biosynthetic mysteries are being solved, and the advent of DNA microarrays have allowed us to now examine global genomic responses to various stimuli at an extraordinarily fast rate. Such an approach with the geosmin-producing *Nostoc punctiforme* ATCC 29133 (PCC 73102), which has just only now been fully re-annotated, has already been examined for global changes in gene expression for hormogonia and heterocyst development (Campbell *et al*., 2007, Campbell *et al*., 2008). The manipulation of this organism to examine geosmin production as part of a global response mechanism is now possible because the hypothetical protein ZP_00109187 has now been corrected (NPUNMOD) and confirmed as a geosmin synthase gene as a result of the works presented here in Chapters 3 and 4. ZP_00109187 has been recently re-annotated to Npun_2756, in line with NPUNMOD, and any further work to propel our understanding of geosmin production and regulation in cyanobacteria would be well served by microarray analysis of this *Nostoc,* and examination of global transcriptional changes occurring with changes in geosmin synthase and geosmin synthesis. Perhaps any increase in geosmin synthase gene expression and geosmin synthesis may simply reflect an increase of isoprenoid activity, whereas a decrease in geosmin synthase gene expression may reflect a decrease in overall isoprenoid synthesis as cell death becomes imminent.

Chapter 6

DEMONSTRATION OF GEOSMIN SYNTHASES IN ENVIRONMENTAL CYANOBACTERIA

Use the key, unlock the door
See what your fate might have in store...
Come explore your dreams' creation
Enter this world of imagination...

Twilight zone, Rush

6.1 INTRODUCTION

The detection of geosmin in waters is not only troublesome from an aesthetic and operational viewpoint, but also an economical problem on several levels. Tracking the source of geosmin producers before, during, and after blooms is required in order to understand the prevailing conditions that have triggered such an event, and also devise strategies to mitigate future episodes. Rather than being reactive to operationally significant geosmin levels through the traditional use of activated carbon, with or without the use of copper sulphate to control blooms, utilities may benefit from becoming more proactive and conduct research into problematic areas to avoid unwanted blooms (or mass occurrence of benthic organisms) if possible.

In any investigation of a geosmin taste and odour episode, attributing the presence of geosmin in a water body involves (Persson, 1983):

- Concurrence of the odour and the organism in the field,
- Isolation of the organism and proof of odour production in the laboratory, and
- Chemical identification and sensory characterisation of the odour compounds.

Whilst the above points may seem obvious, perhaps even trivial, most reported descriptions of taste and odour events assume that the *occurrence* of a predominant organism is *concurrent* with the odour in the field. Whilst this is a safe assumption in several systems, the presence of an odour may not be the result of a predominant organism which is suspected of producing geosmin (e.g. high levels of geosmin when the predominant organism was *Microcystis aeruginosa* (a cyanobacterium not usually suspected of producing geosmin), Personal communication, Peter Baker).

During an investigation of a taste and odour episode, any given system has a list of “usual suspects” that historically cause problems at any given time. It is perhaps a sometimes complacent attitude that causes operators to first look for the same culprit every time there is a spike in geosmin concentration in the water. In many cases this is a safe assumption (for example several of SA Water’s systems have predictable *Anabaena circinalis* episodes), but for unusual or unseasonal geosmin episodes further investigations may be necessary.

Strain variation for the production of taste and odour compounds has been reported (Izaguirre *et al*., 1999). It appears possible, therefore, that there may be strains of cyanobacteria that are capable of producing and excreting taste and compounds at different rates, perhaps the notion of a "super producer" may even be appropriate. One such example of a possible super producer is an *Oscillatoria* sp that produced a humanly detectable MIB odour at 50 cells/mL in culture – an extraordinarily low cell concentration by usual standards (Rashash *et al*., 1996). Additionally, critical population densities (CPD, the density of cells that produces a detectable taste and/or odour in culture) seems to vary for different taste and odour producers; CPD for geosmin from a *Phormidium* sp was 200 cells/mL, whereas for *Anabaena laxa* the CPD was 3,800 cells/mL (Rashash *et al*., 1996). The rationale of using the presence of predominant species, therefore, may present a red herring in investigations of unusual episodes, as it may be the most inconspicuous species that is in fact causing the event, particularly if an organism such as the *Oscillatoria* sp above is initially overlooked as a suspect.

Several scenarios may occur within the investigation of a taste and odour episode. These may include

- the presence of a predominant planktonic organism known to produce geosmin,
- the presence of a predominant planktonic organism that is previously unrecognised to produce geosmin,
- the absence of a predominant planktonic organism (with and without recognised taste and odour producers),
- low levels of planktonic or benthic cyanobacteria that may or may not produce taste and odour compounds.

Tools to answer the question of "what is producing geosmin in our system?" in a cost efficient and timely manner are currently lacking. The cost of GC-MS to screen several isolates, or mixture of isolates, may often be prohibitive or operationally inconvenient, and often the turnaround time to process these samples is too lengthy to make rapid administrative decisions. Moreover, the sensitivity of GC-MS to detect geosmin from a single cell, or a few cells of cyanobacteria, is insufficient to provide a result using current methods.

Investigating the causative organism of the bloom is important for several reasons. Firstly, it is only through pin-pointing the organism responsible that physical and chemical parameters can be correlated to the presence of that organism, a factor that may alert operators in determining prevailing conditions for perhaps another bloom. Secondly, direct identification of a geosmin-producing organism will ensure that the operational problem is in fact treated. For example, if the predominant planktonic organism in a taste and odour episode is not confirmed as being able to produce geosmin, and if there is a benthic geosmin-producing cyanobacterium, then addition of copper sulphate will be to no avail. Similarly, the addition of copper sulphate to a possible actinomycete-related taste and odour episode will again have little effect.

A rapid tool to assist in identifying taste and odour producers is molecular diagnostics. In the clinical setting, molecular diagnostics have been used extensively for the diagnosis of disease (Nielsen *et al*., 2009), measurement of viral load (Sato *et al*., 2008), and demonstration of genetic mutations (Cecener *et al*., 2008). These same diagnostic techniques have slowly begun to permeate into environmental testing. Currently, there are several manufacturers and assays that are marketed towards food health and safety, such as the DuPont Qualicon BAX® system that tests for wide variety of bacteria including *Salmonella* and *E. coli* 0157 (http://www2.dupont.com/Qualicon/en_US/products/BAX_System/index.html). For industrial waters, GeneSystems has released a PCR test kit for the rapid detection of *Legionella* bacteria (http://www.genesystems.fr/en/Our-Environmental-Products/GeneDisc-Legionella/id-menu-70.html). Progress is being made for raw waters, however, as is highlighted by the use of PCR to detect toxic cyanobacteria (Rasmussen et al., 2008, Rantala et al., 2006, Koskenniemi et al., 2007) and *Naegleria fowleri* (Robinson *et al*., 2006), although these assays have not been commercialised and are currently used for in-house purposes only.

In the case of CYN-producing cyanobacteria, PCR assays have been developed to detect *ps* and *pks* genes of the CYN gene cluster (Fergusson and Saint, 2003, Wilson *et al*., 2000). These assays have demonstrated that the detection of these genes corresponds with the ability of the tested cyanobacteria to produce CYN. Taking these assays one step further, a real-time PCR assay was developed and transferred onto a field-specific PCR platform for the purpose of testing raw waters at North Pine Dam, Queensland. Analysis of these waters demonstrated that although there was poor to moderate correlation for quantitative data (r^2 = 0.203 – 0.443), the LC-MS detection of CYN in the waters was matched by the detection of the *pks* target gene for the

assay (Rasmussen *et al*., 2008). Combined, the initial laboratory studies firstly demonstrated the presence of *pks* in several CYN-producing isolates and absence in non-producing isolates, and secondly, the field studies correlated the detection of *pks* gene and CYN as analysed by LC-MS, suggesting that the detection of *pks* in a water sample is diagnostic of the potential toxicity of a water sample containing CYN-producing cyanobacteria.

Using the above rationale, the same approach for assessing a water body for a potential geosmin episode seemed plausible. First and foremost, however, similar genes (or highly conserved segments of genes), were required to be detected in a suite of geosmin-producing cyanobacteria, and not detected in non-producers (much as was the case for CYN-producing cyanobacteria). In the case of the geosmin synthase gene, any successful application of a direct diagnostic test must demonstrate that when the geosmin synthase gene is detected, the isolate produces geosmin. While the utility of this potential application is largely unknown, at the very least, the correlation of a cyanobacteria's ability to produce geosmin with the detection of a geosmin synthase gene by PCR may provide an alternative for the screening of isolates for the purposes of determining their geosmin production status, without the expense and time of GC-MS analyses. In the most desirable of scenarios, quantitative measurement of geosmin synthases in waters may provide an ideal indicator for predicting taste and odour blooms and administering proactive strategies before there is an operational problem detected, or to aid in the resolution of an episode.

6.2 AIMS

The aims of the research presented in this chapter were to

- Screen cyanobacterial isolates for the presence and absence of geosmin synthase based on N-terminal homology,
- Develop a real-time PCR assay for geosmin synthase, and
- Apply the developed real-time PCR assay to assess its potential use in monitoring geosmin synthase levels in River Murray water over a period of time, with a view to guiding future development of this potential screening tool if applicable.

6.3 SPECIFIC MATERIALS AND METHODS

6.3.1 DNA extraction

Unless otherwise indicated, DNA was extracted from isolates as outlined in 2.2.1. Similarly, DNA was extracted from water samples, except that 1 mL of water was centrifuged and resuspended in 180 µL ATL buffer prior to DNA extraction, and the proteinase K treatment was for 3 hours.

6.3.2 PCRs

6.3.2.1 G-PCR1

G-PCR1 was performed as described previously in 3.3.2.5.

6.3.2.2 G-PCR2

A second geosmin synthase PCR (G-PCR2) was developed for real-time PCR screening of cyanobacterial DNA samples. Reactions were first optimized by standard thermal cycling as described for G-PCR1, and then 2.5 µM SYTO9 (Invitrogen) was incorporated into the mastermix before running on a real-time PCR instrument. Primers used for G-PCR2 were: 288AF (5′-AACGACCTGTTCTCCTA-3′), and 288AR (5′-GCTCGATCTCATGTGCC-3′), generating an amplicon of 288 bp. Real-time PCR was performed using a Corbett Research Rotorgene 6000HRM (Corbett Research, Australia) under the same conditions as G-PCR1, except that the final extension step was replaced with a DNA melting analysis from 75 to 95 °C, with data being acquired every degree with a 5 s hold at each step. All data were acquired on the “Green” channel, with excitation at 470 nm and emission at 510 nm.

6.3.3 Extraction of DNA following agarose gel electrophoresis

When required, DNA bands from agarose gel electrophoresis were excised using a sterile scalpel blade and DNA extracted from the agarose plug using the QIAGEN QIAQuick Gel Extraction Kit according to manufacturer’s instructions.

6.3.4 Generation of PCR product for G-PCR2 quantification

G-PCR2 product was amplified from *Nostoc punctiforme* as described in 6.3.2.2. The PCR product was purified and concentrated using the Montage centrifugal filters as described in 2.2.5, and quantified by standard UV spectroscopy. PCR product was serially diluted in sterile

nuclease free water, and 5 µl was added per reaction to generate a standard curve for quantitative analysis.

6.3.5 Collection of water samples from the River Murray, South Australia

Samples were collected from Cobdogla, Cadell, Morgan, and Lock 1, from January to February 2009, during a period of moderate to high geosmin levels (Figures 6-1 and 6-2). A 2 meter sampling rod was immersed into the water at a depth of 1m, and 5 x 1L of water was collected. Samples were aliquoted and analysed for the presence of cyanobacteria (in particular *Anabaena* sp.), and also analysed for geosmin using GC-MS (single analysis only). Levels of geosmin synthase genes in the corresponding samples were determined using quantitative G-PCR2, using the standard curve method and standards generated in 6.3.4, and the comparison of the unknown River Murray sample signal against the standard curve. PCRs were performed on neat, 1:10, and 1:100 serially diluted DNA extracts, with the highest calculated results plotted in graphs as detailed in 6.4.3.3.2.

6.3.6 Reference Strains

The following reference strains were generously provided by George Izaguirre: *Pseudanabaena limnetica* (MIB producer, Castaic Lake, (Izaguirre and Taylor, 1998)), *Anabaena laxa* (CA783, geosmin producer, Castaic Lake,(Rashash *et al*., 1996)), *Phormidium calcicola* (MIB and geosmin producer,LM689, Lake Matthews, (Izaguirre, 1992)), and *Nostoc* sp (SWC701,geosmin producer).



Figure 6-1. Sampling locations, state level (boxed).

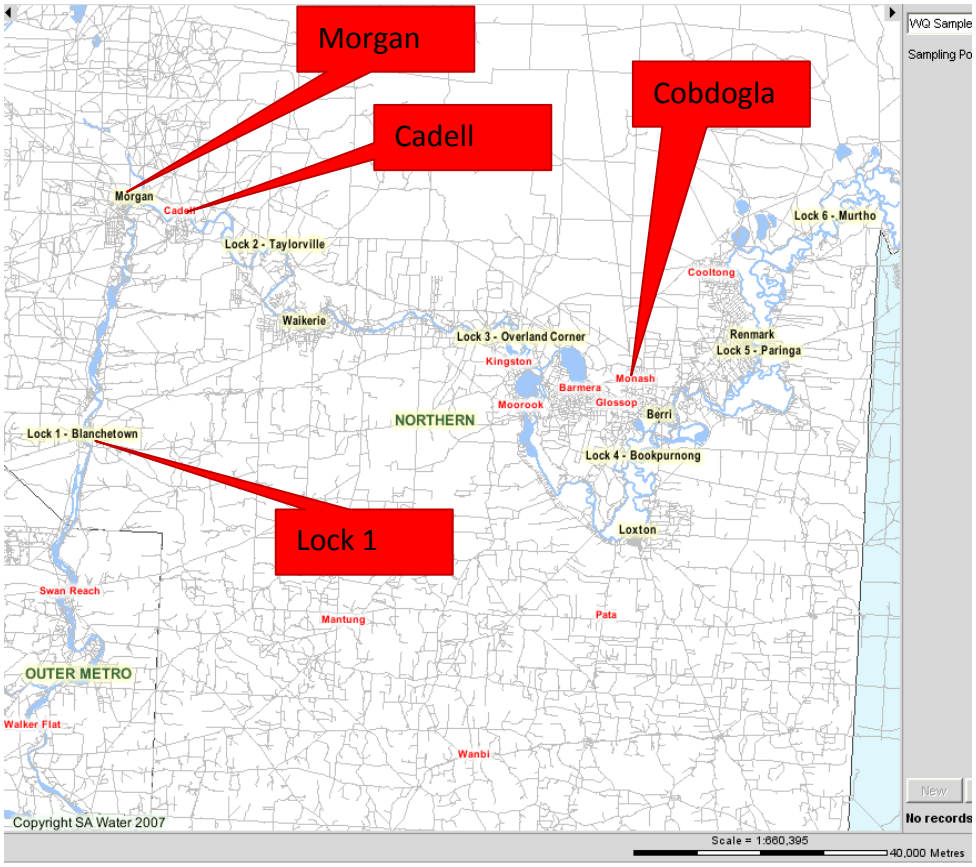


Figure 6-2. Sampling locations in detail.

6.4 RESULTS AND DISCUSSION

6.4.1 Screening of reference strains for geosmin synthase

Nostoc punctiforme and the reference strains were screened by G-PCR1. G-PCR1 assay was designed to flank the highly conserved DDHFLE and NSE motifs in the N-terminal half of geosmin synthase. The N-terminal domain had higher amino acid conservation between *S. coelicolor, S. avermitilis, and N. punctiforme*, and was thus chosen to screen additional isolates with a view to developing a diagnostic PCR.

Figure 6-3 shows the presence of an approximately 743 bp band for *N. punctiforme*, *Nostoc* sp*, Phormidium calcicola*, and *A. laxa*, with a failure to detect a band in the *Pseudanabaena limnetica* isolate. The presence of a band correlated with geosmin production by the isolate as detected by GC-MS, and the absence of a band correlated to the absence of geosmin production.

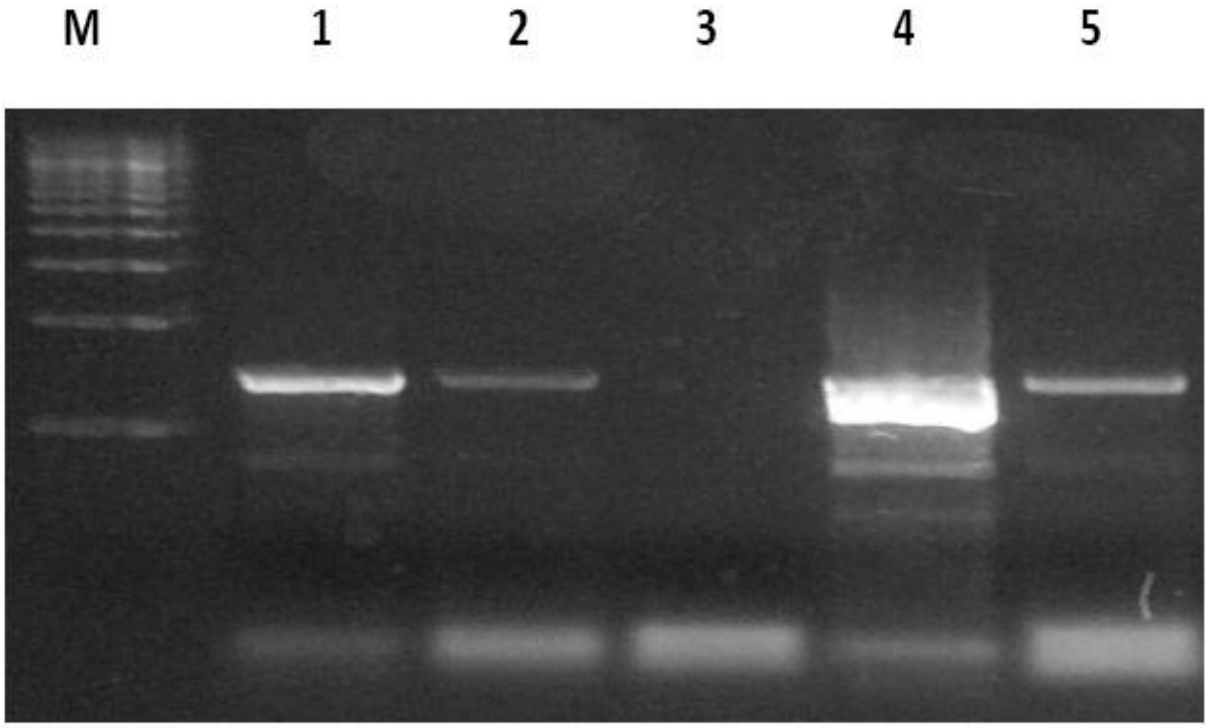


Figure 6-3. Gel electrophoresis demonstrating the presence of putative geosmin synthase homologs in cyanobacteria. (M; 500 bp DNA ladder; 1, *Nostoc punctiforme* ATCC 29133 (geosmin producer); 2, *Nostoc* sp. SWC701 (geosmin producer); 3, *Pseudanabaena limnetica* (geosmin non-producer); 4, *Phormidium calcicola* (geosmin and MIB producer); 5, *Anabaena laxa* (geosmin producer).

Electrophoresis bands of expected size were excised and DNA extracted as described in 6.3.3. Purified DNA was cloned into the TOPO 2.1 vector (see 2.2.6), and the cloned insert sequenced using M13 sequencing primers supplied with the kit. The results in Figure 6-4 show that the translated protein products harbour the anticipated DDHFLE and NSE motifs, with a high similarity to the amino acid sequences of each other and the corresponding *Streptomyces coelicolor* and *Streptomyces avermitilis* protein segments, and are representative of the N-terminal domain of a geosmin synthase (nucleotide sequences available in Appendix 1-6).

```
N.punctiforme       ------------------------------------------------------------
A.laxa              ------------------------------------------------------------
P.calcicola         ------------------------------------------------------------
Nostoc. sp          ------------------------------------------------------------
SCO6073             MTQQPFQLPHFYLPHPARLNPHLDEARAHSTTWAREMGMLEGSGVWEQSDLEAHDYGLLC 60
SAV2163             -MTQPFQLPHFYMPYPARLNPHLDEARAHSTRWARGMGMLEGSGIWEQSDLDAHDYGLLC 59


N.punctiforme       ------------------------FFDDHFLEIYKRTQDMAGAKEYLGRLPMFMPIYPTET 37
A.laxa              ------------------------FFDDHFLEIYKRSQDMVGAKEYLDRLPAFMPIYPRDN 37
P.calcicola         ------------------------FFDDHFLEIYKRTKDMKGAKHYLHRLRAFMPIHSTET 37
Nostoc. sp          ------------------------FFDDHFLELYKRSQDMAGAKKYLDGLPAFMPVQPQET 37
SCO6073             AYTHPDCDGPALSLITDWYVWVFFFDDHFLEKYKRSQDRLAGKAHLDRLPLFMPLDDAAG 120
SAV2163             AYTHPDCDGPALSLITDWYVWVFFFDDHFLETFKRTQDREGGKAYLDRLPLFMPLDLSAP 119
                                            ******** :**::*  ..* :*  *  ***:

N.punctiforme       PPVPTNPVECGLADLWSRTAFTKSVDWRLRFFGSTKNLLEESLWELAYINQDRVANPIEY 97
A.laxa              PLVPTNPVERGLADLWSRTAFTKSVEWRRRFFKSTKNLLDESMWELANINQNRIANPIEY 97
P.calcicola         MPAPTNPVERGLADLWSRTALTKSVEWRVRFSESTKNLLEESLWELANINQNRVSNPIEY 97
Nostoc. sp          PPEPTNAVERGLVDLWDRTIPNASQDWVIRFSESTINLLKESQWELANISQNRVANPIEY 97
SCO6073             MPEPRNPVEAGLADLWTRTVPAMSADWRRRFAVATEHLLNESMWELSNINEGRVANPVEY 180
SAV2163             VPEPENPVEAGLADLWARTVPAMSADWRKRFAVSTEHLLNESLWELSNINEGRIANPVEY 179
                       * *.** **.*** **    * :*  **  :* :**.** ***: *.:.*::**:**

N.punctiforme       IEMRRKVGGAPWSADLVEHAVFIEIPADIASTRPMRVLKDTFADEVHLRNDLFSYQREVE 157
A.laxa              IEMRRKVGGAPWSADLVEHAAFVEVPAKIAATRPMRVLKDTFADGVHLRNDLFSYQREVE 157
P.calcicola         IEMRRKVGGAPWSADLVEHAAFVEVPAQIAATRPMRVLKDTFADGVHLHNDLFSYQREVE 157
Nostoc. sp          IEMRRKVGGAPWSANLVEHAVGADIPAAIAPTRPMRVLKDTFSDGVHLRNDIFSYQREVE 157
SCO6073             IEMRRKVGGAPWSAGLVEYAT-AEVPAAVAGTRPLRVLMETFSDAVHLRNDLFSYQREVE 239
SAV2163             IEMRRKVGGAPWSAGLVEYAT-AEVPAAVAGSRPLRVLMETFSDGVHLRNDLFSYQREVE 238
                    **************.***:*.  ::** :* :**:*** :**:* ***:**:********

N.punctiforme       DEGENANCVLVLERFLNVSTQEAANLTNELLTSRLYQFDNTAVTELPPLFEEYGLDPVAR 217
A.laxa              DEGENSNCVLVIEKFLNVSTQEAANLTNELLNSRLYQFDNTAVTELPSLFEEYGVDPVER 217
P.calcicola         DEGENANCVLVLERFLDVTTQEAANLTNELLSSRLYQFDNTAVTELPPLFEEHGLDPAAR 217
Nostoc. sp          EEGENANCILVLERFLDVSTQEAANLTNDLLTSRVQQFENTFVTELPSLFEEYSLSPDER 217
SCO6073             DEGELSNGVLVLETFFGCTTQEAADLVNDVLTSRLHQFEHTAFTEVPAVALEKGLTPLEV 299
SAV2163             EEGELSNGVLVLETFFGCTTQEAAETVNDILTSRLHQFEHTALTEVPALALEKGLTPPEV 298
                    :*** :* :**:* *:. :*****: .*::*.**: **::* .**:*.:  * .: *

N.punctiforme       VNVLLYIKGLQDWQSGGHEWHMRSSRYMNE------------------------------ 247
A.laxa              VNVLLYIKGLQDWQSGGHEWHMRSSRYMNE------------------------------ 247
P.calcicola         MSVVLYIKGLQDWQSGGHEWHMRSSRYMNK------------------------------ 247
Nostoc. sp          LKVVLYAKGLQDWQSGGHEWHMRSSRYMNE------------------------------ 247
SCO6073             AAVGAYTKGLQDWQSGGHEWHMRSSRYMNKGERPLAGWQALTGPGTSAADVGALLADAVA 359
SAV2163             AAVAAYARGLQDWQSGGHEWHLRSSRYMNEGALSQKRPFGLSAIGTSAADLRGLLADAGA 358
                      *  * :*************:*******:
```

Figure 6-4. Reference strain amino acid sequences from putative geosmin synthases. Essential and characteristic motifs boxed in yellow.

