ResearchOnline@JCU

JAMES COOK UNIVERSITY AUSTRALIA

This file is part of the following work:

Beltran Ramirez, Victor (2010) ***Molecular aspects of the fluorescent protein homologues in Acropora millepora.*** **PhD Thesis, James Cook University.**

Access to this file is available from:

https://doi.org/10.25903/mhdb%2Drb84

Copyright © 2010 Victor Beltran Ramirez

The author has certified to JCU that they have made a reasonable effort to gain permission and acknowledge the owners of any third party copyright material included in this document. If you believe that this is not the case, please email researchonline@jcu.edu.au

MOLECULAR ASPECTS OF THE FLUORESCENT PROTEIN HOMOLOGUES IN *Acropora millepora*

Thesis submitted by
Víctor Hugo Beltrán Ramírez
BSs (Hons) UABCS, MSc UNAM
Feb 2010

For the degree of **Doctor of Philosophy**
In the School of Pharmacy and Molecular Sciences
James Cook University

STATEMENT OF ACCESS

I, the undersigned author of this work, understand that James Cook University will make this thesis available for use within the University Library and, via the Australian Digital Theses network, for use elsewhere.

I understand that, as an unpublished work, a thesis has significant protection under the Copyright Act and;

I do not wish to place any further restriction on access to this work

_____ _____

Signature Date

STATEMENT ON SOURCES DECLARATION

DECLARATION

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Victor Hugo Beltran Ramirez
24th of February 2010

ELECTRONIC COPY

I, the undersigned, the author of this work, declare that the electronic copy of this thesis provided to the James Cook University Library, is an accurate copy of the print thesis submitted, within the limits of the technology available.

_____ _____

Signature Date

ACKNOWLEDGEMENTS

During my PhD, I have experienced many good moments that assisted me to continue, the long and sometimes hard process of learning. Dr. David miller, played an important role in the way I approached this research and the final outcome.

I am grateful with my laboratory mates, who constantly hear my new ideas, technical problems, and during spawning season, helping in an unconditional way to get the most amazing coral babies being raised and “observed” under our evo-devo criteria, thank you all: Chuya, Akira, Hiroki, Lucija, Alejandro, Danielle, Zoe, Brent, Lubna and Nikki.

I’m also want to express my gratitude to the following Drs: Eldon Ball, Madeleine van Oppen, David Hayward, Kerstin Heinman, Jim Burnell, Richard Keene, Alan Baxter, Misha Matz, Jörg Wiedenmann, Michael Anctil and David Yellowlees, each of them were in one way or other responsible of some contribution that enriched my work.

To my wife Griselda, my parents and my siblings who always encourage me during difficult times.

Finally, the Mexican Government through a CONACyT scholarship No.144227 supported my PhD.

ABSTRACT

GFP-like proteins are responsible for some of the most spectacular colours displayed in coral reefs all over the planet. In here, the GFP repertoire from the scleractinian *Acropora millepora* was molecularly and phenotypically characterised in the context of embryonic development.

Phylogenetic analyses demonstrated the existence of two major clades; corresponding to the fluorescent and non-fluorescent chromoprotein genes respectively. The cDNA sequences were identified from a large EST library constructed using the pre settlement stage of *A. millepora* larvae. Both the fluorescent and non-fluorescent chromoprotein genes have highly similar intron-exon structure, thus, indicating a relatively recent common origin. Gene comparisons against the gene bank indicated that the larval GFPs are genetically distinct to their adult homologs.

Representative clones encoding green -amilGFP517 a, b-, red fluorescent -amilRFP602- and two blue chromoproteins -amilCP597 and amilCP601- were expressed in a bacterial system and biochemically characterized. Importantly the recombinant chromoproteins were shown to transform from blue non fluorescence to red fluorescent, after dehydration and further excitation with green/red light.

In situ hybridisation studies indicated that the chromoprotein genes were expressed early in development and predominantly in the endoderm. In contrast, the expression of the genes encoding fluorescent proteins was initiated later and the mRNA was detected mainly in the ectoderm. Although, the high mRNA similarity between GFP and RFP hampered a clear detection of gene expression; the corresponding expression of the proteins was observed to be axially restricted, as in planulae intense green or red fluorescence were associated with the oral and aboral regions and fluorescence patterns were restricted to specific cells. After metamorphosis, the red fluorescence associated with the aboral ectoderm disappeared, but the green signal remained in oral tissue.

Light/dark exposure experiments indicated that embryos reared in darkness showed some reduction of fluorescent-like gene expression whereas chromoprotein expression was unaffected

TABLE OF CONTENTS

STATEMENT OF ACCESS ii
STATEMENT ON SOURCES DECLARATION iii
ACKNOWLEDGEMENTS iv
ABSTRACT v
TABLE OF CONTENTS vi
LIST OF FIGURES ix
LIST OF TABLES x
Chapter 1 1
General Introduction 1
1.1 The colour in cnidarians 1
1.2 GFP-like protein structure and classification 3
1.3 Evolution of the GFP-like proteins 3
1.4 Ecological and biological aspects of GFP-like proteins in corals 5
1.5 Importance of evolutionary developmental studies (Evo-Devo) 6
Chapter 2 8
Material and Methods 8
2.1 Nuclear gene cloning, sequencing and assembling 8
2.1.1 [α-^{32}P] radioactive probe generation by PCR 8
2.1.2 Agarose gel electrophoresis and storage of radioactive probes 9
2.1.3 Competent cell preparation for viral infection 9
2.2 Tittering the genomic libraries 10
2.2.1 Phage library titration 10
2.2.2 Cosmid library titration 10
2.2.3 Acropora genomic Cosmid library amplification and storage 11
2.3 Genomic λ lambda library screening 11
2.3.1 Blotting and marking of membranes 11
2.3.2 Blotting and marking of membranes (Cosmid library) 11
2.3.3 Preparation of the membranes for phage DNA hybridisation 12
2.3.4 Preparation of the membranes for cosmid DNA hybridisation 12
2.4 Hybridisation and DNA recovery of the positive plaques 12
2.4.1 DNA Hybridisation 12
2.4.2 Washing the unbound products and detection 13
2.4.3 Secondary screening 14
2.4.4 DNA purification from the nuclear genomic clones 14
2.4.5 DNA precipitation and storage 15
2.4.6 Restriction enzyme digestions genomic fragments to detect differences 15
2.4.7 Southern blot and subcloning 15
2.4.8 Sequencing and gene assembling 16
2.5 Recombinant protein characterization 17
2.5.1 PCR amplification of the amilGFP517a from the presettlement cDNA library 17
2.5.2 Gel extraction and ligation into PCR-based propagation vector 18
2.5.3 Cloning of the GFP-like proteins from the EST collection 18
2.5.4 PCR-based introduction of restriction sites and subcloning into expression vector 19
2.5.5 Dephosphorylation of the expression vectors 20
2.5.6 Cell strain preparation and transformation 20
2.5.6.1 Competent cells preparation and its storage 20
2.5.6.2 Bacterial transformation with plasmid vectors 21
2.5.7 Screening and protein purification 21

2.5.7.1 Bacterial visual screening for the presence of target insert..... 21
2.5.7.2 Bacterial re-protein purification 22
2.5.7.3 Histidine affinity chromatography 23
2.5.8 Recombinant protein characterization 23
2.5.8.1 Electrophoretic characterization in SDS-PAGE assays 23
2.5.8.2 Spectroscopic characterization of the recombinant proteins..... 24
2.6 *In vivo* and *in situ* detection of the GFP-like genes 25
2.6.1 Coral colony sampling and general embryo manipulation 25
2.6.2 Live fluorescence microscopy and embryo fixation during the coral development 25
2.6.3 Riboprobe creation (DIG labelling protocol)..... 26
2.6.4 Riboprobe fractionation 27
2.6.5 Whole mount *in situ* hybridisation..... 28
2.6.5.1 Coral embryo preparation 28
2.6.6 Whole mount *in situ* hybridisation..... 28
2.6.7 Washing the unbound material form the preparations 29
2.6.8 Detection and documentation of the spatial mRNA signal. 29
2.7 Northern blotting procedures 30
Chapter 3 31
Characterisation of two GFP loci from the coral *Acropora millepora* 31
3.1 Introduction 31
3.1.1 Statement of Goals 32
3.2 Results 33
3.2.1 Gene structure comparison across the phylum Cnidaria 33
3.2.3 Evolutionary relationships between GFP-like genes 39
3.3 Discussion..... 43
3.3.1 *Acropora* GFP-like loci: An ancient gene structure 43
3.3.2 Possible transcriptional GFP regulation 45
3.3.3 Phylogenetic relationships of *Acropora* GFPs 46
3.4 Conclusion 49
Chapter 4 50
Biochemical and spectral analyses of different recombinant GFP-like proteins from the Acropora presettlement cDNA library 50
4.1 Introduction 50
4.1.1 Statement of Goals 52
4.2 Results 53
4.2.1 GFP-like protein diversity observed in larvae ESTs 53
4.2.2 3D representation of the observed diversity in each contig 57
4.2.3 Molecular masses of the proteins..... 62
4.2.4 Spectral characterization of GFP-like recombinant proteins from *Acropora millepora* 64
4.2.4.1 Spectral signature of *A. millepora* green FP (AmilGFP517a, b)..... 64
4.2.4.2 Spectral signature of *A. millepora* red FP (AmilRFP602)..... 66
4.2.4.3 Spectral signature of *A. millepora* CP (AmillCP597)..... 68
4.2.4.4 Spectral signature of *A. millepora* CP (AmilCP601) 68
4.2.4.5 Spectral characterization of the recombinant proteins under extreme pH conditions..... 70
4.3 Discussion..... 72
4.3.1 Amino acid substitutions and the functions and phenotypes of the protein72
4.3.2 Acropora GFP-like protein electrophoretic signature 80
4.3.3 The AmilRFP602 spectral signature: similarities and hypotheses 81
4.3.4 Larval AmilGFP517a, b: The most common emission for a coral GFP-like protein 82
4.3.5 The amilCP products: similarities with other cnidarian proteins and potential FPs 83

4.3.6 Chemical nature of the GFP-like chromophore: hypsochromic pH-dependent shift 84
4.3.7 The 3D structure of GFP and its nearest structural protein neighbour 85
4.4 Conclusions 85
Chapter 5 86
Temporal and spatial mRNA expression and fluorescence analyses of GFP-like genes during *Acropora millepora* embryogenesis 86
5.1 Introduction 86
5.1.1 Statement of Goals 90
5.2 Results 90
5.2.1 The major GFP gene types are differentially expressed in early coral development 90
5.2.2 Fluorescence microscopy of GFP variants 93
5.4 Discussion 96
5.4.1 GFP-like message temporal regulation 96
5.4.2 Nature of the observed colouration in the early stages 96
5.4.3 Green-red signal distribution in planulae 97
5.4.4 Non-endodermal *in situ* staining; no fluorescent product was observed ... 98
5.4.5 GFP-like protein-expressing cell types in *Acropora* 98
5.4.6 Possible interaction of position-governing genes and differential GFP-like protein expression along the anteroposterior axis in *Acropora* embryos 101
5.4.7 Cell surface array 102
5.5 Conclusions 103
Chapter 6 104
General discussion 104
6.1 Differential exon evolution, functionality, origins and regulators of the GFP-like genes. 104
6.2 Protein world 105
6.3 Too much attention to the core 106
6.4 Environmental influences on early embryogenesis: does light controls GFP expression? 106
6.5 Future directions 108
References 109
Appendix 119

LIST OF FIGURES

Figure 1.1 Adult *A. millepora* corals observed under fluorescence (GFP2 filter). 2
Figure 3.1 Cloning of the nuclear GFP-like genes. 35
Figure 3.2 Comparison of the intron position over the amino acid alignment from several sequenced cnidarian GFP-like genes. 36
Figure 3.3 Upstream region (from the start codon) comparison of GFP-like genes from anthozoans. 38
Figure 3.4 Phylogenetic analyses of the GFP-like genes using the neighbour joining algorithm to construct the tree. 41
Figure 3.5 Schematic representation of the *Acropora millepora* GFP-like loci and putative gene neighbours. 47
Figure 3.6: Schematic representation of the *Nematostella vectensis* GFP-like loci. 48
Figure 4.1 Amino acid alignments of all complete clones that belong to the Chromo-like contig. 55
Figure 4.2 Amino acid alignments of all complete clones that belong to the Fluorescent-like contig. 56
Figure 4.3 3D model of the CP recombinant proteins and the substitutions observed in the whole contig. 60
Figure 4.4 3D model representation of fluorescent recombinant proteins, including the observed diversity of amino acid substitutions inside the contig. 61
Figure 4.5 Phenotypic comparison of the expressed recombinant GFP-like proteins. 63
Figure 4.6 Spectral characterization of amilGFP517a and amilGFP517b proteins in solution (PBS, pH 7.2) at room temperature. 65
Figure 4.7 Spectral characterization of recombinant AmilRFP602 in solution (PBS, pH 7.2) at room temperature. 67
Figure 4.8 Spectral characterization of two recombinant CPs measured in solution (PBS, pH 7.2) at room temperature. 69
Figure 4.9 pH-dependent behaviour of the recombinant proteins in solution at room temperature. 71
Figure 4.10 Amino acid alignment from representative GFP homologues across Metazoa. 73
Figure 5.1 Embryological development of *Acropora millepora*. 88
Figure 5.2 *In situ* and *in vitro* time-dependent detection of the GFP-like mRNAs in *Acropora* development. 92
Figure 5.3 *In vivo* observation of the principal morphological and fluorescence characteristics of *Acropora millepora* embryogenesis. 95
Figure 6.1 Spectral characteristics of the recombinant proteins from *Acropora millepora* presettlement library. 119
Figure 6.2 Box shade alignment between larval and adult *Acropora* fluorescent proteins. 120
Figure 6.3 Prawn chip coloured embryos. 120
Figure 6.4 38h-FITC coral embryo 120
Figure 6.5 Red granulations in planula cells. 120
Figure 6.6 Planula section stained against amilFP 120

LIST OF TABLES

Table 1. Primers characteristics; CP correspond to chromoproteins and FP to fluorescence genes. 19

Table 2. Filters used for live fluorescence 26

Table 3. Developmental stages of *Acropora millepora* 28

Table 4. Information related to all GFP-like proteins used to construct the general interphylum evolutionary tree. 42

CHAPTER 1

GENERAL INTRODUCTION

1.1 The colour in cnidarians

Scleractinian corals and many other cnidarians found in tropical reef areas, frequently display intense coloration such as blue, green, pink, red or violet. In addition to the overall coloration of the organism, particular zones of the polyp can be differentially coloured. These vivid colours are almost exclusively due to local expression of particular proteins belonging to the green fluorescent protein (GFP) family. The first member of such proteins was isolated from the hydromedusa *Aequorea victoria* (Prasher et al. 1992). In this jellyfish, the GFP is coupled to the calcium-sensor protein, aequorin (AEQ). In presence of calcium, a photochemical reaction occurs and AEQ emits blue photons, which are absorbed by the GFP chromophore and in turn this energy is re-emitted in the form of intense green fluorescence (Prasher et al. 1992, Wilson & Hastings 1998, Prendergast 2000). In nature, these pulses of light associated with a body part, usually around the mouth and basal part of the tentacles, and are used for predation and protection purposes (Morin 1974).

In contrast to the hydrozoan jellyfish *A. victoria*, the staghorn coral: *Acropora millepora* and other reef building are not bioluminescent organisms but many of its GFP-like proteins have the potential to emit fluorescence when excited with the appropriate light. Further, fluorescence energy transfer or FRET occurs between the batteries of differently coloured proteins (located on both tissue layers but preferentially at the endoderm associated with the symbiont algae), leading to a combinatorial pattern of fluorescence polyp under appropriate conditions (Cox & Salih 2005), *i.e* in some orange/yellow *A. millepora* colonies, each polyp mouth is green fluorescent and the rim of each corallite is also green fluorescent (Figure 1.1a, b). Yellow or green colonies shown green endodermal fluorescence in the coenostum area, and at the tentacle tips (Figure 1.1c). In a red colony showing the polyps tentacles extended, complete endodermal green fluorescence is observed (Figure 1.1d).

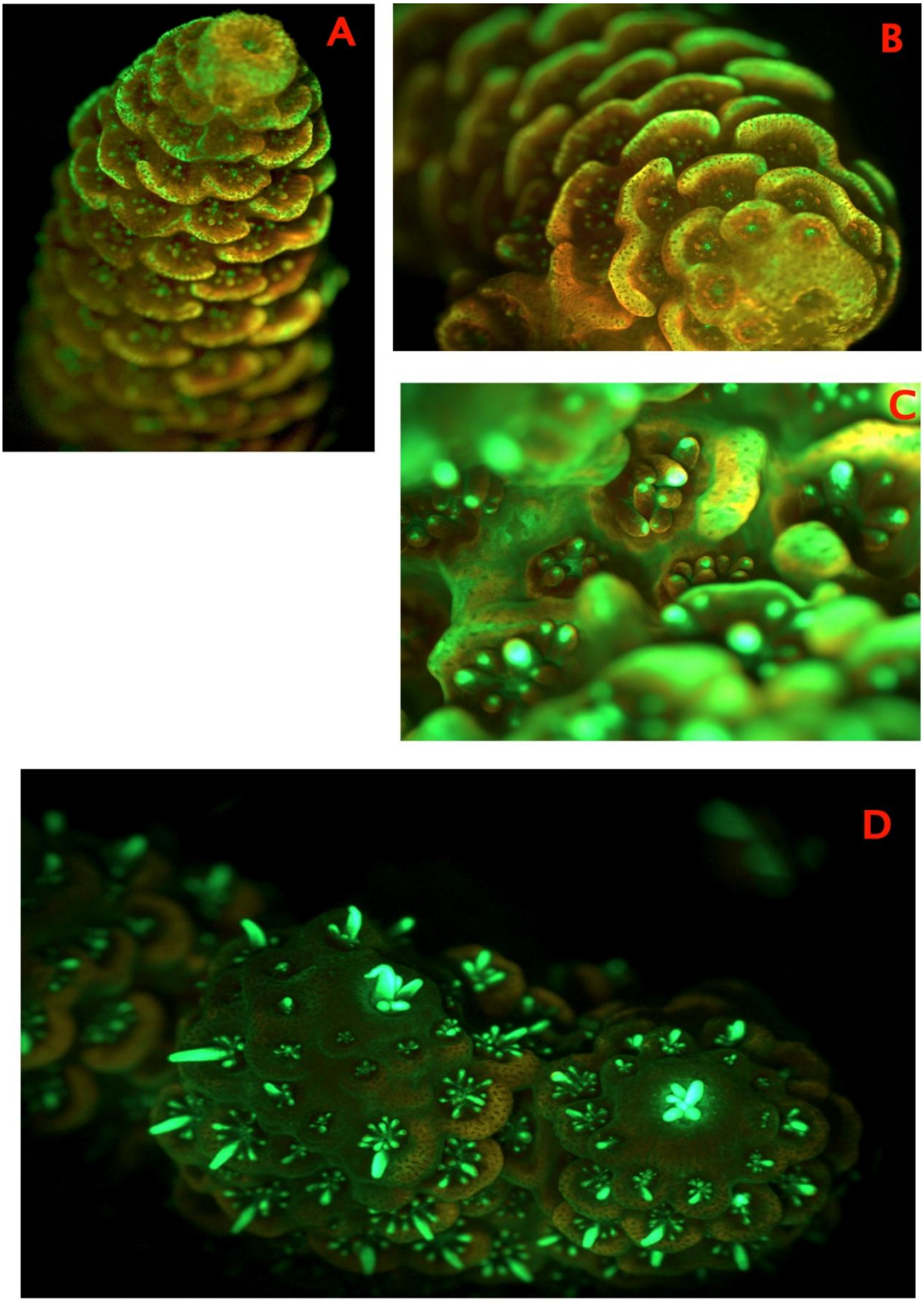

Figure 1.1 Adult *A. millepora* corals observed under fluorescence (GFP2 filter).
A-D) Side, top, close-up, and top view of live coral branches respectively

1.2 GFP-like protein structure and classification

To date, quite a large number of GFP-like proteins have been cloned from anthozoans such as scleractinian corals, soft corals and anemones. These animals have provided the richest source of GFP proteins and GFPs awaken biotechnological interest due to their capacity of auto catalytically form a chromophore by an internal reaction between key amino acids and without any additional reactive but molecular oxygen to complete its biogenesis (Lukyanov et al. 2000, Lukyanov et al. 2005, Bulina et al. 2006). This extraordinary property of self-assembly allows to use GFPs as tags when fused to target proteins, thus, permitting visualisation within the living cell where the fluorescence protein is being expressed, and therefore track *i.e* protein interactions, gene expression etc. This technology has been successfully applied in several model organisms, creating a total revolution in the way that biological processes are observed (van Roessel & Brand 2002, Verkhusha & Lukyanov 2004).

The basic anatomy of GFP is described by eleven beta sheets, folded forming a cylinder-like shape structure, the interior of which is protected from the surrounding environment and locate the chromophore (see three dimension reconstruction of this protein in the chapter 4). According to the chromophore structure, the protein can be classified as fluorescent -*i.e* cyan, green, yellow, and red- and non fluorescent chromoproteins displaying colorations such as purple and blue- (Labas et al. 2002). Some of these proteins, by random and directed mutagenesis, can be transformed from non-fluorescent to fluorescent type, with the advantage of some far-red mutants suitable for multiple labelling with low auto fluorescence (Fradkov et al. 2000, Bulina et al. 2002). Other mutations, in turn, favours the shift of colour fluorescence with the time, acting as a timer (Terskikh et al. 2000). The compact and enclosed beta barrel conformation of GFP-like proteins, confers extended life to the fluorescence when the protein is exposed to drastic changes in temperature or pH. Hydrozoan GFPs, are mainly monomeric whereas in most Anthozoans the protein complex is tetrameric. Each beta barrel weight in the range of 29-33 kDa and composed by 220-240 amino acids (Wiedenmann 2000, Matz et al. 2002, Shkrob et al. 2005).

1.3 Evolution of the GFP-like proteins

The identification of GFP homologues in several bilaterians such as crustacean and amphioxus (Masuda et al. 2006, Deheyn et al. 2007), indicates that the GFP-like

protein family is likely to predate the split of the Cnidaria and the Bilateria. However, the evolutionary history of this family of proteins, is fraught with complications; the phylogenetic analyses at the protein level of GFP-like genes from many organisms have yielded topology trees that are difficult to reconcile with formal taxonomy (Shagin et al. 2004) and that ultimately have suggested multiple origins of red fluorescent and blue chromoproteins. Secondly, the evolutionary scenario complicates further due to the amazing topological similarity found between the GFP-like proteins and the G2F Nidogen domain.

The G2F protein domain has no chromophore and functions as a binding module or laminins, collagens, and other protein components of basement membranes, which together form a thin extracellular layer that supports the nervous system (Ekblom et al. 1994, Kadoya et al. 1997, Kim & Wadsworth 2000, Hopf et al. 2001). The biological relevance of these molecules became apparent when Kim and Wadsworth (2000), using the nematode *Caenorhabditis elegans* as a model, demonstrated that deletion of the *nid-1* gene developed nerves in the wrong position.

The remarkably, similarity between G2F and the GFP-like proteins suggest common ancestry, however it is still unclear whether the G2F domain lost its capacity to form the chromophore and was recruited as binding element or if the GFP-like protein gained the capacity of autocatalytic chromophore biogenesis and latter was recruited to the bioluminescence system, as a secondary emitter (in some marine hydrozoans) or as, still debated, photo-protector/photo-enhancer mechanism in many reef dweller non-bioluminescent cnidarians (Salih et al. 2000).

The recently sequenced genomes of the sea anemone *Nematostella vectensis* and the amphioxus *Branchiostoma floridae* (Baumann et al. 2008), revealed that both organisms express several copies of the GFP-like homologues and also the G2F cnidogen domain. Therefore, further comparative studies dealing with the GFP-like structure and function may lead to a better understanding of the relationships and evolution of the ß-can structures, that ultimately may shed light on the functionality of a putative precursor protein on the Urbilateria common ancestor.