Sequence comparison of *Phormidium calcicola* (representative of the newly sequenced putative geosmin synthases) against deposited sequences in the GenBank database matched several other protein sequences as described in Table 6-1. Characteristic DDHFLE and NSE residues were apparent in all the named sequences. Although the majority of the enzymes detailed have not been subjected to the same functional assays as described for SCO6073, SAV2163, and now the NPUNMOD protein, other uncharacterised terpene synthases or sesquiterpene synthases that are not geosmin synthases can be ruled out because of the presence of the DDHFLE motif that is characteristic of geosmin synthases, as opposed to characteristic FFxxDDxxD (where x can be any amino acid) that are found in pentalene and epi-isozizaene synthases that are also part of the terpene synthase/cyclase group of enzymes (Komatsu *et al*., 2008). Combined, these preliminary results strongly suggested that the detection of N-terminal geosmin synthase homologs, dubbed in some of the BLAST assigned functions as a germacradienol or a germacradienol/germacrene D synthase, was sufficient to screen isolates for the presence of geosmin synthase, and to assess the correlation between the detection of this gene and the isolate's ability to produce geosmin as detected by GC-MS.

Table 6-1. Protein BLAST of *Phormidium calcicola* sequence.

Gene / Protein Reference	Organism/clone	% identity	BLAST assigned function
YP_001866236.1, ACH86110.1, ACH86109.1	*Nostoc punctiforme*, NPUNMOD, NJ2	87	terpene synthase metal-binding domain-containing protein
ABU93239.1	*Phormidium* sp. P2r	81	Sesquiterpene synthase
ABU93238.1	*Phormidium* sp. P2r	79	Sesquiterpene synthase
ZP_01460669.1	*Stigmatella aurantiaca*	72	Germacradienol synthase
YP_483306.1	*Frankia* sp.	71	Terpene synthase
YP_001612078.1	*Sorangium cellosum*	72	Terpene cyclase
YP_002198494.1	*Streptomyces pristinaespiralis*	69	Germacradienol/germacrene D synthase
YP_001107098.1	*Saccharopolyspora ertyhrae*	69	Terpene synthase
YP_001828351.1	*Streptomyces griseus* subsp *griseus*	68	Germacradienol synthase
ABY50951.1	*Streptomyces peucetius*	65	Germacradienol synthase
YP_002178267	*Streptomyces* sp.	66	Germacradienol/germacrene D synthase
YP_002201639	*Streptomyces sviceus*	65	Germacradienol/germacrene D synthase
NP_823339.1	*Streptomyces avermitilis*	64	Germacradienol synthase
NP_630182.1	*Streptomyces coelicolor*	65	Germacradienol/germacrene D synthase

6.4.2 G-PCR2 - real-time PCR assay for geosmin synthase

The initial success of the G-PCR1 assay on reference strains instilled the confidence to pursue the N-terminal domain as a potential diagnostic target for geosmin synthases. A new suite of isolates were analysed by GC-MS for geosmin production, and assayed using a rapid real-time PCR (G-PCR2) that was designed to use primers internal to the sequenced G-PCR1 products. Seventeen such isolates were screened using this PCR assay and analysed in parallel with GC-MS analyses of isolate suspensions. Table 6-2 demonstrates that G-PCR2 was able to detect additional isolates of *Phormidium*, *Anabaena*, and *Geitlerinema* sp, all of which were GC-MS positive for geosmin. Moreover, the real-time G-PCR2 protocol did not amplify DNA from any

GC-MS geosmin-negative isolates such as *Planktothrix*, *Oscillatoria*, and *Pseudanabaena* sp., clearly implying that the presence of a geosmin synthase gene (moreover the detection of the N-terminal domain) is invariably associated with geosmin production by the host. These data also suggested that the expression of geosmin synthase may be constitutive, as there was no event of a GC-MS positive sample in the absence of gene detection. Combined, these results validated the utility of the real-time PCR gene detection as an informative diagnostic assay for cyanobacterial geosmin producers.

Table 6-2. Comparison of G-PCR2 with GCMS geosmin analysis on additional cyanobacterial isolates.

Isolate	Origin	Geosmin production	G-PCR2
***Oscillatoria* LM603**	Lake Mathews, USA	-	-
***Oscillatoria* sp.**	Hope Valley Reservoir, SA	-	-
***Oscillatoria* sp.**	Hope Valley Reservoir, SA	-	-
***Phormidium* sp.**	Happy Valley Reservoir, SA	+	+
***Phormidium* sp.**	Happy Valley Reservoir, SA	+	+
***Phormidium* sp.**	Happy Valley Reservoir, SA	+	+
***Phormidium* sp.**	Happy Valley Boat Ramp, SA	+	+
***Phormidium* sp.**	Happy Valley Reservoir, SA	+	+
***Phormidium* sp.**	Happy Valley Reservoir, SA	+	+
***Planktothrix* sp. DVL1003C**	Diamond Valley Lake, USA	-	-
***Anabaena circinalis* ANA346B**	Myponga Reservoir, SA	+	+
***Pseudanabaena* sp.**	Happy Valley Reservoir, SA	-	-
***Pseudanabaena* sp.**	Happy Valley Reservoir, SA	-	-
***Geitlerinema* sp.**	Little Para Reservoir, SA	+	+
***Phormidium* sp. 007D**	Hope Valley Reservoir, SA	+	+
***Phormidium* sp. 005E**	Hope Valley Reservoir, SA	+	+
***Phormidium* sp. 012G**	Hope Valley Reservoir, SA	+	+

The use of G-PCR2 and associated melting curve analysis has the additional advantage of potentially differentiating geosmin-producing species using a single PCR reaction. In this PCR protocol, the positive *Geitlerinema, Phormidium,* and *Anabaena* species produced different characteristic melting profiles (Figure 6-5). For example, *Anabaena circinalis* gave rise to a

single peak with a melting temperature (*T*m) of 84.5 °C, while *Anabaena laxa* had a more complex melting pattern with two *T*ms; one at 84.5 °C and another at 82.3 °C. A similarly complex melting pattern was seen for all *Phormidium* sp, with *T*ms at 84.5 and 87.5 °C and also for *Geitlerinema* that has *T*ms at 84.5 and 86.5 °C. The use of melting curve analysis for species identification and genotyping is becoming increasingly popular, especially with the availability of *T*m prediction software such as POLAND and MELTSIM (Rasmussen *et al*., 2007). Furthermore, the observation of differential *T*ms from a single primer pair has been used elsewhere to characterize several *Naegleria* isolates in a diagnostic setting (Robinson *et al*., 2006), suggesting that the combination of real-time PCR for detection of geosmin synthase genes, and *T*m analysis, may turn out to be similarly useful if validated further.

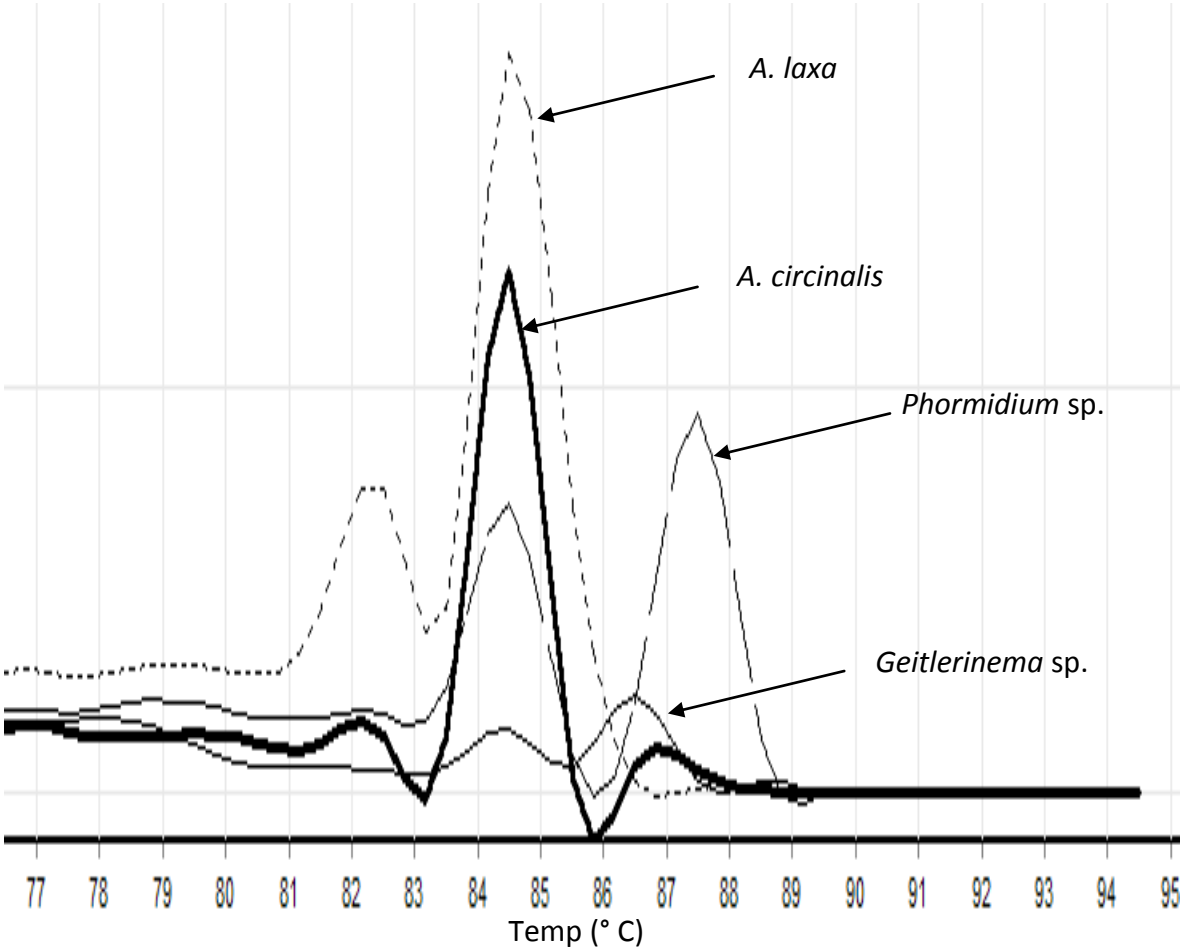


Figure 6-5. Melting curve analysis of G-PCR2 products. *Anabaena circinalis* (thick line), *Anabaena laxa* (dotted), *Phormidium* sp (large dashes), and *Geitlerinema* (thin line) all displaying differentiated melting curve patterns.

6.4.3 Monitoring geosmin synthase levels in the River Murray, South Australia

6.4.3.1 Geosmin levels in sampling locations

The purpose of this study was to investigate the utility of the G-PCR2 assay directly on field samples, and to compare the geosmin synthase data generated against algal counts in conjunction with the detection of geosmin in the water using GC-MS. At the time of this study, geosmin levels ranged from 90 ng/L at Cobdogla, to <2 ng/L at Lock 1 (Figure 6-6) during the sampling period (boxed in red). The sampling of these waters provided an excellent opportunity to determine the utility of G-PCR2. Low levels, as seen at Lock 1 (range of <2 – 6 ng/L), and moderate levels as seen in Cadell (5-34 ng/L) are similar to those seen in reservoir bodies during inactive and active taste and odour episodes, and would therefore provide a valuable indication whether the usefulness of PCR may be realised at operationally significant levels of geosmin concentrations in waters.

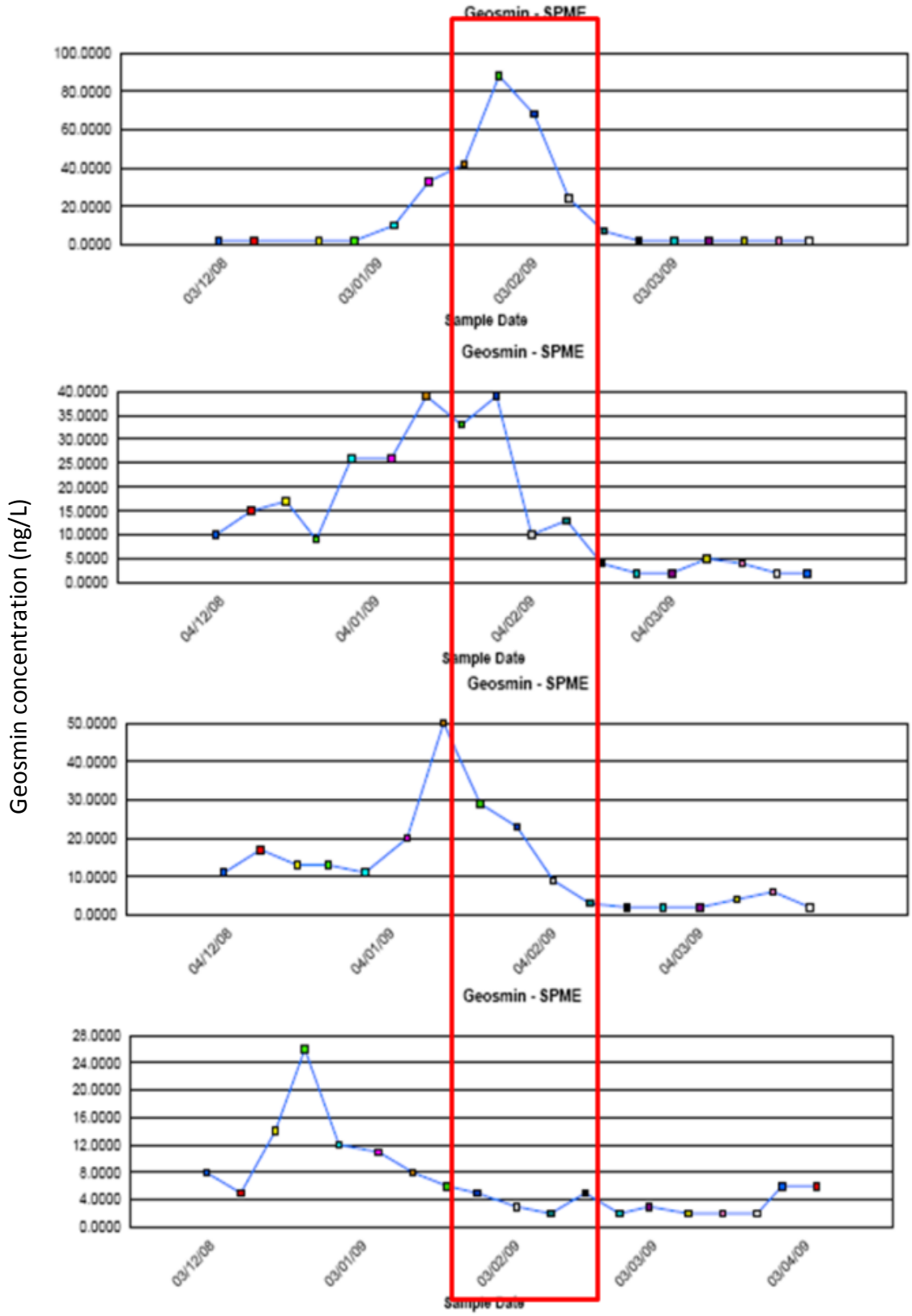


Figure 6-6. Geosmin levels at, from top to bottom, Cobdogla, Cadell, Morgan, and Lock 1 from December 2008 to March 2009. Sampling period is boxed in red.

6.4.3.2 Generation of a geosmin synthase standard curve

The concentration of purified G-PCR2 PCR product was 72 ng/µl, equivalent to 3.22 x 10^9 copies/µl. The PCR product was subsequently serially diluted in nuclease free water and added to the PCR reaction to cover the range from1.61 x 10^8 to 1.61 x 10^1 copies/reaction. All reactions were performed in triplicate and the standard curve generated as below. The standard error bars are shown in red, and the standard deviations ranged from 0.02 – 0.11 C_T, with a correlation coefficient (*r*) of 0.998.

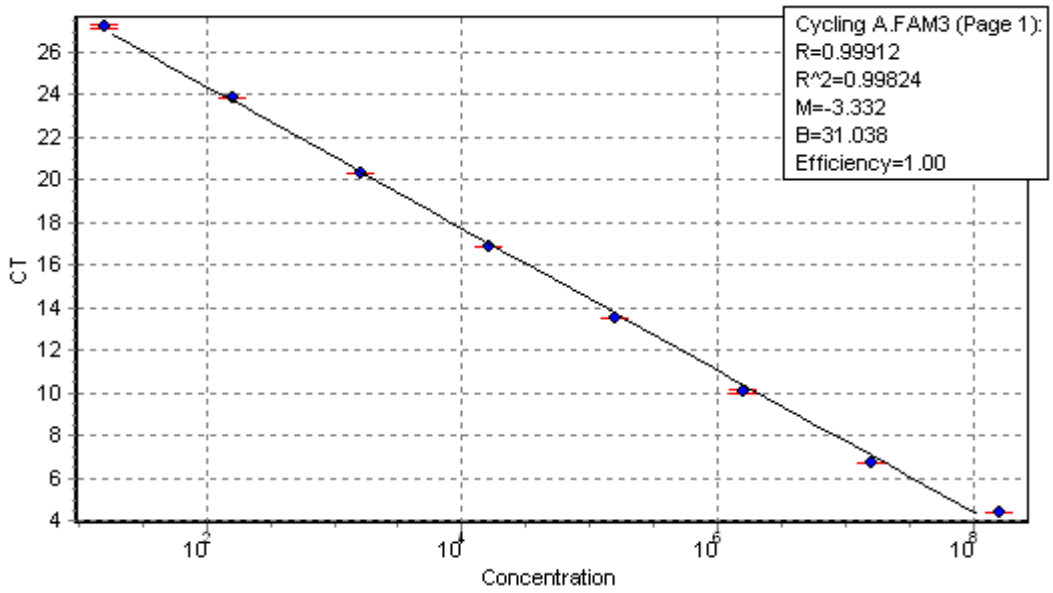


Figure 6-7. Standard curve for the G-PCR2 assay.

6.4.3.3 Analysis of River Murray samples

6.4.3.3.1 Melting curve analysis

The majority of samples analysed were positive for the presence of geosmin synthase as determined by the detection of a specific amplicon (corresponding to an electrophoresis band of 288 bp, data not shown), with an associated melting curve profile and *T*m of 87.5 °C as shown in Figure 6-8. The same *T*m for all samples indicated that there was likely to be only one species of geosmin-producing organism present (confirmed from the microscopy of samples that did not demonstrate other species of usual geosmin-producing organisms other than *A. circinalis*). It is also possible that the *T*m represented a combination of geosmin producing strains of *A. circinalis*, although this would be difficult to verify by microscopy.Givent hat the

controls experiments were similar between experiments, the difference between the *T*ms is likely to be represent genuine a DNA sequence variation from the control *A. circinalis.*

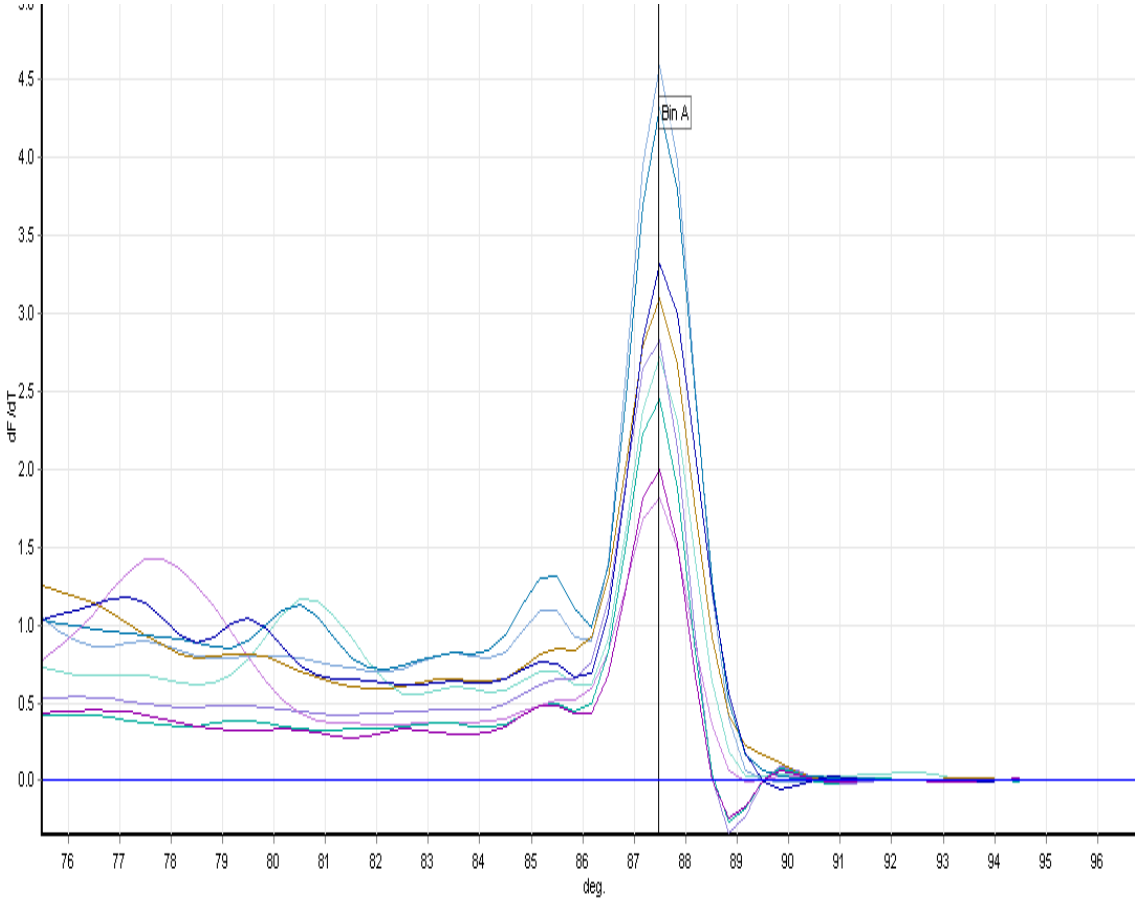


Figure 6-8. Geosmin synthase melting curve profile of positive River Murray samples with a single *T*m of 87.5 °C.

6.4.3.3.2 Quantification of geosmin synthase

Samples that displayed a specific melting profile were quantified against the standard curve generated in 6.4.3.2. Results of the real-time PCR geosmin synthase (copies/mL) and corresponding geosmin levels and *Anabaena* counts are summarised and presented in Figures 6-9 to 6-12.

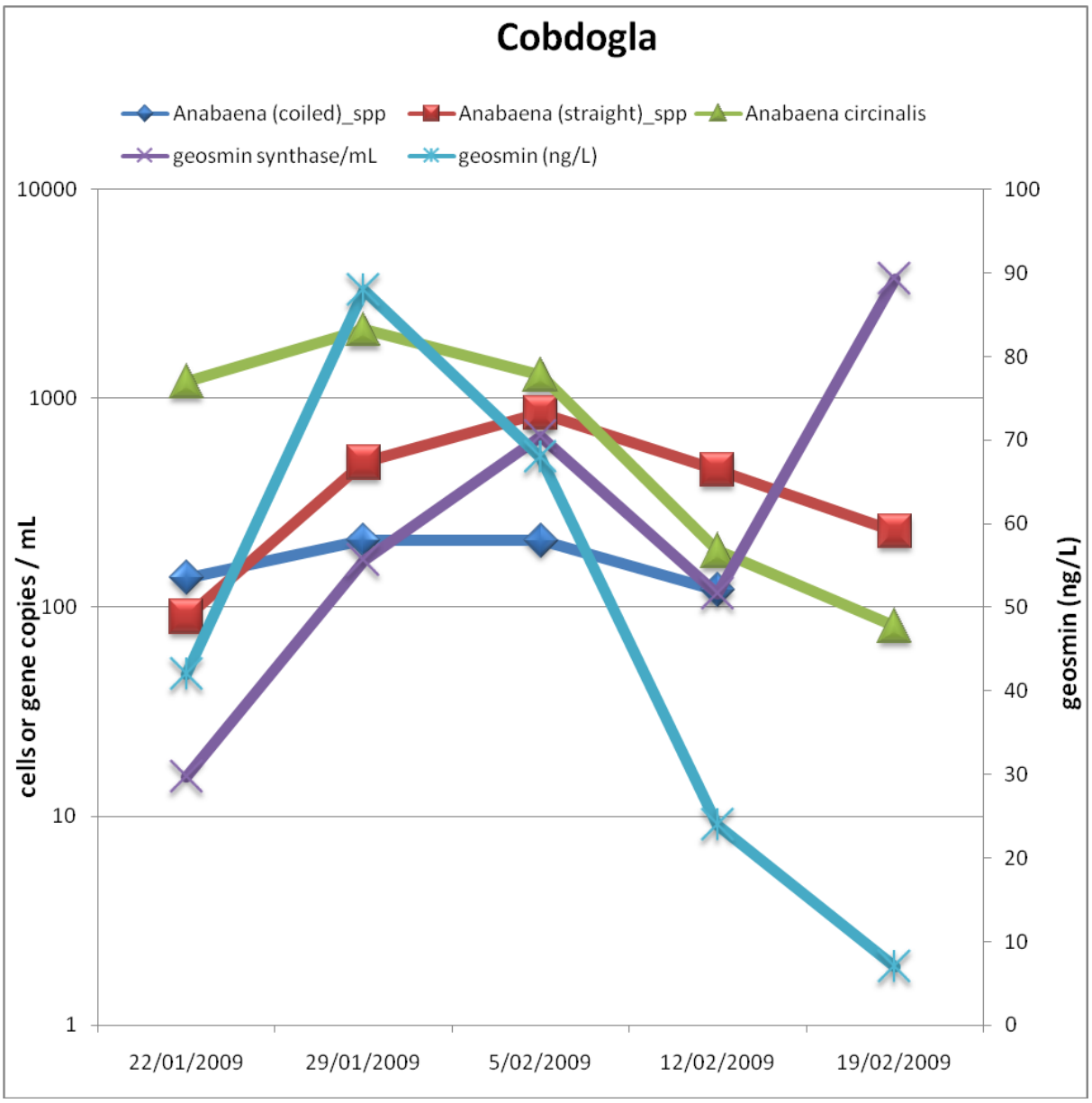


Figure 6-9. Comparison of *Anabaena* counts, geosmin levels, and geosmin synthase copies/mL from Cobdogla.

Data obtained from Cobdogla showed that the trend of geosmin concentraion followed the general trend of the *Anabaena*, rising from 22/1/09 to 29/1/06, then declining to low levels. Examination of the geosmin synthase gene levels does not follow this trend specifically, but the detection of geosmin synthase gene corresponds to the presence of geosmin in the samples, most likely due to the *Anabaena* present. While several other cyanobacteria were present in these samples (e.g. other *Anabaena* sp. and *Aphanizomenon* sp.) the presence of other geosmin-producing cyanobacteria, especially benthic cyanobacteria, were not specifically sought after in this or other samples in this study. *Anabaena circinalis* has historically been responsible for the increase in geosmin levels seen at all the sites investigated in this study, and

it is unlikely that the levels of geosmin in the water are the result of any benthic species, particularly as these waters are relatively turbid.

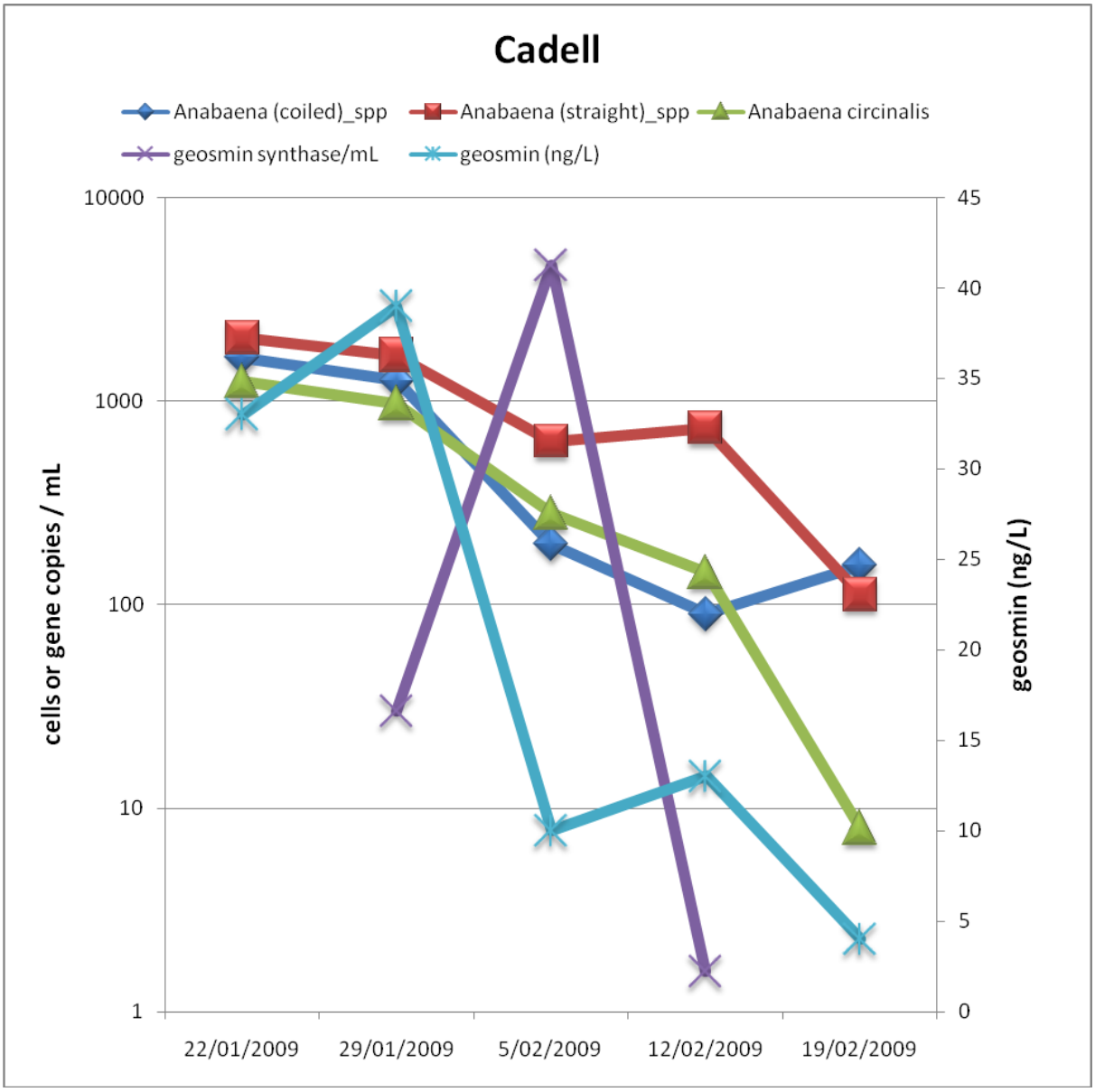


Figure 6-10. Comparison of *Anabaena* counts, geosmin levels, and geosmin synthase copies/mL from Cadell.

When investigating the Cadell location, aside from the failure to detect geosmin synthase genes on 22/1/09 and very low detectable signal on 19/2/09 (calculated equivalent value <1, result not plotted), the general downward trend of geosmin synthase gene levels was similar to the decline in geosmin and *Anabaena* levels. Again, however, the use of G-PCR2 to detect geosmin-producing organisms was demonstrated in waters with operationally significant geosmin concentrations.

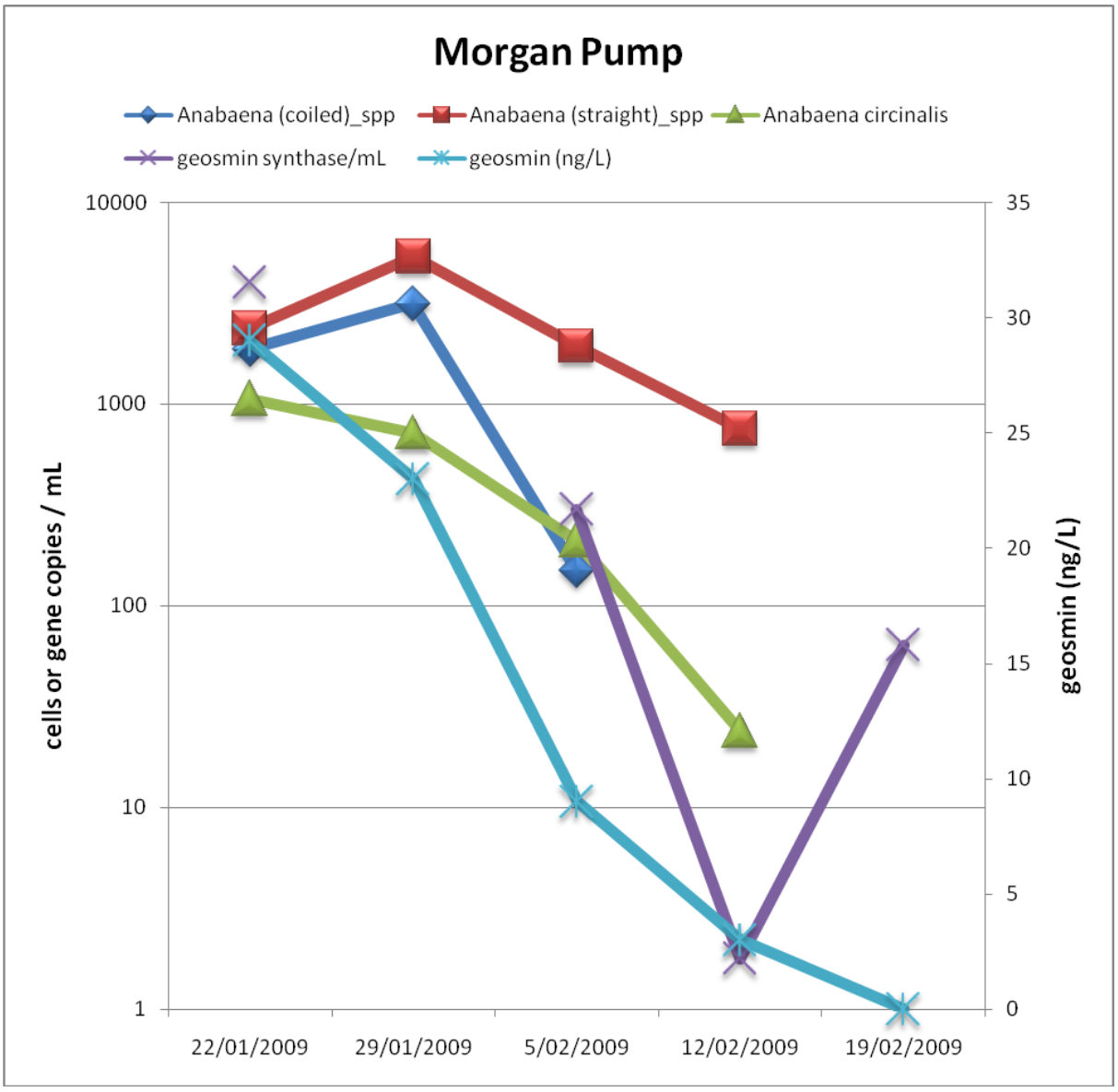


Figure 6-11. Comparison of *Anabaena* counts, geosmin levels, and geosmin synthase copies/mL from Morgan.

Analysis of the samples from Morgan was similar to Cadell, in that apart from the failure to detect geosmin synthase genes on 29/1/09, the general trend follows both geosmin and *Anabaena* levels.

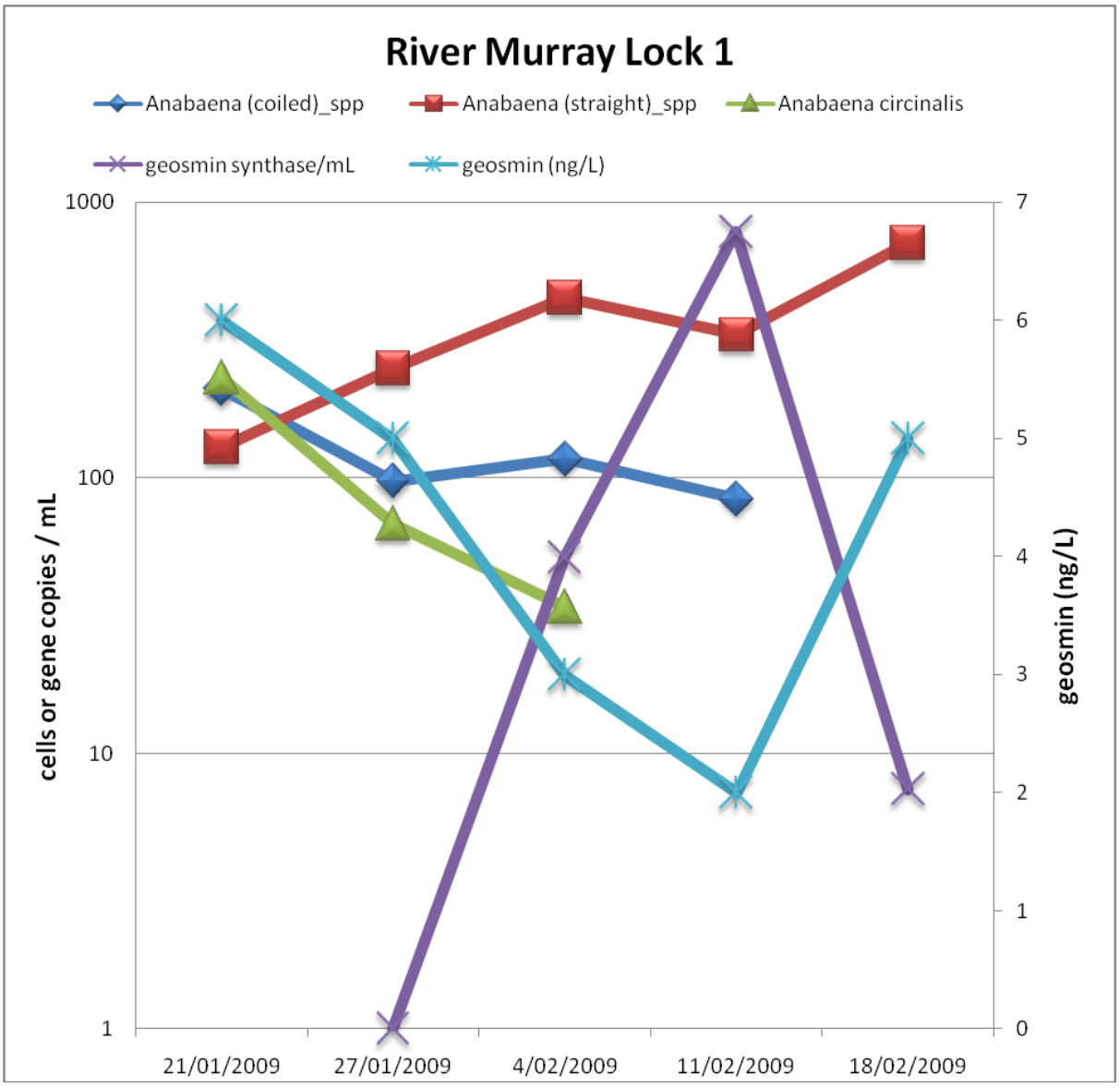


Figure 6-12. Comparison of *Anabaena* counts, geosmin levels, and geosmin synthase copies / mL from Lock 1.

Lock 1 was a site where the levels of geosmin ranged from <2 – 6 ng/L, a range that is at, or close to, the limit of detection of standard GC-MS analyses for detection of geosmin in waters. There were generally lower levels of *Anabaena*, specifically *A. circinalis*, but G-PCR2 was still able to detect the presence of geosmin synthase genes in 4 out of the 5 samples tested.

The apparent poor-moderate correlation between geosmin synthase gene levels and *A. circinalis* (assuming this is the only geosmin producing organism in these samples (historically this is the case at these sampling points)) in all locations may have resulted for several reasons. Firstly, the DNA extraction method may have resulted in the loss of samples through this process as several geosmin synthase levels (copies/mL) were lower than the *A. circinalis* cell count. Assuming a ratio of 1 geosmin synthase gene per cell (as a minimum), the numbers should be at least equal. Several other commercially available DNA extraction kits are available, and a systematic evaluation of the efficiency and quality of DNA extraction from these kits may prove beneficial for the accurate quantification of geosmin synthase gene levels if required. Secondly, there may have been mixed populations of *A. circinalis* strains, some of which produce geosmin and some that do not. If this is the case a general relationship between genes/mL and geosmin/L may not exist, as the assumption that there is a gene-dosage effect has not been investigated for geosmin production (this investigation is now a possibility using the tools arising from this thesis). Recent work conducted in our laboratory to compare toxin gene number with cell number and toxin production have demonstrated complex relationships, where the number of genes per cell and the amount of toxin per cell appears to be related to the growth phase of the cell. The increase in genes per cell is during stationary phase whereas the ratio of genes per cell is lower during lag and logathmic phases; the production of toxin per cell is highest in cells during logaroithmic phase(Monis *et al*, unpublished). It may be the case that the amount of geosmin produced per cell may or may not follow a direct relationship, as demionstrated by recent studies for toxin production, and indeed warrants further investigation.

A third reason for the poor to moderate relationship between geosmin synthase gene / mL is the age of the *A. circinalis* cells themselves. It is unlikely that the population of *Anabaena* will be a) synchronous or b) clonal, resulting in a distribution of cells with varied susceptibility to extraction methods. It is likely that senescent cells are more susceptible to lysis methods than younger more robust cells. Fourthly, we must ask ourselves “should we indeed expect a correlation between genes/mL and geosmin concentration?”. Several unexplained questions still remain, 1) is the expression of geosmin synthase the same in all strains of cyanobacteria, 2) are there different factors that affect the expression of geosmin synthase when *in vitro* or *in enviro*, 3) what is the stability of the enzyme itself, and is this consistent amongst all species and strains, 4) is there post transcription and/or post translational control? These unanswered

questions certainly present scope for work to continue in this area, and may provide clues into why some strains are so called "super-producers" of geosmin (and/or MIB).

From a water treatment perspective, detection at this low level may be operationally significant, as early detection and confirmation of organisms that have the ability to produce geosmin may alert operators to pending problems, and confirm suspicions that suspected cyanobacteria are in fact producing geosmin, that if allowed to bloom, would most certainly cause a taste and odour episode. This in turn may open up several opportunities to exploit infrastructure present to avoid potential blooms of geosmin-producing cyanobacteria, or at the very least, alert operators to perhaps increase surveillance of these waters for problematic organisms. Whilst in the final stages of writing this dissertation, this exact point was highlighted when counts from a water storage in South Australia revealed an unusual morphotype of a "narrow" *Phormidium* sp. at a level of 420 cells/mL. G-PCR2 confirmed that this organism was a geosmin-producer, thus alerting operators that this morphotype may indeed be problematic should it reach sufficient numbers.

Investigations into taste and odour episodes usually focus on planktonic cyanobacteria, and the search for a usual culprit that consistently causes seasonal problems in any given water. While this may be a safe assumption in most cases, sometimes this complacency may cause the focus to be turned away from other possible sources, and elucidation of the actual source of the problem in some cases may remain unanswered or questionable. For example two episodes have been described where the levels of geosmin seen in waters show little consistent relationship with known planktonic cyanobacterial geosmin-producers (Juttner and Watson, 2007). Firstly, in Lake Ontario, geosmin concentration peaked at 200 ng/L, and correlated poorly with the abundance of *Anabaena lammermannii*, and secondly in Tulsa, geosmin peaks of up to 2000 ng/L correlated poorly to the normally dominant *Anabaena circinalis*. The authors continue to point out that specialised microscopists are required to correctly identity changing taxa of cyanobacteria, and wrong identification may lead to the assumption that incorrectly classified cyanobacteria, usually suspected of being geosmin producers, may not be geosmin-producing cyanobacteria at all. In such a case, if planktonic cyanobacteria present were geosmin-producers, the G-PCR2 assay would be beneficial in confirming that a species present, perhaps not necessarily the dominant species, can produce geosmin; moreover, it may demonstrate that the dominant species is <u>not</u> a geosmin producer.

Because the sensitivity of PCR is such that it could detect 1-10 cells per PCR reaction, an organism can be isolated from surrounding cyanobacteria by micro-pipetting, and G-PCR2 may be used to detect geosmin synthase and confirm or disprove that it is a geosmin-producer. Similarly, this method can be applied to all species within a sample, and geosmin-producing organisms identified by fusing microscopy and geosmin synthase PCR. Such a method would be invaluable to researchers, and extremely cost and time-effective when compared to isolating and culturing the cyanobacteria, and testing isolates for geosmin by GC-MS.

The application for detecting geosmin-producing cyanobacteria is not limited to potable or source waters. Several fresh water aquaculture farms suffer from high levels of geosmin produced from cyanobacteria in fish farms (Martin *et al*., 1991, Van der Ploeg and Boyd, 1991, Schrader and Dennis, 2005). Geosmin taints the flesh of the fish, making them unfit for sale, or requiring them to be purged in a fresh pond at considerable expense and time of the farmer. The impact of geosmin in these farms is economically significant, as the loss of fish and additional treatment required has been estimated to be approximately 60 million dollars (Schrader *et al*., 2003). Again, early detection of geosmin-producing cyanobacteria using the presence of the geosmin synthase gene would seem desirable, thus allow mitigating strategies to be deployed in this important industry also.

6.5 SUMMARY

In this chapter, the presence of geosmin synthase homologs via G-PCR1 and G-PCR2 (using the highly conserved N-terminal domain as a target) was demonstrated in geosmin-producing cyanobacteria. The geosmin synthase gene was detected in several cyanobacterial genera (*Anabaena*, *Geitlerinema*, *Phormidium*, and *Nostoc*) suggesting the utilisation of these assays may be beneficial for a) confirming the presence of geosmin-producing cyanobacteria in culture, and b) confirming the presence of geosmin-producing cyanobacteria in waters.

Application of the real-time assay, G-PCR2, to River Murray waters showed that the assay detected geosmin synthase genes in waters where *Anabaena* sp. was present. The levels of *Anabaena* sp also correlated with the concentration of geosmin in these waters, suggesting a strong relationship between the two, as has been the case historically in these waters. Although the numbers of geosmin synthase genes/mL showed poor to moderate correlation with *Anabaena* cell counts and geosmin concentration, the utility of this assay to detect low levels of

geosmin-producing *Anabaena,* as demonstrated here, is likely to be operationally significant. Nonetheless, further modification and validation of this assay will be necessary before implementing this technology as a routine operational tool. Firstly, a robust DNA extraction method needs to be evaluated, particularly to determine the losses during the extraction process and ability to deal with PCR inhibiting substances such as humic and fulvic substances commonly found in waters. Secondly, it would be desirable for a larger suite of isolates to be screened with G-PCR1 to establish a database of the N-terminal sequences of geosmin synthases. By doing so, the reliability of G-PCR2 can be assessed across a broader range of species and genera, and, furthermore, whether melt curve analysis using G-PCR2 will be useful in discrimination of geosmin-producing organisms. Thirdly, an internal control for each PCR reaction should be incorporated into any routine application of G-PCR2.