1.4 Ecological and biological aspects of GFP-like proteins in corals

The fascination for the study of coloration in reef cnidarians was promoted by the introduction of diving portable UV/blue lamps. Since then a number of ecological studies have used the GFP proteins as markers. For example, Rinkevich & Loya, (1985) found that there are not apparent differences in growth rates between morphotypes of *Stylophora pistillata*, however the purple-red morphotype, is more aggressive than the yellow morph. The yellow is eventually excluded by competition with the purple-red morph. Similarly, to the adult coral population studies; the uncomplicated quantitation of fluorescence has been used in early polyp stages to infer population dynamics of corals (Baird et al. 2006).

Other relevant work includes the combination of ecological and genetic studies over different colour morphs in the massive coral *Montastraea cavernosa,* the results suggested that the same basic gene repertoire may ultimately give rise to differences in the final colour morphology of the colony (Kelmanson & Matz 2003). In addition, some apparent correlations have been found between GFP spectral characteristics and depth habitat of the *M. cavernosa* coral colony (Kao et al. 2007).

The first note on the fluorescence of scleractinian larva was reported from the Red sea Coral *Stylophora pistillata*. This work reported that green fluorescence is associated with the mouth tissue, and mentions its putative photo protective role in adult corals (Rinkevich & Loya 1979). Recent work on the mushroom coral *Fungia scutaria,* shows that GFP-like proteins are up-regulated when the embryo is approaching to the planula stage and that the spatial distribution of the green/red) fluorescence is mainly restricted to endodermal cavity and (green) mouth tissue (Hollingsworth et al. 2006).

Since its early description in reef cnidarians until today, the primary context in which biological roles for GFPs have been discussed is in terms of their potential for either enhancing symbiont photosynthesis or protecting symbionts from excess light. Photo-enhancing has been suggested, specially where the bearing organism is living under low light conditions, providing to the dinoflagellate symbionts photosynthesis-quality photons (Kawaguti 1969). Photo protection is invoked for those organisms exposed to intense solar radiation, for example on the reef flat; the proteins may act as shields, absorbing high energy photons and transforming them to low energy ones or heat, hence potentially reducing photo inhibition, which could otherwise lead to

coral bleaching due to the disruption of the symbiosis (Salih et al. 1998, Salih et al. 2000, Salih et al. 2004). Despite the appeal of these hypotheses, they are supported only by weak and equivocal experimental evidence (Mazel & Fuchs 2003, Mazel et al. 2003), leaving open the biological roles of these proteins.

1.5 Importance of evolutionary developmental studies (Evo-Devo)

Comparison of gene expression data through development and among organisms has provided some important insights into the origins and evolution of certain traits. For instance, the detection of mesoderm specific genes during cnidarian diploblastic development has led to a much better understanding of the original roles of these genes in cell proliferation, that predated functions in specifying the third germ layer in triploblasts. In a similar way, the present work deals with temporal expression of GFP-like genes during coral embryogenesis and later development in an attempt to better understand the association of the formation of a new individual and the spatial variation of the FP genes. Finally, as homologous genes have recently been described in several bilaterians, the present work may serve as platform to compare the dynamics of GFP-like gene expression between the coral and bilaterian organisms.

1.6 Scope of this study

This research aimed to study the characteristics of the GFP-like protein repertoire in *Acropora millepora*, in terms of gene structure, protein characteristics, and expression patterns during development. The overall goal were to elucidate in a model system, the defining features of these molecules and to explain why are they so abundant and diverse in corals. Basic information could lead to better understanding of the significance of colour diversity on coral reefs and to gain insights into to possible biological functions of individual proteins

This thesis is divided in six chapters: General introduction, materials and methods, three data chapters (Chapters 3 to 5), and a last section devoted to general discussion and further directions.

The experimental chapter three reports the analyses of intron-exon structure, transcription factor binding site analyses and phylogenetics of nuclear loci encoding two GFP homologues. The chapter four report the biochemical and biophysical characterization of selected proteins derived from the Expression Sequence Tag (EST) collection from early stages of coral development, emphasising in the description of the evolutionary protein “hot spots”, and its relation with the current phenotype. Chapter five describes the spatio/temporal expression of the GFP mRNAs and proteins during coral development in order to determine the time of onset of expression and the type of cells that express the message.

CHAPTER 2

MATERIAL AND METHODS

2.1 Nuclear gene cloning, sequencing and assembling

2.1.1 [α-^{32}P] radioactive probe generation by PCR

This section details the generation of radioactive probes by Polymerase Chain Reaction (PCR) using a method modified from (Saiki et al. 1988). A non-fluorescent chromoprotein and a fluorescent protein cDNA clones were chosen due to their abundant allelic variation within the *A. millepora* cDNA library and those clones are as follows: **AmilCP597** -***Acropora millepora*** **C**hromo**P**rotein absorbance maxima at **597**nm- and **AmilGFP517a** –***A. mil****lepora* **G**reen **F**luorescent **P**rotein emission maxima at **517**nm-. Amplicons derivated from pre settlement *A. millepora* cDNA library were then sub cloned into pGEMT (Promega) and named 2:2 and 52c2 respectively. Note: sub-cloning of 2:2 was performed by L. Tomljenovic

To clone the chromoprotein amilCP and amilFP nuclear genes, radioactive PCR reactions were prepared as follow; in a 500 µl PCR tube, approximately 180 ng (1 µl) of plasmid DNA was added to a ice cold mixture containing: 5 µl of 10X Taq polymerase buffer, 1 µl of 0.1 mM dATP, 1 µl of 10 mM dTTP, 1 µl of 10 mM dGTP, 1 µl of 10 mM dCTP, 4 µl of 25 mM $MgCl_2$, 1 µl of 20 µM correspondent gene specific forward and backward primers, 0.5 µl of Taq DNA polymerase (10 U/µl, promega), 5 µl of [α^{32}P] dATP, and 28.5 µl of PCR-grade water.

For probe generation of amilCP, the AmilCP597; clone 2:2 cDNA was used in conjunction with the forward f1 primer (5'-GGAGAGCAGACAGAGAAGC–3') and the backward f2 primer (5'- GGAATCATTGGTGACAGTACA–3') . The [α^{32}P] labelled probe of amilFP was achieved using the clone 52c2 (AmilGFP517a) as template, the forward f3 primer 5'-CACTGTTATATCTGCACTG-3', and the backward primer f4 5'-GGAAACAGACCATCAATCTGT-3'.

PCR control reactions for both amilCP and amilFP nuclear genes were carried out without the addition of [α^{32}P] dATP.

PCR reactions were thermo-cycled under the following profile (designed to amplify a fragment of approximately 220 bp fragment): Initial denaturation of 3 minutes at 95ºC, then 32 cycles of denaturation at 95 ºC for 45 seconds, annealing at 50 ºC for 45 seconds, and polymerisation at 72 ºC for 1 minute, then a final one long step of 3 minutes at 72 ºC and finally the reaction was cool down to 4 ºC. The PCR products were stored at -20 ºC until need.

2.1.2 Agarose gel electrophoresis and storage of radioactive probes

Agarose gel electrophoresis described by (Sambrook & Russell 2001) was a regular procedure not only to determine expected amplicons from library screenings, but also for fragments separation after restriction enzyme digestions, verification of the quality of genomic and plasmid DNA extractions, or during Southern assays (Southern 1975).

PCR amplifications using [α^{32}P] nucleotides were not resolved by electrophoresis assay. Instead, the control reactions, were used to indirectly confirm the success of the amplification. In this case, 10 µl of the PCR control reaction were mixed with 3 µl of loading “Orange G” buffer (0.2 % w/v Orange G, 40 % w/v Sucrose) then loaded into 1 % TAE (40mM Tris-acetate, 1 mM EDTA pH 8.0) agarose gel (pre casted with 0.4 µg/ml of Ethidium Bromide), run at 85 volts for approximately 45 minutes and finally visualised/photographed under UV light (365 nm). Upon confirmation of a single product of approximately 220 bp, the radioactive product was clean up (Quiagen) from the excess of primers and radioactive unincorporated nucleotides.
Radioactive probes were then resuspended in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0) and keep at -20 ºC until required.

2.1.3 Competent cell preparation for viral infection

Nuclear genomic fragments corresponding to the amilCP gene were obtained by DNA hybridisation. from a genomic library constructed using λ-gem11 vector (Promega) .

The strain LE392 of *E. coli* was used to propagate the phages, Infected bacteria were incubated overnight at 37ºC into LB-agar media (10 g Tryptone, 5 g yeast

extract, 10 g NaCl, bacto Agar 15 g and water to a litre, pH 7.5), then single colonies were picked and inoculated into 10 ml of LB media and incubated overnight.

One millilitre of the overnight culture was inoculated into 50 ml LB media containing 10 mM $MgSO_4$, 0.2 % Maltose and incubated for 1.5 hours. Bacteria harvested by centrifugation at $4X10^3$ rpm 10 minutes. Cell pellet was resuspended into an ice cold solution consisting of 10 mM $MgSO_4$, to the volume required to reach 2.0 of absorbance at 600 nm. 100 µl of these competent cells were used as a host for the lambda phage carrying the genomic library. These competent cells were viable for up to 6 days at 4 ºC.

2.2 Tittering the genomic libraries

2.2.1 Phage library titration

Approximately 350 $x10^3$ plaque forming units (*pfu*) were screened with the radioactive probe. For this purpose, the *Acropora millepora* genomic library tittered by serial dilutions in the range of 10^1 to 10^7, using 10 µl of the master library into 90 µl of SM buffer (Per Litre: 5.8 g NaCl; 2 g $MgSO_4.6H_20$).

100 µl of fresh *E. coli* suspension were infected with the desired λ phage dilutions and mixed by pipetting. Viral infection was carried at 37 ºC for 10 minutes, then gently mixed with 3ml of melted top agarose and plated on top of pre-warmed 90 mm LB-agar plates (per litre: 10 g of Tryptone, 5 g NaCl, 7 g agarose, $MgSO_4$ to final concentration of 10 mM), Plates were incubated overnight at 37ºC. Visible phage plaques were counted. According to the dilution and the total volume of the master library, it was estimated the volume necessary to reach a density of 3.5 x 10^5 *pfu* was calculated.

2.2.2 Cosmid library titration

Amplified library was tittered as described before for the cloning of the nuclear CP gene, plated onto 90 mm LB-ampicillin plates, and incubated overnight at 37 ºC. Once the efficiency of the library was determined the desired number of colonies were obtained due to the storage capacity of the cosmid vector is about 3 to 5 time bigger to lambda vector it was decided to plate 9X 10^4 *cfu* for the screening using the FP probe.

2.2.3 Acropora genomic Cosmid library amplification and storage

The FP gene was identified from the genomic library constructed using a cosmid vector (Super-Cos, Invitrogen). Since the cosmid vector contain ampicillin resistance the library was amplified with antibiotic enriched media due to this enhances the chances to detect low represented clones.

Briefly, 125 ul of master library sterilized 600 ml flasks (X4) containing 50 ml of LB-ampicillin [50 mg/ml] and incubated at 37ºC, 150 rpm and for approximately 6 hours. Bacteria cells were harvested by centrifugation at $4X10^3$ rpm and 4 ºC for 15 minutes and finally, resuspended in 20 ml of chilled LB-ampicillin. 300 µl of freshly made bacterial suspension were used to titter the amplified genomic library, Bacteria sample was also cryopreserved by adding 4 ml of sterile glycerol, mixed, aliquoted and preserved at -80ºC for further use.

2.3 Genomic λ lambda library screening

2.3.1 Blotting and marking of membranes

Desired density of phages were spread with top agarose using six 180 mm petri-dish as described by (Ausubel et al. 1996); Nylon membrane discs (137 mm, Hybond-C Extra, Amersham Biosciences), were overlain on each plate and blotted for 1 minute Plates were marked with non-symmetrically side pin-holes distributed and recording the plate identity. Upon detection of positive clones. The holes in the membranes gaps served as alignment guide.

2.3.2 Blotting and marking of membranes (Cosmid library)

$3X10^4$ colony forming units (*cfu*) were spread on top of nylon membranes that were overlaid in LB-agar plates (180 mm). Bacteria grew at 37 ºC for an overnight period.

Bacterial growth was stopped by placing the plates in a 4 ºC incubation for 2 hours. A copy of the master plate was made by carefully removing the membrane from the plate and placed over (bacterial colonies facing upwards) a sterilized 3M paper with a sheet of clean 3 mm glass underneath. A new nylon membrane was placed exactly on top of the one which contained the colonies, then an overlaid of 3M paper and a

glass sheet afterwards. The two membranes were pressed against each other while the edges were perforated using a sterilized needle. Copies were placed over a new LB-ampicillin plates (bacteria facing upwards). The master membranes were returned to their respective plates and stored at 4 ºC until need, while the copies were incubated 4 hours at 37 ºC then also stored at 4 ºC.

2.3.3 Preparation of the membranes for phage DNA hybridisation

After the blotting, the DNA was fixed to the nylon membranes using the alkali method (Benton & Davis 1977). Membranes were exposed (DNA facing up) to a denaturing solution (0.2 M NaOH; 1.5 M NaCl) for a lapse of 3 minutes followed by other 3 minutes into neutralisation solution (0.5 M Tris.Cl (pH 7.4); 1.5M NaCl; 1 mM EDTA), and another 5 minutes in a two rinses with a 2X SSC solution [20X; 3 M NaCl, 0.3 M Na_3-citrate. 2 H_2O, adjusted to pH 7.0]). The membranes were air-dried and finally the phage DNA was cross linked by 3 minutes exposure at UV (362 nm) light in a commercial transilluminator.

2.3.4 Preparation of the membranes for cosmid DNA hybridisation

Nylon membranes were treated as follow: 2 minutes on top of pre-wetted 3M paper with 0.5 M NaOH, 3 minutes on top of a 3M paper pre-wetted with 1 M Tris-HCl, pH7.6, then 3 minutes on top of a 3M paper pre-wetted with 1M Tris HCl (pH 7.6), 1.5M NaCl, and ultimately, two rinses with the last buffer in order to remove bacterial debris. The membranes were air-dried over a 3M paper and DNA cross-linked by exposing them (DNA facing downwards) at UV light (362 nm) in a commercial transilluminator for 3 minutes each.

2.4 Hybridisation and DNA recovery of the positive plaques

2.4.1 DNA Hybridisation

The technique for DNA hybridisation of both lambda and cosmid DNA fixed in nylon membranes was the same but the radioactive probes were specific to amilCP or amilFP respectively

To reduce background and detect the positive clones from the population of the blotted plaques/colonies, the membranes were pre-hybridised with rocking at 65 ºC and for 1 hour into approximately 40 ml of hybridisation solution (5X SSPE, 5X Denhart's solution, 0.5 % (w/v) SDS). Following incubation, pre-hybridisation solution was replaced by the same volume of fresh hybridisation solution containing 50 μl of previously denatured radioactive DNA probe. Hybridizing membranes were placed in a sealed container and incubated overnight with constant rocking at 65 ºC.

2.4.2 Washing the unbound products and detection

Subsequently to DNA hybridisation, the radioactive liquid was discarded and the membranes were washed twice with 50 ml of low stringency washing buffer (2X SSC, 0.1% SDS), at 50 ºC for 15 minutes then for washed for 10 minutes at 55 ºC using 50 ml of medium stringency washing buffer (1X SSC, 0.1% SDS), and a final 15 minutes wash at 65 ºC using 50 ml of high stringency washing buffer (0.1X SSC, 0,1% SDS). Radioactivity was measured with a Geiger counter to ensure that the noise of the unbound material disappeared, leaving behind the signal of the positive clones.

The membranes were wrapped with plastic film, then attached to 3M paper and subsequently a sensible screen was overlaid, this assemble was inserted inside a cassette. Clear radioactivity signal was detected after a lapse of 4 hours at room temperature.

Positive clones were detected using a chemo-luminescent detector (Phosphor Imager ®, GE-healthcare). The radioactive signal was captured and the image enhanced by edition-visualization software, and then printed on a 3A-size paper. The positive clones were identified by aligning the specific side-holes of individual plate-membranes and comparing it with the images of the respective plate/master-membrane. Detected positive cores were resuspended in 200 μl of SM buffer with the addition of one drop of chloroform. Samples were stored at 4 ºC.

Note: The picked *E.coli* colonies from the cosmid library were inoculated into 1 ml of LB-ampicillin and they grew at 37 ºC for 3 hours then stored at 4 ºC until needed.

2.4.3 Secondary screening

Due to the high density of phages/bacteria from the primary screening, a second round of screening was carried to achieve lower density (approximately *pfu*/*cfu* per 90 mm plate respectively), and thus ensuring single positive clone isolation. .Single clones were expanded and either stored or used for DNA extraction and further sequencing.

2.4.4 DNA purification from the nuclear genomic clones

After the overnight expansion of single clones in 50 ml of phage media. 50 μl of chloroform was added to prevent bacteria contamination, samples were placed at 4ºC.

The phage DNA was extracted using midiprep-lambda Quiagen kit following the procedures of the manufacturer. Alternatively, phage DNA was purified as described in (Lockett 1990). In brief, cell debris was cleaned from the solution containing phages by centrifugation at $4X10^3$ rpm for 10 minutes. The phage capsids were lysed and precipitated from the preparation by incubating the samples with proteinase K to final concentration of 100 μg/ml for approximately 1 hour at 56ºC, followed by the addition of polyethylene glycol 8000 (10 % w/v). The sample was placed in an ice bath for 3 hours and finally spined down at $1X10^5$ rpm for 30 minutes. The supernatant was decanted in new tubes and DNA extracted twice with equal volume of phenol/chloroform and twice with equal volume of chloroform. DNA was recovered, concentrated (section 2.4.5) and stored in appropriate conditions.

Positive and single cosmid clones in *E.coli* grew in10 ml of LB-ampicillin at 37 ºC overnight, then DNA extracted using commercial midiprep-plasmid Quiagen kit following the instruction of the manufacturer. The eluted DNA was measured and stored at -20 ºC.

2.4.5 DNA precipitation and storage

DNA precipitation described by Sambrook & Russell (2001) was a routine procedure not only for recovery and concentration of phage DNA but also for DNA from plasmids, excised bands from gels, or low yield PCR products etc.

Phage DNA was precipitated from the aquose phase by the addition of 2 volumes of ice cold 100 % v/v ethanol and 0.1 volumes of 3 M of sodium acetate (pH 5.2). This mixture was incubated 45 minutes at -20 ºC, spined down at $1.3X10^5$ rpm for 30 minutes and supernatant discarded by decantation. Residual salts were removed by the addition of 500 μl of 70 % v/v ethanol. The DNA pellet was recovered by 10 minutes centrifugation at $1.3X10^5$ rpm, supernatant discarded and the tube air-dried. The DNA was finally resuspended in TE or ddH2O. The protein-free DNA obtained was measured by its absorbance at 260 nm, and stored at -20 ºC until required.

2.4.6 Restriction enzyme digestions genomic fragments to detect differences

Restriction enzyme digestion assays were conducted as per Sambrook & Russell (2001) and it was an habitual procedure not only for differentiate between genomic clones but also to achieve clean GFP-like cDNA fragments for subsequent cloning e into expression vectors.

Before sequencing, 20 μl restriction enzyme reactions were assembled using approximately 1 μg of phage/cosmid DNA from each clone into total reaction, 10 units of either or both *Bam HI, Hind III, and XbaI* (Promega). According to the vectors map, the first two enzymes target the multi cloning region, thus releasing the desired fragment and determining he molecular size. In all cases, the reactions were incubated at 37 ºC for 3 hours then, the products were resolved through agarose gel electrophoresis. Gels were stained with ethidium bromide and visualized under UV light.

2.4.7 Southern blot and subcloning

The southern assay, performed as described in Sambrook & Russell (2001) modified from Southern (1975), was chosen because at the moment of the amilCP fragment sequencing, some regions appeared unreadable due to the high content of guanine

and cytosine. Consequently, the southern blot aimed to detect fragments that cross-reacted with the original probe used to screen the library, and further those fragments were subcloned and sequenced.

Genomic DNA was digested and DNA fragmentation was resolved by agarose electrophoresis. The gel was washed once with distilled water for 20 minutes with continuos movement. The water was replaced by denaturing solution (1.5 M NaCl, 0.5 M NaOH) in gentle agitation for 45 minutes. The gel was rinsed with water and neutralized with a solution containing: 1 M Tris HCl pH 7.4, 1.5 M NaCl, 1 mM EDTA and for a lapse of 30 minutes and with movement. With a clean scissors, a piece of Nylon (Hybon-C) paper of the size of the gel was excised and poured over the gel, the gel itself was resting over 3 layers of 3M (whatman) paper pre-wetted with transfer buffer (10X SSC). Over the nylon paper, two layers of 3M paper and 10X the thickness of the gel of paper towels, a sheet of glass and a weight were accommodated. The DNA was transferred to the paper overnight and finally DNA was cross-linked to the membrane for 3 minutes under UV light.

The DNA Hybridisation with a radioactive probe was performed as described earlier. After several washes and detection time in the range of 3-12 hours (using phosphor-imager screens (Molecular Dynamics). Positive bands were selected/excised from the agarose gel, and DNA recovered using Qiagen gel extraction kit and subcloned into pBluescipt (invitrogen) for sequencing.

2.4.8 Sequencing and gene assembling

The sequencing of the genomic fragments was initiated using the vector priming sites by dedeoxy chain terminator procedure (Sanger et al. 1977) at the JCU sequencing facility. Upon sequence analyses, specific primers were designed to continue the sequencing process by primer walking.

The sequence of amilCP gene was assembled (Sequencher 4.8), and the intron/exon boundaries were used to design amilFP-based primers targeting those spliced donor-acceptor sites. The designed primers were used in a long-PCR procedure using as template 100 ng of the cosmid DNA clone in each case. The profile of the amplification was adjusted to amplify up to 5 Kb. The PCR products

were compared for its relative size using a standard agarose gel electrophoresis and subsequently cloned into pGEMT vector (Promega).

2.5 Recombinant protein characterization

2.5.1 PCR amplification of the amilGFP517a from the presettlement cDNA library

The AmillGFP517a (pGEMT clone 52c2) was the only GFP that was cloned pulled out by PCR from the pre-settlement cDNA library. A combination of T7 universal primer and a gene specific primer targeting the first 6 amino acids from the FP and CP contig was used for PCR amplification.

The cDNA encoding the remaining three of the proteins selected for bacterial expression were chosen among the approximately 9000 EST tags conforming the *A.millepora* EST collection. The selection criteria was based in the amino acid alignments given in Chapter 3 (See figure 3.1 and 3.2 for details), the abundance of transcript representation as well as the nucleotide/protein substitutions that were considered to have influence in the chromophore formation/maturation

The amilGFP517a cDNA was isolated from the cDNA library using approximately 100 ng of denatured DNA. The PCR mixture included: 2.5 µl of 10X Taq Polymerase buffer (Promega), 2.5 µl of 2 mM dNTP mix [final 200µM], 2 µl of 25 mM $MgCl_2$ [final 2 mM], 1 µl of 20 µM [final 800 nM] forward “FP1” (5’-GACTTGGTGGCTATAGAGA-3’) primer, 1 µl of 20 µM [final 800 nM] T7 (5’-GACTTGGTGGCTATAGAGA-3’) reverse primer, 0.1 µl of 10 Units/µl Taq DNA Polymerase (Promega) [final 0.04 U], and approximately 15 µl of PCR-grade water to a total volume reaction of 25 µl. The PCR reaction was submitted under a profile designed to amplify a fragment of approximately 680-700 bp fragment: Initial denaturation of 3 minutes at 95 ºC, then 32 cycles of denaturation at 95 ºC for 45 seconds, annealing at 50 ºC for 45 seconds, and polymerisation at 72 ºC for 1 minute, extension for 3 minutes at 72 ºC and finally the reaction was cool down to 4 ºC. The PCR product was confirmed by standard agarose gel electrophoresis and stored at -20 ºC.

2.5.2 Gel extraction and ligation into PCR-based propagation vector

20 µl of the PCR reaction was mixed with 3 µl of loading “Orange G” buffer (0.2 % w/v Orange G, 40 % w/v Sucrose) then loaded into 1 % TAE (40 mM Tris-acetate, 1 mM EDTA pH8.0) agarose gel (pre-casted with 0.4 µg/ml of Ethidium Bromide), ran at 85 V for approximately 45 minutes and visualised/photographed under UV light (365 nm).

The 700 bp band was excised from the gel using a clean stain-steel blade, and DNA extracted using Qiagen gel extraction Kit following the instructions of the manufacturer. The eluted DNA from the PCR band was measured for its concentration at 260 nm and stored at -20ºC until needed.

For the ligation of the PCR product into pGEMT vector (Promega), approximately 300-500 ng of band-extracted DNA was added to an ice cold mixture containing 1 µl of 10X T4 ligase buffer, 1 µl of T4 ligase (10 U/µl, Promega) [final 1.0 U], 1 µl of 50 ng/µl vector pGEMT [final 5 ng/µl] and DNAse free water to a total volume of 10 µl. The assembled reaction then was mixed by pipetting and incubated overnight at 4ºC. These ligations were used for transformation into NM522 *E. coli* competent cells. A single positive clone named 52C2 was isolated and sequenced, and its DNA used as a template for the generation of the restriction sites at the edges of the ORF region.

2.5.3 Cloning of the GFP-like proteins from the EST collection

The plasmid DNA was kindly provided by Dr David Hayward (Australia National University) by the specific request using the EST clone identity for that purpose. The plasmid DNA from each clone was blotted by Dr Hayward onto 3M papers, wrapped in kitchen plastic and send to me by mail. DNA was recovered from the paper with 20µl of Mq H_2O and incubated overnight at 4 ºC. The eluted plasmid was used to transform into NM522 cells by standard heat shock. Plasmid DNA was recovered and a glycerol stock from single colony transformed bacteria was stored at -80 for further work.

2.5.4 PCR-based introduction of restriction sites and subcloning into expression vector

A non-directional strategy for subcloning amil**FP** or amil**CP**-based was elected. PCR primers were designed to contain either *Bam* HI or *Sal*I restriction sites. Primer sequences are given further ahead in the text in the form of table.

The following EST clones were used: C015-a8 (re-named AmilCP601), C012-C9 (re named AmilGFP517b), C018-g5 (AmilRFP602), and pGEMT clones 52c2 (AmilGFP517a), and 2:2 (AmilCP597).

Approximately 100 ng of the required plasmid DNA was added to a chilled 25 µl PCR reaction mixture containing 2.5 µl of 10 X Taq Polymerase buffer, 2.5 µl of 2 mM dNTP mix [final 200 µM], 2 µl of 25 mM $MgCl_2$ [final 2 mM], 0.1 µl of 10 Units/µl Taq DNA Polymerase (Promega) [final 0.04 U], 1 µl of 20 µM of forward primer and backward primers. See the table bellow. Each primer was used to a final primer concentration in the reaction of 800 nM.

Table 1 Primers characteristics; CP correspond to chromoproteins and FP to fluorescence genes.

Forward ID	Sequence
CP-BamHI	5'-CGCGGATCCAGTGTGATCGCTAAAGAAATG-3'
FP-BamHI	5'-GCGGATCCG CTCTGTCAAAGCACGGTCT-3'
CP-SalI-	5'- GCGTCGACCTATGCAACCAAAG GTTTTCG-3'
FP-SalI	5'-GGGTCGACGATGGCTCTGTCAAAGCACGGT-3'
Backward ID	**Sequence**
CP-BamHI	5'-GCGGATCCC TAGGCGACCAAAGGTTTGC-3'
FP-BamHI	5'-GCGGATCCTCAAGCCAATG CAGATCGA-3'
CP-SalI-	5'-GCGTCGACCTATGCAACCAAAGGTTTTCG-3',
FP- SalI	5'-GCGTCGACTCAAGCCAATGCAGATCGAGA-3'

PCR reactions were thermally treated using Hybad and Eppendorf thermocyclers under the a profile designed to amplify a fragment of approximately 680-700bp fragment: Initial denaturation of 3 minutes at 95 ºC, then 32 cycles of denaturation at 95 ºC for 45 seconds, annealing at 50 ºC for 45 seconds, and polymerisation at 72 ºC for 1 minute, an extension of 3 minutes at 72 ºC and finally the reaction was cool down to 4ºC. Approximately 20 µl of each PCR reaction were resolved by agarose gel electrophoresis. Upon confirmation of amplicon and size, the DNA bands were excised from the gel extracted (QIAquick, Quiagen) and measured using OD 260 nm.

The PCR fragments were cloned into pGEMT vector as described before for the clone 52c2. Once each of the desired cDNAs were inserted into pGEMT, confirmation of insert present and size was carried by restriction assays; The plasmids were subjected to restriction enzyme digestion in which approximately 1 µg of plasmid DNA were added to an a ice cold mixture containing: 2 µl of 10X restriction enzyme buffer (Promega), 1 µl of the restriction enzyme (*Bam H*I or *Sal*I depending on the clone) [10 U/µl] and DNAse free water to 20 µl, mixed by pipetting and incubated for 2.5-3 hours at 37 ºC and then digested samples were resolved by agarose electrophoreses an the desired band purified (QIAQuik Quiagen). The fragments were stored at -20ºC until need for direct ligation into previously digested and dephosphorylated expression (pQE, Qiagen) vector.

2.5.5 Dephosphorylation of the expression vectors

Incompatibility of the restriction enzyme buffers in double digestion required that the expression vector had to be opened with one enzyme then dephosphorylated to prevent the re-ligation. For this reason, pQE30, pQE32 (Quiagen) and pProEX-B (Invitrogen) vectors were first digested with *Bam H*I or *Sal*I (approximately 1.5 µg of vector with 12 U of enzyme) as described above. After the 2.5 hour incubation, the entire 20 µl reactions were added to a mixture of 20 µl of 10X CIAP buffer, 5 µl of 10 U/µl Calf alkaline Phosphatase [final 0.25 U] and 155 µl of DNAse free water, to a total 200 µl reaction, and further incubated at 37 ºC for 45 minutes. The reaction was cleaned using commercial PCR cleaning column procedures. The eluted dephosphorylated vector then was measured for its concentration and mixed in appropriate vector:insert ratios to be used in DNA ligation

2.5.6 Cell strain preparation and transformation

2.5.6.1 Competent cells preparation and its storage

Glycerol stocks of *E. coli* NM522 and BL21 strains were used for plasmid propagation and competent bacterial cells this research were produced using $CaCl_2$, according with the methodology in Sambrook & Russel (2001).

In short, 5ml of single colony overnight liquid culture, were inoculated into pre-warmed 500 ml of LB media for about 4 to 5 hours until the OD reached 0.3.

Bacterial cells were harvested by centrifugation at $4X10^3$ rpm 10 minutes and the pellet resuspended with 30 ml of chilled solution containing: 80 mM $MgCl_2$ and 10mM $CaCl_2$, a second round of centrifugation was done and finally resuspended into 2 ml of ice cold 10 mM $CaCl_2$. The cell suspension was incubated at 4 ºC overnight to increase the cell competency. Aliquots of competent cells (50 µl) in a solution containing DMSO were snap frozen with liquid nitrogen and finally cryopreserved at -80ºC stored until required.

2.5.6.2 Bacterial transformation with plasmid vectors

Transformations using cryopreserved competent cells were allowed to thaw at room temperature and incubated for 10 minutes on ice, then transformed as described by Sambrook & Russell (2001), using the entire reaction of ligation. 500 µl of LB media were added to transformed cells samples and incubated for 1.5 hours at 37ºC. Cells were pelleted in a top bench micro centrifuge at maximum speed for 1 minute, LB-media decanted and cells resuspended into fresh 80 µl LB and ultimately plated onto 90 mm LB-agar plates containing 100 mg/ml of ampicillin, 20 mg/ml of Xgal and 1.0 mM IPTG (BL21 cells were plated without Xgal) and incubated overnight at 37 ºC.

2.5.7 Screening and protein purification

2.5.7.1 Bacterial visual screening for the presence of target insert

Overnight 37ºC incubated plates of bacteria containing pGEMT plasmids were placed at 4 ºC for about 1 to 2 hours to allow colour development; for blue-white selection and in addition, to stop the bacterial growth. Three to four white bacterial colonies were piked from each plate and inoculated into 5 ml of fresh LB-ampicillin media, and then incubated at 37 ºC for 8 hours. Plasmid DNA was extracted from 4 ml culture (QIAprep Spin Miniprep Kit, Qiagen). Upon quality and quantity of DNA was calculated by optic density at 260 nm and electrophoresis, the samples were sequenced, to confirm DNA identity. A fraction of the O/N bacterial cultures was cryopreserved by the addition of 150 µl of sterile glycerol, mixed by pipetting and snap frozen in liquid nitrogen and stored at -80ºC until needed.

In the case of transformations using expression-based plasmids containing the ORF of the GFP-like proteins, after the initial overnight incubation at 37 ºC, the plates were transferred to a dark container at 20 ºC for a period of 4-5 days, then each plate was visually screened under a epifluorescent dissection microscope (LEICA, MZ FL III, Wetzlar, Germany) equipped with a 100 W Hg lamp and sets of 550/40 nm excitation filter (green light), and 450/40 nm excitation filter (blue light). The positive coloured/fluorescent colonies were resuspended into 6 ml of fresh LB-ampicillin media, overnight incubated at 37 ºC and plasmid DNA extracted/sequenced from 4 ml culture as described above. 1 ml culture was cryopreserved as described above, and the left one millilitre culture was expanded into fresh media as detailed below.

2.5.7.2 Bacterial re-protein purification

Expression/purification of re-proteins was performed as described in the QIAexpressionist handbook (Qiagen) under the non denaturant protocol, modified from (Janknecht et al. 1991).

Habitually, 1 ml of overnight culture was inoculated into 1.5 L autoclaved sterilized flask (wrapped with aluminium foil) with 200 ml of LB-ampicillin, incubated with shaking (120 rpm) for approximately 2 hours at 37 ºC then protein expression induced with the addition of IPTG to a final concentration of 1 mM and the temperature and motion adjusted to 20 ºC and 80 rpm, respectively for approximately 4-5 days. After this long incubation the bacteria were harvested by centrifugation into 50 ml falcon tubes at $4X10^3$ rpm for 15 minutes. The cell pellet was resuspended in 10 ml ice cold PBS pH 7.2 (137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4 \cdot 7H_2O$, 1.4 mM KH_2PO_4), and incubated one hour on ice with the addition of lysozyme to a final concentration of 1mg/ml. Cells were disrupted by 3 to 6 sonication strokes with intervals of 10 seconds. In order to dissociate inclusion bodies preparations included a final concentration of 1.8 % Triton X-100 (Sigma) to the sonicated cell-buffer mixture and incubated at room temperature for 1 hour with slow shaking.

Cell lysates were transferred to 30 ml centrifuge tubes and spined down at $2X10^4$ rpm for 20 minutes. In some occasions due to the fluorescent and chromoproteins were clearly visible in the pellets after this last centrifugation, a second round of sonication and triton X-100 incubation were performed until the vast majority of the protein turned soluble.

2.5.7.3 Histidine affinity chromatography

A two step strategy was used for recombinant protein purification by an initial batch interaction, followed by column chromatography.

Sonicated cell lysates containing the re-proteins were transferred to 50 ml falcon tubes containing 2 ml of Ni-NTA super flow resin (Qiagen). The affinity matrix and recombinant protein interacted for a period an hour on ice and at constant 60-80 rpm.

The second step included gravity-flow chromatography. Lysate/Ni-NTA agarose bead slurry was loaded into disposable bottom sealed chromatographic columns (Bio-Rad). Once the matrix settled down, the bottom seals were opened to clear unwanted solution by flow by gravity, -Collecting part of this flow through as well as the posterior washes to monitor of the purification processes-. The columns were washed twice to remove non retained proteins with 10 ml of PBS, then with 10 ml of low stringency wash (PBS, 5 mM Imidazole), followed by 10 ml of medium stringency wash (PBS, 10 mM Imidazole), and a final 10 ml of high stringency wash (PBS, 20 mM imidazole). Due to the color of the protein allowed monitoring trough the purification process

Histidine-tagged proteins were eluted with PBS containing 250 mM imidazole; 500 µl fractions were collected and the visually most colored elution reveled the most concentrated purified proteins fractions, Bradford assay (Bradford 1976) and standard curve using bovine serum albumin (Promega) were used to calculate protein concentration .

2.5.8 Recombinant protein characterization

2.5.8.1 Electrophoretic characterization in SDS-PAGE assays

analysis of recombinant protein by SDS-polyacrylamide gel electrophoresis was performed as described by Laemmli (1970), using Mini-PROTEAN II electrophoretic chambers (BioRad). Approximately 35 µg of recombinant protein were mixed in a 1:1 ration with protein loading buffer (100 mM Tris.Cl pH 6.8; 4 % SDS; 0.2 % bromophenol blue; 20 % glycerol; 200 mM DTT) denatured for 5 minutes at 95ºC and loaded into 13 % SDS-PAGE gel. Protein migrated in a glycine based buffer (3 g

Tris, 14.4 g Glycine 1 g SDS, and distilled water to 1 L) at 100 V for 1 hour. The protein gel was stained with coomasie blue solution (40 % Methanol, 10 % Glacial Acetic Acid, and 0.25 g Brilliant Blue R-250 per 100 ml of solution) for 3 hours, fixed and de-stained with several changes of a solution made of 40 % Methanol and 10 % Glacial Acetic Acid, until the background coloration disappeared, and finally photographed on the transilluminator.

2.5.8.2 Spectroscopic characterization of the recombinant proteins.

Spectral measurements were taken at room temperature with approximate 0.35 µg/µl of re-protein. 250 to 700 nm absorbance scans with a resolution of 0.5 nm were done by triplicate at room temperature (measured using DU-650 Beckman spectrophotometer, USA).

The fluorescence scans were acquired from a LS-50B, Perkin Elmer, luminescence Spectrometer and were also made at 0.5 nm of resolution at room temperature. Excitation scans were done in 350 to 700 nm emissions collected at 580 (for the case of AmilGFP517a-b), and 650 nm (for the case of AmilRFP602), with 15-10 excitation-emission slit path in nanometres. Emission spectra data was taken exiting the molecules at 470 and 537 nm.

Resulted data reading values were processed with Excel (Microsoft). In order to compare between spectral signatures, its respective peaks in the visible region were used to normalize the spectra. The average of the three replicates per value were used to calculate the actual spectral curve. The fourth derivate of the dependent variable (absorbance, excitation, emission) with respect of the energy of the photon (wave length in nanometres) was calculated following the method purposed by (Butler & Hopkins 1970, Terao et al. 1985). Briefly, the relativised data was background reduced by applying the average between values in 5 nm ranges. This smoothed values were used to calculate the ^{1th} derivate with a three nanometre increment. For the calculation of the 4th derivate, same 3 nanometre increment was chose (using in this case the data coming from the first derivate calculation). The difference was divided by three and multiplied by -4. The fourth derivate values were incorporated into the original spectral signature representation as secondary Y axis values as a function of wavelength.

2.6 *In vivo* and *in situ* detection of the GFP-like genes

2.6.1 Coral colony sampling and general embryo manipulation

Coral spawning was followed during the 2004 mass spawning in Magnetic Island located at 19° 9’ 25.28’’ S, 146° 51’ 53.02’’ E.

Three days before the predicted spawning day several exploratory diving or snorkelling campaigns were performed along the local reef looking for orange, red and yellow *Acropora millepora* morphotypes. The selected colonies were tagged with a buoy then sampled by chiselling and moved near the shore to be monitored for seedling signals. During the afternoon of the predicted day, the colonies were transferred to small pools on the beach filled with 0.5 µm-filtered seawater. Five colonies of each colour-morph were placed together into these pools then when the spawning occurs; the bundles (small spherical aggregations of egg and sperm) were transferred to different 60 litres containers to allow the fertilization and development take place.

Two hours after fertilization the first cell cleavage occurs, then the embryos were washed twice in filtered sea water and let them develop in 190 mm petri dishes (approximately 1000 embryos per plate) always taking care no mixing the colour offspring. The next morning, the embryos were moved to laboratory conditions at 24 °C, 10:14 light-dark (50 µmol quanta cm^{-2} sec^{-1}) conditions. The embryos were continuously changed with fresh filtered seawater every 12 hours, then once reached the planulae stage sea water changes were once every two days.

2.6.2 Live fluorescence microscopy and embryo fixation during the coral development

During all the developmental process, series of sampling at specific time points were made. Live material were photographed and filmed in a epifluorescence microscope Olympus (Accu-Scope 3016) equipped with a 100 W lamp and the following set of filters:

Table 2 Filters used for live fluorescence

Filter	Specifications/applications
WIB	Designed to detect the FITC/PI fluorochromes, the filter has the windows of 460-490 nm and 515 and up for the excitation and the emission photon characteristics
FITC	designed to detect green signal with 490 ± 20 and 510 ± 20 nm of excitation and emission windows
TR	designed to detect far red fluorescence with 590 ± 20 and 610 ± 20 nm of excitation and emission windows
Cy3	designed to detect red/orange fluorescence signal with 550 ± 20 and 560 ± 20 nm excitation and emission windows
DAPI	from the chemical 4,6-diamidino-2-phenylindole-2-HCl, (designed to detect blue fluorescence signal with 350 ± 20 and 460 ± 20 nm of excitation and emission windows

To observe under the epifluorescent microscope the coral embryos, approximately 20 to 30 organisms were concentrated into 100 µl of seawater in a single concave glass slide, then slowly half of the water was retired by pipetting and replaced with 50 µl of 0.5 uM $MgCl_2$. The added salts allowed the relaxation of the embryos, which specially at the planula stage, the active swimming greatly impeded the proper focusing of the objets. Embryos were allowed to relax after approximately 2-5 minutes then the excess of water-buffer was adjusted by pipetting prior the incorporation of the top cover slip taking care of don't leave air bubbles in the preparation. From these preparations, short movies (20 seconds) were taken on MPEG format, then using iMovie HD software (Apple), single frame pictures were selected, and further edited in Photoshop (Adobe) when required.

Coral embryos form representative developmental stages were fixed for *in situ* assays (15 minutes, rocking at 60 rpm) in 4 % formaldehyde in Millipore filtered sea water buffered with Hepes pH 8.0, then washed twice with Hepes-sea water buffer and stored at 4 °C for up to 6 days and finally dehydrated by several washes of increasing percentage of methanol and stored at -20 ºC until need. For quantitative RNA procedures, approximately 0.2 volumes of the cryotube was filled with living material from important developmental stages. Excess of seawater was pipetted out and the tubes preserved under liquid nitrogen until need.

2.6.3 Riboprobe creation (DIG labelling protocol)

The following protocol is a modification made by Dr. Dave Hayward (ANU) from Stratagene manual. The plastic wear (tubes, tips) and the solutions were RNAse free. The labelling reactions were assembled (in a total volume of 20 µl) as followed: One microgram of plasmid DNA (pBuescript) digested with *Xba* 1 (Promega) corresponding to the EST clones C008e4 (AmilCP) and C012c9 (AmilFP) was added

to each tube containing a mixture of 4 µl 5X reverse transcriptase buffer, 2 µl DIG-NTP mix, 1 µl RNAsin RNAse inhibitor, 1.7 µl of 100 µM DTT, 1.5 µl T7 RNA Polymerase, and water to 20 µl.

The reactions were incubated at 37 ºC for 3 hours. After the incubation, each reaction was stopped with the addition of 2 µl of 0.2M EDTA. At this point a 2 µl sample was taken from each tube to be analysed by agarose gel (1%) electrophoresis. The rest of the reaction was precipitated by adding 2.2 µl of 3 M NaOAc pH 5.2 and 50 µl of 100% ethanol then leaved overnight at -20 ºC. Cold-precipitated reactions were spined down in a normal top bench centrifuge at $1.3X10^4$ rpm for 25 minutes, liquid discarded by pipetting, vacuum-dried the RNA pellet, water resuspended in 50 µl and stored at -20 ºC until required.

2.6.4 Riboprobe fractionation

The *in situ* detection is based on the antisense-generated probe and organism sense RNA interaction, the probabilities of hybridisation are directly related with the size of the probe, the smaller the higher the efficiency of the molecule to travel across the tissues and hybridise with the sense mRNA. For this purpose the RNA probe was hydrolysed into small pieces OF approximately 250 bp. The hydrolysis was made by addition of 5.5 µl of 0.2 M Carbonate buffer to 50 µl of each of the riboprobes then incubated at 60 ºC for a calculated time (necessary to obtain fragments of about 0.25 Kb) according to the following formula:

$$\text{Time (minutes)} = (Lo - Lf)/(K)(Lo)(Lf)$$

where

Lo = starting length (in Kb)

Lf = final length (in Kb, 0.25 is recommended)

K = 0.11 (strand excision constant in Kb/minutes)

Because both GFP-like mRNA's were almost the same length (approximately 0.7 Kb in length) the value for the time of incubation was 22.8 minutes.

After the hydrolysis, the probes were precipitated with the addition of 2µl of 3 M NaOAc pH 5.2, 2 µl tRNA [10 mg/ml] (Sigma), and 150 µl of 100 % ethanol. The reaction was incubated overnight at -20 ºC, spined down at 13200 rpm for 25 minutes, liquid discarded by pipetting, vacuum-semidried and finally was resuspended in 80 µl of probe buffer (50 % formamide, 50 % TE, 0.1 % Tween-20) and stored at -20ºC until needed.

2.6.5 Whole mount *in situ* hybridisation.

2.6.5.1 Coral embryo preparation

A pool of approximately 30-50 embryos from each developmental stage (See Table 3) were treated in different tubes (2 ml eppendorf RNAse free), and the solutions (800 µl each time) used for the preparation of the embryos were directly added and remove from the same tube using RNAse-free tips, taking care of not damaging or pulling off the embryos each time.

Table 3 Developmental stages of *Acropora millepora*

Developmental stage	Period of time after fertilization (in hours)
Prawn chip	11 – 13
Donut	18 – 24
Sphere to pre pear	30 – 35
Pear	50 – 64
Planula	72- 88
Post-settlement	96 hours

The methanol-stored embryos were moved from -20 ºC to room temperature, and then partially hydrated with a 5 minutes wash into 70 % methanol. During this process an small proportion of embryos from all stages were dissected under the stereo-microscope. The sections and whole embryos were further hydrated into 50% methanol for 5 minutes then rinsed into PBS pH 7.2 (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4 . $7H_2O$, 1.4 mM KH_2PO_4) and PBT buffers (1x PBS, 0.1 % Tween 20) 5 minutes each. Finally embryos were incubated overnight into RIPA buffer (150 mM NaCl, 1 % Nonidet-P40, 0.5 % Na deoxycholate, 0.1 % SDS, 1 mM EDTA, 50 mM Tris pH 8.0).

2.6.6 Whole mount *in situ* hybridisation

RIPA treated embryos were rinsed twice into PBS buffer, then dehydrated by passing them (approximately 5 minutes each rinse and repeating twice the final 100 % rinse) through increasing concentrations of ethanol (50, 70, 90, and 100 %). In the last dehydration step, the volume of 100 % ethanol was reduced to 250 µl then 250µl of xylene was added and mixed by tube inversion and incubated at room temperature

for 10 minutes. The mixture xylene-ethanol was replaced by pure xylene and let it stand for 3 to 4 hours at room temperature. Following this process, the embryos were fairly clear, so special care was taken to avoid pull them off.