Given the demonstrated mobility of diagnostic PCR testing in environmental samples, it is foreseeable that monitoring for geosmin synthase genes may follow suit. With the ease and speed of PCR, a cost efficient and timely tool is now available for investigation into taste and odours – either directly from a water sample, or directly from an isolate. At the very least, PCR may now be used to screen isolates to confirm their ability to produce geosmin using laboratory based PCR instruments, rather than rely upon GC-MS, thus allowing higher throughput, allowing confirmation of isolates from less material than needed for GC-MS, and also reducing costs. It also allows scope to mobilise this technology into the field, using robust and field-specific PCR instrumentation for an even quicker answer. With additional work to refine and optimise DNA extraction techniques, monitoring of waters for geosmin synthase genes may provide a powerful predictive tool for the occurrence of taste and odour episodes. At any given water storage facility the developed technology may be used to a) predict what preventative measures may be necessary in the short term; b) track the emergence of taste and odour producers in the longer term; and c) anticipate what trends may occur in the future. Although proof of concept has been demonstrated here, the routine implementation of such an assay will of course require a detailed and systematic investigation of waters over a longer period of time than studied here, but the proposed utility of this tool may be of great interest to catchment managers.

By identifying and sequencing the genes involved in geosmin production, this work has provided a fundamental step forward for developing a more exhaustive and robust PCR assay,

and will allow any newer nucleic acid-based detection system (microarrays, oligoprobe-based nanotechnology *etc*), and alternative biochemical test such as ELISA, to be realised. Such options for monitoring waters have not yet been explored, but it is exciting to imagine the utilisation of some of these state-of-the-art techniques for the detection of geosmin synthase genes in our waters.

Chapter 7

SUMMARY AND DISCUSSION

From the point of ignition

To the final drive

The point of a journey

Is not to arrive

Anything can happen.....

Prime Mover, Rush

7.1 SUMMARY

Consumer confidence in water utilities is undermined when the supplied water is either visually unappealing or the taste and/or odour is unpalatable. One of the most common complaints to water utilities regarding water quality is that of an earthy taste and odour caused by the presence of geosmin; the associated political and social costs of a severe taste and odour event can be great. Water treatment and the provision of safe, clean, and taste-free water have made significant advances over the last few decades. As utilities become more cost and environmentally aware, efforts are now being directed into better management techniques at the source and catchment areas to prevent taste and odour episodes, rather than treat them when they arise. Fundamental scientific questions often need to be answered to provide a better understanding of the issues at hand, and to generate tools to detect potential problems before they occur. In this respect, this dissertation has detailed the fundamental mechanism involved in geosmin synthesis in cyanobacteria at the molecular level, and has provided a novel and rapid tool to determine the likelihood of a cyanobacterium becoming (or being) a taste and odour problem.

It is only in the last few years, and through the course of this work, that our knowledge of geosmin biosynthesis in organisms has matured. Relatively little work in this field has been performed since the 1980s, but when the complete genome of *Streptomyces coelicolor* A3(2) became available, our knowledge rapidly advanced. After some initial confusion and the postulation of several alternative pathways of geosmin synthesis in different bacteria, it was demonstrated that SCO6073 in *Streptomyces coelicolor* A3(2) was both necessary and sufficient to synthesise geosmin. SCO6073 encoded a unique bi-functional terpene synthase made up of two similar simple terpene synthase halves; the N-terminal domain was shown to catalyse the conversion of FPP to germacradienol and germacrene D, whilst the C-terminal domain completed the conversion of germacradienol to geosmin.

In Chapter 3, a N-terminal gene fragment homologous to SCO6073 was identified in *Nostoc punctiforme* ATCC 29133 (PCC 73102), *npun02003620,* encoding for the hypothetical protein ZP_00109187. Whilst this protein had good similarity to the SCO6073 N-terminal domain, it lacked similarity in the C-terminal domain; firstly it lacked an essential NSE triad, and secondly, it was some 100 aa shorter than SCO6073. Protein ZP_00109187 was shown to catalyse the conversion of FPP to germacradienol and germacrene D, without any geosmin

production. A vital sequencing error was discovered in the deposited *npun02003620* sequence, and once corrected, the extended sequence harboured a C-terminal NSE triad and was similar in length to SCO6073 and other putative geosmin synthases. The resultant protein from the corrected sequence, dubbed NPUNMOD, was investigated in Chapter 4 and shown to be a fully functional geosmin synthase, the first reported geosmin synthase from a cyanobacterium.

The demonstration of this bi-functional geosmin synthase in cyanobacteria confirmed a common precursor and common pathway for geosmin synthesis in *Streptomyces* and *Nostoc*. The earlier assumptions alluding to different possible mechanisms for geosmin synthesis therefore remains an unlikely prospect. Moreover, given the conservation of the MEP pathway in eubacteria, it is unlikely there is any other alternative mechanism for isoprenoid or geosmin synthesis. Current evidence suggests that the presence of geosmin synthases in eukaryotic organisms is likely to be similar, except that a phylogenetically divergent suite of isoprenoid genes is present.

Chapters 3 and 4 conclusively demonstrated the "how" of geosmin synthesis in cyanobacteria. An insight into the "why or when" was the purpose of gene expression studies in Chapter 5. Previous studies had attempted to gain an understanding of geosmin production by monitoring physical and chemical parameters, such as chl *a* and geosmin concentration, with the technology available at that time. With the findings presented in this dissertation, the examination of geosmin synthase gene expression in cyanobacteria became possible. An investigation into gene expression was performed with *Anabaena circinalis* by measuring geosmin synthase gene expression as well as geosmin produced from the organisms in response to light conditions, at a constant temperature. The experimental data suggested that the expression of geosmin synthase is constitutive and is a function of growth. Of particular interest was the detection of increasing intracellular geosmin concentrations when apparent gene expression was significantly down regulated (during stationary phase and in the absence of light), suggesting that the geosmin synthase enzyme continues to be functional for several days after translation.

In Chapter 6, the presence of putative geosmin synthases was demonstrated in several geosmin-producing cyanobacteria including *Anabaena, Nostoc, Phormidium*, and *Geitlerinema,* using PCRs directed to highly conserved DDHFLE and NSE motifs in the N-

terminal domain, and sequence data for these PCR fragments was determined. Furthermore, geosmin synthase was not detected in several strains of *Oscillatoria, Pseudanabaena*, and *Planktothrix* that did not produce geosmin. Together, these results suggest that cyanobacteria with an intact geosmin synthase gene always produce geosmin and that cyanobacteria that do not produce geosmin will not have this gene present. Throughout the work presented in this dissertation, there was no instance where a geosmin synthase homolog was detected in a cyanobacterium that does not produce geosmin, nor was there any example of geosmin-producing organisms that did not have a detectable geosmin synthase gene by PCR. The G-PCR2 real-time PCR shows promise as a screening tool for the detection, and possible species identification, of geosmin-producing organisms using DNA melting curve analysis. When G-PCR2 was used to monitor an event during which increases in geosmin concentration occurred in a section of the River Murray, the quantification of geosmin synthase genes directly from the water samples moderately correlated with corresponding geosmin concentration and *Anabaena circinalis* levels. The utility of G-PCR2, therefore, shows promise as an adjunct to current monitoring regimes.

7.2 DISCUSSION AND FUTURE WORKS

"You can't teach an old dog new tricks". We are fortunate that in the water industry, even though there are some old dogs (with all due respect!), new tricks are welcomed; indeed new tricks are needed. Such new tricks include utilising the abundance of molecular tools now available to applied scientists. At the moment, the momentum for implementing these new methodologies is beginning to reach a critical mass, and it is hoped by this author that molecular tools will become almost routine in the near future. Such is the case in the clinical microbiology laboratory setting, where definitive and diagnostic PCRs are used to diagnose disease and monitor treatment efficacy, and to ascertain the likelihood of success of various treatments.

A new trick realised by the work presented in this dissertation is the detection of the geosmin synthase gene in cyanobacteria using real-time PCR. Although the use of this technique for routine monitoring or predicting the emergence of blooms is foreseeable in the future, it lends itself as being immediately applicable as part of a suite of routine tools to examine waters when required to investigate taste and odour related issues. The use of real-time PCR for "diagnosis" of geosmin-producing cyanobacteria is advantageous for several

reasons; it is a mobile technology, it is relatively cost efficient, and it is rapid. Screening of waters/isolates/filaments/cells can be done in bulk, and there are now automated additions to several real-time PCR platforms that are "walk away" in nature, requiring only sample loading and result interpretation.

A proposed method for the use of PCR in taste and odour investigations is presented in Figure 7-1. The current mentality for investigating taste and odour episodes involves looking for the usual suspect cyanobacteria by microscopy in any given water body at a time of increased geosmin concentration in the corresponding water sample. Current mentality also dictates that the most prevalent organism is likely to be the cause of a taste and odour issue, but this is not always the case (depicted in Figure 7-1, yellow box, negative PCR result with a positive GC-MS suggests that uni-algal species evident in microscopy is not a geosmin-producing organism). The use of real-time PCR can be used as an adjunct to microscopy, and can be performed directly on water samples to determine the potential of organisms present to produce geosmin, or directly from cyanobacterial cells micro-manipulated from the same sample to determine *exactly* which cyanobacterium is a geosmin producer.

If, in a uni-algal water sample, there was a positive geosmin synthase PCR in the absence of a detectable GC-MS positive geosmin result, this would present us with several questions. Is the GC-MS result correct? Is the geosmin synthase PCR result correct? Whilst we would like to believe there are seldom errors in either analyses, or errors in sample receipt or processing, the combined results provide conflicting interpretations in regards to the status of the water body. It is pertinent, therefore, that repeat testing or investigations into this discrepancy occur. It may be the case that the GC-MS result was erroneous, and a potential taste and odour event may have been missed, if it was not for the accompanying geosmin synthase PCR suggesting otherwise. Figure 7-1 describes several of the possible scenarios and suggests interpretations/courses of actions that may be required when introducing geosmin synthase PCR as an adjunct to current monitoring programs. Although there are likely to be other possible scenarios not detailed in Figure 7-1, it is hoped that this diagram serves as template for the interpretation of results, if introducing PCR into the modern day water testing laboratory.

One may suggest that actinomycetes potentially stuck to the mucilage of cyanobacteria may result in a false positive interpretation of a positive G-PCR2, where in fact the actinomycete

is a geosmin producer and the cyanobacterium in question is not. While this is theoretically possible (probably more so in benthic cyanobacteria and their close association with sediments where actinomycetes reside), there are no published reports describing the prevalence of this association, or possible impact on analyses. As new geosmin synthase sequence data from bacteria and cyanobacteria become more plentiful, PCRs will become more refined, more stringent, and more likely capable of distinguishing between cyanobacterial and actinomycete geosmin synthases. More specifically, given that actinomycetes generally have a higher G+C DNA content than cyanobacteria, it should be possible to resolve any differences in geosmin synthases by targeting conserved regions with defined differences in codon usage with regions of the gene. The gathering of such sequence data, therefore, should be a focus for future work to avoid possible questions relating to the source of the detected geosmin synthase gene, if required.

Although there is interest to adopt these new tricks, there is a reluctance amongst many utilities or testing laboratories to implement these newer technologies. This reluctance is two-fold; the first is related to the cost of validating newer techniques and purchasing capital equipment, the second is related to moving the goal post for operationally significant parameters on which to act. Currently for SA Water, critical alert levels for *Anabaena circinalis* in reservoirs are set at 2,000 cells/mL. When this value is exceeded, an increased monitoring event is triggered, and the organisation is on standby for a bloom response. The question operators pose is how would a reported geosmin synthase count of 2,000 geosmin synthase genes/mL affect these alert levels? How should utilities react to this level? What does this level *mean*? Unfortunately, there is no simple answer. The lack of simplicity to make a newly reported parameter an operational trigger is concerning to all stakeholders involved. Even more concerning is the task involved with redefining operational triggers to align with current technologies, trends, and practices. If newer technologies are to be implemented, and used to produce data that provides operationally significant trigger levels, efforts must be made to trial and validate these technologies in parallel to current methodologies. It would be germane to all water industry personnel to embrace these technologies, and trial them for what they are worth – as PCR in the environmental management setting is truly an unknown entity, with enormous potential to rapidly detect and quantify nuisance organisms and allow timely remedial actions. Granted, there may be sensitivity issues that require investigations into sample concentration methodologies, as detection of clinical pathogens is usually at high numbers compared to the water industry,

where the detection of 1 organism per 100 mL or per litre may be required. However, if the clinical pathology environment is able to use PCR as a life or death decision making tool, surely its use in water management is achievable in the near future?

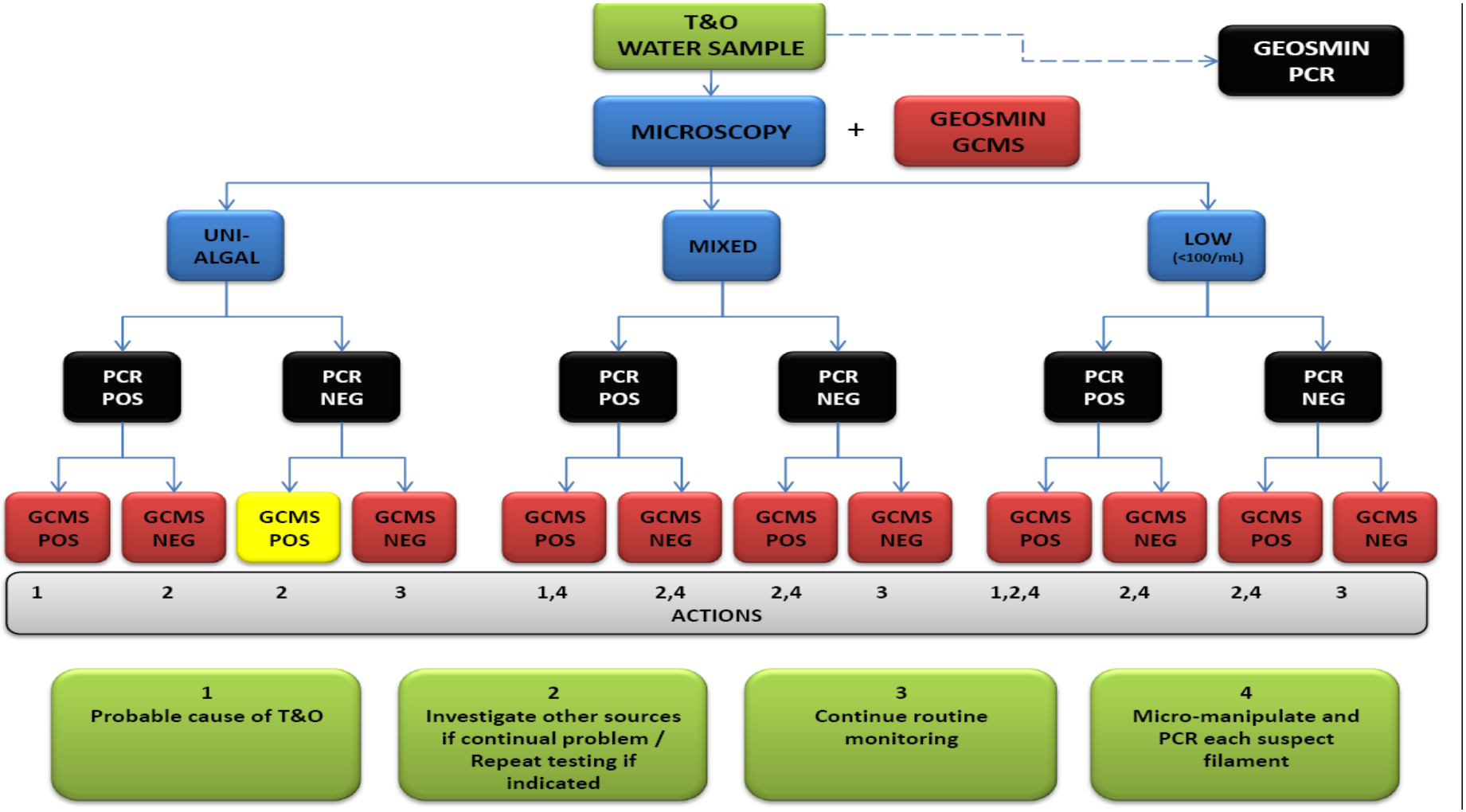
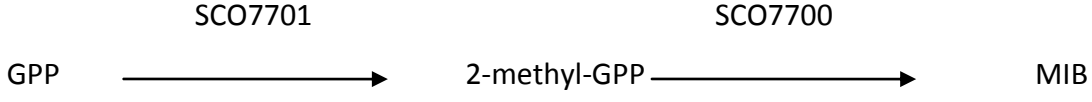


Figure 7-1. Schematic for the proposed inclusion of geosmin synthase PCR in monitoring of waters in T&O events

Further work directed towards the examination of the expression of the geosmin synthase gene and synthesis of the protein would present an opportunity to explore the "why or when" of geosmin synthesis. One way of exploring this question is through the use of deletion mutants of geosmin synthase in *Nostoc punctiforme* PCC 73102 (ATCC 29133) and comparison with wild type under various conditions, as has been previously described (Soule *et al*., 2007). By subjecting both strains to similar conditions, the role of geosmin synthesis in the survival (long term or short term) may become apparent when employing a systematic investigation into exploring the response of mutants when subjected to challenging stimuli.

A priority for future research should be microarray investigation of *Nostoc punctiforme* ATCC 29133 (PCC 73102) under conditions to test geosmin production hypotheses such as the effect of light and temperature, or available (and/or limiting) nitrogen and phosphorous. Microarray analyses on *Nostoc* will provide a relatively rapid assessment of global transcription patterns under defined experimental conditions, and will allow genome-wide observations of the response of geosmin synthase gene expression with respect to other regulatory systems. Moreover, it will allow direct comparison of geosmin synthase gene expression as a function of other isoprenoid genes, and will determine if there is a co-ordinated regulation of isoprenoid synthesis in response to stimuli, or conditions conducive to cell senescence.

A natural extension of the work presented here is to determine if the genes responsible for MIB production in *Streptomyces* are similar in cyanobacteria. Only recently have advances been made in this field with the finding that SCO7701 (a C-methlytransferase) and SCO7700 (a terpene synthase) form a 2 gene operon responsible for the conversion of GPP to MIB via the reaction below (Komatsu *et al*., 2008, Wang and Cane, 2008),



Given the confirmation of a similar biosynthetic pathway for geosmin synthesis in both *Streptomyces* and cyanobacteria, it is highly likely that a similar mechanism for MIB synthesis would be present in cyanobacteria. Currently, however, there is no sequenced MIB-producing cyanobacterium, making initial observations into this hypothesis extremely

difficult. Nonetheless, a concerted effort to explore this hypothesis would similarly be extremely useful for the water industry.

Perhaps one of the greatest questions in taste and odour biology is “what is the function of geosmin”, and continues to invoke imagination and speculation amongst academics, researchers, and industry people alike. What function does geosmin have? While there is no direct evidence for the role of geosmin produced by cyanobacteria, in other instances it is hypothesised that geosmin may act as a land mass attractant for marine *Anguilla* eels, which prefer to breed in waters with decreased salinity levels brought about by terrestrial run offs where fresh water and soil, and hence geosmin, are deposited at an estuary’s edge (Tosi and Sola, 1993). It has also been suggested that geosmin produced by actinomycetes in desert waters may act as an attractant to camels, which are attracted to the earthy smell of geosmin, and thus propagate actinomycete spores in their further travels (Personal communication, Sir David Hopwood to David Cane).

With water perhaps being our most valuable modern day commodity, additional tools in the taste and odour “toolbox” (such as geosmin/MIB/toxin gene PCRs) would be a valuable resource to proactively manage our waters so that future taste and odour and toxic episodes are minimised. The findings of the research presented in this dissertation provide a pivotal point from which to expand and explore the use of some of the new tricks described for use in monitoring of waters, and to broaden our knowledge into the “why” of geosmin production by cyanobacteria. The production of secondary metabolites by cyanobacteria is metabolically burdensome, and it is unlikely that such efforts from a cell would be without need - understanding the “why” of geosmin production may answer this obstinate question, and aid water utilities in the management of this persistent problem in the future.

BIBLIOGRAPHY

ADAMS, D. G. & CARR, N. G. (1981) The developmental biology of heterocyst and akinete formation in cyanobacteria. *Crit Rev Microbiol,* 9**,** 45-100.

ADWG (2004) Australian Drinking Water Guidelines. *National Health and Medical Research Council*.

AGGER, S. A., LOPEZ-GALLEGO, F., HOYE, T. R. & SCHMIDT-DANNERT, C. (2008) Identification of sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc sp*. strain PCC 7120. *J Bacteriol,* 190**,** 6084-96.

AJIKUMAR, P. K., TYO, K., CARLSEN, S., MUCHA, O., PHON, T. H. & STEPHANOPOULOS, G. (2008) Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol Pharm,* 5**,** 167-90.

AL-THUKAIR, A. A., ABED, R. M. & MOHAMED, L. (2007) Microbial community of cyanobacteria mats in the intertidal zone of oil-polluted coast of Saudi Arabia. *Mar Pollut Bull,* 54**,** 173-9.

ALLAKHVERDIEV, S. I., KLIMOV, V. V. & HAGEMANN, M. (2005) Cellular energization protects the photosynthetic machinery against salt-induced inactivation in *Synechococcus*. *Biochim Biophys Acta,* 1708**,** 201-8.

ALLAKHVERDIEV, S. I. & MURATA, N. (2008) Salt stress inhibits photosystems II and I in cyanobacteria. *Photosynth Res,* 98**,** 529-539.

APHA (1998) *Standard Methods for the Examination of Water and Wastewater,* Washington DC, American Public Health Association, American Water Works Association, and Water Environment Federation.

BAKER, R. J., SUFFET, I. H., ANSELME, C. & MALLEVIALLE, J. (1973) Evaluation of water treatment methods for removal of taste and odor causing compounds from drinking water. *AWWA Water Quality Technology Conference.*

BEHETS, J., DECLERCK, P., DELAEDT, Y., VERELST, L. & OLLEVIER, F. (2007) A duplex real-time PCR assay for the quantitative detection of *Naegleria fowleri* in water samples. *Water Res,* 41**,** 118-26.

BENTLEY, R. & MEGANATHAN, R. (1981) Geosmin and methylisoborneol biosynthesis in streptomycetes. Evidence for an isoprenoid pathway and its absence in non-differentiating isolates. *FEBS Lett,* 125**,** 220-2.

BERG, H., ZIEGLER, K., PIOTUKH, K., BAIER, K., LOCKAU, W. & VOLKMER-ENGERT, R. (2000) Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-poly-L-aspartic

acid (cyanophycin): mechanism of the cyanophycin synthetase reaction studied with synthetic primers. *Eur J Biochem,* 267**,** 5561-70.

BINDER, B. J. & CHISHOLM, S. W. (1995) Cell Cycle Regulation in Marine *Synechococcus sp*. Strains. *Appl Environ Microbiol,* 61**,** 708-717.

BLEVINS, W. T., SCHRADER, K. K. & SAADOUN, I. (1995) Comparative physiology of geosmin production by *Streptomyces halstedii* and *Anabaena* sp. *Wat Sci Tech,* 31**,** 127-133.

BLYTH, S. (1980) Palm Island mystery disease. *Med. J. Aust,* 2**,** 40-42.

BOISON, G., STEINGEN, C., STAL, L. J. & BOTHE, H. (2006) The rice field cyanobacteria *Anabaena azotica* and *Anabaena sp*. CH1 express vanadium-dependent nitrogenase. *Arch Microbiol,* 186**,** 367-76.