Hydration was achieved as follow: one wash through 50:50 Ethanol:Xylene, three times 100 % ethanol, then 25:75, 50:50 75:25 of PBT:ethanol mixtures, respectively and finally three times into PBT (5 minutes washes each on a rotator). The volume of PBT was adjusted to 250 µl and 250 µl of *in situ* hybridisation solution (50 % Formamide, 4 x SSC, 50 µg/ml Heparin, 1X Denhardt's solution, 5 % Dextran sulphate, 0.1 % Tween-20, 500 µg/ml denatured salmon sperm DNA) was added, mixed and incubated at room temperature for 15 minutes then replaced by 250 µl of 100 % hybridisation solution, incubated also at room temperature for other 15 minutes, and for the pre hybridization process the solution was replaced with new 400 µl of hybridisation solution then incubated-rotating at 56.5 ºC for 2-3 hours. Finally the volume was adjusted to 250 µl and the probe was added-mixed (2µl of hydrolysed-probe into 20µl of hybridisation solution) and incubated without movement for approximately 48 hours at 56.5 ºC.

2.6.7 Washing the unbound material form the preparations

After the hybridisation process, the embryos were washed extensively (3 x 20 minutes each) using pre-warmed (hybridisation temperature) *in situ*-wash solution (50 % Formamide, 4 x SSC, 0.1 % Tween) then leave them rotating at hybridisation temperature in the oven overnight. For the antibody reaction, a 30 minutes wash (*in situ* wash solution) at room temperature was made then replaced with 50:50 *in situ* wash:PBT for 15 minutes followed by six consecutive (15 minutes each) washes into PBT. AntiDIG-AP was diluted 1:1600 into PBT then added to each tube then incubated at room temperature with rotation for 2 hours. Embryos were 30 minutes washed in PBT, and a finally fresh PBT was replaced to each tube and incubated overnight at 4ºC.

2.6.8 Detection and documentation of the spatial mRNA signal.

For the developing reactions; embryos were washed in PBT six times, 30 minutes each followed by two (5 minutes each) washes into NTMT buffer (100 mM NaCl, 5 mM MgCl2, 0.1 % (v/v) Tween-20, 100 mM Tris pH 9.5) then replaced with

NBT/BCIP (Alkaline phosphatase substrate buffer, Kit IV, Vector) or BM purple (Alkaline phosphatase Substrate (Roche)), prepared according to the manufacturer and let them develop by rotating in the dark for about 25-30 minutes. Color develop was stopped changing the embryos into PBT. At this point, embryos were observed and photographed under the stereo-microscope (LEICA, Wetzlar, Germany) equipped with a 100 W Hg lamp and sets of 550/40 nm excitation filter (green light), and 450/40 nm excitation filter (blue light).

To enhance the clearness of samples the PBT was replaced with 70 % glycerol (rotating 2-3 days at 4 ºC) and finally stored at 4 ºC. After the glycerol treatment a second round of more detailed photographs were taken.

2.7 Northern blotting procedures

Total RNA from six developmentally important stages was isolated by the guanidine thiocyanate/phenol:chloroform technique. Coral tissue was mixed with RNAwiz (ambion) then using liquid nitrogen grounded with mortar and pestle. Aliquots of tRNA were used to isolate only the mRNA using polyATract mRNA Isolation System IV (Promega) following the manufacturer instructions. Approximately 2 µg of poly (A) + RNA from each stage was denatured in the presence of 2.2% formaldehyde then loaded in a (pre-stained with ethidium bromide) agarose (1 %) gel in 1X Mops buffer. The samples were separated at 100 V for 60-90 minutes. After the separation, the gel was blotted on Nylon membranes (Hybond-C Extra, Amersham Biosciences), then UV fixed by exposure on the transilluminator for 2 minutes.
For the RNA detection, membranes were hybridised either with CP or FP radioactive probes (same as used to clone the nuclear genes) as described in the section 2.4.

CHAPTER 3

CHARACTERISATION OF TWO GFP LOCI FROM THE CORAL *ACROPORA MILLEPORA*

3.1 Introduction

Coelenterate bioluminescence has always been of scientific interest to serve in clarifying the mechanisms by which light is produced by an organism. Early works showed that a combination of two protein extracts, a photo protein and a green fluorescent protein (GFP) from several *Aequorea* jellyfish were able to reproduce green light emissions *in vitro* (Cormier & Totter 1964). Similar results were obtained from protein extracts from the sea pansy *Renilla* (Morin 1974, Ward & Cormier 1975, 1979). Although the absorbance of both GFPs was different, the emission maximum was identical. Such differences were attributed to the protein environment surrounding the same chromophoric structure, but the extent of the evolutionary relationship between those proteins has remained unclear.

In 1992, Prasher and colleagues cloned the first GFP gene from *Aequorea* but were unable to reproduce a heterologous recombinant protein, although they detailed the hydrozoan fluorescent protein (FP) gene structure. The jellyfish gene spanned 2.6 kb and was interrupted by three introns, two of which were along the coding region and the other in the 5'-UTR. Then, in 1994, Chalfie *et al.* were able to clone and express the GFP cDNA into bacterial and eukaryotic living systems. This achievement was a breakthrough in biology, opening the possibility to track target proteins fused with GFP, without the requirement of a cofactor for fluorescence.

The years following the successful GFP cloning and consequent recombinant expression were devoted to elucidating and modifying the fluorescence characteristics to produce a palette of tags for multiple detection. However, after several mutational attempts, no red fluorescent protein (RFP) was observed until two groups in 1999 succeeded in cloning a homologue from non-bioluminescent cnidarians (Matz *et al*. 1999, Wiedenmann 2000). In the next few years, the number

of GFP-like proteins cloned mainly from reef cnidarians had increased dramatically, enabling the classification and investigation of phylogenetic relationships. However, due to the relatively small number of amino acids involved in chromophore synthesis, the results of conventional phylogenetic analyses within the GFP family were unclear, revealing, for example, multiple independent origins of red variants (Shagin *et al*. 2004).

In an attempt to better understand the evolutionary relationships within the GFP family, the gene structures of two GFP-like genes corresponding to a fluorescent and nonfluorescent chromoprotein locus from *Acropora* were determined and compared against GFP-like genes belonging to the scleractinian *Montastraea*, the soft coral *Discosoma* (Carter et al. 2004), the starlet sea anemone *Nematostella* and the jellyfish *Aequorea* (Prasher *et al*. 1992). Moreover, whilst this research was being prepared for submission, a large-scale GFP-like genomic sequence from the cephalochordate *Amphioxus* became available (Baumann *et al*. 2008). Although a direct comparison of the cnidarian and the bilaterian GFP structure will not be addressed here, features such as intron number and position will be discussed.

Using the GFP-like region immediately upstream of the transcription site initiation as a point of reference, genomic data from three cnidarian species were screened for transcription factor (TF)-binding sites. This comparison was expected to reveal similar TFs across different species. Although experimental confirmation is required to verify whether such shared TFs are able to bind and perhaps control the GFP-like expression, the assay represents a good starting point. Finally, phylogenetic analyses were performed to “localise” the *Acropora* GFP genes into the scleractinian lineage and interphylum using cnidarian and bilaterian molecules.

3.1.1 Statement of Goals

The main aim in characterising the structure of genes representing the two major gene GFP-like variants is to clarify evolutionary relationships between these and with the genes encoding other related proteins. It is well established that common features of intron/exon organization are conserved and reflect evolutionary relationships, with an overall trend towards intron loss in many lineages. During evolutionary time, gene structure often becomes simplified, usually by loss of introns towards the 3’ end of the gene. By comparing the organization of the *Acropora* GFP-like genes it is possible to infer relationships with the ancestral GFP locus. Using the

structural information for the anthozoan GFP-like loci, a series of comparative analysis were carried over the corresponding regions of non-coding DNA in search for common transcription factor binding sites. The purpose of comparing the GFP-like promoters is primarily to understand which genes could be involved in the regulation of GFP-like transcription. Although the putative binding elements identified in this study must be verified experimentally, this study will provide the basis for further exploration into the transcriptional regulation of the GFP-like loci spatially, temporally and in response of external stimuli.

Another important aim of this work was to clarify relationships of the Acropora GFP-like genes through the use of phylogenetic analysis. Although some aspects of these analyses were inconclusive due to the limitations of the approach, the results highlight the clear difference between GFP variants expressed in larvae and adult, the which functional significance of which is as yet unclear.

3.2 Results

3.2.1 Gene structure comparison across the phylum Cnidaria

To clarify relationships between the *Acropora* GFP genes, we characterised the intron/exon structures of representatives from the two major clades. Both genes were cloned using conventional PCR-radioactive labelled fragments targeting the third exon as a probe (Figure 3.1A). In the case of amiCP, the fragment containing the gene was cloned from an array of 350,000 phages in a lambda genomic library, whereas in the case of amilFP, the fragment containing the gene was cloned from an array of 90,000 *cfu* in a cosmid library; 17 and 6 positive clones were detected from the primary screening for amilCP and amilFP, respectively. Several single positive clones were digested with *Eco*RI and *Bam*HI to identify differences in banding patterns and estimate the sizes of the inserts (Figure 3.1B). When possible, the digestions were blotted against the same probe to identify fragments for subcloning (Figure 3.1C).

Assembly of the amilCP gene was accomplished using a combination of primer walking from the edges of the 12-kb (amilCP clone 123) lambda fragment, as well as by subcloning *Xba*I- and *Bam*HI-digested fragments into pBluescript and subsequent sequencing and assembling via sequence overlap. The 5’-end of the corresponding cDNA was located next to the T7 promoting region of the lambda clone.

Given that the amilFP (clone cos9.4) fragment was approximately 32 kb, assembly of the gene was accomplished using the amilCP spliced donor–acceptor sites as a model to generate gene-specific primers for use in a long-range PCR reaction. The produced fragments were cloned into pGEMT and sequenced. The 5'-end of the respective cDNA of the amilFP was located approximately 6000 bp from the T3 promoter region of the cosmid clone.

Once the linear fragments were obtained for both genes, the splicing sites were determined by manual alignment of the cDNAs (amilCP597 and amilGFP517a) against the genomic fragment (Figure 3.5A and B for amilCP and amilFP, respectively). Both *Acropora* GFP-like genes were interrupted by four introns in almost identical positions (Figure 3.2, red arrowheads). Although the introns sizes differed significantly, the phases were conserved between the genes (Figure 3.2, right and left numerals in parenthesis). The classic polyadenylation signal AATAAA was observed 85 nucleotides downstream of the FP ORF gene, whereas the less common ATTAAA signal was located 109 nucleotides downstream of the CP gene (data not shown).

When comparing the *Acropora* GFP loci with other cnidarian GFP-like genes cloned from the scleractinian *Montastraea faveolata* (GenBank Accession No. **ABC68475**), the corallimorpharian *Discosoma* sp. (GenBank Accession No. **ABC68474;** Carter *et al*. 2004), the actinarian *Nematostella vectensis* (Figure 3.6) and the hydrozoan *Aequorea victoria (*GenBank Accession No*.* **P42212**; Prasher *et al*. 1992), the first two exons were very similar in all anthozoans, whereas the remaining exons were similar in the scleractinians/actinarians. It appears that the last introns of the corallimorpharian GFP-like loci "DspX" were lost during the evolution of this gene. In all but the second intron of the corallimorpharian GFP-like loci, the classical splicing site 5'-GU-AG-3' had changed to 5'-GA-AG-3'. The jellyfish gene structure showed two introns, one near the chromophore tripeptide site (Figure 3.2, green arrowheads), thus differing considerably from those of the anthozoans. Intron sizes differed dramatically between genes; however, intron 2 appeared to display less variation in length, showing a size range of 624 nucleotides (Figure 3.2).

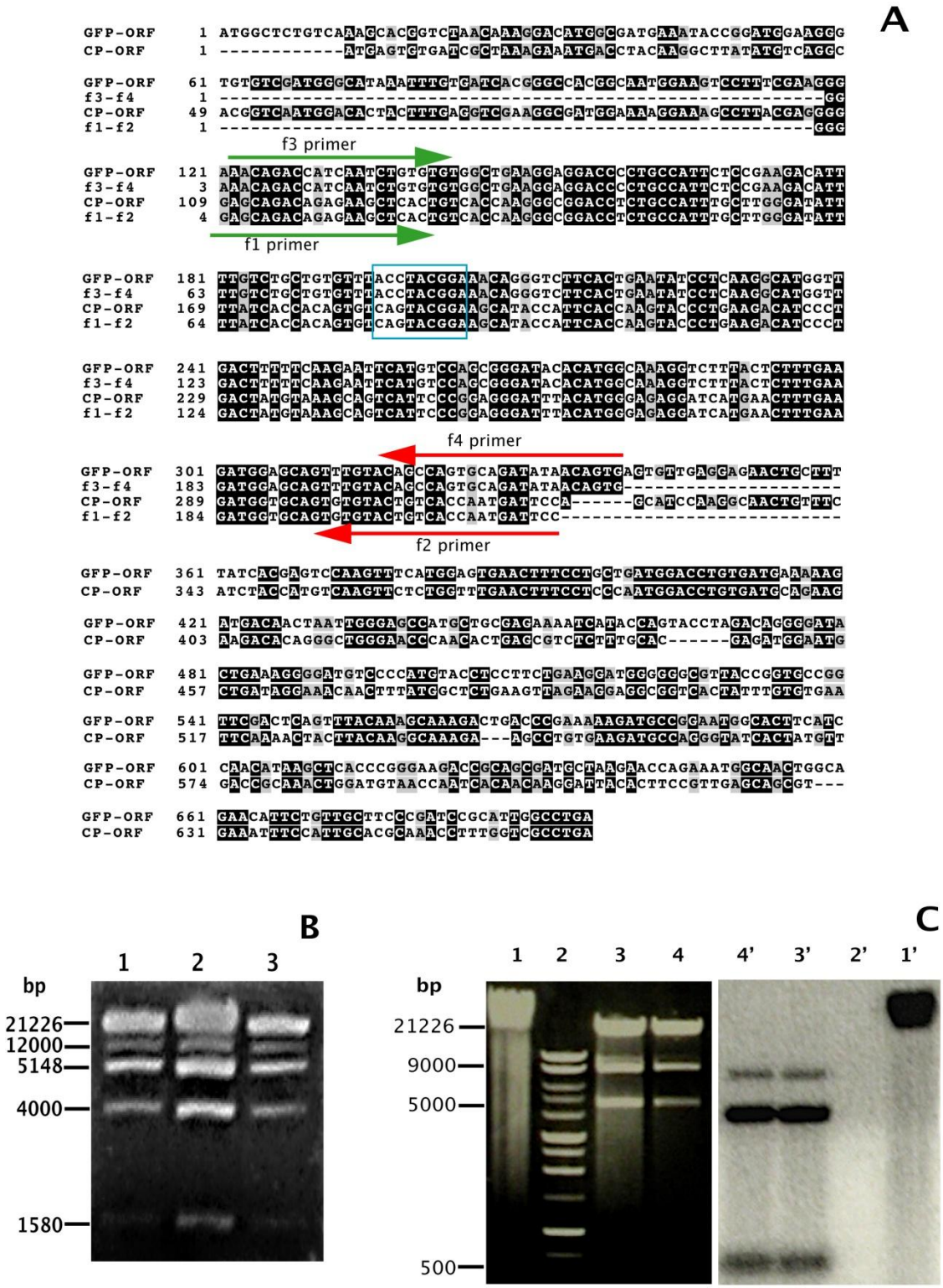

Figure 3.1 Cloning of the nuclear GFP-like genes.
(A) ORF nucleotide sequence comparison of the cDNAs applied as template to generate the probe used to clone the nuclear genes. GFP (amilGFP517a) and CP (amilCP597) correspond to the fluorescent and nonfluorescent phenotype of the protein product. The arrows indicate the actual priming site and direction of transcription in a PCR reaction. The primers f1-f2 and f3-f4 amplify the CP and FP gene templates, respectively. The codons that encode the chromophore structure are listed inside the blue box. (B) Results of the restriction enzyme digestion with EcoR1 over randomly selected clones from the Cp screening. (C) Southern blot analyses of two independent clones digested with BamH1. Lanes on gel: 1, uncut clone; 2, DNA ladder; 3 and 4, clones 123 and 313 (AmilCP), respectively. Lanes on membrane (mirror image): 4'–1' correspond to the lane numbers on the agarose gel.

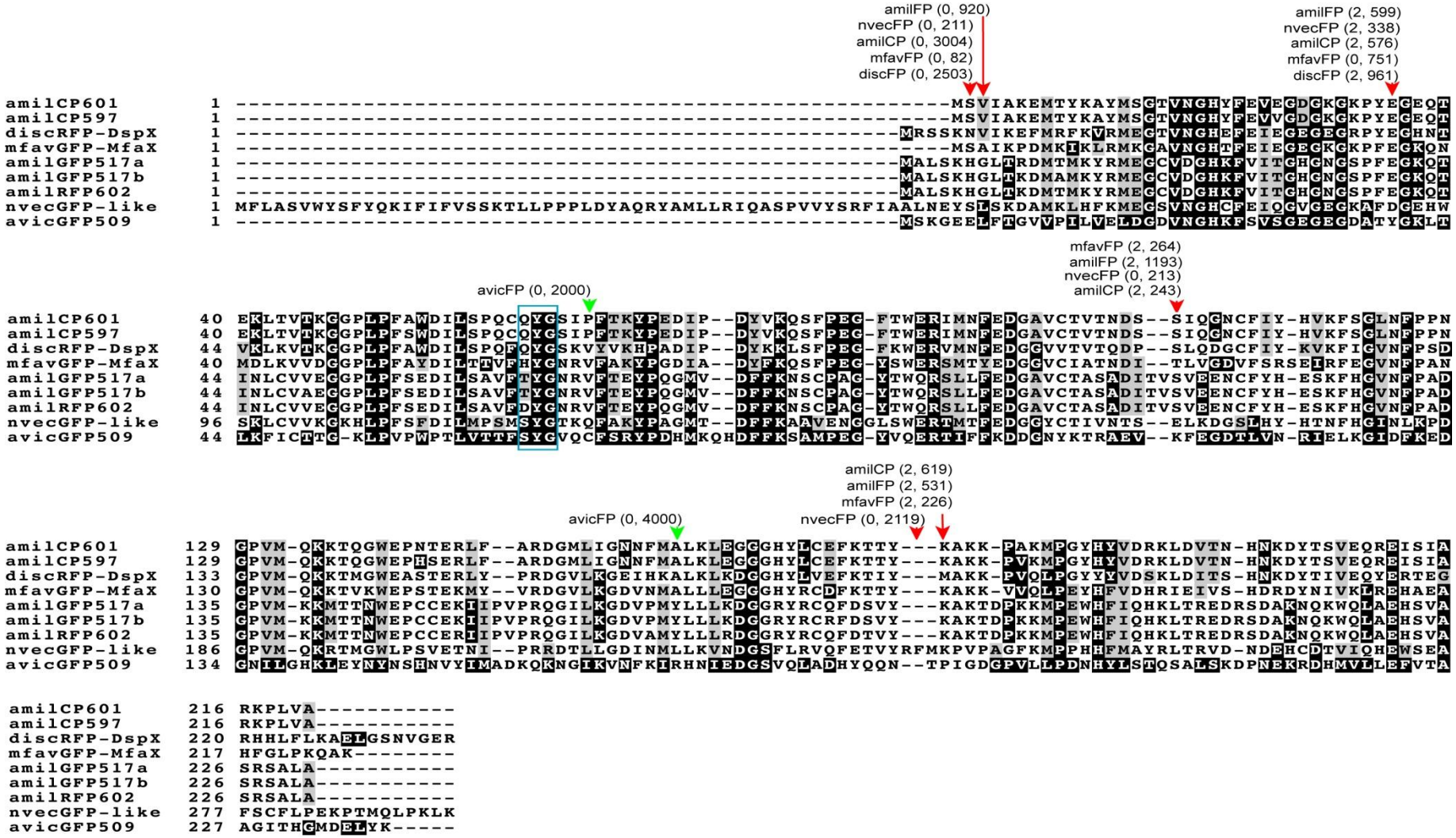

Figure 3.2 Comparison of the intron position over the amino acid alignment from several sequenced cnidarian GFP-like genes. Coloured downward pointing arrows at the top of the alignment represent the positions, and in parenthesis, the phase and size (in nucleotides) of the introns from each gene depending on the cnidarian class. The abbreviations nvecFP, amilFP, amilCP, mfavFP, discFP and avicFP represent *Nematostella vectensis, Acropora millepora, Montastraea faveolata, Discosoma* sp. and *Aequorea victoria* fluorescent protein (FP) and chromoprotein (CP), respectively. Inside the blue box is the tripeptide motif responsible for the chromophore formation. Boxes were made using a 0.2 fraction of sequences that agree for shading. Shading is relative to the avicGFP509.

3.2.2 Comparison of *cis* regulatory regions

Genomic *cis* regulatory regions are responsible for proper spatiotemporal gene transcription. These discrete small DNA regions will bind to specific TFs/proteins. Given that the GFP-like gene structure is essentially the same across the Anthozoa, the possibility exists that aspects of transcriptional regulation might also be conserved. To test this idea, upstream regions from the start codon of the GFP-like genes from the symbiotic corals *A. millepora, M. faveolata* and the sea anemone *N. vectensis* were compared using an online TF search engine (www.ifti.org/cgi-bin/ifti/Tfsitescan.pl). In general, the analyses detected two common groups of sequences of six to seven nucleotides each interspersed along the DNA fragments (Figure 3.3A, coloured nucleotides). The more abundant of these sequences belonged to the FOX family (Figure 3.3b, blue boxes) with up to three sites observed in the amilCP and nvecFP, whereas the other belonged to the Nkx family (Figure 3.3b, red boxes). *In situ* assays of these TF homologues in some anthozoans have revealed strong specific signals in the oral region and endoderm of the planula (de Jong *et al.* 2006). Green fluorescence was also observed in living coral larvae in the mouth tissue as well blue “chromoprotein” endodermal colouration (see results, Chapter 5).

A

```
-301 GCAAATGTGG ATGTCGGAGC TGAAATGCCC TCTGGCCCAA GTGAACCGAC CAAAAAGGTT
-241 CCCGCTGGCT GCTGAATAAT AATATATCAC TTTCACAACA TTAGATTACG TAACGATAAG
-181 TATTGTGTCA TACACGTCGT TGACCTTTTC CCATAAATGG TCCTATTGAC GCCATTAATT
-151 TGGTCTGTTA TCTGTTGGAT ATCTTTTAAA ACACGACTAA TAAAAGAGGT GTACTGGGTG
 -91 AGGCGTGCAT TCACATCAGC GTCTCTTTCG CGGAAGGAGC TACTTAAGAA ACGTTTGCGA
 -31 ATCGTTGTTC TGCGCTACTT AGTGTCAATA TG ...amilCP ORF
```

```
-301 TTGTCATGCA AGAAACGCAG TAGAGTTACT CCAAAAAGAT AAGAAACAGA AAGTCACGTT
-241 TTTCTCAACC CGAAGATTGC AACCAACAAA TGAATGGGTT TAATCAGTAA GGTATTATTA
-181 ATTTTTGCAG TAAGGTATTA CGTTTTTTAT AAAGGACTTT CGGCTTATGG TTGGCTCAAA
-151 TGGATTTTAG GTGTGACATT CAGGTTTAAA ACAGCACTCT CCGCGCTGAG GCCCCATTCA
 -91 CCCAGGTGAT TTGGAAGAGA GCAGATCGTG AAAAAACAAG AGCTAGCTAA GAGCCGTAAG
 -31 GCTGATATCT CACTTTACGT TTACCATCAT G ...mfavFP ORF
```

```
-301 GAAAAACAGA CAGCCTGCCG CAGATTAGCG GTCCTTTTTG GTCTCATGAC CTTGATAAAC
-241 AGATTTATTC CAGTGAGTAT GTATCATATT AAGCTGTCTA AAATCTAATT AGTTTGAAAG
-181 CCCAAGTGTT TGTCTCTTGC ATTTTTTATT GTCTTCGGAG TATAATAAAA AGCGCCCTCT
-151 CAGTAACTTT TCAGAAGTGT CTACTAGCGC GCCAAGCACT ATTGCACCAA GTCCTGATTG
 -91 AGAGGTTGTC TTCAAAGGCC AGTAACAAAA ACACAATGGT AATAGTATTT GTCCTTATTT
 -31 GTGTGCATGA CAATTGACTG AATTCGGTAT G ...nvecFP ORF
```

B

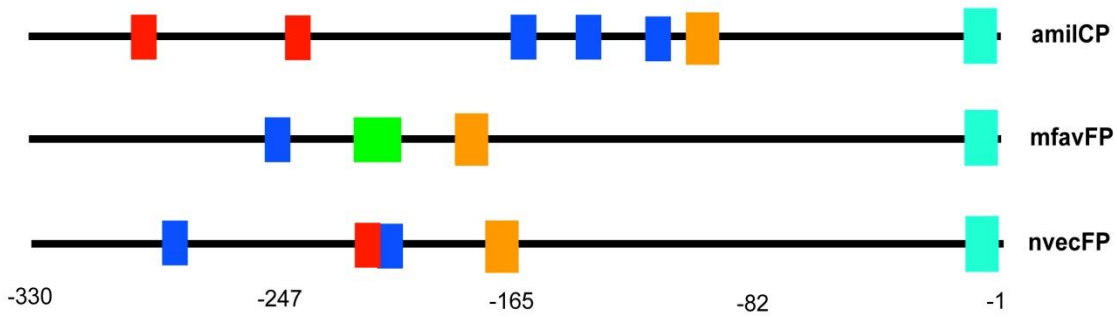

Figure 3.3 Upstream region (from the start codon) comparison of GFP-like genes from anthozoans.
(A) Nucleotide sequence of the 330-bp uptream region of A. millepora, M. faveolata and N. vectensis GFP-like genes showing conserved binding sites along the fragment. The cyan rectangle represents the Kozak sequence, the arrowhead indicates a purine residue characteristic of the binding site, the orange boxes represent the TATA-box consensus sequence and the blue, green and red rectangles represent the binding sequence for the transcription factors FOX, HNF and Nkx, respectively. (B) Diagrammatic representation of the sequences from above. The solid rectangles represent the sequence motifs encased in the rectangle of the same colour. The horizontal lines represent DNA strands.