BOUCHER, Y. & DOOLITTLE, W. F. (2000) The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol Microbiol,* 37**,** 703-16.

BOWMER, K. H., PADOVAN, A., OLIVER, R. L., KORTH, W. & GANF, G. G. (1992) Physiology of Geosmin production by *Anabaena circinalis* isolated from the Murrumbidgee river, Australia. *Wat Sci Tech,* 25**,** 259-267.

BRASIER, M. D., GREEN, O. R., JEPHCOAT, A. P., KLEPPE, A. K., VAN KRANENDONK, M. J., LINDSAY, J. F., STEELE, A. & GRASSINEAU, N. V. (2002) Questioning the evidence for Earth's oldest fossils. *Nature,* 416**,** 76-81.

BROCK, T. D. (1976) Halophilic-blue-green algae. *Arch Microbiol,* 107**,** 109-11.

BRUCE, D., WESTERHOFF, P. & BRAWLEY-CHESWORTH, A. (2002) Removal of 2-methylisoborneol and geosmin in surface water treatment plants in Arizona. *Journal of Water Supply Research and Technology-Aqua,* 51**,** 183-197.

BURCH, M. D., BROOKES, J. D., TARRANT, P. & DELLAVERDE, P. (2000) Mixing it up with blue-green algae. *Water,* 21**,** 21-22.

BURLINGAME, G. A., DANN, R. M. & BROCK, G. L. (1986) A case study of geosmin in Philadelphia's water. *Journal AWWA***,** 56-61.

CAMPBELL, E. L., CHRISTMAN, H. & MEEKS, J. C. (2008) DNA microarray comparisons of plant factor- and nitrogen deprivation-induced hormogonia reveal decision-making transcriptional regulation patterns in *Nostoc punctiforme*. *J Bacteriol,* 190**,** 7382-91.

CAMPBELL, E. L., SUMMERS, M. L., CHRISTMAN, H., MARTIN, M. E. & MEEKS, J. C. (2007) Global gene expression patterns of *Nostoc punctiforme* in steady-state dinitrogen-grown heterocyst-containing cultures and at single time points during the differentiation of akinetes and hormogonia. *J Bacteriol,* 189**,** 5247-56.

CANE, D. E., HE, X., KOBAYASHI, S., OMURA, S. & IKEDA, H. (2006) Geosmin biosynthesis in *Streptomyces avermitilis*. Molecular cloning, expression, and mechanistic study of the germacradienol/geosmin synthase. *J Antibiot (Tokyo),* 59**,** 471-9.

CANE, D. E. & WATT, R. M. (2003) Expression and mechanistic analysis of a germacradienol synthase from *Streptomyces coelicolor* implicated in geosmin biosynthesis. *Proc Natl Acad Sci U S A,* 100**,** 1547-51.

CASTENHOLZ, R. W. (2001) Oxygenic photosynthetic bacteria. IN BOONE, D. R. & CASTENHOLZ, R. W. (Eds.) *Bergey's Manual of Systematic Bacteriology.* 2nd ed. New York, Springer-Verlag.

CASTENHOLZ, R. W. & WATERBURY, J. B. (1989) *Bergey's Manual of Systematic Bacteriology,* Baltimore, Williams and Wilkins.

CECENER, G., TUNCA, B., EGELI, U., KARADAG, M., VATAN, O., UZASLAN, E. & TOLUNAY, S. (2008) Mutation analysis of the FHIT gene in bronchoscopic specimens from patients with suspected lung cancer. *Tumori,* 94**,** 845-8.

CHINTALAPATI, S., PRAKASH, J. S., SINGH, A. K., OHTANI, S., SUZUKI, I., MURATA, N. & SHIVAJI, S. (2007) Desaturase genes in a psychrotolerant *Nostoc sp*. are constitutively expressed at low temperature. *Biochem Biophys Res Commun,* 362**,** 81-7.

CHOE, S. E., BOUTROS, M., MICHELSON, A. M., CHURCH, G. M. & HALFON, M. S. (2005) Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biol,* 6**,** R16.

CHRISTIANSON, D. W. (2006) Structural biology and chemistry of the terpenoid cyclases. *Chem Rev,* 106**,** 3412-42.

CODD, G. A., AZEVEDO, S. M. F. O., BAGCHI, S. N., BURCH, M. D., W.W., C., HARDING, W. R., KAYA, K. & UTKILEN, H. C. (2005) CYANONET. A Global Network for Cyanobacterial Bloom and Toxin Risk Management - Initial Situation Assessment and recommendations. *Technical Documents in Hydrology,* 76.

COOK, D., NEWCOMBE, G. & SZTAJNBOK, P. (2001) The application of powdered activated carbon for MIB and geosmin removal: predicting PAC doses in four raw waters. *Water Res,* 35**,** 1325-33.

DEWETTINCK, T., VAN HOUTTE, E., GEENENS, D., VAN HEGE, K. & VERSTRAETE, W. (2001) HACCP (Hazard Analysis and Critical Control Points) to guarantee safe water reuse and drinking water production--a case study. *Wat Sci Tec,* 43**,** 31-8.

DICKSCHAT, J. S., BODE, H. B., MAHMUD, T., MULLER, R. & SCHULZ, S. (2005) A novel type of geosmin biosynthesis in myxobacteria. *J Org Chem,* 70**,** 5174-82.

DICKSCHAT, J. S., NAWRATH, T., THIEL, V., KUNZE, B., MULLER, R. & SCHULZ, S. (2007) Biosynthesis of the off-flavor 2-methylisoborneol by the myxobacterium *Nannocystis exedens*. *Angew Chem Int Ed Engl,* 46**,** 8287-90.

DISCH, A., SCHWENDER, J., MULLER, C., LICHTENTHALER, H. K. & ROHMER, M. (1998) Distribution of the mevalonate and glyceraldehyde phosphate/pyruvate pathways for isoprenoid biosynthesis in unicellular algae and the cyanobacterium *Synechocystis* PCC 6714. *Biochem J,* 333 (Pt 2)**,** 381-8.

DONOVAN, E., UNICE, K., ROBERTS, J. D., HARRIS, M. & FINLEY, B. (2008) Risk of gastrointestinal disease associated with exposure to pathogens in the water of the Lower Passaic River. *Appl Environ Microbiol,* 74**,** 994-1003.

EISENBERG, J. N., SETO, E. Y., OLIVIERI, A. W. & SPEAR, R. C. (1996) Quantifying water pathogen risk in an epidemiological framework. *Risk Anal,* 16**,** 549-63.

ERSHOV, Y. V., GANTT, R. R., CUNNINGHAM JR, F. X., JR. & GANTT, E. (2002) Isoprenoid biosynthesis in *Synechocystis sp*. strain PCC6803 is stimulated by compounds of the pentose phosphate cycle but not by pyruvate or deoxyxylulose-5-phosphate. *J Bacteriol,* 184**,** 5045-51.

FERGUSSON, K. M. & SAINT, C. P. (2003) Multiplex PCR assay for *Cylindrospermopsis raciborskii* and cylindrospermopsin-producing cyanobacteria. *Environ Toxicol,* 18**,** 120-5.

FERNANDEZ-VALIENTE, E., QUESADA, A., HOWARD-WILLIAMS, C. & HAWES, I. (2001) N2-Fixation in cyanobacterial mats from ponds on the McMurdo Ice Shelf, Antarctica. *Microb Ecol,* 42**,** 338-349.

FITZGERALD, G. P. (1966) Use of potassium permanganate for control of problem algae. *J. Am. Water Works Assoc,* 58**,** 609-614.

FITZGERALD, G. P. & FAUST, S. L. (1963) Factors affecting the algicidal and algistatic properties of copper. *Appl Microbiol,* 11**,** 345-51.

FLESCH, G. & ROHMER, M. (1988) Prokaryotic hopanoids: the biosynthesis of the bacteriohopane skeleton. Formation of isoprenic units from two distinct acetate pools and a novel type of carbon/carbon linkage between a triterpene and D-ribose. *Eur J Biochem,* 175**,** 405-11.

FRISVAD, J. C., SAMSON, R. A., RASSING, B. R., VAN DER HORST, M. I., VAN RIJN, F. T. & STARK, J. (1997) *Penicillium discolor*, a new species from cheese, nuts and vegetables. *Antonie Van Leeuwenhoek,* 72**,** 119-26.

GAGNE, F., RIDAL, J., BLAISE, C. & BROWNLEE, B. (1999) Toxicological effects of geosmin and 2-methylisoborneol on rainbow trout hepatocytes. *Bulletin of Environmental Contamination and Toxicology,* 63**,** 174-180.

GERBER, N. N. (1979) Volatile substances from actinomycetes: their role in the odor pollution of water. *CRC Crit Rev Microbiol,* 7**,** 191-214.

GERBER, N. N. & LECHEVALIER, H. A. (1965) Geosmin, an earthly-smelling substance isolated from actinomycetes. *Appl Microbiol,* 13**,** 935-8.

GIULIETTI, A., OVERBERGH, L., VALCKX, D., DECALLONNE, B., BOUILLON, R. & MATHIEU, C. (2001) An overview of real-time quantitative PCR: applications to quantify cytokine gene expression. *Methods,* 25**,** 386-401.

GOWRISHANKAR, G., WINZEN, R., BOLLIG, F., GHEBREMEDHIN, B., REDICH, N., RITTER, B., RESCH, K., KRACHT, M. & HOLTMANN, H. (2005) Inhibition of mRNA deadenylation and degradation by ultraviolet light. *Biol Chem,* 386**,** 1287-93.

GUST, B., CHALLIS, G. L., FOWLER, K., KIESER, T. & CHATER, K. F. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A,* 100**,** 1541-6.

HAAG, A. L. (2007) Algae bloom again. *Nature,* 447**,** 520-1.

HAYES, S. J., HAYES, K. & ROBINSON, B. S. (1991) Geosmin as an odorous metabolite in cultures of a free-living amoeba, *Vannella* species (Gymnamoebia, Vannellidae). *J Protozool,* 38**,** 44-47.

HEDGES, S. B., CHEN, H., KUMAR, S., WANG, D. Y., THOMPSON, A. S. & WATANABE, H. (2001) A genomic timescale for the origin of eukaryotes. *BMC Evol Biol,* 1**,** 4.

HIGUCHI, R., FOCKLER, C., DOLLINGER, G. & WATSON, R. (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y),* 11**,** 1026-30.

HIHARA, Y., SONOIKE, K., KANEHISA, M. & IKEUCHI, M. (2003) DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechocystis sp*. strain PCC 6803. *J Bacteriol,* 185**,** 1719-25.

HO, L., HOEFEL, D., BOCK, F., SAINT, C. P. & NEWCOMBE, G. (2007) Biodegradation rates of 2-methylisoborneol (MIB) and geosmin through sand filters and in bioreactors. *Chemosphere,* 66**,** 2210-8.

HO, L., NEWCOMBE, G. & CROUE, J. P. (2002) Influence of the character of NOM on the ozonation of MIB and geosmin. *Water Res,* 36**,** 511-8.

HONG, S., CHA, M. & KIM, I. (2007) A glutaredoxin-fused thiol peroxidase acts an important player in hydrogen peroxide detoxification in late-phased growth of *Anabeana* sp. PCC7120. *Arch Bio Phys,* 475**,** 42-49.

HORWICH, A. L., FENTON, W. A., CHAPMAN, E. & FARR, G. W. (2007) Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol,* 23**,** 115-45.

HOWARD, G., PEDLEY, S. & TIBATEMWA, S. (2006) Quantitative microbial risk assessment to estimate health risks attributable to water supply: can the technique be applied in developing countries with limited data? *J Water Health,* 4**,** 49-65.

IRISARRI, P., GONNET, S. & MONZA, J. (2001) Cyanobacteria in Uruguayan rice fields: diversity, nitrogen fixing ability and tolerance to herbicides and combined nitrogen. *J Biotechnol,* 91**,** 95-103.

IZAGUIRRE, G. (1992) A copper-tolerant *Phormidium* species from lake Matthews, California, that produces 2-methylisoborneol and geosmin. *Wat Sci Tech,* 25**,** 217-223.

IZAGUIRRE, G., HWANG, C. J., KRASNER, S. W. & MCGUIRE, M. J. (1982) Geosmin and 2-methylisoborneol from Cyanobacteria in three water supply systems. *Appl Environ Microbiol,* 43**,** 708-714.

IZAGUIRRE, G. & TAYLOR, W. D. (1998) A *Pseudanabaena* species from Castaic Lake, California, that produces 2-methylisoborneol. *Water Research,* 32**,** 1673-1677.

IZAGUIRRE, G. & TAYLOR, W. D. (2004) A guide to geosmin- and MIB-producing cyanobacteria in the United States. *Wat Sci Tec,* 49**,** 19-24.

IZAGUIRRE, G., TAYLOR, W. D. & PASEK, J. (1999) Off-flavor problems in two reservoirs, associated with planktonic *Pseudanabaena* species. *Wat Sci Tec,* 40**,** 85-90.

JENSEN, S. E., ANDERS, C. L., GOATCHER, L. J., PERLEY, T., KENEFICK, S. & HRUDEY, S. E. (1994) Actinomycetes as a factor in odour problems affecting drinking water from the North Saskatchewan River. *Water Res,* 28**,** 1393-1401.

JIANG, J. & CANE, D. E. (2008) Geosmin biosynthesis. Mechanism of the fragmentation-rearrangement in the conversion of germacradienol to geosmin. *J Am Chem Soc,* 130**,** 428-9.

JIANG, J., HE, X. & CANE, D. E. (2006) Geosmin Biosynthesis. *Streptomyces coelicolor* Germacradienol/Germacrene D Synthase Converts Farnesyl Diphosphate to Geosmin. *J Am Chem Soc,* 128**,** 8128-9.

JIANG, J., HE, X. & CANE, D. E. (2007) Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme. *Nat Chem Biol,* 3**,** 711-5.

JUTTNER, F. & WATSON, S. B. (2007) Biochemical and ecological control of geosmin and 2-methylisoborneol in source waters. *Appl Environ Microbiol,* 73**,** 4395-406.

KASTOVSKA, K., ELSTER, J., STIBAL, M. & SANTRUCKOVA, H. (2005) Microbial assemblages in soil microbial succession after glacial retreat in Svalbard (high arctic). *Microb Ecol,* 50**,** 396-407.

KAY, R. A. (1991) Microalgae as food and supplement. *Crit Rev Food Sci Nutr,* 30**,** 555-73.

KIM, J. H., GLICK, R. E. & MELIS, A. (1993a) Dynamics of Photosystem Stoichiometry Adjustment by Light Quality in Chloroplasts. *Plant Physiol,* 102**,** 181-190.

KIM, M., CHRISTOPHER, D. A. & MULLET, J. E. (1993b) Direct evidence for selective modulation of *psbA, rpoA, rbcL* and 16S RNA stability during barley chloroplast development. *Plant Mol Biol,* 22**,** 447-63.

KLAUSEN, C., NICOLAISEN, M. H., STROBEL, B. W., WARNECKE, F., NIELSEN, J. L. & JORGENSEN, N. O. (2005) Abundance of actinobacteria and production of geosmin and 2-methylisoborneol in Danish streams and fish ponds. *FEMS Microbiol Ecol,* 52**,** 265-78.

KOJIMA, K. & NAKAMOTO, H. (2005) Post-transcriptional control of the cyanobacterial hspA heat-shock induction. *Biochem Biophys Res Commun,* 331**,** 583-8.

KOMATSU, M., TSUDA, M., OMURA, S., OIKAWA, H. & IKEDA, H. (2008) Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. *Proc Natl Acad Sci U S A,* 105**,** 7422-7.

KOSKENNIEMI, K., LYRA, C., RAJANIEMI-WACKLIN, P., JOKELA, J. & SIVONEN, K. (2007) Quantitative real-time PCR detection of toxic *Nodularia* cyanobacteria in the Baltic Sea. *Appl Environ Microbiol,* 73**,** 2173-9.

LALEZARY, S., PIRBAZARI, M. & MCGUIRE, M. (1986) Evaluating activated carbons for removing low concentrations of taste and odour producing organics. *J. Am. Water Works Assoc,* 78**,** 76-82.

LAM, L. T., PICKERAL, O. K., PENG, A. C., ROSENWALD, A., HURT, E. M., GILTNANE, J. M., AVERETT, L. M., ZHAO, H., DAVIS, R. E., SATHYAMOORTHY, M., WAHL, L. M., HARRIS, E. D., MIKOVITS, J. A., MONKS, A. P., HOLLINGSHEAD, M. G., SAUSVILLE, E. A. & STAUDT, L. M. (2001) Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. *Genome Biol,* 2**,** 1-11.

LANGE, B. M., RUJAN, T., MARTIN, W. & CROTEAU, R. (2000) Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A,* 97**,** 13172-7.

LAYTON, A., MCKAY, L., WILLIAMS, D., GARRETT, V., GENTRY, R. & SAYLER, G. (2006) Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. *Appl Environ Microbiol,* 72**,** 4214-24.

LIVAK, K. J. & SCHMITTGEN, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.

LLOYD, S. W. & GRIMM, C. C. (1999) Analysis of 2-methylisoborneol and geosmin in catfish by microwave distillation--solid-phase microextraction. *J Agric Food Chem,* 47**,** 164-9.

LU, G., EDWARDS, C. G., FELLMAN, J. K., MATTINSON, D. S. & NAVAZIO, J. (2003) Biosynthetic origin of geosmin in red beets (*Beta vulgaris* L.). *J Agric Food Chem,* 51**,** 1026-9.

LUDWIG, F., MEDGER, A., BORNICK, H., OPITZ, M., LANG, K., GOTTFERT, M. & ROSKE, I. (2007) Identification and expression analyses of putative sesquiterpene synthase genes in *Phormidium sp*. and prevalence of geoA-like genes in a drinking water reservoir. *Appl Environ Microbiol,* 73**,** 6988-93.

MAO, F., LEUNG, W. Y. & XIN, X. (2007) Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnol,* 7**,** 76.

MARTIN, J. F., IZAGUIRRE, G. & WATERSTRAT, P. (1991) A planktonic *Oscillatoria* species from Mississippi catfish ponds that produces the off-flavour compound 2-methylisoborneol. *Water Res,* 25**,** 1447-1451.

MCGUIRE, M. J. (1995) Off-flavour as the consumers measure of drinking water safety. *Wat Sci Tech,* 31**,** 1-8.

MCKNIGHT, D. (1981) Chemical and biological processes controlling the response of a freshwater ecosystem to copper stress: a field study of the $CuSO_4$ treatment of Mill Pond Reservoir, Burlington, Massachusetts. *Limnol. Oceanogr,* 26.

MOHAMED, Z. A. (2007) First report of toxic *Cylindrospermopsis raciborskii* and *Raphidiopsis mediterranea* (Cyanoprokaryota) in Egyptian fresh waters. *FEMS Microbiol Ecol,* 59**,** 749-61.

MOHAMED, Z. A. (2008) Toxic cyanobacteria and cyanotoxins in public hot springs in Saudi Arabia. *Toxicon,* 51**,** 17-27.

MONIS, P. T. & GIGLIO, S. (2006) Nucleic acid amplification-based techniques for pathogen detection and identification. *Infect Genet Evol,* 6**,** 2-12.

MOUCHET, P. & BONNELYE, V. (1998) Solving algae problems: French expertise and world-wide applications. *Journal of Water Services Research and Technology-Aqua,* 47**,** 125-141.

MULLIS, K. B. & FALOONA, F. A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol,* 155**,** 335-50.

NAES, H., AARNES, H., UTKILEN, H. C., NILSEN, S. & SKULBERG, O. M. (1985) Effect of photon fluence rate and specific growth rate on geosmin production of the cyanobacterium *Oscillatoria brevis* (Kutz.) Gom. *Appl Environ Microbiol,* 49**,** 1538-1540.

NAES, H., UTKILEN, H. C. & POST, A. F. (1989) Geosmin production in the cyanopbacterium *Oscillatoria brevis*. *Arch Microbiol,* 151**,** 407-410.

NAWRATH, T., DICKSCHAT, J. S., MULLER, R., JIANG, J., CANE, D. E. & SCHULZ, S. (2008) Identification of (8S,9S,10S)-8,10-dimethyl-1-octalin, a key intermediate in the biosynthesis of geosmin in bacteria. *J Am Chem Soc,* 130**,** 430-1.

NAYAK, A. K. & ROSE, J. B. (2007) Detection of *Helicobacter pylori* in sewage and water using a new quantitative PCR method with SYBR green. *J Appl Microbiol,* 103**,** 1931-41.

NGAZOA, E. S., FLISS, I. & JEAN, J. (2008) Quantitative study of persistence of human norovirus genome in water using TaqMan real-time RT-PCR. *J Appl Microbiol,* 104**,** 707-15.

NIELSEN, U. S., KNUDSEN, J. B., PEDERSEN, L. N. & MOLLER, J. K. (2009) *Neisseria gonorrhoeae* endocarditis confirmed by nucleic acid amplification assays performed on aortic valve tissue. *J Clin Microbiol,* 47**,** 865-7.

NILSSON, M., RASMUSSEN, U. & BERGMAN, B. (2005) Competition among symbiotic cyanobacterial *Nostoc* strains forming artificial associations with rice (*Oryza sativa*). *FEMS Microbiol Lett,* 245**,** 139-44.

PERSSON, P. E. (1983) Off-flavours in aquatic ecosystems – an introduction. *Wat Sci Tec,* 15**,** 1-11.

PRASANNA, R., KUMAR, V., KUMAR, S., YADAV, A. K., TRIPATHI, U., SINGH, A. K., JAIN, M. C., GUPTA, P., SINGH, P. K. & SETHUNATHAN, N. (2002) Methane production in rice soil is inhibited by cyanobacteria. *Microbiol Res,* 157**,** 1-6.

PROTEAU, P. J. (1998) Biosynthesis of phytol in the cyanobacterium *Synechocystis sp*. UTEX 2470: utilization of the non-mevalonate pathway. *J Nat Prod,* 61**,** 841-3.

QIN, G., GU, H., MA, L., PENG, Y., DENG, X. W., CHEN, Z. & QU, L. J. (2007) Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in *Arabidopsis* by impairing chlorophyll, carotenoid, and gibberellin biosynthesis. *Cell Res,* 17**,** 471-82.

RABOUILLE, S., THEBAULT, J. M. & SALENCON, M. J. (2003) Simulation of carbon reserve dynamics in *Microcystis* and its influence on vertical migration with Yoyo model. *C R Biol,* 326**,** 349-61.

RAMASWAMY, A. V., SORRELS, C. M. & GERWICK, W. H. (2007) Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod,* 70**,** 1977-86.

RANTALA, A., RAJANIEMI-WACKLIN, P., LYRA, C., LEPISTO, L., RINTALA, J., MANKIEWICZ-BOCZEK, J. & SIVONEN, K. (2006) Detection of microcystin-producing cyanobacteria in Finnish lakes with genus-specific microcystin synthetase gene E (*mcyE*) PCR and associations with environmental factors. *Appl Environ Microbiol,* 72**,** 6101-10.

RASHASH, D. M. C., HOEHM, R. C., DIETRICH, A. M., GRIZZARD, T. J. & PARKER, B. C. (1996) *Identification and Control of Odorous Algal Metabolites,* Denver, CO, AWWA Research Foundation and American Water Works Association.

RASMUSSEN, J. P., GIGLIO, S., MONIS, P. T., CAMPBELL, R. J. & SAINT, C. P. (2008) Development and field testing of a real-time PCR assay for cylindrospermopsin-producing cyanobacteria. *J Appl Microbiol,* 104**,** 1503-15.

RASMUSSEN, J. P., SAINT, C. P. & MONIS, P. T. (2007) Use of DNA melting simulation software for in silico diagnostic assay design: targeting regions with complex melting curves and confirmation by real-time PCR using intercalating dyes. *BMC Bioinformatics,* 8**,** 107.

RAVEN, J. A. (2002) The evolution of of cyanobacterial symbioses. IN RAVEN, J. A. (Ed.) *Proceedings of the Royal Irish Academy.*

REITH, F., LENGKE, M. F., FALCONER, D., CRAW, D. & SOUTHAM, G. (2007) The geomicrobiology of gold. *ISME J,* 1**,** 567-84.