3.2.3 Evolutionary relationships between GFP-like genes

According to current knowledge, the GFP-like genes evolved with the emergence of metazoans, expanding greatly in diversity in the early linage of cnidarians with, for example, many bioluminescent medusae or colourful anemone/corals. More recently, a few bilaterian species such as copepods and amphioxus have shown GFP-like homologues in their genomes (Bauman et al 2008). Genes coming from amphioxus were publicly available by the time this thesis was submitted, therefore not included in the interphylum phylogenetic analyses. However, it is expected that the tree topology presented here will not change drastically as all bilaterian-based sequences clustered together as observed in the present work, see below.

To clarify patterns of gene relationships that correspond to this extraordinary diversification, conducting a phylogenetic analyses is necessary; however, this conventional approach, especially in the GFP-like family, has failed to resolve issues such as multiple origins of red variants along different lineages or apparent clustering of distantly related species (*e.g.*, corallimorpharians and scleractinians). Many of these inconsistencies may be attributable to the lack of knowledge regarding the biological role of variations and the selective pressures under which variations occur. Despite some irregular scenarios in the reconstructed evolutionary pathways, until the specific functions of the red and green fluorescent or chromoproteins are revealed, conventional phylogenetic analyses have the power to outline general trends; in particular, when only one taxon, in this case *A. millepora*, is compared against several other taxa, the interpretation of the phylogeny is simplified.
Several *Acropora* larval GFP-like clones were used to compare the relationships within scleractinians and the rest of the metazoans.

The coral-derived neighbour-joining (NJ) tree (Figure 3.4A) places proteins from members of the suborder Faviina, for example, *Montastraea, Favia* and *Trachyphyllia*, at the base of the tree. Going inside the branches, two major clades are observed constituted by the chromo and FPs (Figure 3.4A, blue and green rectangles, respectively). *Acropora hyacinthus* CP was at the base of all CPs, whereas *Montipora* sp. GFP was found to be the closest fluorescent-relative to the CP clade. Note that the larval-derived *A. millepora* CPs were more similar to their homologue in *A. palmata* than to their own adult-derived sequence (*A. millepora*-CP-matz), which cluster with *Acropora tenuis* and *Goniopora tenuidens* CPs.

In the case of the FP clade, sequences derived from this work formed part of the basal clade (as in the case of larval CPs), whilst adult-derived sequences were different from their larval counterpart, for example, AmillGFP-matz and the AmillGFPa and b.

In the more comprehensive analyses (Figure 3.4B), sequences derived from this work were also clearly resolved into two major clades (Figure 3.4B). In the case of the CP-like genes, the analyses indicated Scleractinia as being a sister group and corallimorpharian RFPs as the closest relatives. In contrast, the *Acropora* FPs were more similar to Zoanthidea-like genes rather than their coral counterparts. In general, Actinaria, Bilateria and Hydrozoa-“related” sequences formed well-defined clusters, whereas Octocorallia GFPs were found at the very base of the whole tree.

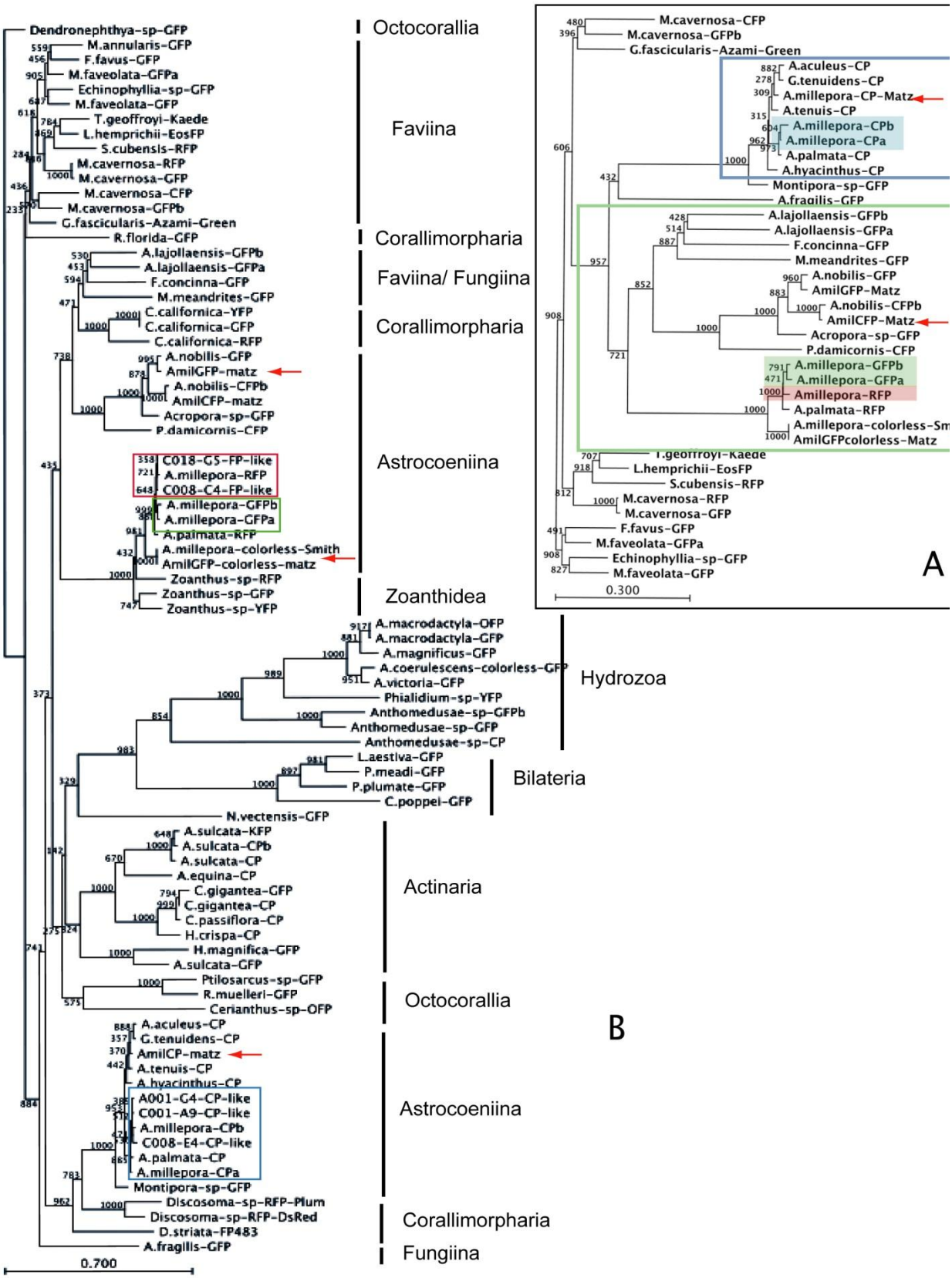

Figure 3.4 Phylogenetic analyses of the GFP-like genes using the neighbour joining algorithm to construct the tree.

(A) Scleractinia-based analyses. (B) General interphylum comparison. Red arrowheads point to sequences derived from *Acropora millepora* adult tissue. In both trees, the sequences derived from this work were depicted by coloured rectangles depending of the colour of the protein product. Numbers in each node represent the proportion of 1000 bootstrap iterations. Scale bars represent the number of substitutions per site. Sequence accession number and reference are given in Table 4.

Table 4. Information related to all GFP-like proteins used to construct the general interphylum evolutionary tree.

Organism sp/suborder-class	Protein name	NCBI No.	Reference	Emi/Abs max (nm)
Acropora millepora /Astrocoeniina	AmilCFP-Matz	AAU06849	Direct Submission	487
Acropora millepora /Astrocoeniina	AmilGFP-Matz	AAU06846	Direct Submission	515
Acropora millepora /Astrocoeniina	AmilGFPcolorless-Matz	AAT77753	Direct Submission	none
Favia favus /Faviina	F.favus-GFP	AB193294.1	Tsutsui et al., 2005	515
Lobophyllia hemprichii /Faviina	L.hemprichii-EosFP	AAV54099	Wiedenmann et al 2003	516-581
Echinophyllia sp/Faviina	Echinophyllia-sp-GFP	BAD72874	Ando et al., 2004	503
Trachyphyllia geoffroyi/ Faviina	T.geoffroyi-Kaede	BAC20344	Ando et al., 2002	519-582/509-572
Agaricia fragilis /Fungiina	A.fragilis-GFP	AAK71331	Direct Submission	undetermined
Galaxea fascicularis /Faviina	G.fascicularis-Azami-Green	BAD52001	Karasawa et al., 2003	492
Scolymia cubensis/Faviina	S.cubensis-RFP	AAU06843	Direct Submission	576
Montipora sp/Astrocoeniina	Montipora-sp-GFP	BAE93225	Kogure et al., 2006	520/440
Acropora millepora /Astrocoeniina	Amillepora-RFP	pendent	This work	610/563
Acropora millepora /Astrocoeniina	A.millepora-GFPa	pendent	This work	515/504
Acropora millepora /Astrocoeniina	A.millepora-GFPb	pendent	This work	515/505
Acropora millepora /Astrocoeniina	A.millepora-colorless-Smith	AAT77753	Direct Submission	none
Acropora millepora /Astrocoeniina	A.millepora-CPa	pendent	This work	597
Goniopora tenuidens /Fungiina	G.tenuidens-CP	Q95P04	Gurskaya et al., 2001	571
Acropora millepora /Astrocoeniina	A.millepora-CPb	pendent	This work	601
Acropora millepora /Astrocoeniina	A.millepora-CP-Matz	AAU06854	Direct Submission	588
Acropora aculeus/Astrocoeniina	A.aculeus-CP	AAU06856	Direct Submission	580
Acropora hyacinthus/ Astrocoeniina	A.hyacinthus-CP	AAU06855	Direct Submission	580
Acropora tenuis/Astrocoeniina	A.tenuis-CP	AAU06853	Direct Submission	580
Montastraea cavernosa/Faviina	M.cavernosa-GFP	AAQ90465	Sun et al., 2004	432/477
Meandrina meandrites /Faviina	M.meandrites-GFP	AAO00732	Sun et al., 2004	398/505
Montastraea cavernosa/ Faviina	M.cavernosa-CFP	AAL17905	Sun et al., 2004	432/477
Montastraea faveolata/ Faviina	M.faveolata-GFP	ABC68475	Carter et al., 2004	undetermined
Acropora sp/Astrocoeniina	Acropora-sp-GFP	BAD24723	Karasawa et al., 2004	472/497
Fungia concinna /Fungiina	F.concinna-GFP	BAD24721	Karasawa et al., 2004	547/562
Astrangia lajollaensis /Faviina	A.lajollaensis-GFPa	AAS18272	Direct Submission	undetermined
Astrangia lajollaensis /Faviina	A.lajollaensis-GFPb	AAS18271	Direct Submission	undetermined
Pocillopora damicornis/ Astrocoeniina	P.damicornis-CFP	AAU4450	Carter et al., 2004	492
Montastraea faveolata/ Faviina	M.faveolata-GFPa	AAU4449	Carter et al., 2004	508
Montastraea cavernosa/ Faviina	M.cavernosa-GFPb	AAU04448	Carter et al., 2004	502
Acropora nobilis/ Astrocoeniina	A.nobilis-CFPb	AAU06851	Direct Submission	485
Montastraea cavernosa/ Faviina	M.cavernosa-RFP	AAO61598	Kelmanson and Matz, 2003	582
Acropora palmata/ Astrocoeniina	A.palmata-CP	pendent	Medina-A.palmata-ESTproject	undetermined
Acropora palmata/ Astrocoeniina	A.palmata-RFP	pendent	Medina-A.palmata-ESTproject	undetermined
Corynactis californica /Corallimorpharia	C.californica-GFP	AAZ67343	Direct Submission	undetermined
Corynactis californica /Corallimorpharia	C.californica-RFP	AAZ14787	Direct Submission	undetermined
Dendronephthya sp/Octocorallia	Dendronephthya-sp-GFP	AAM10625	Labas et al., 2002	508
Actinia equina /Actinaria	A.equina-CP	ABA41382	Shkrob et al., 2005	597
Anemonia sulcata /Actinaria	A.sulcata-CP	AAG41206	Wiedenmann et al 2000	562
Condylactis passiflora /Actinaria	C.passiflora-CP	Q95W11	Gurskaya et al., 2001	571
Anemonia sulcata /Actinaria	A.sulcata-GFP	AAG41205	Wiedenmann et al 2000	499
Anemonia sulcata /Actinaria	A.sulcata-CPb	AAG02385	Lukyanov et al., 2000	568
Pontellina plumata /Artropoda	P.plumate-GFP	AAQ01183	Shagin et al., 2004	502
Condylactis gigantea/Actinaria	C.gigantea-GFP	AAK71343	Direct Submission	undetermined
Zoanthus sp/Zoanthidea	Zoanthus-sp-YFP	Q9U6Y4	Matz et al., 1999	538
Zoanthus sp/Zoanthidea	Zoanthus-sp-GFP	Q9U6Y5	Matz et al., 1999	506
Discosoma sp/Corallimorpharia	Discosoma-sp-RFP-DsRed	Q9U6Y8	Matz et al., 1999	583
Discosoma striata /Corallimorpharia	D.striata-CFP	Q9U6Y7	Matz et al., 1999	483
Anemonia majano/Actiniaria	A.majano-CFP	Q9U6Y6	Matz et al., 1999	486
Condylactis gigantea /Actiniaria	C.gigantea-CP	AAL27537	Gurskaya et al., 2001	571
Heteractis crispa /Actiniaria	H.crispa-CP	Q95W85	Gurskaya et al., 2001	578
Anthomedusae sp/Hydroida	Anthomedusae-sp-CP	AAR85352	Shagin et al., 2004	572
Aequorea coerulescens /Hydroida	A.coerulescens-colorless-GFP	AAN41637	Gurskaya et al., 2003	none
Cerianthus sp/Ceriantipatharia	Cerianthus-sp-OFP	AAP55761	Direct Submission	undetermined
Anthomedusae sp/Hydroida	Anthomedusae-sp-RFP-KillerRed	AAY40168	Bulina et al., 2006	610
Anemonia sulcata /Actinaria	A.sulcata-KFP	AAO43187	Chudakov et al., 2003	600
Renilla muelleri /Octocorallia	R.muelleri-GFP	AAG54098	Direct Submission	undetermined
Ptilosarcus sp/Octocorallia	Ptilosarcus-sp-GFP	AAG54097	Direct Submission	undetermined
Corynactis californica /Corallimorpharia	C.californica-YFP	AAZ67343	Direct Submission	undetermined
Corynactis californica /Corallimorpharia	C.californica-RFP	AAZ67342	Direct Submission	undetermined
Heteractis magnifica /Actiniaria	H.magnifica-GFP	AAO16871	Tu et al., 2003	510
Ricordea florida /Corallimorpharia	R.florida-GFP	AAU06844	Direct Submission	517
Montastraea annularis /Faviina	M.annularis-GFP	AAK71332	Direct Submission	undetermined
Zoanthus sp/Zoanthidea	Zoanthus-sp-RFP	AAL23574	Labas et al., 2002	576
Aequorea macrodactyla /Hydroida	A.macrodactyla-GFP	AAL33918	Luo et al., 2006	502
Aequorea macrodactyla /Hydroida	A.macrodactyla-OFP	AAL33917	Luo et al., 2006	523
Chiridius poppei /Arthropoda	C.poppei-GFP	BAE78442	Masuda et al., 2006	517
Discosoma sp/Corallimorpharia	Discosoma-sp-RFP-Plum	AAV65487	Wang et al., 2004	649
Anthomedusae sp/Hydroida	Anthomedusae-sp-GFP	AAR85351	Shagin et al., 2004	504
Anthomedusae sp/Hydroida	Anthomedusae-sp-GFPb	AAR85351	Shagin et al., 2004	495
Phialidium s sp/Hydroida	Phialidium-sp-YFP	AAR85349	Shagin et al., 2004	537
Pontella meadi/Arthropoda	P.meadi-GFP	AAQ01187	Shagin et al., 2004	502
Labidocera aestiva /Arthropoda	L.aestiva-GFP	AAQ01185	Shagin et al., 2004	506
Nematostella vectensis /Actinaria	N.vectensis-GFP	pendent	WGS-Nematostella project	undetermined
Aldersladia magnificus /Hydroida	A.magnificus-GFP	pendent	Avila-Soria et al., in prep	509/398
Aequorea victoria /Hydroida	A.victoria-GFP	AAA27722	Prasher et al., 1992	504
Acropora millepora /Astrocoeniina	AmilRFP-Matz	AAU06852	Direct Submission	593
Entacmaea quadricolor /Actiniaria	E.quadricolor-RFP	AAN05449	Wiedenmann et al., 2002	611

3.3 Discussion

3.3.1 *Acropora* GFP-like loci: An ancient gene structure

The present work details the structure of two nuclear genes encoding for two products belonging to the fluorescent and non-fluorescent protein family. It is clear from the EST collection that there are more than these two alleles, however it was out of scope of this thesis to quantify the total number of genes in the acropora genome. Although the intron number of certain non-sequenced alleles could diverge (as for the amphioxus GFP gene array, see below) from the structural description presented here, the following evolutionary discussions summarised some common trends amongst the available data, however, it is possible that in the future, when the whole set of GFP-like genes from Acropora and the other organisms mentioned here, that a more comprehensive picture of the evolution of the GFP-like family will be possible.

Sequencing of the two *Acropora* GFP-like loci revealed very similar/identical intron/exon structures: four introns interrupted the gene (Figure 3.2). This fact, together with the structural similarity of both proteins (Figure 3.4; Chapter 4), argues for a common origin. This similarity in the intron position appears not to be exclusive to the *Acropora* GFP-like genes when compared against its homologues in *M. faveolata* (a common massive Caribbean coral) and the actinarian *N. vectensis*, which implies that this signature was present in the common Anthozoa-like ancestor. The soft coral *Discosoma* sp. GFP-like gene, however, is interrupted by only two introns instead of four, splitting the gene in the similar first positions of the coral/anemone loci. In this context, the phylogenetically broader gene structure comparison argues for an intron-rich scenario for genes belonging to ancient lineages, with a common trend to reduce its number (intron loss) according to the time of origin of the new bearing taxa. This effect can be amplified with a possibly different ecology/physiology that promotes its strong divergence (Rogozin *et al.* 2003, 2005, Sverdlov *et al.* 2004). In the present work, for example, Actinaria represents one of the more ancient cnidarian lineages dating its origins from 560 to 720 million years (Myr) (Chen 2002); Scleractinia originates at some point approximately 250 Myr ago and Corallimorpharia represents the most recent (Anthozoa, Hexacorallia) form, dating its origins to about 130 Myr ago (Chen *et al.* 2002, Medina *et al.* 2006). The gene structure comparison of the GFP gene amongst

the anthozoan cnidarians appears to follow the “normal” intron loss (time–taxa-derived) toward the 3’-end of the gene in the case of the corallimorpharian GFP-like gene.

In relation to the discrepancy in the GFP-like gene structure between jellyfish and anemones/corals, the reasons that could explain this difference are based on the general agreement based on anatomical and genetic evidence in considering anthozoans to be more primitive than hydrozoans (Bridge *et al.* 1992, 1995, Collins 2002, Collins *et al.* 2006). The anthozoan-based GFP-like gene has retained its primitive structure, changed basically through time in number rather than in position of the intron; the hydrozoan homologue, however, has experienced major gene organization, considering that since the time of the first evolved jellyfish (bearing the GFP-like gene) from an anthozoan-like ancestor, the deviation in the original gene organization was amplified by the new adult planktonic lifestyle and bioluminescence acquisition, which implies different selective forces acting on its evolution.

Despite the bilaterian genes not being aligned against those of cnidarians, the amphioxus genome yielded a total of 12 different GFP-like alleles having an intron average equal to 3.41 per gene, with the two first introns at a similar position on the coral/anemone loci and the rest scattered on the second-third of the molecule. Some of the genes, however, contain only the first two introns, whilst others have up to five introns (Baumann *et al.* 2008). The intron-rich structure and the partial alignment of some of them against the anthozoan genes argue for a close relationship. Indeed, none of the sequenced amphioxus are bioluminescent (Deheyn *et al.* 2007), the same as in the several scleractinian species from which GFPs have been cloned. Thus, the biological role may be also related. Finally, a current notion of gene loss within the ecdysozoan lineage (Miller *et al.* 2007, Putnam *et al.* 2007) could predict different arrangements and complexity for the GFP homologue in marine crustaceans (Shagin *et al.* 2004, Masuda *et al.* 2006).

Sequencing of the CP locus suggested the possibility of tandem organisation of the GFP loci. Indirect evidence, based on Southern blotting (Figure 3.1C) and a complete assembled gene with a fraction of another, both present in the same lambda clone, points to an organisation type “head to tail” with approximately 2.8 kb between each unit. The possibility of tandem-arrayed GFP-like genes provides starting material for the evolution of new genes, allowing an increase in genome complexity (Graham 1995). In *Acropora*, this type of arrangement has been found in

many homeobox and nuclear receptor genes (Grasso *et al.* 2001, Hislop 2003, Hislop *et al.* 2005). In other cnidarians, such as *Nematostella*, *snail* and *mox* genes also were duplicated (Martindale *et al.* 2004), and recently, the genome draft of the starlet anemone revealed five copies of FP genes close to neurotoxin domains (Putnam *et al.* 2007).

3.3.2 Possible transcriptional GFP regulation

By comparing the immediate upstream DNA regions from equivalent genes of different anthozoan species, *cis* regulatory elements (or TF sequences) can be identified. These regions are protected from random drift during evolutionary time periods (Iwama & Gojobori 2004). Regulatory sequences normally lie outside the coding sequences and are intermixed in a highly variable sequence background. The detection of TF sequences in cnidarian genomes using a bilaterian weight matrix is not unusual (Thomsen *et al.* 2004) and points out that the TF proteins may retain their binding specificity during the course of evolution.