RIPPKA, R., DURUELLES, J., WATERBURY, J. B., HERDMAN, M. & STAINER, R. J. (1979) Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *J Gen Microbiol,* 111**,** 1-61.

ROBINSON, B. S., MONIS, P. T. & DOBSON, P. J. (2006) Rapid, sensitive, and discriminating identification of *Naegleria* spp. by real-time PCR and melting-curve analysis. *Appl Environ Microbiol,* 72**,** 5857-63.

ROESELERS, G., NORRIS, T. B., CASTENHOLZ, R. W., RYSGAARD, S., GLUD, R. N., KUHL, M. & MUYZER, G. (2007) Diversity of phototrophic bacteria in microbial mats from Arctic hot springs (Greenland). *Environ Microbiol,* 9**,** 26-38.

SAADOUN, I. M., SCHRADER, K. K. & BLEVINS, W. T. (2001) Environmental and nutritional factors affecting geosmin synthesis by *Anabaena sp*. *Water Res,* 35**,** 1209-18.

SAHMI, M., NICOLA, E. S. & PRICE, C. A. (2006) Hormonal regulation of cytochrome P450 aromatase mRNA stability in non-luteinizing bovine granulosa cells in vitro. *J Endocrinol,* 190**,** 107-15.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989) *Molecular Cloning A Laboratory Manual,* NY, Cold Spring Harbor Laboratory Press.

SATO, T., FUJIEDA, M., TANAKA, E., MIYAMURA, M., CHIKAMOTO, H., HISANO, M., AKIOKA, Y., ISHIURA, Y., DOHNO, S., MAEDA, A., HATTORI, M. & WAKIGUCHI, H. (2008) Monitoring of Epstein-Barr virus load and antibody in pediatric renal transplant patients. *Pediatr Int,* 50**,** 454-8.

SCHAFER, L., SANDMANN, M., WOITSCH, S. & SANDMANN, G. (2006) Coordinate up-regulation of carotenoid biosynthesis as a response to light stress in *Synechococcus* PCC7942. *Plant Cell Environ,* 29**,** 1349-56.

SCHINDLER, D. W., HECKY, R. E., FINDLAY, D. L., STAINTON, M. P., PARKER, B. R., PATERSON, M. J., BEATY, K. G., LYNG, M. & KASIAN, S. E. (2008) Eutrophication of lakes cannot be controlled by reducing nitrogen input: results of a 37-year whole-ecosystem experiment. *Proc Natl Acad Sci U S A,* 105**,** 11254-8.

SCHNURER, J., OLSSON, J. & BORJESSON, T. (1999) Fungal volatiles as indicators of food and feeds spoilage. *Fungal Genet Biol,* 27**,** 209-17.

SCHOPF, J. W., KUDRYAVTSEV, A. B., AGRESTI, D. G., WDOWIAK, T. J. & CZAJA, A. D. (2002) Laser--Raman imagery of Earth's earliest fossils. *Nature,* 416**,** 73-6.

SCHOPF, J. W. & PACKER, B. M. (1987) Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science,* 237**,** 70-3.

SCHRADER, K. K. & DENNIS, M. E. (2005) Cyanobacteria and earthy/musty compounds found in commercial catfish (*Ictalurus punctatus*) ponds in the Mississippi Delta and Mississippi--Alabama Blackland Prairie. *Water Res,* 39**,** 2807-14.

SCHRADER, K. K., NANAYAKKARA, N. P., TUCKER, C. S., RIMANDO, A. M., GANZERA, M. & SCHANEBERG, B. T. (2003) Novel derivatives of 9,10-anthraquinone are selective algicides against the musty-odor cyanobacterium *Oscillatoria perornata*. *Appl Environ Microbiol,* 69**,** 5319-27.

SCHULZ, S., FUHLENDORFF, J. & REICHENBACH, H. (2004) Identification and synthesis of volatiles released by the myxobacterium *Chondromyces crocatus*. *Tetrahedron,* 60**,** 3863-3872.

SETO, H., WATANABE, H. & FURIHATA, K. (1996) Simultaneous operation of the mevalonate and non-mevalonate pathways in the biosynthesis of isopentenyl diphosphate in *Streptomyces aeriouvifer*. *Tetrahedron Letters,* 37**,** 7979-7982.

SHAW, G. R., SUKENIK, A., LIVNE, A., CHISWELL, R. K., SMITH, M. J., SEAWRIGHT, A. A., NORRIS, R. L., EAGLESHAM, G. K. & MOORE, M. R. (1999) Blooms of the cylindrospermopsin containing cyanobacterium, *Aphanizomenon ovalisporum* (Forti), in newly constructed lakes, Queensland, Australia. *Environmental Toxicology,* 14**,** 167-177.

SKLENAR, K. S. & HORNE, A. J. (1999) Horizontal distribution of geosmin in a reservoir before and after copper treatment. *Wat Sci Tec,* 40**,** 229-237.

SMART, L. B. & MCINTOSH, L. (1991) Expression of photosynthesis genes in the cyanobacterium *Synechocystis sp*. PCC 6803: *psaA-psaB* and *psbA* transcripts accumulate in dark-grown cells. *Plant Mol Biol,* 17**,** 959-71.

SMITH, V. H., SIEBER-DENLINGER, J., DENOYELLES, F., CAMPBELL, S., PAN, S., RANDTKE, S. J., BLAIN, G. T. & STRASSER, V. A. (2002) Managing taste and odor problems in a eutrophic drinking water reservoir. *Lake and Reservoir Management,* 18**,** 319-323.

SOULE, T., STOUT, V., SWINGLEY, W. D., MEEKS, J. C. & GARCIA-PICHEL, F. (2007) Molecular genetics and genomic analysis of scytonemin biosynthesis in *Nostoc punctiforme* ATCC 29133. *J Bacteriol,* 189**,** 4465-72.

STANIER, R. Y. & COHEN-BAZIRE, G. (1977) Phototrophic prokaryotes: the cyanobacteria. *Annu Rev Microbiol,* 31**,** 225-74.

STRYER, L. (1981) *Biochemistry,* New York, W.H. Freeman and Company.

SUGIURA, N., IWAMI, N., INAMORI, Y., NISHIMURA, O. & SUDO, R. (1998) Significance of attached cyanobacteria relevant to the occurrence of musty odor in Lake Kasumigaura. *Water Research,* 32**,** 3549-3554.

TAN, L. T. (2007) Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry,* 68**,** 954-79.

TANAKA, D., TANAKA, S., YAMASHIRO, Y. & NAKAMURA, S. (2008) Distribution of oil-degrading bacteria in coastal seawater, Toyama Bay, Japan. *Environ Toxicol*.

TAYLOR, W. D., LOSEE, R. F., TOROBIN, M., IZAGUIRRE, G. & KHIARI, D. (2006) *Early Warning and Management of Surface Water Taste-and-Odor Events,* Denver, AWWA Research Foundation.

THOMAS, A. D., SAKER, M. L., NORTON, J. H. & OLSEN, R. D. (1998) Cyanobacterium *Cylindrospermopsis raciborskii* as a probable cause of death in cattle in northern Queensland. *Aust Vet J,* 76**,** 592-4.

THOMPSON, J. D., GIBSON, T. J., PLEWNIAK, F., JEANMOUGIN, F. & HIGGINS, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res,* 25**,** 4876-82.

TICE, M. M. & LOWE, D. R. (2006) Hydrogen-based carbon fixationin the earliest known photosynthetic organisms. *Geology,* 34**,** 37-40.

TOSI, L. & SOLA, C. (1993) Role of geosmin, a typical inland water odour, in guiding glass eel *Anguilla anguilla* (L.) migration. *Ethology,* 95**,** 177-185.

UTKILEN, H. C. & FROSHAUG, M. (1992) Geosmin production and excretion in a planktonic and benthic *Oscillatoria*. *Wat Sci Tech,* 25**,** 199-206.

VAN BREEMEN, L. W. C. A., DITS, J. S. & KETELAARS, H. A. M. (1992) Production and reduction of geosmin and 2-methylisoborneol during storage of river water in deep reservoirs. *Wat Sci Tech,* 25**,** 233-240.

VAN DER PLOEG, M. & BOYD, C. E. (1991) Geosmin production by Cyanobacteria (Blue-green algae) in fish ponds at Auburn, Alabama. *Journal of the World Aquaculture Society,* 22**,** 207-216.

VAN DER PLOEG, M., TUCKER, C. S. & BOYD, C. E. (1992) Geosmin and 2-methylisoborneol prodcution by cyanobacteria in fish ponds in the southeastern United States. *Wat Sci Tech,* 25**,** 283-290.

VAN LIEVERLOO, J. H., BLOKKER, E. J. & MEDEMA, G. (2007) Quantitative microbial risk assessment of distributed drinking water using faecal indicator incidence and concentrations. *J Water Health,* 5 Suppl 1**,** 131-49.

WALERON, M., WALERON, K., VINCENT, W. F. & WILMOTTE, A. (2007) Allochthonous inputs of riverine picocyanobacteria to coastal waters in the Arctic Ocean. *FEMS Microbiol Ecol,* 59**,** 356-65.

WALKER, C. & GREER, L. (1992) The Economic Costs Associated with Lost Recreation Benefits due to Blue-Green Algae in New South Wales: Three case studies. *Final report of the New South Wales Blue-green Algal Task force, Department of Water Resources***,** 33-39.

WALSBY, A. E. (1972) Gas-filled structures providing buoyancy in photosynthetic organisms. *Symp Soc Exp Biol,* 26**,** 233-50.

WALSBY, A. E., SCHANZ, F. & SCHMID, M. (2006) The Burgundy-blood phenomenon: a model of buoyancy change explains autumnal waterblooms by *Planktothrix rubescens* in Lake Zurich. *New Phytol,* 169**,** 109-22.

WANG, C. M. & CANE, D. E. (2008) Biochemistry and molecular genetics of the biosynthesis of the earthy odorant methylisoborneol in *Streptomyces coelicolor*. *J Am Chem Soc,* 130**,** 8908-9.

WATSON, S. B. (2003) Cyanobacterial and eukaryotic algal odour compounds: signals or by-products? A review of their biological activity. *Phycologia,* 42**,** 332-350.

WESTRELL, T., SCHONNING, C., STENSTROM, T. A. & ASHBOLT, N. J. (2004) QMRA (quantitative microbial risk assessment) and HACCP (hazard analysis and critical control points) for management of pathogens in wastewater and sewage sludge treatment and reuse. *Wat Sci Tec,* 50**,** 23-30.

WILSON, K. M., SCHEMBRI, M. A., BAKER, P. D. & SAINT, C. P. (2000) Molecular characterization of the toxic cyanobacterium *Cylindrospermopsis raciborskii* and design of a species-specific PCR. *Appl Environ Microbiol,* 66**,** 332-8.

WU, J. T. & JUTTNER, F. (1988) Effect of environmental factors on geosmin production by *Fischerella muscicola*. *Wat Sci Tec,* 20**,** 143-148.

WU, J. T., MA, P. I. & CHOU, T. L. (1991) Variation of geosmin content in *Anabaena* cells and its relation to nitrogen utilization. *Arch Microbiol,* 157**,** 66-9.

YANG, E., VAN NIMWEGEN, E., ZAVOLAN, M., RAJEWSKY, N., SCHROEDER, M., MAGNASCO, M. & DARNELL, J. E., JR. (2003) Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Res,* 13**,** 1863-72.

YOUNG, J. C., AGASHE, V. R., SIEGERS, K. & HARTL, F. U. (2004) Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol,* 5**,** 781-91.

ZHANG, T., LI, L., SONG, L. & CHEN, W. (2009) Effects of temperature and light on the growth and geosmin production of *Lyngbya kuetzingii* (Cyanophyta). *J Appl Phycol,* 21**,** 279-285.

ZIMBA, P. V., DIONIGI, C. P. & MILLIE, D. F. (1999) Evaluating the relationship between photopigment synthesis and 2-methylisoborneol accumulation in cyanobacteria. *Journal of Phycology,* 35**,** 1422-1429.

APPENDIX 1

DNA and Amino Acid Sequences and Alignments

```
NJ2             ATGCAACCCTTTGAACTGCCAGAATTTTACATGCCTTGGCCCGCGCGGCTGAATCCAAAC 60
NPUN02003620    ATGCAACCCTTTGAACTGCCAGAATTTTACATGCCTTGGCCCGCGCGGCTGAATCCAAAC 60
                ************************************************************

NJ2             CTGGAAGCGGCGCGATCGCATTCCAAAGCCTGGGCTTACCAGATGGGGATACTTGGCTCA 120
NPUN02003620    CTGGAAGCGGCGCGATCGCATTCCAAAGCCTGGGCTTACCAGATGGGGATACTTGGCTCA 120
                ************************************************************

NJ2             AAAGAGGAAGCGGAAAGCTCGGTTATCTGGGACGAGCGCACATTCGATGCTCACGACTAC 180
NPUN02003620    AAAGAGGAAGCGGAAAGCTCGGTTATCTGGGACGAGCGCACATTCGATGCTCACGACTAC 180
                ************************************************************

NJ2             GCCTTGCTTTGCTCGTATACCCATCCAGATGCACCAGGAACGGAGCTTGACCTAGTGACC 240
NPUN02003620    GCCTTGCTTTGCTCGTATACCCATCCAGATGCACCAGGAACGGAGCTTGACCTAGTGACC 240
                ************************************************************

NJ2             GACTGGTATGTTTGGGTGTTCTTCTTCGACGATCACTTCCTTGAAATCTATAAGCGTACC 300
NPUN02003620    GACTGGTATGTTTGGGTGTTCTTCTTCGACGATCACTTCCTTGAAATCTATAAGCGTACC 300
                ************************************************************

NJ2             CAAGACATGGCTGGGGCGAAGGAGTATCTCGGCAGATTACCAATGTTCATGCCAATTTAC 360
NPUN02003620    CAAGACATGGCTGGGGCGAAGGAGTATCTCGGCAGATTACCAATGTTCATGCCAATTTAC 360
                ************************************************************

NJ2             CCCACTGAAACCCCCCCAGTACCTACCAACCCCGTGGAGTGCGGCTTGGCCGACCTGTGG 420
NPUN02003620    CCCACTGAAACCCCCCCAGTACCTACCAACCCCGTGGAGTGCGGCTTGGCCGACCTGTGG 420
                ************************************************************

NJ2             TCTCGCACCGCATTTACTAAATCTGTGGACTGGCGACTTCGATTCTTCGAGAGTACCAAG 480
NPUN02003620    TCTCGCACCGCATTTACTAAATCTGTGGACTGGCGACTTCGATTCTTCGAGAGTACCAAG 480
                ************************************************************

NJ2             AACCTCCTGGAAGAGTCTCTGTGGGAACTCGCCAATATCAACCAGGATCGAGTCGCTAAC 540
NPUN02003620    AACCTCCTGGAAGAGTCTCTGTGGGAACTCGCCAATATCAACCAGGATCGAGTCGCTAAC 540
                ************************************************************

NJ2             CCCATCGAATACATCGAAATGCGCCGCAAGGTTGGTGGCGCTCCGTGGTCAGCGGATCTC 600
NPUN02003620    CCCATCGAATACATCGAAATGCGCCGCAAGGTTGGTGGCGCTCCGTGGTCAGCGGATCTC 600
                ************************************************************

NJ2             GTCGAACACGCCGTGTTTATAGAGATTCCGGCTGACATTGCCTCGACTCGGCCAATGCGT 660
NPUN02003620    GTCGAACACGCCGTGTTTATAGAGATTCCGGCTGACATTGCCTCGACTCGGCCAATGCGT 660
                ************************************************************

NJ2             GTGCTAAAAGACACGTTTGCCGATGGAGTGCATTTACGCAACGACCTCTTCTCCTACCAG 720
NPUN02003620    GTGCTAAAAGACACGTTTGCCGATGGAGTGCATTTACGCAACGACCTCTTCTCCTACCAG 720
                ************************************************************

NJ2             AGAGAAGTGGAAGATGAAGGCGAAAATGCCAACTGTGTGTTGGTCTTGGAGCGCTTCTTA 780
NPUN02003620    AGAGAAGTGGAAGATGAAGGCGAAAATGCCAACTGTGTGTTGGTCTTGGAGCGCTTCTTA 780
                ************************************************************

NJ2             AATGTGAGTACCCAGGAGGCAGCTAACCTCACCAACGAACTGCTGACCTCGCGGTTATAC 840
NPUN02003620    AATGTGAGTACCCAGGAGGCAGCTAACCTCACCAACGAACTGCTGACCTCGCGGTTATAC 840
                ************************************************************

NJ2             CAGTTTGATAACACTGCCGTCACCGAGTTGCCCCCCCCTCTTTGAGGAGTATGGACTAGAC 900
NPUN02003620    CAGTTTGATAACACTGCCGTCACCGAGTTGCCCCCCCCTCTTTGAGGAGTATGGACTAGAC 900
                ************************************************************

NJ2             CCAGTGGCTCGTGTGAACGTTCTCCTTTATATCAAAGGACTTCAGGACTGGCAGTCGGGC 960
NPUN02003620    CCAGTGGCTCGTGTGAACGTTCTCCTTTATATCAAAGGACTTCAGGACTGGCAGTCGGGC 960
                ************************************************************

NJ2             GGTCACGAGTGGCACATGCGATCAAGCCGTTACATGAACAAGGGAGGAGACAATTCTCCA 1020
NPUN02003620    GGTCACGAGTGGCACATGCGATCAAGCCGTTACATGAACAAGGGAGGAGACAATTCTCCA 1020
                ************************************************************

NJ2             ACATCTACGGTTCTGGGAGGACCCACAGGGCTAGGCACATCGGCTGCGCGGATTGAATCG 1080
NPUN02003620    ACATCTACGGTTCTGGGAGGACCCACAGGGCTAGGCACATCGGCTGCGCGGATTGAATCG 1080
                ************************************************************
```

```
NJ2                 TTATATGCGGCCTTGGGTTTAGGAAGAATTAAAAGCTTCACTCACGTTCCATACCAGCCT 1140
NPUN02003620        TTATATGCGGCCTTGGGTTTAGGAAGAATTAAAAGCTTCACTCACGTTCCATACCAGCCT 1140
                    ************************************************************

NJ2                 GTGGGACCAGTGACCCTGCCGAAGTTTTACATGCCGTTCACTACAAGTTTGAATCCTCAT 1200
NPUN02003620        GTGGGACCAGTGACCCTGCCGAAGTTTTACATGCCGTTCACTACAAGTTTGAATCCTCAT 1200
                    ************************************************************

NJ2                 CTCAATGCCGCACGGAAGCATTCTAAGGAATGGGCGCGCCAGATGGGGATGCTGGAATCA 1260
NPUN02003620        CTCAATGCCGCACGGAAGCATTCTAAGGAATGGGCGCGCCAGATGGGGATGCTGGAATCA 1260
                    ************************************************************

NJ2                 CTACCTGGGATTCCTGATGCCGTCATCTGGGATGACCACAAGTTTGATGTTGCCGACGTA 1320
NPUN02003620        CTACCTGGGATTCCTGATGCCGTCATCTGGGATGACCACAAGTTTGATGTTGCCGACGTA 1320
                    ************************************************************

NJ2                 GCTCTCTGCGGTGCGTTGATCCATCCGAATGGGTCTGGTCTTGAACTGAATCTGACAGCG 1380
NPUN02003620        GCTCTCTGCGGTGCGTTGATCCATCCGAATGGGTCTGGTCTTGAACTGAATCTGACAGCG 1380
                    ************************************************************

NJ2                 TGCTGGCTTGTTTGGGGAACCTATGCCGACGATTACTTCCCGGCGCTCTACGGGAATAAC 1440
NPUN02003620        TGCTGGCTTGTTTGGGGAACCTATGCCGACGATTACTTCCCGGCGCTCTACGGGAATAAC 1440
                    ************************************************************

NJ2                 CGCAACATGGCTGGTGCGAAAGTATTCAACGCTCGACTGTCGGCGTTCATGCCTCTGGAT 1500
NPUN02003620        CGCAACATGGCTGGTGCGAAAGTATTCAACGCTCGACTGTCGGCGTTCATGCCTCTGGAT 1500
                    ************************************************************

NJ2                 GACTCCACCCCCAGCGAGGTTCCGACCAATCCAGTGGAAGCGGGCTTGGCAGATATTTGG 1560
NPUN02003620        GACTCCACCCCCAGCGAGGTTCCGACCAATCCAGTGGAAGCGGGCTTGGCAGATATTTGG 1560
                    ************************************************************

NJ2                 TCTCGCACAGCTGGTCCCATGTCTGCCAACGCACGGACTCAGTTCCGCCGCGCAATCCAG 1620
NPUN02003620        TCTCGCACAGCTGGTCCCATGTCTGCCAACGCACGGACTCAGTTCCGCCGCGCAATCCAG 1620
                    ************************************************************

NJ2                 GATATGACTGACAGTTGGGTGTGGGAACTCGCCAACCAGATCCAAAATCGAATTCCTGAC 1680
NPUN02003620        GATATGACTGACAGTTGGGTGTGGGAACTCGCCAACCAGATCCAAAATCGAATTCCTGAC 1680
                    ************************************************************

NJ2                 CCGATAGATTATGTTGAGATGCGCCGTAAGACCTTTGGCTCGGATCTGACCATGAGCCTG 1740
NPUN02003620        CCGATAGATTATGTTGAGATGCGCCGTAAGACCTTTGGCTCGGATCTGACCATGAGCCTG 1740
                    ************************************************************

NJ2                 TCGCGACTAGCTCAGGGGAGCGAGATCCCGCAGGAGATTTACCGCACCCGAACGATGCGA 1800
NPUN02003620        TCGCGACTAGCTCAGGGGAGCGAGATCCCGCAGGAGATTTACCGCACCCGAACGATGCGA 1800
                    ************************************************************

NJ2                 TCGCTCGATAATTCGGCCGCCGACTTCGCCTGTTTAACCAACGATAT-CTTTTCCTATCA 1859
NPUN02003620        TCGCTCGATAATTCGGCCGCCGACTTCGCCTGTTTAACCAACGATATTCTTTTCCTATCA 1860
                    ********************************************** ************

NJ2                 GAAAGAAATCGAATTCGAGGGCGATA------- 1885
NPUN02003620        GAAAGAAATCGAATTCGAGGGCGAAATCCATAA 1893
                    ************************ *
```

Appendix 1-1. Multiple alignment of experimentally derived NJ2 and *npun02003620* demonstrating an additional "T" nucleotide boxed in yellow. ("*" indicates exact sequence match, ":", indicates moderate match, and "." Indicates a low match.)

```
SCO6073          MTQQPFQLPHFYLPHPARLNPHLDEARAHSTTWAREMGML------EGSGVWEQSDLEAH
54SAV2163        -MTQPFQLPHFYMPYPARLNPHLDEARAHSTRWARGMGML------EGSGIWEQSDLDAH 53
ZP_00109187      --MQPFELPEFYMPWPARLNPNLEAARSHSKAWAYQMGILGSKEEAESSVIWDERTFDAH 58
NJ2              --MQPFELPEFYMPWPARLNPNLEAARSHSKAWAYQMGILGSKEEAESSVIWDERTFDAH 58
                    ***:**.**:* ******:*: **:**. **  **:*      *.* :*::  ::**

SCO6073          DYGLLCAYTHPDCDGPALSLITDWYVWVFFFDDHFLEKYKRSQDRLAGKAHLDRLPLFMP 114
SAV2163          DYGLLCAYTHPDCDGPALSLITDWYVWVFFFDDHFLETFKRTQDREGGKAYLDRLPLFMP 113
ZP_00109187      DYALLCSYTHPDAPGTELDLVTDWYVWVFFFDDHFLEIYKRTQDMAGAKEYLGRLPMFMP 118
NJ2              DYALLCSYTHPDAPGTELDLVTDWYVWVFFFDDHFLEIYKRTQDMAGAKEYLGRLPMFMP 118
                 **.***:*****. *. *.*:***************** :**:**  ..* :*.***:***

SCO6073          LDDAAGMPEPRNPVEAGLADLWTRTVPAMSADWRRRFAVATEHLLNESMWELSNINEGRV 174
SAV2163          LDLSAPVPEPENPVEAGLADLWARTVPAMSADWRKRFAVSTEHLLNESLWELSNINEGRI 173
ZP_00109187      IYPTETPPVPTNPVECGLADLWSRTAFTKSVDWRLRFFESTKNLLEESLWELANINQDRV 178
NJ2              IYPTETPPVPTNPVECGLADLWSRTAFTKSVDWRLRFFESTKNLLEESLWELANINQDRV 178
                 :  :   * * ****.******:**. : *.*** **  :*::**:**:***:***:.*:

SCO6073          ANPVEYIEMRRKVGGAPWSAGLVEYAT-AEVPAAVAGTRPLRVLMETFSDAVHLRNDLFS 233
SAV2163          ANPVEYIEMRRKVGGAPWSAGLVEYAT-AEVPAAVAGSRPLRVLMETFSDGVHLRNDLFS 232
ZP_00109187      ANPIEYIEMRRKVGGAPWSADLVEHAVFIEIPADIASTRPMRVLKDTFADGVHLRNDLFS 238
NJ2              ANPIEYIEMRRKVGGAPWSADLVEHAVFIEIPADIASTRPMRVLKDTFADGVHLRNDLFS 238
                 ***:****************.***:*.  *:** :*.:**:*** :**:*.*********

SCO6073          YQREVEDEGELSNGVLVLETFFGCTTQEAADLVNDVLTSRLHQFEHTAFTEVPAVALEKG 293
SAV2163          YQREVEEEGELSNGVLVLETFFGCTTQEAAETVNDILTSRLHQFEHTALTEVPALALEKG 292
ZP_00109187      YQREVEDEGENANCVLVLERFLNVSTQEAANLTNELLTSRLYQFDNTAVTELPPLFEEYG 298
NJ2              YQREVEDEGENANCVLVLERFLNVSTQEAANLTNELLTSRLYQFDNTAVTELPPLFEEYG 298
                 ******:*** :* ***** *:. :*****: .*::*****:**::**.**:*.:  * *

SCO6073          LTPLEVAAVGAYTKGLQDWQSGGHEWHMRSSRYMNKG---ERPLAGWQALTGPGTSAADV 350
SAV2163          LTPPEVAAVAAYARGLQDWQSGGHEWHLRSSRYMNEG---ALSQKRPFGLSAIGTSAADL 349
ZP_00109187      LDPVARVNVLLYIKGLQDWQSGGHEWHMRSSRYMNKGGDNSPTSTVLGGPTGLGTSAARI 358
NJ2              LDPVARVNVLLYIKGLQDWQSGGHEWHMRSSRYMNKGGDNSPTSTVLGGPTGLGTSAARI 358
                 * *   . *  * :**************:*******:*     .     . :. ***** :

SCO6073          GALLADAVAQRARSYTYVPFQKVGPSVIPDIRMPYPLELSPALDGARRHLSEWCREMGIL 410
SAV2163          RGLLADAGAERLRRYTHVPFQKVGPSRIPDFHMPFQVELSPHLEGARARLTPWMHSTGML 409
ZP_00109187      ESLYAALGLGRIKSFTHVPYQPVGPVTLPKFYMPFTTSLNPHLNAARKHSKEWARQMGML 418
NJ2              ESLYAALGLGRIKSFTHVPYQPVGPVTLPKFYMPFTTSLNPHLNAARKHSKEWARQMGML 418
                  .* *     * : :*:**:* ***  :*.: **:  .*.* *:.** : . * :. *:*

SCO6073          SEG-------VWDEDKLESCDLPLCAAGLDPDATQDQLDLASGWLAFGTYGDDYYPLVYG 463
SAV2163          QEG-------VWDEDKLTAYDLPLCSAGLDPDATPDELDLSSRWLAWGTYGDDYYPMVFG 462
ZP_00109187      ESLPGIPDAVIWDDHKFDVADVALCGALIHPNGSGLELNLTACWLVWGTYADDYFPALYG 478
NJ2              ESLPGIPDAVIWDDHKFDVADVALCGALIHPNGSGLELNLTACWLVWGTYADDYFPALYG 478
                 ..        :**:.*:   *:.**.* :.*:.:  :*:*:: **.:***.***:* ::*

SCO6073          HRRDLAAARLTTTRLSDCMPLDG-EPVPPPGNAMERSLIDLWVRTTAGMTPEERRPLKKA 522
SAV2163          PRRDLAAAKLCTRRLSACMPVDG-EEVPAPVNGMERGLIDLWAITTAEMTPDERRTFRAS 521
ZP_00109187      NNRNMAGAKVFNARLSAFMPLDDSTPSEVPTNPVEAGLADIWSRTAGPMSANARTQFRRA 538
NJ2              NNRNMAGAKVFNARLSAFMPLDDSTPSEVPTNPVEAGLADIWSRTAGPMSANARTQFRRA 538
                  .*::*.*:: . ***  **:*.      * * :* .* *:*  *:. *:.: *  :: :

SCO6073          VDDMTEAWLWELSNQIQNRVPDPVDYLEMRRATFGSDLTLGLCRAGHGPAVPPEVYRSGP 582
SAV2163          VDVMTESWVWELSNQLQHRIPDPIDYLEMRRATFGADLTLSLCRVGHGPKVPPEIYRSGP 581
ZP_00109187      IQDMTDSWVWELANQIQNRIPDPIDYVEMRRKTFGSDLTMSLSRLAQGSEIPQEIYRTRT 598
NJ2              IQDMTDSWVWELANQIQNRIPDPIDYVEMRRKTFGSDLTMSLSRLAQGSEIPQEIYRTRT 598
                 :: **::*:***:**:*:*:***:**:**** ***:***:.*.* .:*. :* *:**: .

SCO6073          VRSLENAAIDYACLLNDVFSYQKEIEYEGEIHNAVLVVQNFFGVDYPAALGVVQDLMNQR 642
SAV2163          VRSLENAAVDYGMLINDVFSYQKEIEYEGEVHNAILVVQNFFGCDYPTALGVINDLMTQR 641
ZP_00109187      MRSLDNSAADFACLTNDILFLSERNRIRGRNP---------------------------- 630
NJ2              MRSLDNSAADFACLTNDIFSYQKEIEFEGI---------------------------- 628
                 :***:*:* *:. * **::  .:. . .*
```

```
SCO6073          MRQFEHVVAHELPVVYDDFQLSEEARTVMRGYVTDLQNWMAGILNWHRNVPRYKAEYLAG 702
SAV2163          MHQFEHVAAHELPLLYKDFKLPQEVRDIMDGYVVELQNWMSGILKWHQDCHRYGAADLAR 701
ZP_00109187      ------------------------------------------------------------
NJ2              ------------------------------------------------------------

SCO6073          RTHGFLPDRIPAPPVPRSSPALTH 726
SAV2163          RAHGFVPDRAPSAPFTAWAAPVAR 725
ZP_00109187      -------------------------
NJ2              -------------------------
```

Appendix 1-2. Multiple alignment of amino acid sequences of SCO6073 from *S. coelicolor*, SAV2163 from *S. avermitilis*, ZP_00109187 from *Nostoc punctiforme*, and experimentally derived NJ2. The relieved frame shift from the additional "T" nucleotide in protein product ZP_00109187 unveils a previously masked NSE triad that is apparent in experimental NJ2.

```
NJ2             --MQPFELPEFYMPWPARLNPNLEAARSHSKAWAYQMGILGSKEEAESSVIWDERTFDAH
NPUNMOD         MVMQPFELPEFYMPWPARLNPNLEAARSHSKAWAYQMGILGSKEEAESSVIWDERTFDAH 60
ZP_00109187     --MQPFELPEFYMPWPARLNPNLEAARSHSKAWAYQMGILGSKEEAESSVIWDERTFDAH 58
SCO6073         MTQQPFQLPHFYLPHPARLNPHLDEARAHSTTWAREMGML------EGSGVWEQSDLEAH 54
SAV2163         -MTQPFQLPHFYMPYPARLNPHLDEARAHSTRWARGMGML------EGSGIWEQSDLDAH 53
                   ***:**.**:* ******:*: **:**. **  **:*      *.* :*::  ::**

NJ2             DYALLCSYTHPDAPGTELDLVTDWYVWVFFFDDHFLEIYKRTQDMAGAKEYLGRLPMFMP 118
NPUNMOD         DYALLCSYTHPDAPGTELDLVTDWYVWVFFFDDHFLEIYKRTQDMAGAKEYLGRLPMFMP 120
ZP_00109187     DYALLCSYTHPDAPGTELDLVTDWYVWVFFFDDHFLEIYKRTQDMAGAKEYLGRLPMFMP 118
SCO6073         DYGLLCAYTHPDCDGPALSLITDWYVWVFFFDDHFLEKYKRSQDRLAGKAHLDRLPLFMP 114
SAV2163         DYGLLCAYTHPDCDGPALSLITDWYVWVFFFDDHFLETFKRTQDREGGKAYLDRLPLFMP 113
                **.***:*****. *. *.*:**************** :**:**  ..* :*.***:***

NJ2             IYPTETPPVPTNPVECGLADLWSRTAFTKSVDWRLRFFESTKNLLEESLWELANINQDRV 178
NPUNMOD         IYPTETPPVPTNPVECGLADLWSRTAFTKSVDWRLRFFESTKNLLEESLWELANINQDRV 180
ZP_00109187     IYPTETPPVPTNPVECGLADLWSRTAFTKSVDWRLRFFESTKNLLEESLWELANINQDRV 178
SCO6073         LDDAAGMPEPRNPVEAGLADLWTRTVPAMSADWRRRFAVATEHLLNESMWELSNINEGRV 174
SAV2163         LDLSAPVPEPENPVEAGLADLWARTVPAMSADWRKRFAVSTEHLLNESLWELSNINEGRI 173
                :  :   * * ****.******:**. : *.*** **  :*::**:**:***:***:.*:

NJ2             ANPIEYIEMRRKVGGAPWSADLVEHAVFIEIPADIASTRPMRVLKDTFADGVHLRNDLFS 238
NPUNMOD         ANPIEYIEMRRKVGGAPWSADLVEHAVFIEIPADIASTRPMRVLKDTFADGVHLRNDLFS 240
ZP_00109187     ANPIEYIEMRRKVGGAPWSADLVEHAVFIEIPADIASTRPMRVLKDTFADGVHLRNDLFS 238
SCO6073         ANPVEYIEMRRKVGGAPWSAGLVEYAT-AEVPAAVAGTRPLRVLMETFSDAVHLRNDLFS 233
SAV2163         ANPVEYIEMRRKVGGAPWSAGLVEYAT-AEVPAAVAGSRPLRVLMETFSDGVHLRNDLFS 232
                ***:***************.***:*.  *:** :*.:**:*** :**:*.*********

NJ2             YQREVEDEGENANCVLVLERFLNVSTQEAANLTNELLTSRLYQFDNTAVTELPPLFEEYG 298
NPUNMOD         YQREVEDEGENANCVLVLERFLNVSTQEAANLTNELLTSRLYQFDNTAVTELPPLFEEYG 300
ZP_00109187     YQREVEDEGENANCVLVLERFLNVSTQEAANLTNELLTSRLYQFDNTAVTELPPLFEEYG 298
SCO6073         YQREVEDEGELSNGVLVLETFFGCTTQEAADLVNDVLTSRLHQFEHTAFTEVPAVALEKG 293
SAV2163         YQREVEEEGELSNGVLVLETFFGCTTQEAAETVNDILTSRLHQFEHTALTEVPALALEKG 292
                ******:*** :* ***** *:. :*****: .*::*****:**::**.**:*.:  * *

NJ2             LDPVARVNVLLYIKGLQDWQSGGHEWHMRSSRYMNKGGDNSPTSTVLGGPTGLGTSAARI 358
NPUNMOD         LDPVARVNVLLYIKGLQDWQSGGHEWHMRSSRYMNKGGDNSPTSTVLGGPTGLGTSAARI 360
ZP_00109187     LDPVARVNVLLYIKGLQDWQSGGHEWHMRSSRYMNKGGDNSPTSTVLGGPTGLGTSAARI 358
SCO6073         LTPLEVAAVGAYTKGLQDWQSGGHEWHMRSSRYMNKG---ERPLAGWQALTGPGTSAADV 350
SAV2163         LTPPEVAAVAAYARGLQDWQSGGHEWHLRSSRYMNEG---ALSQKRPFGLSAIGTSAADL 349
                * *   . *  * :**************:*******:*     .     . :. ***** :

NJ2             ESLYAALGLGRIKSFTHVPYQPVGPVTLPKFYMPFTTSLNPHLNAARKHSKEWARQMGML 418
NPUNMOD         ESLYAALGLGRIKSFTHVPYQPVGPVTLPKFYMPFTTSLNPHLNAARKHSKEWARQMGML 420
ZP_00109187     ESLYAALGLGRIKSFTHVPYQPVGPVTLPKFYMPFTTSLNPHLNAARKHSKEWARQMGML 418
SCO6073         GALLADAVAQRARSYTYVPFQKVGPSVIPDIRMPYPLELSPALDGARRHLSEWCREMGIL 410
SAV2163         RGLLADAGAERLRRYTHVPFQKVGPSRIPDFHMPFQVELSPHLEGARARLTPWMHSTGML 409
                 .* *     * : :*:**:* ***  :*.: **:  .*.* *:.** : . * :. *:*

NJ2             ESLPGIPDAVIWDDHKFDVADVALCGALIHPNGSGLELNLTACWLVWGTYADDYFPALYG 478
NPUNMOD         ESLPGIPDAVIWDDHKFDVADVALCGALIHPNGSGLELNLTACWLVWGTYADDYFPALYG 480
ZP_00109187     ESLPGIPDAVIWDDHKFDVADVALCGALIHPNGSGLELNLTACWLVWGTYADDYFPALYG 478
SCO6073         SEG-------VWDEDKLESCDLPLCAAGLDPDATQDQLDLASGWLAFGTYGDDYYPLVYG 463
SAV2163         QEG-------VWDEDKLTAYDLPLCSAGLDPDATPDELDLSSRWLAWGTYGDDYYPMVFG 462
                ..        :**:.*:   *:.**.* :.*:.:  :*:*:: **.:***.***:* ::*

NJ2             NNRNMAGAKVFNARLSAFMPLDDSTPSEVPTNPVEAGLADIWSRTAGPMSANARTQFRRA 538
NPUNMOD         NNRNMAGAKVFNARLSAFMPLDDSTPSEVPTNPVEAGLADIWSRTAGPMSANARTQFRRA 540
ZP_00109187     NNRNMAGAKVFNARLSAFMPLDDSTPSEVPTNPVEAGLADIWSRTAGPMSANARTQFRRA 538
SCO6073         HRRDLAAARLTTTRLSDCMPLDG-EPVPPPGNAMERSLIDLWVRTTAGMTPEERRPLKKA 522
SAV2163         PRRDLAAAKLCTRRLSACMPVDG-EEVPAPVNGMERGLIDLWAITTAEMTPDERRTFRAS 521
                 .*::*.*:: . ***  **:*.      * * :* .* *:*  *:. *:.: *  :: :

NJ2             IQDMTDSWVWELANQIQNRIPDPIDYVEMRRKTFGSDLTMSLSRLAQGSEIPQEIYRTRT 598
NPUNMOD         IQDMTDSWVWELANQIQNRIPDPIDYVEMRRKTFGSDLTMSLSRLAQGSEIPQEIYRTRT 600
ZP_00109187     IQDMTDSWVWELANQIQNRIPDPIDYVEMRRKTFGSDLTMSLSRLAQGSEIPQEIYRTRT 598
SCO6073         VDDMTEAWLWELSNQIQNRVPDPVDYLEMRRATFGSDLTLGLCRAGHGPAVPPEVYRSGP 582
SAV2163         VDVMTESWVWELSNQLQHRIPDPIDYLEMRRATFGADLTLSLCRVGHGPKVPPEIYRSGP 581
                :: **::*:***:**:*:*:***:**:**** ***:***:.*.* .:*. :* *:**: .
```

```
NJ2             MRSLDNSAADFACLTNDIFSYQKEIEFEGI------------------------------ 628
NPUNMOD         MRSLDNSAADFACLTNDIFSYQKEIEFEGEIHNCVLVVQNFLNCDLPQAVEVVNNLMTSR 660
ZP_00109187     MRSLDNSAADFACLTNDILFLSERNRIRGRNP---------------------------- 630
SCO6073         VRSLENAAIDYACLLNDVFSYQKEIEYEGEIHNAVLVVQNFFGVDYPAALGVVQDLMNQR 642
SAV2163         VRSLENAAVDYGMLINDVFSYQKEIEYEGEVHNAILVVQNFFGCDYPTALGVINDLMTQR 641
                :***:*:* *:. * **::  .:. . .*

NJ2             ------------------------------------------------------------
NPUNMOD         ALQFQLIVATELPVLFDDFDLDASTREKLLGYVKKLEQWMCGVLKWHITVDRYKEFELRN 720
ZP_00109187     ------------------------------------------------------------
SCO6073         MRQFEHVVAHELPVVYDDFQLSEEARTVMRGYVTDLQNWMAGILNWHRNVPRYKAEYLAG 702
SAV2163         MHQFEHVAAHELPLLYKDFKLPQEVRDIMDGYVVELQNWMSGILKWHQDCHRYGAADLAR 701


NJ2             ------------------------------------
NPUNMOD         SLAGRLLSGPRGLGTSARRIGSLIGQGSLKSLLGQ 755
ZP_00109187     ------------------------------------
SCO6073         RTHGFLPDRIPAPPVPRSSPALTH----------- 726
```