The phylogenetic footprint analyses showed that sequences that bind to members of the FOX and Nkx were present in three different anthozoans (Figure 3). These TFs have been observed to be involved in cnidarian foot and axial patterning (Siebert *et al.* 2005). *Acropora* and *Nematostella* expression analyses of these TFs have shown specific mouth and endodermal localisation. GFP-like mRNA and protein are also observed in the mouth and endodermal tissue (see Chapter 5). The apparent TF and FP tissue overlapping may suggest that the spatial arrangement of the FPs in *Acropora* may be under the regulation of these TFs. Although the information observed in the upstream region needs to be corroborated by experimental analyses, those TFs are anticipated to affect the GFP expression. Furthermore, TFs have been found to be sequence-specific, and thus the spatial binding-site arrangement differences observed amongst the upstream GFP genes may account for specific positional gene expression between corals and anemones.

3.3.3 Phylogenetic relationships of *Acropora* GFPs

The interphylum evolutionary tree locates an octocoral FP sequence in the basal node and clearly distinguishes the hydrozoan and bilaterian sequences as separate clades (Figure 3.4B), consistent with previous studies based on less comprehensive coral data sets (Labas *et al.* 2002, Shagin *et al.* 2004). However, in the GFP-based coral tree, Astrangia, Fungia and Meandrina FPs form a solid cluster (see Figure 3.4A), which correlates with 16S and mitochondrial DNA data used to investigate phylogenetic relationships within the Scleractinia (Romano & Palumbi 1997, Medina *et al.* 2006). Furthermore, *Lobophyllia hemprichii* “EosFP” and *Favia favus* GFP are clustered with the rest of the favid corals, nearly at the base of the trees. This trend is also supported by morphological and 28S and 16S genes, indicating a robust clade and suggesting that the common ancestor dates at the origins of the Scleractinia (Romano & Palumbi 1997, Daly *et al.* 2003, Medina *et al.* 2006).

In both the interphylum- and the coral-based phylogeny, adult and larval GFPs from *A. millepora* were located apart. In a closer look to the alignments shows that most of the substitutions faces outside the *ß* barrel (Figure 6.2, Appendix). Moreover, this protein divergence is also reflected in their spectroscopic characteristics (see Chapter 4). Whether the amino acid differences, and therefore whether the different optical features between adult and larval GFPs promote some benefit to the organism or could have implications in their usage during the life of the coral, still remains unclear and warrants elucidation.

Finally, despite the abundance of anthozoan and hydrozoan FPs, the intraphyla evolutionary tree may have major gaps; for example, no scyphozoan or cubozoan GFP-like FPs (if they exist) have been identified. The evolution of the GFP-like genes, at least in the early branches, is far from being understood, although within the Scleractinia, this “marker” may be useful in clarifying the complex colour phenomena and their origins. A more accurate reconstruction of these genes will be possible as more sequence data become available.

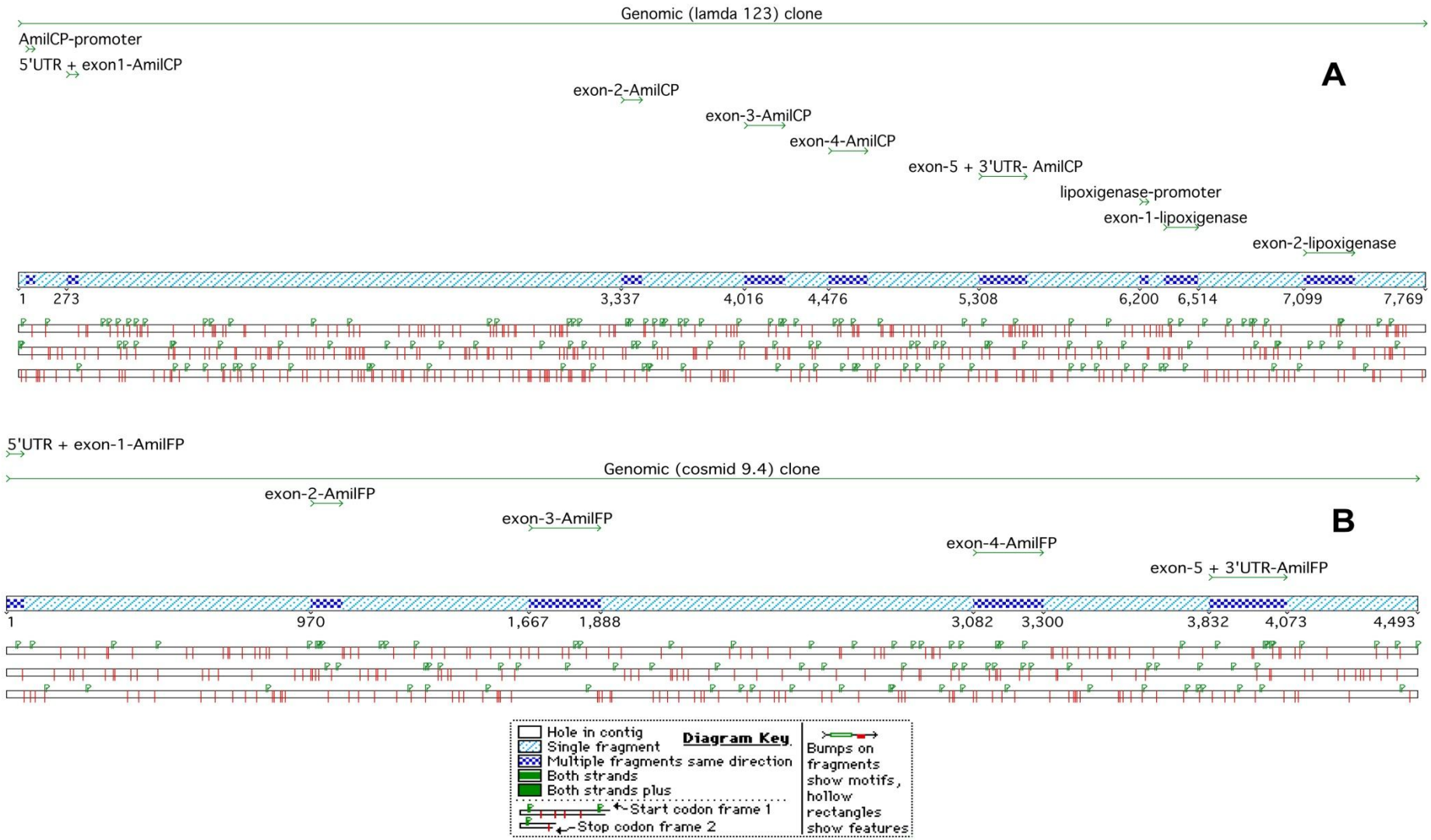

Figure 3.5 Schematic representation of the *Acropora millepora* GFP-like loci and putative gene neighbours.
(A) *Acropora millepora* chromoprotein (AmilCP) intron/exon structure followed by the lipoxygenase-like locus along lambda clone (123) from the genomic library. (B) *Acropora millepora* fluorescent protein (AmilFP) gene organisation, sequenced from a cosmid clone (9.4) in the genomic library. The numbers under the shaded bar represent the length of the fragment in base pairs. The figure was produced using the software Sequencher 4.2.1 and exported to Illustrator for further elucidation. Complete sequences are given in the Appendix.

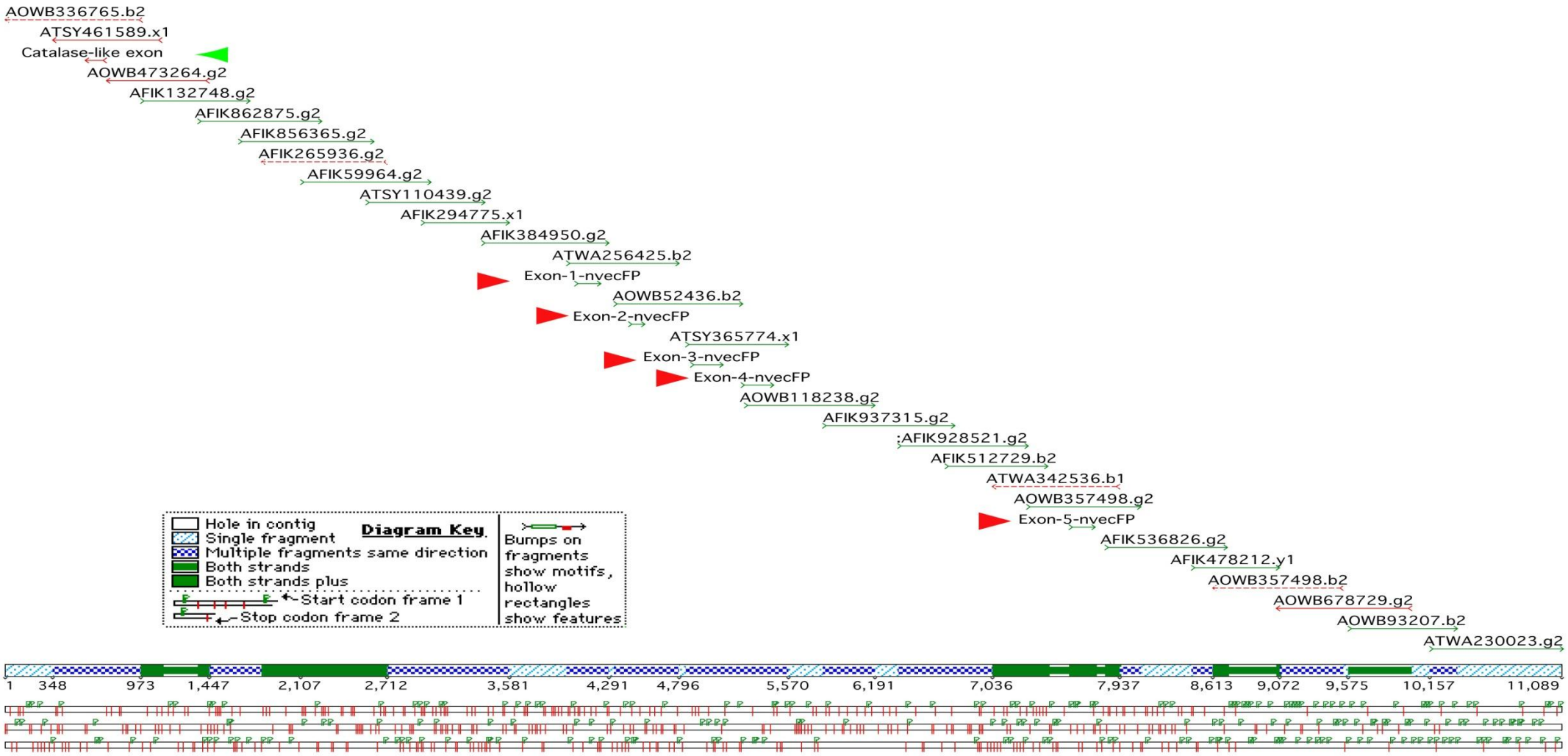

Figure 3.6: Schematic representation of the *Nematostella vectensis* GFP-like loci.
The figure was exported using the software package Sequencher 4.2.1. The accession numbers of each of the constitutive sequences are depicted on top of the respective fragments. Numbers under the shaded bar represent the length of the fragment in base pairs. The exons of the nvecFP gene are represented by arrows. The green arrow represents a catalase-like exon. The complete sequence is detailed in the Appendix.

3.4 Conclusion

Based on gene-specific PCR over the lambda and cosmid clones of 12 and 32 kb long, respectively; two *Acropora* GFP-like nuclear gene sequences were elucidated, revealing similar intron/exon sequence structures, possible tandem-arrayed organization and a lack of linkage.

Phylogenetic analyses clearly separated the amilCP and amilFP genes in two well specified clades, whereas adult derived sequences seemed to differ from their larval counterparts.

Similar patterns of transcription binding sites were found upstream of three different GFP-like genes, suggesting possible implications in regulatory expression.

CHAPTER 4

BIOCHEMICAL AND SPECTRAL ANALYSES OF DIFFERENT RECOMBINANT GFP-LIKE PROTEINS FROM THE ACROPORA PRESETTLEMENT CDNA LIBRARY

4.1 Introduction

GFP-like proteins represent a diverse group of molecules found mainly in Hydrozoa and Anthozoa cnidarians and a few marine bilaterians (Shagin et al. 2004, Masuda et al. 2006, Mocz 2007). This diversity can be broadly categorized into different fluorescent (green, yellow, and red) and non-fluorescent (purple-blue) colour phenotypes (Labas et al. 2002). The first GFP-like protein was isolated and cloned from the bioluminescent hydrozoan jellyfish *Aquorea victoria* (Prasher et al. 1992, Prasher 1995, Niwa et al. 1996, Ormo et al. 1996). In the jellyfish, the GFP (avGFP) is coupled with a photoprotein called Aequorin and is bioluminescent. The photoprotein binds a luciferin-like molecule called coelenterezine that, upon oxidation, releases a photon. The oxidation of coelenterezine is mediated by conformational changes of Aequorin that are induced by the binding of Ca^{2+} ions released by the cell. The light produced from this reaction is blue (490 nm); however, when photons are absorbed by the avGFP chromophore (p-hydroxybenzyl dineimidazolidinone), the phenoxide group acts as a donor to the imidazolidinone acceptor, and, by a charge-transfer mechanism, light is reemitted as fluorescence in the green spectral region (Morin 1974).

The GFP-like protein is composed of 11 β-sheets that fold to form a cylinder 30 Å in circumference and 40 Å in height (Tsien 1998, Kummer 2000). An α-helix bearing the tripeptide “XYG” motif (positions 63–65) runs through the centre of the cylinder and is responsible for the fluorescent properties of the protein. The tertiary arrangement of the protein is responsible for its green fluorescence. A series of torsional adjustments that occur during protein folding relocate the carboxyl carbon of one residue close to the amino group of the other; then, by oxygen-dependent cyclization and dehydrogenation reactions, an imidazolin-5-one heterocyclic ring systems is formed

(Cubitt et al. 1995, Barondeau et al. 2003). For the red fluorescence to occur (e.g., in corallimorpharian *Discosoma* DsRed), an additional oxygen-dependent dehydrogenation that extends the “green-like” structure by two electron-withdrawing double bonds is required (Baird et al. 2000, Fradkov et al. 2000, Gross et al. 2000). In certain scleractinean GFP-like proteins, the red fluorescence is acquired by photon-mediated fragmentation of the protein backbone, resulting in the acquisition of the extended red-like chromophore (Pakhomov et al. 2004, Oswald et al. 2007). The behaviour of these chromophores allows for the tracking of desired objects in the cell. For example, cellular organelles labelled with GFP-like proteins that are exposed to ultraviolet light emit in the red region of the spectrum (Lukyanov et al. 2005). The yellow fluorescent protein (FP), cloned from a sea anemone (*Zoanthus* sp.), has a chromophoric arrangement in which a transiently appearing DsRed-like acylamine reacts with the terminal group of L66 to form a new six-membered ring that accounts for the yellow shift (Remington et al. 2005). Reactions similar to those that occur with the red chromophore also take place in non-fluorescent GFP-like chromoproteins (CPs); however, the non-coplanar conformation of the light-absorbing molecule results in an inability of the protein to emit light (Fradkov et al. 2000, Dove et al. 2001, Bulina et al. 2002, Verkhusha et al. 2004, Chan et al. 2006). Nevertheless, intense green light can still promote red fluorescence from supposed non-fluorescent CPs. CPs with the capability to fluoresce upon radiation are called kindling FPs (Chudakov et al. 2003).

The protein barrel is often a monomer in hydrozoan GFP-like molecules; however, most non-bioluminescent anthozoan GFPs are tetrameric. Their aggregation behaviour is mediated by the presence of two (hydrophobic and hydrophilic) interfaces formed by the interaction of the external side chains of different monomers (Dove et al. 2001, Campbell et al. 2002, Martynov et al. 2003, Gurskaya et al. 2006). In the biological context, aggregation has been related to light-trapping mechanisms, such as in reef organisms exposed to intense sunlight (Salih et al. 2004, Andresen et al. 2005). In the laboratory, however, this natural tendency for oligomerization represents a problem when the proteins are to be used as a tag (Campbell et al. 2002, Gurskaya et al. 2006).

Although much is known about GFP-like proteins in adult coral, there are no data concerning the actions and roles of these proteins in the early developmental stages. This is important because age-dependent changes in gene expression and GFP-like

protein composition could relate to the acquisition of symbionts in the free-living stage and in metamorphosis.

This chapter discusses the cloning and characterization of some GFP-like proteins from Acropora in the pre- and postsettlement stages. The proteins were characterized by their spectral properties (i.e., absorbance, excitation, and emission), molecular masses, and responses to changes in pH. Data acquired from these studies were compared to information found in the literature for all GFP-like molecules, and their similarities and differences are discussed.

4.1.1 Statement of Goals

By characterising several GFP-like variants it is possible to explain at the molecular level how certain mutations could be reflected into some measurable phenotypic characteristic such as differences in optical or solubility properties. With such information it is possible in principle not only to estimate the potential colour variation in the species, but also to explore the link between (non-spectroscopic) properties of the proteins and their possible roles in the biology of the coral. With respect to colour variation, the present study focussed mainly on larval material. Despite the fact that some of the proteins expressed in adult corals are spectroscopically similar to those expressed in larvae, the amino acid sequences are quite different but the functional significance of this is unclear. In addition, some of the mutant recombinant proteins produced had altered solubility relative to the parental protein. Some of the natural variants detected apparently differed with respect to oligomerization pattern. Other mutations not clearly implicated in spectroscopic or aggregative properties were analysed in a search for protein-protein binding patches potentially involved in heterologous protein-protein interactions.

Pure protein preparations were subjected to pH gradients in order to investigate the chemical characteristics of the chromophore. This information is critical for understanding the nature of the chromophore in the *A.millepora* proteins, and enables comparison with published information. This study also sheds some light on biophysical aspects of the GFP-like proteins and may have biotechnological applications.

4.2 Results

Results pertaining to the diversity of the expressed sequence tag (EST) clones present in the larval cDNA library and the phenotypic variations associated with the fluorescence or colour of each protein are found in this section. Three-dimensional (3D) models of some of the proteins were used to understand how “hot spots” in protein structure confer spectral characteristics of GFP-like proteins. In addition, characteristics such as molecular mass and response to pH are described for each cloned GFP-like protein.

4.2.1 GFP-like protein diversity observed in larvae ESTs

Information on the abundance/diversity of the GFP-like EST sequences was derived from an extensive array of up to 9000 sequences from three developmentally important stages: the prawn chip, planula or presettlement stage, and postsettlement stages that occur just after metamorphosis. Most of the sequences were derived from the planula stage library, although others came from the postsettlement library. No complete transcripts were detected at the prawn chip stage. All detected GFP-like sequences were grouped into two main contigs according to their nucleotide sequence composition. No other contigs containing a GFP-like sequence were observed in the EST collection. After visual comparison, each of the two contigs was revealed to contain fluorescent-like and non-fluorescent, CP-like genes.

The primary structures of the CPs recorded in the EST collections are compared in **Figure 4.1**. The non-fluorescent CP contig record showed 17 transcripts that coded for proteins 221 amino acids in length. Amino acid alignments of the different CPs were performed relative to the most abundant clone. On top of each line, the specific mutation is denoted with the relative proportion in parentheses. The alignment showed an overall variation of 5% in this molecule, with only 2–12 variable substitutions.

Figure 4.2 shows the primary structures of FPs. The FP contig contained a total of 31 clones, each with different frequencies. The length of the transcripts in all cases was 231 amino acids. The analysis was performed relative to the most abundant clone so that the substitutions were easy to detect. On the top of each line, the specific amino acid substitution is presented, and in parentheses is the proportion of that substitution. The amino acids responsible for fluorophore formation or those

involved in final autocatalytic fluorescence/colour development are highlighted. The alignment showed a 6% variation in the molecule, with only 4 of the 13 variable positions important for FP proportion.

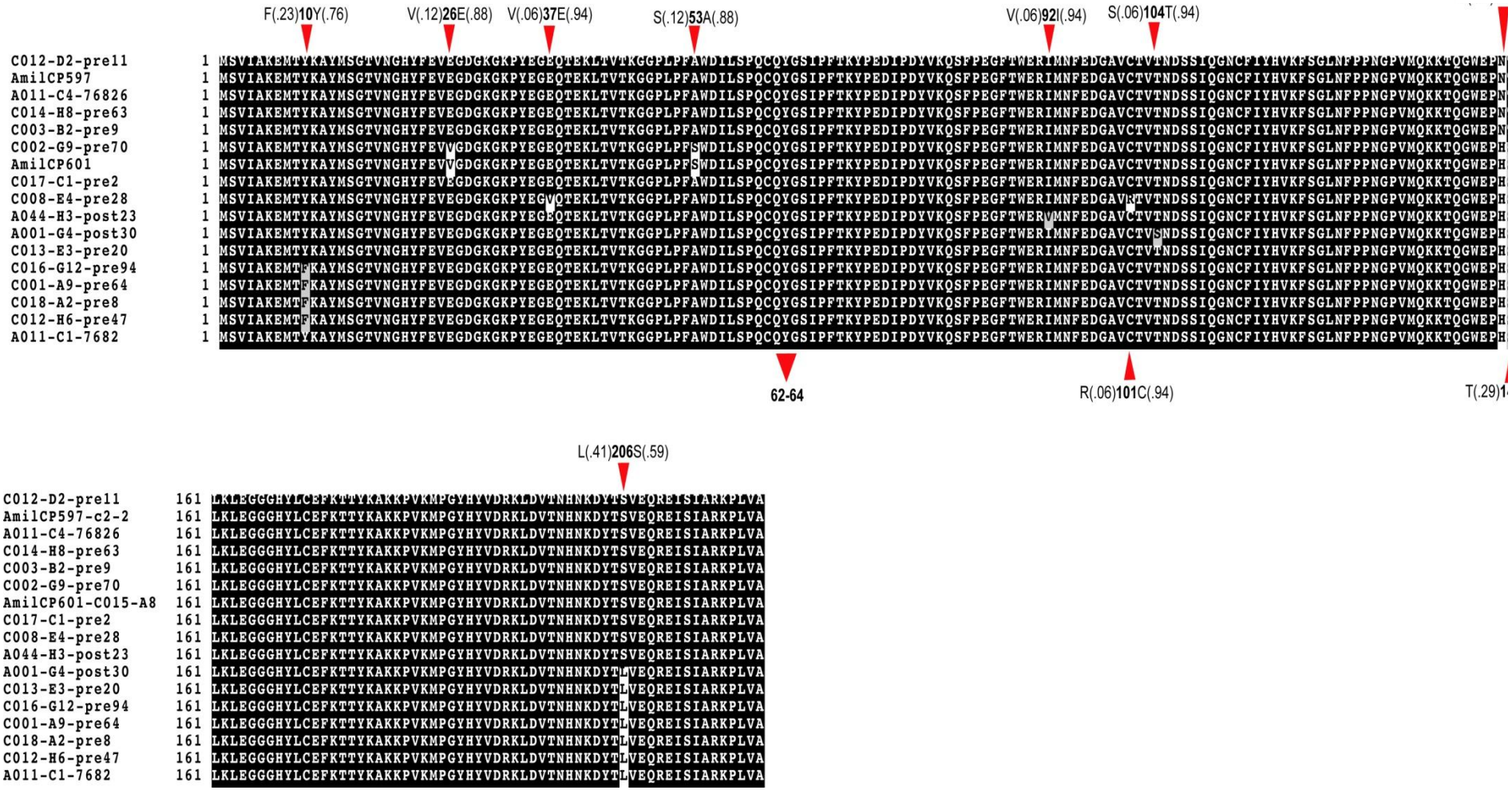

Figure 4.1 Amino acid alignments of all complete clones that belong to the Chromo-like contig.
The names on the left represent the assigned identification for the EST clone and also include two bacterially expressed CPs that have distinctive absorbance maxima. The arrows point to the substitutions. Information concerning the proportion of each amino acid is shown in parentheses, and the number in the polypeptide chain is shown in bold type. The alignment was made relative to the clone amilGFP597, which was most abundant in the contig.

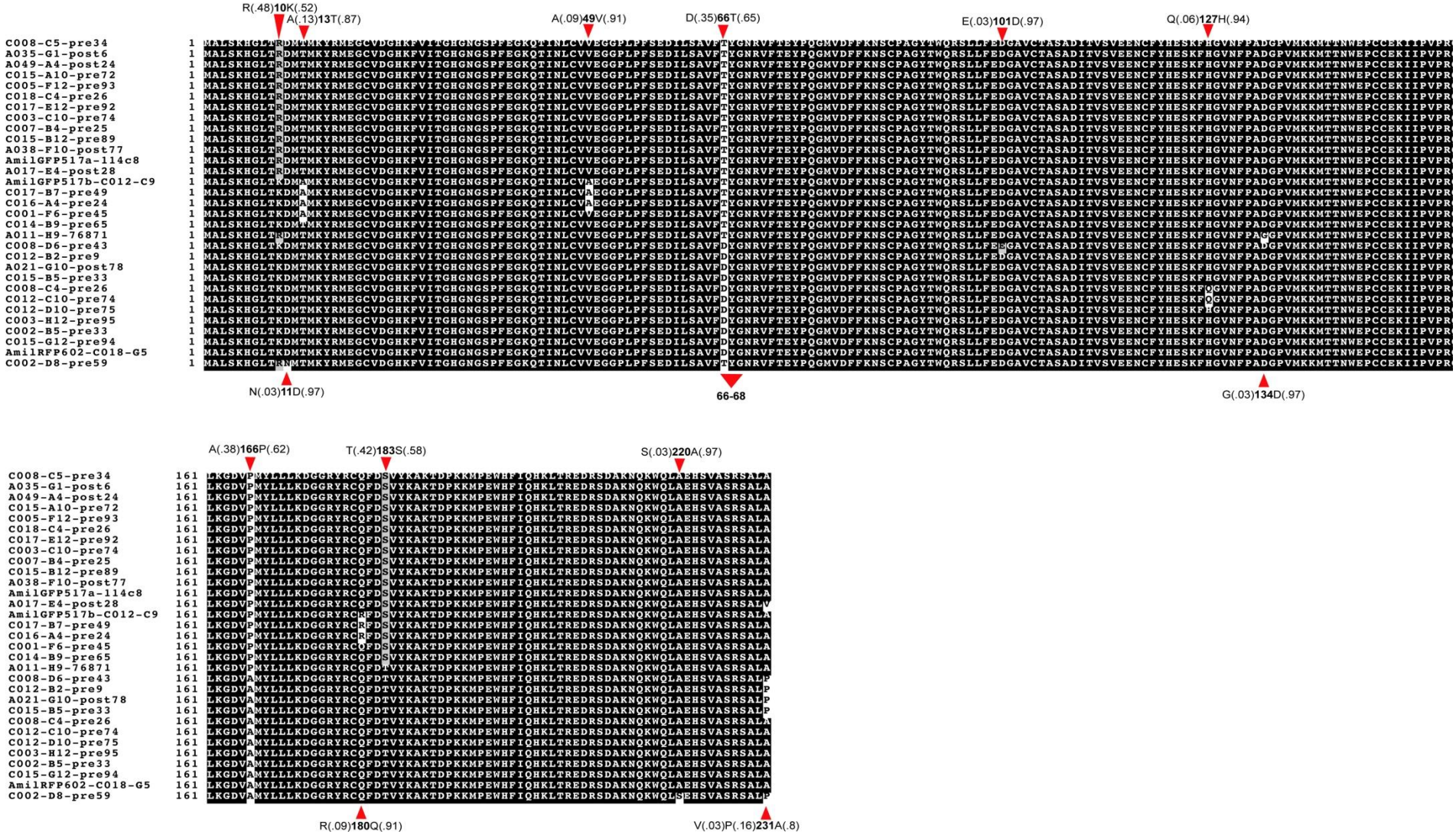

Figure 4.2 Amino acid alignments of all complete clones that belong to the Fluorescent-like contig.
The names on the left represent the assigned identification for the EST clone. Also included are two bacterially expressed CPs with distinctive absorbance maxima. The arrows point to the substitutions. Information concerning the proportion of each amino acid is in parentheses, and the number in the polypeptide chain is in bold type. The alignment was performed relative to the clone amilGFP597, which was the most abundant in the contig.

4.2.2 3D representation of the observed diversity in each contig

Several crystals of FPs have been produced, allowing for the construction of internet sites such as http://www.cbs.dtu.dk/services/CPHmodels/ that map atomic 3D arrangements of FPs from submitted amino acid sequences. For GFP-like proteins, all crystallized proteins have been shown thus far to possess identical β-can structures. Hence, it is reasonable to assume that the predictions made by the program are correct. The program, however, is incapable of reproducing the actual shape of the chromophore inside the β-barrel; the linear sequence of the three peptides that constitute the fluorophore are modified post-translationally, which regulates the tertiary structure of the protein to form the planar chromophore structure. Nevertheless, the models are important because they allow for comparisons of protein shapes and the distribution of specific residues of GFP-like proteins. In this study, the online engine was used to retrieve the protein model in .pdb format using Swiss-Pdb Viewer software (GlaxoSmithKline), and the appropriate modifications to the model were made.

Figure 4.3 shows the sequences of two phenotypically different clones, AmilCP597 and AmilCP601. Also shown is the total variation inside the CP contig from other non-expressed GFP-like ESTs. Both proteins display differences in the consecutive dipeptide N142H and T143S. The simulation shows that these substitutions are located at similar locations with respect to the chromophore; however, only the side chain that faces the internal β-can structure is in position N142H. Another substitution, V26G, is visible at the opposite side of the cylinder. The side chain of this amino acid faces the exterior. In the basal portion of the α-helix, a S53A substitution is observed in the two proteins. A substitution not recorded in these phenotypes but that was part of the contig variation is F10Y, which is a moderately abundant substitution. V37E is located close to the N-C termini, near the loop between two β-sheets and the side chain that faces outside the protein core.

Three variable sites (V92I, R101C, and S104T) were observed in the EST clones opposite to this substitution and just below the chromophore. Only R101C faces inward; the other two substitutions expose the side chain to the exterior. In this latter protein region, but closer to the N-C termini, the R151G and L154I substitutions expose their side chains to the exterior. Finally, the substitution L206S, located close to the base of the last β-sheet, faces outside the core of the protein.

Figure 4.4 includes the sequences of three FP clones categorized into two main phenotypes: the AmilGFP515a, b and the AmilRFP610. The figure also shows the diversity observed in the entire FP contig for the RFP model. The first position (close to the N-terminus) with variation was R10K; the side chain faces outside the Beta-can and is distributed between green and red variants. A13T, located in proximity to the last substitution, is observed in one of the green phenotypes (Figure 4.4B). The side chain faces the exterior of the protein. A49V was also a substitution observed in the green phenotype. The substitution faces outside the protein core; it has no apparent influence in the fluorescence phenotype because V49 is distributed in both the red and green variants.

Perhaps one of the most important substitutions for chromophore formation or variation is D66T, located just in the centre of the molecule as a part of the tripeptide motif. T66 was observed in both green FPs, and D66 was found only in the red FP. A166P was observed only in the RFP (Figure 4.4C); both GFPs have P166, although the side chain faces the exterior of the protein. R180Q and S183T are two substitutions close to the 166 position. R180 was observed only in Amil515b; the other green phenotype and the red phenotype carry Q180. The side chain faces outside the protein core.

Although S183T was present in both green phenotypes, the red phenotypes showed T183 at this position. The side chain faces the interior of the protein. Judging by its relative closeness to the chromophore, it is possible that these positions have some effect on the fluorescence phenotype. The rest of the substitutions were minor (in abundance) and were not expressed in the bacteria-based system. In the figure, however, the position on the RFP model represents the overall variation in the molecule.

The N11D substitution occurred in only one clone in the proximity of the first major variable site, R10K, which was between both coloured phenotypes. The side chain faces the exterior of the protein. E101D was also represented by one clone; it was located at the base of the cylinder, opposite to the N-C termini, with the side chain facing the interior of the protein.

Q127H was observed in two clones, presumably belonging to the red-like FPs. The substitution G134D, observed in one clone, was located in the middle of one of the

lids of the β-sheet barrel opposite the N-C termini. The amino acid side chain faces the exterior of the protein cylinder.

Another isolated substitution is S220A, which was located near the C-terminus. It is part of the last β-sheet and faces outside of the protein. The last amino acid of the protein also showed variability; P/V231A substitutions occurred, with P231 representing 16% and V231 representing 3% of the population. These amino acids do not appear to form part of the cylinder body.

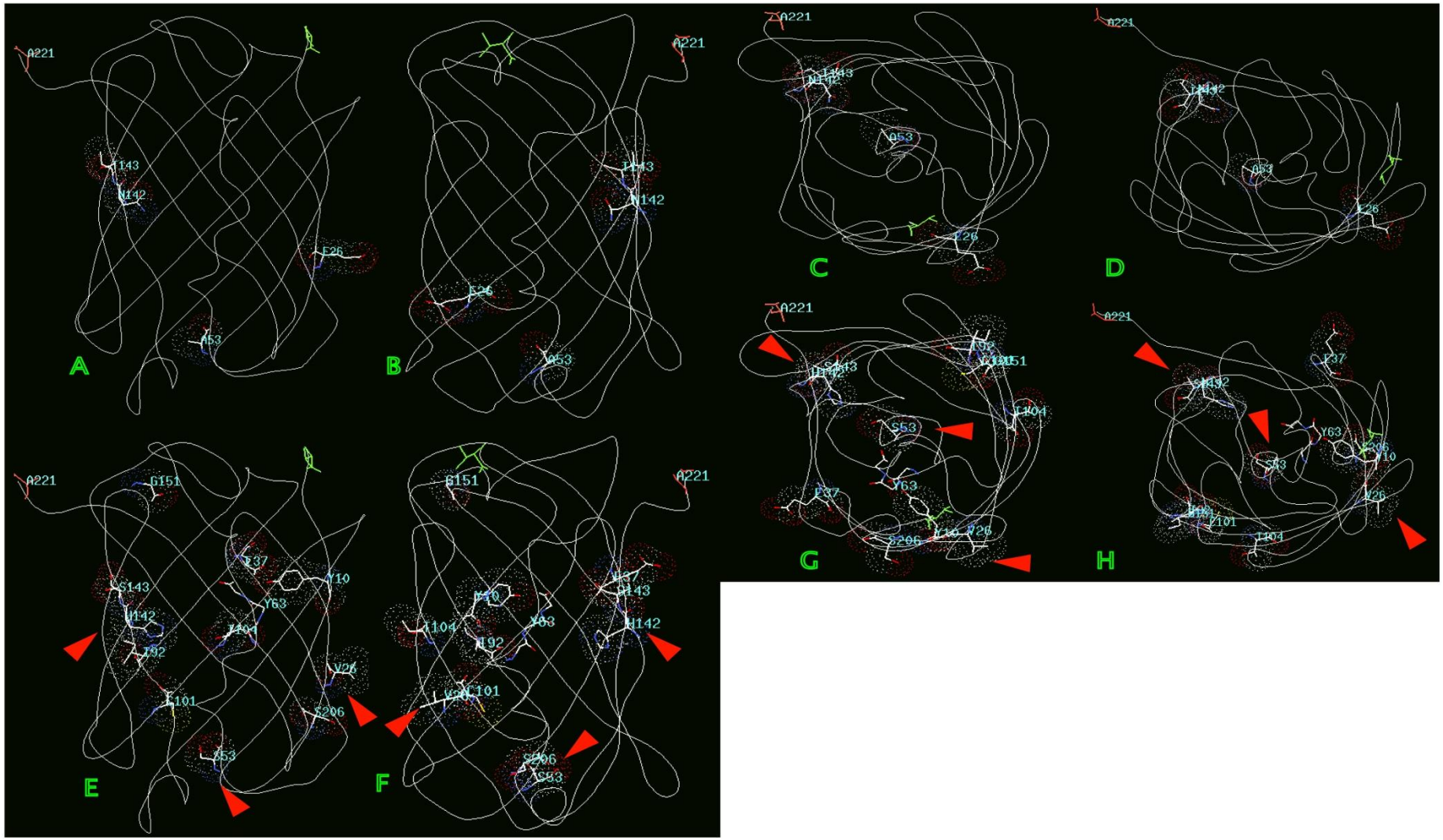

Figure 4.3 3D model of the CP recombinant proteins and the substitutions observed in the whole contig.
A–D and E–H represent amilCP601 and amilCP597, respectively, viewed laterally and from the top. The amino acids indicated by arrows are the substitutions that were compared with the other clone. Other amino acids that are not indicated by arrows represent the overall variation observed in the contig**.**

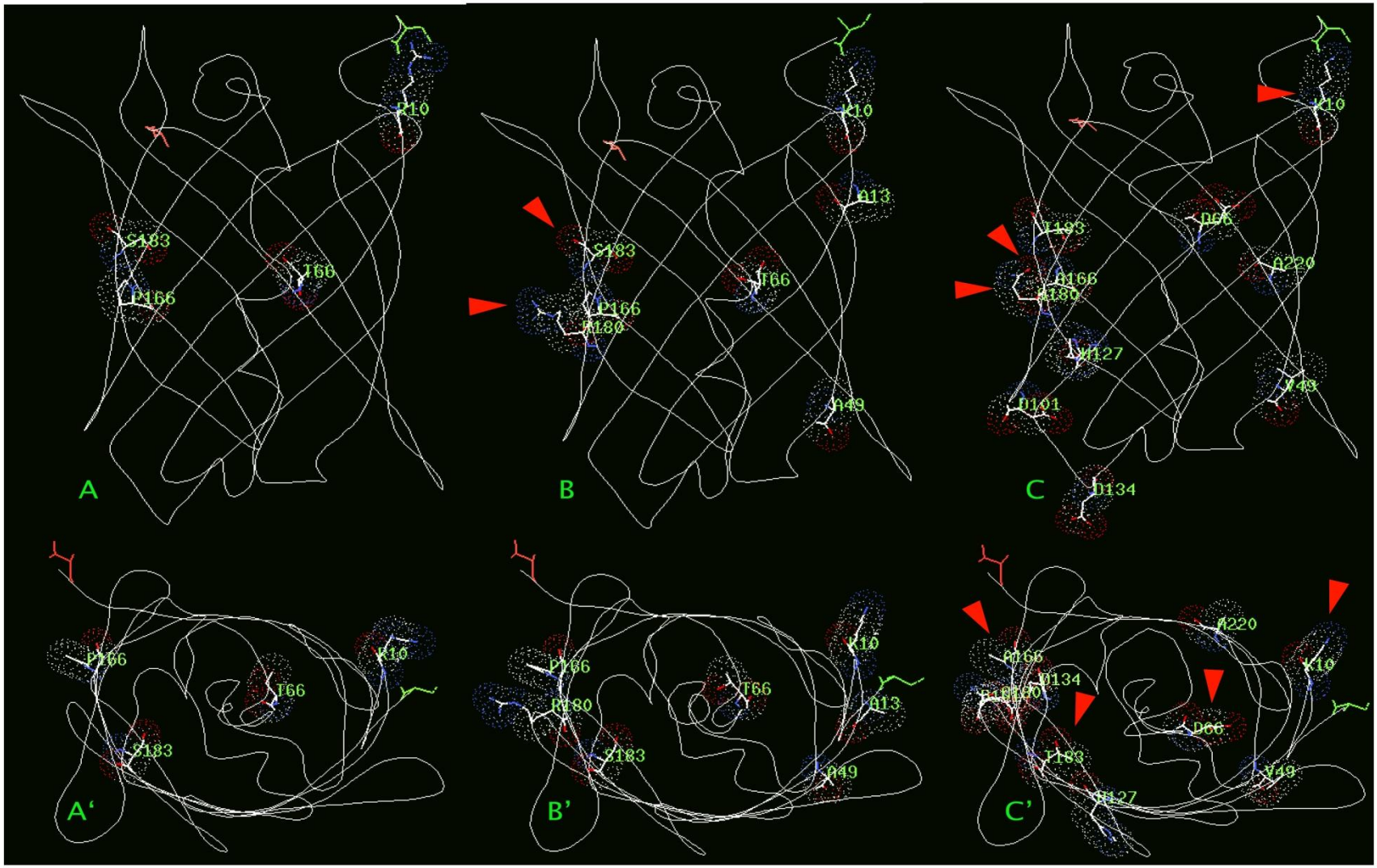

Figure 4.4 3D model representation of fluorescent recombinant proteins, including the observed diversity of amino acid substitutions inside the contig.

A-A', B-B', and C-C' represent the side and top views of the protein from the AmilGFP517a (clone 52c2), AmilGFP517b (clone c012c9), and AmilRFP602 (clone c018g5) clones, respectively. The amino acids indicated by the arrows represent the substitutions that were compared with the other clones. Other amino acids that are not indicated by arrows represent the variation observed in the whole contig.

4.2.3 Molecular masses of the proteins

To confirm the theoretical molecular mass of the proteins predicted by the amino acids sequence, identical amounts of the recombinant proteins were separated by 13% SDS-PAGE, and the proteins were visualized after Coomassie Blue staining. For AmilGFP517b (EST clone c012-c9), a major band at 29 kDa was flanked by two minor bands at 26 and 35 kDa (Figure 4.5A, lane 1). The AmilGFP517a (clone 114c8) banding signature (Figure 4.5A, lane 3) was almost identical to its sibling green fluorescent clone; however, the 26 kDa band was not observed for this recombinant protein. The AmilRFP602 protein migrated to the 29 kDa position (Figure 4.5A, lane 2), and a faint band was observed at ~10–12 kDa. The AmilCP597 and AmilCP601 CPs (Figure 4.5A, lanes 4 and 5, respectively) migrated primarily to 27 kDa; three faint bands at 47, 90, and 120 kDa were also observed.

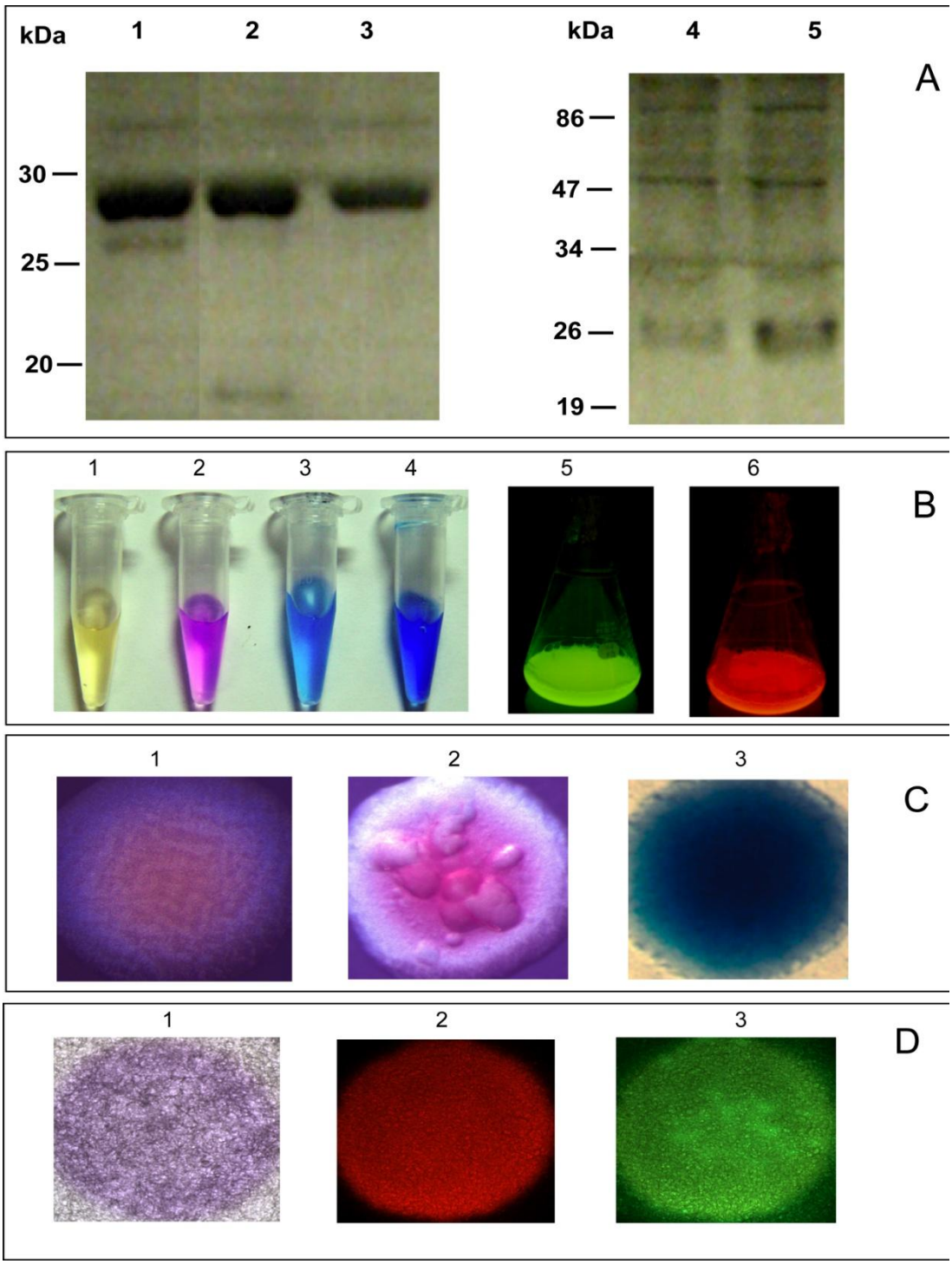

Figure 4.5 Phenotypic comparison of the expressed recombinant GFP-like proteins.

A) Coomassie blue–stained GFP-like recombinant proteins after separation on 13% SDS polyacrylamide gels: 1) Amil517b (EST clone c012c9); 2) AmilRFP602 (EST clone c018-g5); 3) AmilGFP517a (clone 114c8); 4) AmilCP597 (clone 2:2); 5) AmilCP601 (EST Clone c015-a8). The relative molecular mass is displayed at the left of each figure in daltons. B) Visual comparison under normal light conditions of the eluted pure proteins in solution. Also shown is the bacterial culture expressing each protein excited with UV light: 1 and 5) AmilGFP517 under normal and UV light; 2 and 6) AmilRFP602 under normal and UV light; 3 and 4) AmilCP597 and AmilCP601 observed with normal light; C) *E. coli* expressing the recombinant proteins. 1–3) AmilGFP, RFP, and CP, respectively; D) Fluorescence of the CP after being blotted on 3M paper. 1–3) CP observed under normal light; the GREEN filter (designed to observe red fluorescence) and GFP2 filter (designed to observe the green fluorescence) were used.

4.2.4 Spectral characterization of GFP-like recombinant proteins from *Acropora millepora*

This section deals with the spectral signature of the recombinant proteins in solution. Scans were taken across the visible region to obtain absorbance, excitation, and emission signals. In all graphic representations, the X axis represents the wavelength in nanometres, and the principal Y axis represents the value in absorbance and intensity (excitation and emission) relative to the main peak. The secondary Y axis is the fourth derivate analyses of the data (absorbance, excitation, and emission in arbitrary units) with respect to the wavelength. This line overlaps the spectral data and can be used to identify the internal peaks under the curve.

4.2.4.1 Spectral signature of *A. millepora* green FP (AmilGFP517a, b)

Two different proteins that had green fluorescence were cloned. Despite the amino acid differences between them, they had identical emission characteristics. However, one of the clones (114c8, amilGFP517a) showed distinct hydrophobic characteristics during purification. After clarification of the cell homogenate, it was difficult to suspend AmilGFP517a in the aqueous phase for separation by affinity chromatography.

The green FPs (AmilGFP517a, b; clone 114c8 and EST C012-c9 clone) at a concentration of 0.35 µg/µl appeared orange-pink to the naked eye (Figure 4.5B, tube 1). The absorbance spectra showed two double component curves with peaks at 506 and 505 nm and shoulders at 477 and 478 nm for AmilGFP517a and AmilGFP517b, respectively (Figures 4.6A and 4.6D, open orange circles).