Appendix 1-3. Amino acid multiple alignment with experimental NJ2; hypothetical *Nostoc punctiforme* geosmin synthase NPUNMOD; predicted ZP_00109187 from *Nostoc punctiforme*, SCO6073, and SAV 2163

```
NPM2_58         ATGGTTATGCAACCCTTTGAACTGCCAGAATTTTACATGCCTTGGCCCGCGCGGCTGAAT
NPUNMOD         ------ATGCAACCCTTTGAACTGCCAGAATTTTACATGCCTTGGCCCGCGCGGCTGAAT 54
                      **************************************************

NPM2_58         CCAAACCTGGAAGCGGCGCGATCGCATTCCAAAGCCTGGGCTTACCAGATGGGGATACTT 120
NPUNMOD         CCAAACCTGGAAGCGGCGCGATCGCATTCCAAAGCCTGGGCTTACCAGATGGGGATACTT 114
                ************************************************************

NPM2_58         GGCTCAAAAGAGGAAGCGGAAAGCTCGGTTATCTGGGACGAGCGCACATTCGATGCTCAC 180
NPUNMOD         GGCTCAAAAGAGGAAGCGGAAAGCTCGGTTATCTGGGACGAGCGCACATTCGATGCTCAC 174
                ************************************************************

NPM2_58         GACTACGCCTTGCTTTGCTCGTATACCCATCCAGATGCACCAGGAACGGAGCTTGACCTA 240
NPUNMOD         GACTACGCCTTGCTTTGCTCGTATACCCATCCAGATGCACCAGGAACGGAGCTTGACCTA 234
                ************************************************************

NPM2_58         GTGACCGACTGGTATGTTTGGGTGTTCTTCTTCGACGATCACTTCCTTGAAATCTATAAG 300
NPUNMOD         GTGACCGACTGGTATGTTTGGGTGTTCTTCTTCGACGATCACTTCCTTGAAATCTATAAG 294
                ************************************************************

NPM2_58         CGTACCCAAGACATGGCTGGGGCGAAGGAGTATCTCGGCAGATTACCAATGTTCATGCCA 360
NPUNMOD         CGTACCCAAGACATGGCTGGGGCGAAGGAGTATCTCGGCAGATTACCAATGTTCATGCCA 354
                ************************************************************

NPM2_58         ATTTACCCCACTGAAACCCCCCCCAGTACCTACCAACCCCGTGGAGTGCGGCTTGGCCGAC 420
NPUNMOD         ATTTACCCCACTGAAACCCCCCCCAGTACCTACCAACCCCGTGGAGTGCGGCTTGGCCGAC 414
                ************************************************************

NPM2_58         CTGTGGTCTCGCACCGCATTTACTAAATCTGTGGACTGGCGACTTCGATTCTTCGAGAGT 480
NPUNMOD         CTGTGGTCTCGCACCGCATTTACTAAATCTGTGGACTGGCGACTTCGATTCTTCGAGAGT 474
                ************************************************************

NPM2_58         ACCAAGAACCTCCTGGAAGAGTCTCTGTGGGAACTCGCCAATATCAACCAGGATCGAGTC 540
NPUNMOD         ACCAAGAACCTCCTGGAAGAGTCTCTGTGGGAACTCGCCAATATCAACCAGGATCGAGTC 534
                ************************************************************

NPM2_58         GCTAACCCCATCGAATACATCGAAATGCGCCGCAAGGTTGGTGGCGCTCCGTGGTCAGCG 600
NPUNMOD         GCTAACCCCATCGAATACATCGAAATGCGCCGCAAGGTTGGTGGCGCTCCGTGGTCAGCG 594
                ************************************************************

NPM2_58         GATCTCGTCGAACACGCCGTGTTTATAGAGATTCCGGCTGACATTGCCTCGACTCGGCCA 660
NPUNMOD         GATCTCGTCGAACACGCCGTGTTTATAGAGATTCCGGCTGACATTGCCTCGACTCGGCCA 654
                ************************************************************

NPM2_58         ATGCGTGTGCTAAAAGACACGTTTGCCGATGGAGTGCATTTACGCAACGACCTCTTCTCC 720
NPUNMOD         ATGCGTGTGCTAAAAGACACGTTTGCCGATGGAGTGCATTTACGCAACGACCTCTTCTCC 714
                ************************************************************

NPM2_58         TACCAGAGAGAAGTGGAAGATGAAGGCGAAAATGCCAACTGTGTGTTGGTCTTGGAGCGC 780
NPUNMOD         TACCAGAGAGAAGTGGAAGATGAAGGCGAAAATGCCAACTGTGTGTTGGTCTTGGAGCGC 774
                ************************************************************

NPM2_58         TTCTTAAATGTGAGTACCCAGGAGGCAGCTAACCTCACCAACGAACTGCTGACCTCGCGG 840
NPUNMOD         TTCTTAAATGTGAGTACCCAGGAGGCAGCTAACCTCACCAACGAACTGCTGACCTCGCGG 834
                ************************************************************

NPM2_58         TTATACCAGTTTGATAACACTGCCGTCACCGAGTTGCCCCCCCTCTTTGAGGAGTATGGA 900
NPUNMOD         TTATACCAGTTTGATAACACTGCCGTCACCGAGTTGCCCCCCCTCTTTGAGGAGTATGGA 894
                ************************************************************

NPM2_58         CTAGACCCAGTGGCTCGTGTGAACGTTCTCCTTTATATCAAAGGACTTCAGGACTGGCAG 960
NPUNMOD         CTAGACCCAGTGGCTCGTGTGAACGTTCTCCTTTATATCAAAGGACTTCAGGACTGGCAG 954
                ************************************************************

NPM2_58         TCGGGCGGTCACGAGTGGCACATGCGATCAAGCCGTTACATGAACAAGGGAGGAGACAAT 1020
NPUNMOD         TCGGGCGGTCACGAGTGGCACATGCGATCAAGCCGTTACATGAACAAGGGAGGAGACAAT 1014
                ************************************************************
```

```
NPM2_58         TCTCCAACATCTACGGTTCTGGGAGGACCCACAGGGCTAGGCACATCGGCTGCGCGGATT 1080
NPUNMOD         TCTCCAACATCTACGGTTCTGGGAGGACCCACAGGGCTAGGCACATCGGCTGCGCGGATT 1074
                ************************************************************

NPM2_58         GAATCGTTATATGCGGCCTTGGGTTTAGGAAGAATTAAAAGCTTCACTCACGTTCCATAC 1140
NPUNMOD         GAATCGTTATATGCGGCCTTGGGTTTAGGAAGAATTAAAAGCTTCACTCACGTTCCATAC 1134
                ************************************************************

NPM2_58         CAGCCTGTGGGACCAGTGACCCTGCCGAAGTTTTACATGCCGTTCACTACAAGTTTGAAT 1200
NPUNMOD         CAGCCTGTGGGACCAGTGACCCTGCCGAAGTTTTACATGCCGTTCACTACAAGTTTGAAT 1194
                ************************************************************

NPM2_58         CCTCATCTCAATGCCGCACGGAAGCATTCTAAGGAATGGGCGCGCCAGATGGGGATGCTG 1260
NPUNMOD         CCTCATCTCAATGCCGCACGGAAGCATTCTAAGGAATGGGCGCGCCAGATGGGGATGCTG 1254
                ************************************************************

NPM2_58         GAATCACTACCTGGGATTCCTGATGCCGTCATCTGGGATGACCACAAGTTTGATGTTGCC 1320
NPUNMOD         GAATCACTACCTGGGATTCCTGATGCCGTCATCTGGGATGACCACAAGTTTGATGTTGCC 1314
                ************************************************************

NPM2_58         GACGTAGCTCTCTGCGGTGCGTTGATCCATCCGAATGGGTCTGGTCTTGAACTGAATCTG 1380
NPUNMOD         GACGTAGCTCTCTGCGGTGCGTTGATCCATCCGAATGGGTCTGGTCTTGAACTGAATCTG 1374
                ************************************************************

NPM2_58         ACAGCGTGCTGGCTTGTTTGGGGAACCTATGCCGACGATTACTTCCCGGCGCTCTACGGG 1440
NPUNMOD         ACAGCGTGCTGGCTTGTTTGGGGAACCTATGCCGACGATTACTTCCCGGCGCTCTACGGG 1434
                ************************************************************

NPM2_58         AATAACCGCAACATGGCTGGTGCGAAAGTATTCAACGCTCGACTGTCGGCGTTCATGCCT 1500
NPUNMOD         AATAACCGCAACATGGCTGGTGCGAAAGTATTCAACGCTCGACTGTCGGCGTTCATGCCT 1494
                ************************************************************

NPM2_58         CTGGATGACTCCACCCCCAGCGAGGTTCCGACCAATCCAGTGGAAGCGGGCTTGGCAGAT 1560
NPUNMOD         CTGGATGACTCCACCCCCAGCGAGGTTCCGACCAATCCAGTGGAAGCGGGCTTGGCAGAT 1554
                ************************************************************

NPM2_58         ATTTGGTCTCGCACAGCTGGTCCCATGTCTGCCAACGCACGGACTCAGTTCCGCCGCGCA 1620
NPUNMOD         ATTTGGTCTCGCACAGCTGGTCCCATGTCTGCCAACGCACGGACTCAGTTCCGCCGCGCA 1614
                ************************************************************

NPM2_58         ATCCAGGATATGACTGACAGTTGGGTGTGGGAACTCGCCAACCAGATCCAAAATCGAATT 1680
NPUNMOD         ATCCAGGATATGACTGACAGTTGGGTGTGGGAACTCGCCAACCAGATCCAAAATCGAATT 1674
                ************************************************************

NPM2_58         CCTGACCCGATAGATTATGTTGAGATGCGCCGTAAGACCTTTGGCTCGGATCTGACCATG 1740
NPUNMOD         CCTGACCCGATAGATTATGTTGAGATGCGCCGTAAGACCTTTGGCTCGGATCTGACCATG 1734
                ************************************************************

NPM2_58         AGCCTGTCGCGACTAGCTCAGGGGAGCGAGATCCCGCAGGAGATTTACCGCACCCGAACG 1800
NPUNMOD         AGCCTGTCGCGACTAGCTCAGGGGAGCGAGATCCCGCAGGAGATTTACCGCACCCGAACG 1794
                ************************************************************

NPM2_58         ATGCGATCGCTCGATAATTCGGCCGCCGACTTCGCCTGTTTAACCAACGATATCTTTTCC 1860
NPUNMOD         ATGCGATCGCTCGATAATTCGGCCGCCGACTTCGCCTGTTTAACCAACGATTTCTTTTCC 1854
                ********************************************** ********

NPM2_58         TATCAGAAAGAAATCGAATTCGAGGGCGAAATCCATAACTGTGTGCTGGTCGTTCAGAAT 1920
NPUNMOD         TATCAGAAAGAAATCGAATTCGAGGGCGAAATCCATAACTGTGTGCTGGTCGTTCAGAAT 1914
                ************************************************************

NPM2_58         TTCCTCAACTGCGATTTGCCGCAGGCCGTCGAAGTTGTCAACAACCTGATGACCTCTAGG 1980
NPUNMOD         TTCCTCAACTGCGATTTGCCGCAGGCCGTCGAAAGTGTCAACAACCTGATGACCTCTAGG 1974
                ********************************  ************************

NPM2_58         GCGCTCCAGTTTCAACTCATCGTCGCCACCGAACTGCCAGTTCTTTTCGATGATTTCGAC 2040
NPUNMOD         GCGCTCCAGTTTCAACTCATCGTCGCCACCGAACTGCCAGTTCTTTTCGATGATTTCGAC 2034
                ************************************************************

NPM2_58         CTGGATGCAAGTACCCGCGAGAAACTGCTCGGATACGTCAAGAAACTAGAGCAGTGGATG 2100
NPUNMOD         CTGGATGCAAGTACCCGCGAGAAACTGCTCGGATACGTCAAGAAACTAGAGCAGTGGATG 2094
                ************************************************************
```

```
NPM2_58         TGCGGCGTCCTCAAGTGGCATATAACGGTAGACCGCTATAAGGAATTTGAATTGCGTAAT 2160
NPUNMOD         TGCGGCGTCCTCAAGTGGCATATAACGGTAGACCGCTATAAGGAATTTGAATTGCGTAAT 2154
                ************************************************************

NPM2_58         TCGTTAGCAGGGCGGCTACTTAGCGGTCCCAGAGGGCTGGGTACTTCAGCTAGGCGTATT 2220
NPUNMOD         TCGTTAGCAGGGCGGCTACTTAGCGGTCCCAGAGGGCTGGGTACTTCAGCTAGGCGTATT 2214
                ************************************************************

NPM2_58         GGATCGTTGATCGGTCAGGGGAGTCTCAAGTCATTACTCGGTCAATTACTCGAGCACCAC 2280
NPUNMOD         GGATCGTTGATCGGTCAGGGGAGTCTCAAGTCATTACTCGGTCAATGA------------ 2262
                ******************************************** *

NPM2_58         CACCACCACCACTGA 2295
NPUNMOD         --------------- 
```

Appendix 1-4. DNA sequence multiple alignment of experimental NMP2_58 and predicted NPUNMOD protein.

```
NPM2_58          MVMQPFELPEFYMPWPARLNPNLEAARSHSKAWAYQMGILGSKEEAESSVIWDERTFDAH 60
NPUNMOD          --MQPFELPEFYMPWPARLNPNLEAARSHSKAWAYQMGILGSKEEAESSVIWDERTFDAH 58
                   ************************************************************

NPM2_58          DYALLCSYTHPDAPGTELDLVTDWYVWVFFFDDHFLEIYKRTQDMAGAKEYLGRLPMFMP 120
NPUNMOD          DYALLCSYTHPDAPGTELDLVTDWYVWVFFFDDHFLEIYKRTQDMAGAKEYLGRLPMFMP 118
                 ************************************************************

NPM2_58          IYPTETPPVPTNPVECGLADLWSRTAFTKSVDWRLRFFESTKNLLEESLWELANINQDRV 180
NPUNMOD          IYPTETPPVPTNPVECGLADLWSRTAFTKSVDWRLRFFESTKNLLEESLWELANINQDRV 178
                 ************************************************************

NPM2_58          ANPIEYIEMRRKVGGAPWSADLVEHAVFIEIPADIASTRPMRVLKDTFADGVHLRNDLFS 240
NPUNMOD          ANPIEYIEMRRKVGGAPWSADLVEHAVFIEIPADIASTRPMRVLKDTFADGVHLRNDLFS 238
                 ************************************************************

NPM2_58          YQREVEDEGENANCVLVLERFLNVSTQEAANLTNELLTSRLYQFDNTAVTELPPLFEEYG 300
NPUNMOD          YQREVEDEGENANCVLVLERFLNVSTQEAANLTNELLTSRLYQFDNTAVTELPPLFEEYG 298
                 ************************************************************

NPM2_58          LDPVARVNVLLYIKGLQDWQSGGHEWHMRSSRYMNKGGDNSPTSTVLGGPTGLGTSAARI 360
NPUNMOD          LDPVARVNVLLYIKGLQDWQSGGHEWHMRSSRYMNKGGDNSPTSTVLGGPTGLGTSAARI 358
                 ************************************************************

NPM2_58          ESLYAALGLGRIKSFTHVPYQPVGPVTLPKFYMPFTTSLNPHLNAARKHSKEWARQMGML 420
NPUNMOD          ESLYAALGLGRIKSFTHVPYQPVGPVTLPKFYMPFTTSLNPHLNAARKHSKEWARQMGML 418
                 ************************************************************

NPM2_58          ESLPGIPDAVIWDDHKFDVADVALCGALIHPNGSGLELNLTACWLVWGTYADDYFPALYG 480
NPUNMOD          ESLPGIPDAVIWDDHKFDVADVALCGALIHPNGSGLELNLTACWLVWGTYADDYFPALYG 478
                 ************************************************************

NPM2_58          NNRNMAGAKVFNARLSAFMPLDDSTPSEVPTNPVEAGLADIWSRTAGPMSANARTQFRRA 540
NPUNMOD          NNRNMAGAKVFNARLSAFMPLDDSTPSEVPTNPVEAGLADIWSRTAGPMSANARTQFRRA 538
                 ************************************************************

NPM2_58          IQDMTDSWVWELANQIQNRIPDPIDYVEMRRKTFGSDLTMSLSRLAQGSEIPQEIYRTRT 600
NPUNMOD          IQDMTDSWVWELANQIQNRIPDPIDYVEMRRKTFGSDLTMSLSRLAQGSEIPQEIYRTRT 598
                 ************************************************************

NPM2_58          MRSLDNSAADFACLTNDIFSYQKEIEFEGEIHNCVLVVQNFLNCDLPQAVEVVNNLMTSR 660
NPUNMOD          MRSLDNSAADFACLTNDIFSYQKEIEFEGEIHNCVLVVQNFLNCDLPQAVESVNNLMTSR 658
                 ********************************************************* ********

NPM2_58          ALQFQLIVATELPVLFDDFDLDASTREKLLGYVKKLEQWMCGVLKWHITVDRYKEFELRN 720
NPUNMOD          ALQFQLIVATELPVLFDDFDLDASTREKLLGYVKKLEQWMCGVLKWHITVDRYKEFELRN 718
                 ************************************************************

NPM2_58          SLAGRLLSGPRGLGTSARRIGSLIGQGSLKSLLGQLLEHHHHHH 764
NPUNMOD          SLAGRLLSGPRGLGTSARRIGSLIGQGSLKSLLGQ--------- 753
                 ************************************
```

Appendix 1-5. Amino acids sequence multiple alignment of experimental NMP2_58 and predicted NPUNMOD protein.

```
P.calcicola         TTCTTCGACGACCACTTCCTTGAAATCTATAAGCGTACCCAAGACATGGCTGGGGCGAAG 60
N.punctiforme       TTCTTCGACGACCACTTCCTTGAAATCTATAAGCGTACCCAAGACATGGCTGGGGCGAAG 60
A.laxa              TTCTTCGACGATCACTTCCTGGAAATCTATAAACGCAGTCAGGATATGGTTGGGGCGAAG 60
Nostoc.sp           TTCTTCGACGATCACTTCCTTGAGCTTTACAAGCGTAGCCAAGATATGGCTGGGGCTAAG 60
SCO6073             TTCTTCGACGACCACTTCCTGGAGAAGTACAAACGCAGCCAGGACCGCCTCGCCGGCAAG 60
SAV2163             TTCTTCGACGACCACTTCCTGGAGACGTTCAAGCGCACCCAGGACCGCGAGGGCGGCAAG 60
                    *********** ******** **    *  ** ** *  ** **       *  *  ***

P.calcicola         GAGTATCTCGGCAGATTACCAATGTTCATGCCAATTTACCCCACTGAAACCCCCCCAGTA 120
N.punctiforme       GAGTATCTCGGCAGATTACCAATGTTCATGCCAATTTACCCCACTGAAACCCCCCCAGTA 120
A.laxa              GAGTATCTTGACCGACTCCCCGCATTTATGCCGATTTATCCCAGGGACAACCCTCTCGTT 120
Nostoc.sp           AAGTACCTCGATGGACTGCCAGCGTTTATGCCTGTACAGCCACAAGAAACGCCGCCAGAA 120
SCO6073             GCCCACCTGGACCGGCTCCCGCTGTTCATGCCGCTCGACGACGCCGCCGGGATGCCCGAG 120
SAV2163             GCCTATCTGGACCGGCTGCCGCTCTTCATGCCGCTCGACCTGTCGGCACCCGTTCCGGAG 120
                        * ** *   *  * **    ** *****  *  *       *       *  *

P.calcicola         CCTACCAACCCCGTGGAGTGCGGCTTGGCCGACCTGTGGTCTCGCACCGCATTTACTAAA 180
N.punctiforme       CCTACCAACCCCGTGGAGTGCGGCTTGGCCGACCTGTGGTCTCGCACCGCATTTACTAAA 180
A.laxa              CCCACCAACCCAGTAGAGCGCGGTTTGGCTGACTTGTGGTCTCGCACCGCATTTACTAAG 180
Nostoc.sp           CCCACCAATGCCGTAGAACGGGGTTTAGTAGACCTCTGGGATCGCACCATCCCCAATGCA 180
SCO6073             CCGCGGAACCCGGTGGAGGCCGGACTCGCCGACCTGTGGACCCGCACGGTGCCCGCGATG 180
SAV2163             CCGGAGAATCCGGTCGAGGCGGGCCTCGCCGACCTGTGGGCGCGCACGGTGCCCGCGATG 180
                    **    **  * ** **    **  * *  *** * ***   *****

P.calcicola         TCTGTGGACTGGCGGCTTCGATTCTTCGGGAGTACCAAGAACCTCCTGGAAGAGTCTCTG 240
N.punctiforme       TCTGTGGACTGGCGGCTTCGATTCTTCGGGAGTACCAAGAACCTCCTGGAAGAGTCTCTG 240
A.laxa              TCTGTGGAATGGCGGCGACGATTCTTTAAAAGTACCAAAAATCTTTTAGATGAGTCAATG 240
Nostoc.sp           TCCCAAGATTGGGTAATCCGGTTTTCTGAGAGTACCATCAATCTTCTCAAAGAATCTCAG 240
SCO6073             TCGGCCGACTGGCGCCGCCGCTTCGCCGTCGCCACCGAGCACCTCCTCAACGAGTCCATG 240
SAV2163             TCGGCGGACTGGCGCAAGCGGTTCGCCGTCTCGACCGAGCACCTGCTCAACGAGTCGCTG 240
                    **    ** ***      ** **          ***    * **  *  * ** **   *

P.calcicola         TGGGAACTCGCCTACATCAACCAGGATCGAGTCGCTAACCCCATCGAATACATCGAAATG 300
N.punctiforme       TGGGAACTCGCCTACATCAACCAGGATCGAGTCGCTAACCCCATCGAATACATCGAAATG 300
A.laxa              TGGGAACTGGCCAACATCAATCAAAATCGAATTGCTAACCCCATCGAATACATTGAGATG 300
Nostoc.sp           TGGGAACTCGCCAATATCAGCCAAAATCGAGTTGCTAACCCGATTGAGTACATTGAGATG 300
SCO6073             TGGGAGCTGTCCAACATCAACGAGGGGCGGGTCGCCAACCCGGTCGAGTACATCGAGATG 300
SAV2163             TGGGAGCTGTCCAACATCAACGAGGGGCGGATCGCCAACCCCGTCGAGTACATCGAGATG 300
                    ***** **  ** * ****   *    **  * ** *****  * ** ***** ** ***

P.calcicola         CGCCGCAAGGTTGGTGGCGCTCCGTGGTCAGCGGATCTCGTCGAACACGCCGTGTTTATA 360
N.punctiforme       CGCCGCAAGGTTGGTGGCGCTCCGTGGTCAGCGGATCTCGTCGAACACGCCGTGTTTATA 360
A.laxa              CGGCGTAAAGTTGGTGGCGCACCCTGGTCAGCCGATCTGGTGGAACACGCCGCGTTCGTG 360
Nostoc.sp           CGCCGTAAGGTGGGGGGAGCGCCCTGGTCAGCTAACCTAGTGGAACACGCGGTTGGGGCA 360
SCO6073             CGCCGCAAGGTCGGCGGCGCCCCGTGGTCGGCCGGGCTCGTGGAGTACGCGAC---CGCC 357
SAV2163             CGCCGCAAGGTCGGCGGAGCCCCCTGGTCGGCGGGGCTCGTCGAGTACGCGAC---CGCC 357
                    ** ** ** ** ** ** ** ** ***** **    ** ** **  ****

P.calcicola         GAGATTCCGGCTGACATTGCCTCGACTCGGCCAATGCGTGTGCTAAAAGACACGTTTGCC 420
N.punctiforme       GAGATTCCGGCTGACATTGCCTCGACTCGGCCAATGCGTGTGCTAAAAGACACGTTTGCC 420
A.laxa              GAAGTCCCGGCTAAAATTGCGGCAACTAGACCAATGCGGGTTTTGAAAGACACATTTGCT 420
Nostoc.sp           GACATTCCGGCTGCGATCGCCCCGACCCGACCAATGCGTGTCCTCAAAGACACTTTTTCT 420
SCO6073             GAGGTGCCCGCCGCCGTCGCCGGGACCAGGCCGCTCAGGGTGCTGATGGAGACGTTCTCC 417
SAV2163             GAAGTGCCGGCCGCCGTCGCCGGGTCGAGGCCGCTGCGGGTGCTGATGGAGACCTTCTCC 417
                    **  * ** **     * **     *  * **  *  * **  * *  ** ** **  *

P.calcicola         GATGAAGTGCATTTACGCAACGACCTCTTCTCCTACCAGAGAGAAGTGGAAGATGAAGGC 480
N.punctiforme       GATGAAGTGCATTTACGCAACGACCTCTTCTCCTACCAGAGAGAAGTGGAAGATGAAGGC 480
A.laxa              GATGGAGTACATCTCCGCAACGACCTATTCTCCTATCAAAGAGAAGTGGAAGATGAAGGT 480
Nostoc.sp           GATGGGGTACATCTGCGAAATGATATCTTCTCTTACCAAAGAGAAGTGGAAGAAGAAGGG 480
SCO6073             GACGCCGTGCACCTGCGCAACGACCTCTTCTCCTACCAGCGCGAGGTCGAGGACGAGGGC 477
SAV2163             GACGGGGTGCACCTGCGCAACGACCTGTTCTCGTACCAGCGGGAGGTCGAGGAGGAGGGT 477
                    ** *  ** **  * ** ** **  * ***** ** **  * ** ** ** ** ** **

P.calcicola         GAAAATGCCAACTGTGTGTTGGTCTTGGAGCGCTTCTTAAATGTGAGTACCCAGGAGGCA 540
N.punctiforme       GAAAATGCCAACTGTGTGTTGGTCTTGGAGCGCTTCTTAAATGTGAGTACCCAGGAGGCA 540
A.laxa              GAAAATTCTAATTGTGTGCTTGTAATTGAAAAATTCTTGAATGTAAGTACCCAAGAGGCG 540
Nostoc.sp           GAAAATGCTAACTGTATCCTGGTTTTAGAACGATTCCTGGATGTCAGTACCCAAGAGGCG 540
SCO6073             GAGCTGAGCAACGGGGTGCTGGTGTTGGAGACCTTCTTCGGCTGCACCACCCAGGAGGCC 537
SAV2163             GAGCTGAGCAACGGCGTCCTCGTCCTGGAGACCTTCTTCGGCTGTACGACGCAGGAGGCC 537
                    **       **  *  *  * **  * **    *** *       *  ** ** *****
```

```
P.calcicola         GCTAACCTCACCAACGAACTGCTGACCTCGCGGTTATACCAGTTTGATAACACTGCCGTC 600
N.punctiforme       GCTAACCTCACCAACGAACTGCTGACCTCGCGGTTATACCAGTTTGATAACACTGCCGTC 600
A.laxa              GCTAACCTCACTAACGAACTACTCAACTCCCGTTTATATCAGTTTGATAACACTGCTGTC 600
Nostoc.sp           GCCAATCTCACTAACGACTTACTAACATCTAGGGTGCAACAGTTCGAGAACACCTTTGTC 600
SCO6073             GCCGACCTGGTCAACGACGTCCTCACCTCGCGGCTGCACCAGTTCGAGCACACCGCGTTC 597
SAV2163             GCCGAGACCGTCAACGACATCCTCACCTCGCGGCTCCACCAGTTCGAGCACACGGCGCTC 597
                    **  *       *****  * ** *  **  *  *  * ***** **  ****     **

P.calcicola         ACCGAGTTGCCCCCCCCTCTTTGAGGAGTATGGACTAGACCCAGTGGCTCGTGTGAACGTT 660
N.punctiforme       ACCGAGTTGCCCCCCCCTCTTTGAGGAGTATGGACTAGACCCAGTGGCTCGTGTGAACGTT 660
A.laxa              ACCGAGTTGCCCCTCCCTTTTTGAGGAGTATGGAGTAGACCCAGTAGAGCGTGTGAATGTT 660
Nostoc.sp           ACTGAGCTTCCTTCTTTGTTTGAGGAATATAGTCTGAGTCCCGATGAGCGTCTCAAAGTA 660
SCO6073             ACCGAGGTGCCCGCCGTCGCCCTGGAGAAGGGCCTGACCCCGTTGGAGGTCGCCGCCGTC 657
SAV2163             ACCGAAGTGCCCGCGCTGGCCCTGGAGAAAGGCCTCACCCCACCCGAGGTCGCGGCGGTC 657
                    ** **  * **  *  *      ***  *  *  *    **    *           **

P.calcicola         CTCCTTTATATCAAAGGACTTCAGGACTGGCAGTCGGGCGGTCACGAGTGGCACATGCGA 720
N.punctiforme       CTCCTTTATATCAAAGGACTTCAGGACTGGCAGTCGGGCGGTCACGAGTGGCACATGCGA 720
A.laxa              CTCCTTTACATCAAAGGACTTCAGGATTGGCAATCTGGCGGTCATGAGTGGCACATGAGA 720
Nostoc.sp           GTCCTATATGCCAAAGGACTTCAGGACTGGCAGTCGGGGGGTCATGAGTGGCACATGAGA 720
SCO6073             GGCGCGTACACGAAGGGCCTCCAGGACTGGCAGTCCGGCGGCCACGAGTGGCACATGCGT 717
SAV2163             GCGGCGTACGCGCGGGGACTTCAGGACTGGCAGTCGGGCGGCCACGAATGGCATCTGAGG 717
                          **       ** ** ***** ***** ** ** ** ** ** *****  ** *

P.calcicola         TCAAGCCGTTACATGAACGAGGG 743
N.punctiforme       TCAAGCCGTTACATGAACGAGGG 743
A.laxa              TCAAGCCGTTACATGAACGAGGG 743
Nostoc.sp           TCGAGCCGTTACATGAACGAGGG 743
SCO6073             TCCAGCCGCTACATGAACAAGGG 740
SAV2163             TCCAGCCGCTACATGAACGAGGG 740
                    ** ***** ********* ****
```

Appendix 1-6. Reference strain DNA sequences from putative geosmin synthases determined from G-PCR1.