The excitation spectra of these two proteins were characterized by peaks at 518 and 517 nm and shoulders around 490 nm for the AmilGFP517a and AmilGFP517b proteins, respectively (Figures 4.6B and 4.6E, blue filled circles). Their emission spectra had a major peak at 517 nm and a shoulder at 550 nm (Figures 4.6C and 4.6F, green circles).

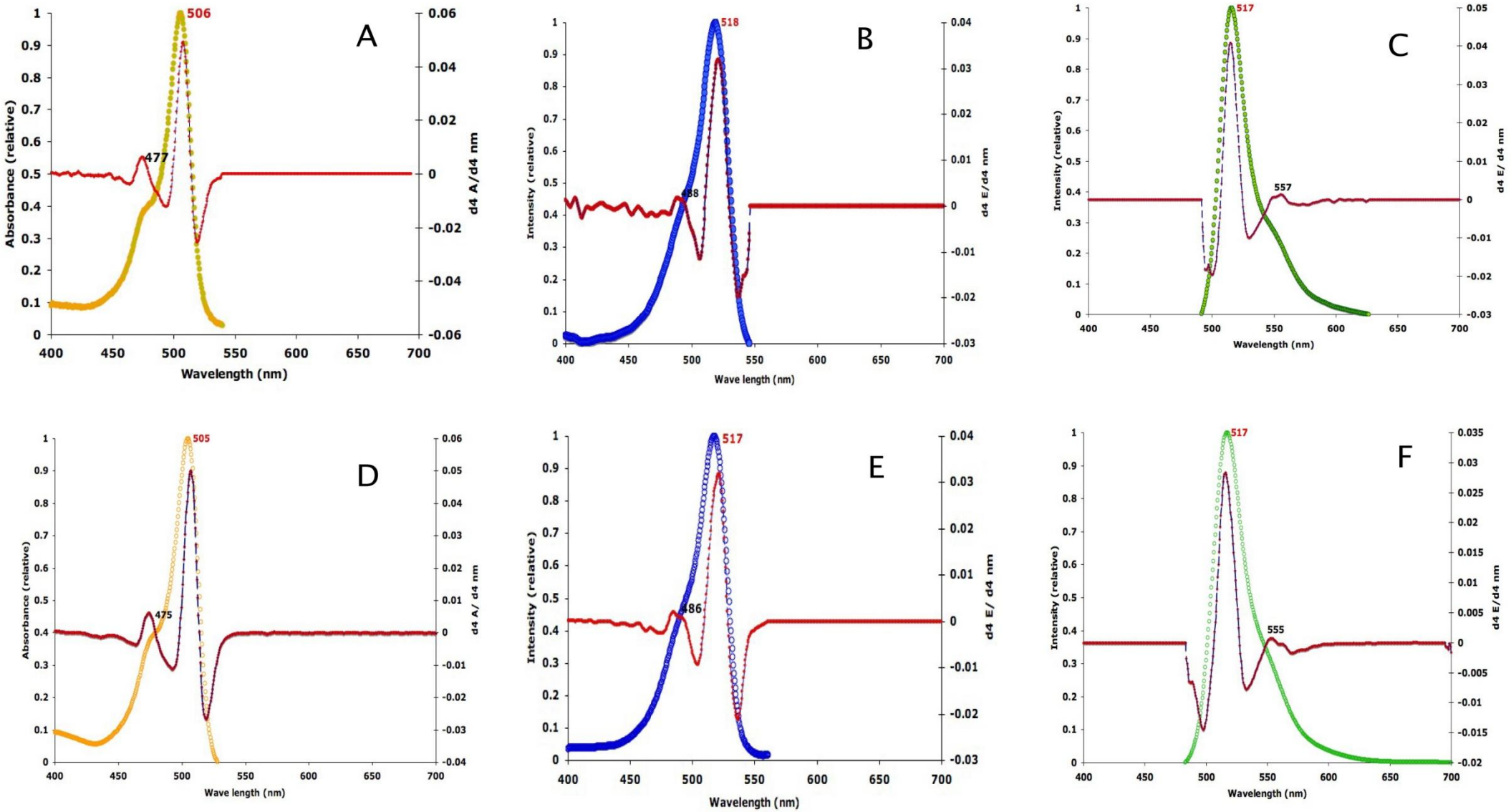

Figure 4.6 Spectral characterization of amilGFP517a and amilGFP517b proteins in solution (PBS, pH 7.2) at room temperature.
All spectra were normalized relative to the main peak. The values are expressed as units on the principal y axis. The secondary y axis represents the fourth derivate value of the dependent variable with respect to the energy of the photon and is represented as the red line overlapping the spectra. A) Absorbance spectrum of amilGFP517a, denoted by "●". B) Excitation spectrum of amilGFP517a, denoted by "●". C) Emission spectrum of amilGFP517a denoted by "●". D) Absorbance spectrum of amilGFP517b, denoted by "○". E) Excitation spectrum of amilGFP517b, denoted by "○". F) Emission spectrum of amilGFP517b, denoted by "○". The numbers in the graphs represent the values of the peaks to the left in nanometres.

4.2.4.2 Spectral signature of *A. millepora* red FP (AmilRFP602)

The AmilRFP602 (EST C004h7, pQE clone vblg274c) recombinant protein (0.35 µg protein/µl) appeared red-violet to the naked eye (Figure 4.5B, tube 2). The bacterial colony, which normally is creamy white in colour, was carmine-red in the centre of the colony when expressing RFP (Figure 4.5C, panel 2). The absorbance spectrum of this protein (Figure 4.7A) spanned from 450 to 580 nm and was characterized by a main peak and a shoulder at 568 and 529 nm, respectively.

The excitation spectrum of the RFP (Figure 4.7B) spanned from 500 to 620 nm. The spectral curve was defined by two main components, a peak at 584 and a shoulder at 540 nm. The emission signal of this RFP spanned from 580 to 700 nm. The curve was defined by a single peak at 602 nm; however, results from derivate analyses showed heterogeneity under the curve and predicted up to three components, which explained peaks with maxima at 582, 591, and 616 nm, respectively.

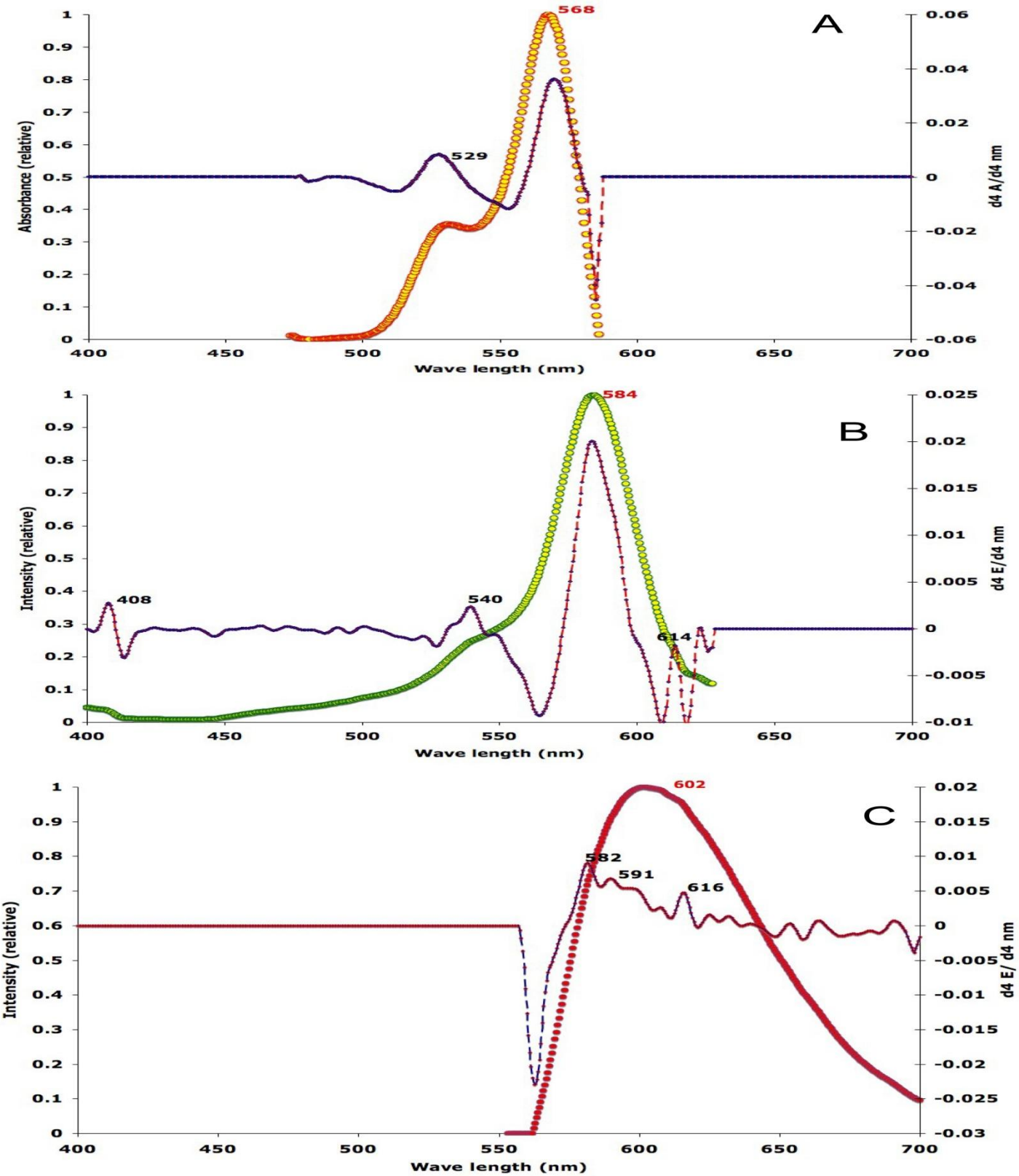

Figure 4.7 Spectral characterization of recombinant AmilRFP602 in solution (PBS, pH 7.2) at room temperature.

All spectra were normalized to the main peak. The values are expressed as units on the principal y axis. The secondary y axis represents the fourth derivate value of the dependent variable with respect to the energy of the photon and is represented by the red line that overlaps the spectra. A) Absorbance spectrum denoted by " " B) Excitation spectrum denoted by " ". C) Emission spectrum denoted by " ". The numbers in the graphs represent the values of the peaks to the left in nanometres.

4.2.4.3 Spectral signature of *A. millepora* CP (AmillCP597)

The AmillCP597 (clone 287-pQE, clone 2:2-pGMT) recombinant protein (0.35 µg/µl) appeared blue to the naked eye (Figure 4.5B, tube 3). When expressing this protein, the bacterial colony (Figure 4.5C, panel 3) had a deep blue centre, characteristic of the blue/white selection cloning-assay, which is based on the reaction of the β-lactosidase enzyme with the substrate X-gal.

The absorbance spectrum (Figure 4.8A) spanned from 460 to 650 nm, defined by a peak and a shoulder at 597 and 563 nm, respectively. Results from derivate analyses showed a small contribution at 460 nm, and the shoulder had two components at 555 and 563 nm. The main peak was heterogeneous in composition and was defined by two peaks at 602 and 610 nm.

4.2.4.4 Spectral signature of *A. millepora* CP (AmilCP601)

The AmillCP601 (clone 281, EST clone c015-a8) protein (0.35 µg/µl) appeared almost identical to its CP597 counterpart (Figure 4.5B, tube 4), and the bacterial colony also appeared deep blue (Figure 4.5C, panel 3).

The absorbance spectrum (Figure 4.8B) spanned from 460 to 650 nm, defined by a peak and a shoulder at 601 and 566 nm, respectively. Results from derivate analyses show a small contribution at 461 nm, the shoulder defined by two components at 555 and 566 nm. The main peak appeared to be heterogeneous in composition and had two spectral peaks at 605 and 612nm.

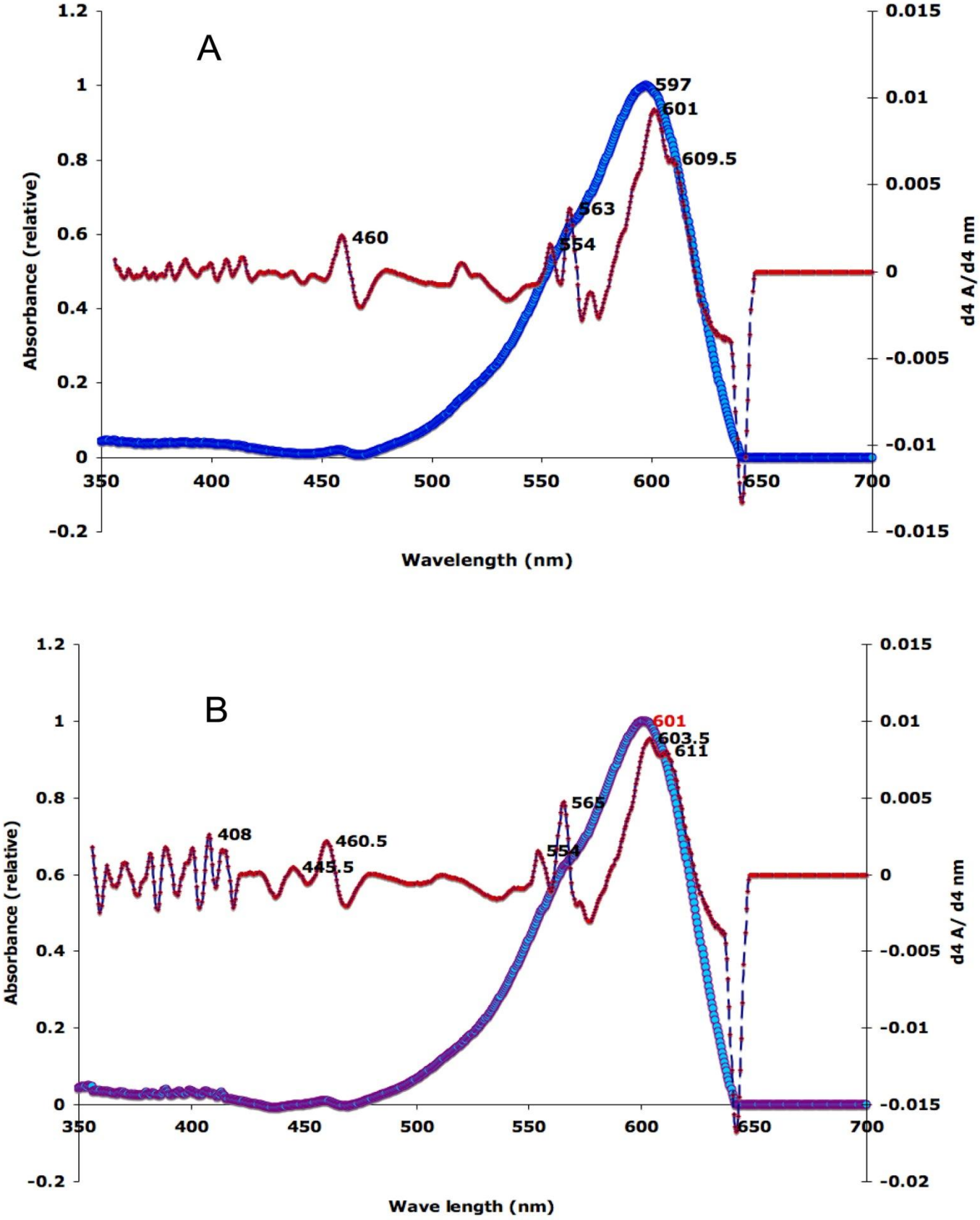

Figure 4.8. Spectral characterization of two recombinant CPs measured in solution (PBS, pH 7.2) at room temperature.

The spectra were normalized to the main peak. The values are expressed as units on the principal y axis. The secondary y axis represents the fourth derivate value of the dependent variable with respect to the energy of the photon and is represented by the red line that overlaps the spectra. A) Absorbance spectrum from the protein AmilCP597 (clone 2:2) denoted by "●". B) Absorbance spectrum from the protein AmilCP601 (EST clone c015-a8) denoted by "●". The numbers in the graphs represent the value of the peaks to the left in nanometres.

4.2.4.5 Spectral characterization of the recombinant proteins under extreme pH conditions

To analyse chromophoric solvatation under acidic or basic conditions, a series of absorbance scans over the entire visible regions was performed using the same amount of recombinant protein at pH values of 3–12. Similar to “untreated” sample spectral characterizations, the x axis represents the wavelength in nanometres, and the y axis represents the absorbance in relative units.

The AmillGFP517b protein (pQE clone 167m1, EST c012-c9) normal absorbance spectrum was characterized by a main peak at 506 nm and a shoulder at ~470 nm (Figure 4.9B, black rhombus). Under acidic conditions, the protein showed a similar spectral shape (Figure 4.9B, blue triangles); the shoulder was maintained and there was a slight shift of 3 nm in the main peak. However, a new peak appeared in the blue region at ~390 nm under acidic conditions. In alkaline conditions, the protein had an absorbance maximum at 447 nm, and this peak was bell shaped (Figure 4.9B red circles).

The absorbance of the AmillRFP602 protein (pQE vblg274, EST c004-h7) under normal conditions was defined by a main peak, shoulder, and small peak at 568, 529, and 460 nm, respectively (Figure 4.9C, black rhombus). Under acidic conditions, the protein (Figure 4.9C, blue triangles) had different spectral characteristics, and the absorbance maxima shifted toward the UV-blue region. The main peak occurred at 380 nm, and a UV-shifted shoulder and two minor components appeared at 354, 455, and 556 nm, respectively. Alkaline conditions (Figure 4.9C, red circles) also affected the spectral characteristics of the protein. The spectral peaks moved toward the blue region but to a lesser extent compared to under the acidic conditions. The spectrum exhibited a single bell-shaped curve with an absorbance maximum at 454 nm and a small peak at 556 nm.

The AmillCP601 protein (pQE clone 281, EST c015-a8) normal absorbance spectrum was characterized by a maximum peak at 601 nm and a shoulder and smaller component at 566 and 462 nm, respectively (Figure 4.9A, black rhombus). The spectral shape did not change under acidic conditions (Figure 4.9A, blue triangles); however, there were 4-nm blue shifts in the main peak, shoulder, and small peak. In contrast, the spectral signature of the protein under alkaline conditions showed a more dramatic blue shift. Basic conditions resulted in a shift in the absorbance maximum to 452 nm; the peak appeared to be a single component and was bell shaped (Figure 4.9A, red circles). A small peak was observed at 557 nm.

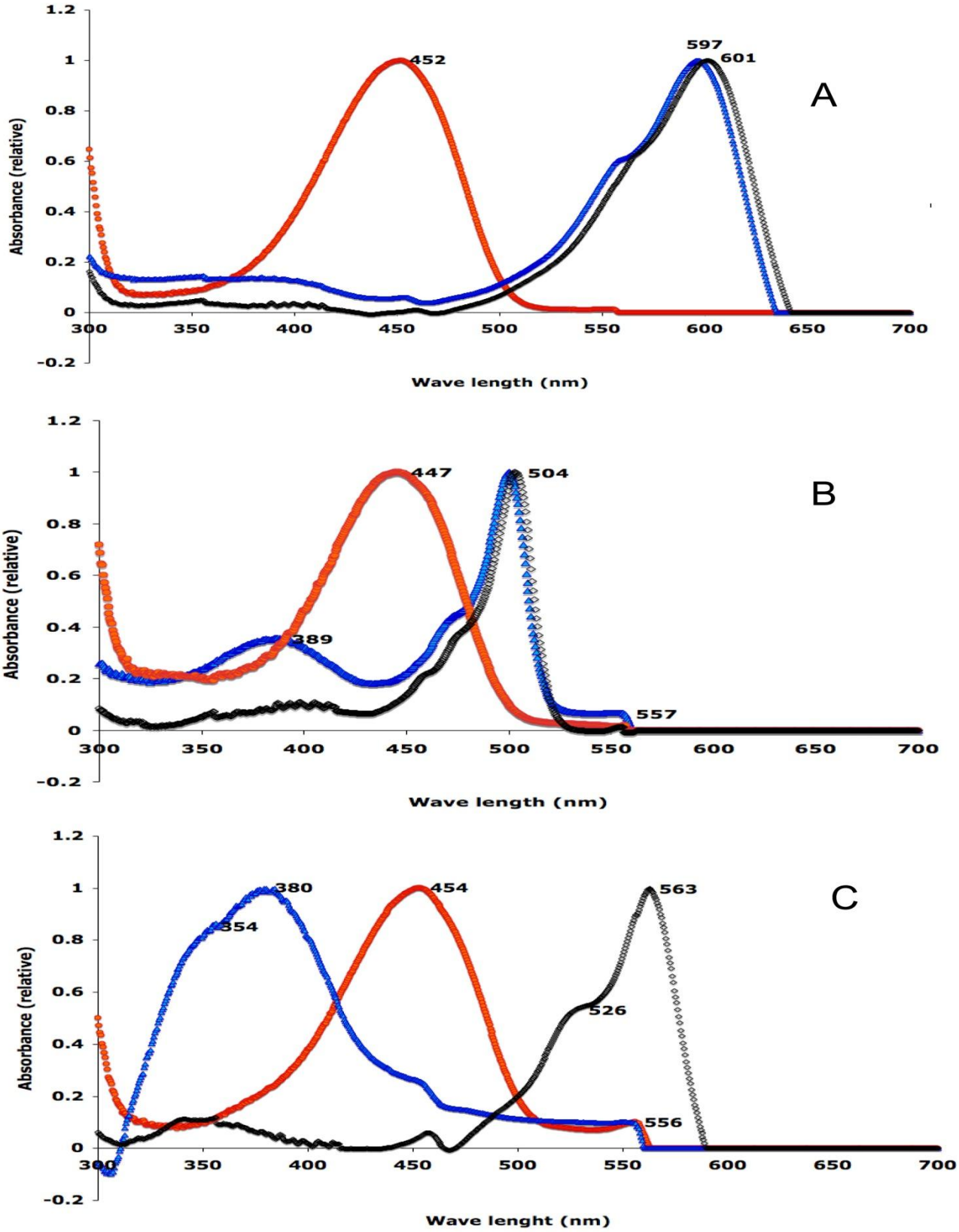

Figure 4.9 pH-dependent behaviour of the recombinant proteins in solution at room temperature.

The spectra were normalized to the main peak. The values are expressed as units on the principal y axis. In all spectra, " ◇ ", " ▲ ", and " ● " represent absorbance data collected under normal (pH 7.2), acidic (pH 4.0), and alkaline (pH 12) conditions, respectively. A) AmilCP601; B) AmilGFP515b; C) AmilRFP601. The numbers in the graphs represent the value of the peaks to the left in nanometres.

4.3 Discussion

4.3.1 Amino acid substitutions and the functions and phenotypes of the protein

Scattered substitutions in the primary structure of the proteins were observed in amino acid alignments of similar clones inside each contig (Figures 4.1 and 4.2). The expressed clones also revealed substitutions that determined their fluorescent phenotype, e.g., the position H142N of the AmilCP601 and AmillCP597 proteins (Figure 4.3 and 4.10). The imidazole ring of histidine allows it to act as either a proton donor or acceptor at physiological pH. Hence, it is frequently found in the reactive centre of enzymes. This position is also variable in several other GFP-like proteins and, to some extent, is related to the phenotype of the protein. For example, in the hydrozoan GFP and several fluorescent anthozoan-related molecules, this position is occupied by histidine and serine, respectively (Prasher et al. 1992, Labas et al. 2002, Gurskaya et al. 2003, Tu et al. 2003, Karasawa et al. 2004, Sun et al. 2004, Luo et al. 2006), whereas in non-fluorescent CPs from hydrozoans and anthozoans, this position is mainly occupied by threonine, alanine, or cysteine (Gurskaya et al. 2001, Shagin et al. 2004). The importance of this position for generating fluorescent mutants from CPs cloned from sea anemones and corals was assessed by changing it to serine. The mutation generates a far-red (615–645 nm) FP that differentiates it from non-fluorescent ones (Lukyanov et al. 2000, Gurskaya et al. 2001).

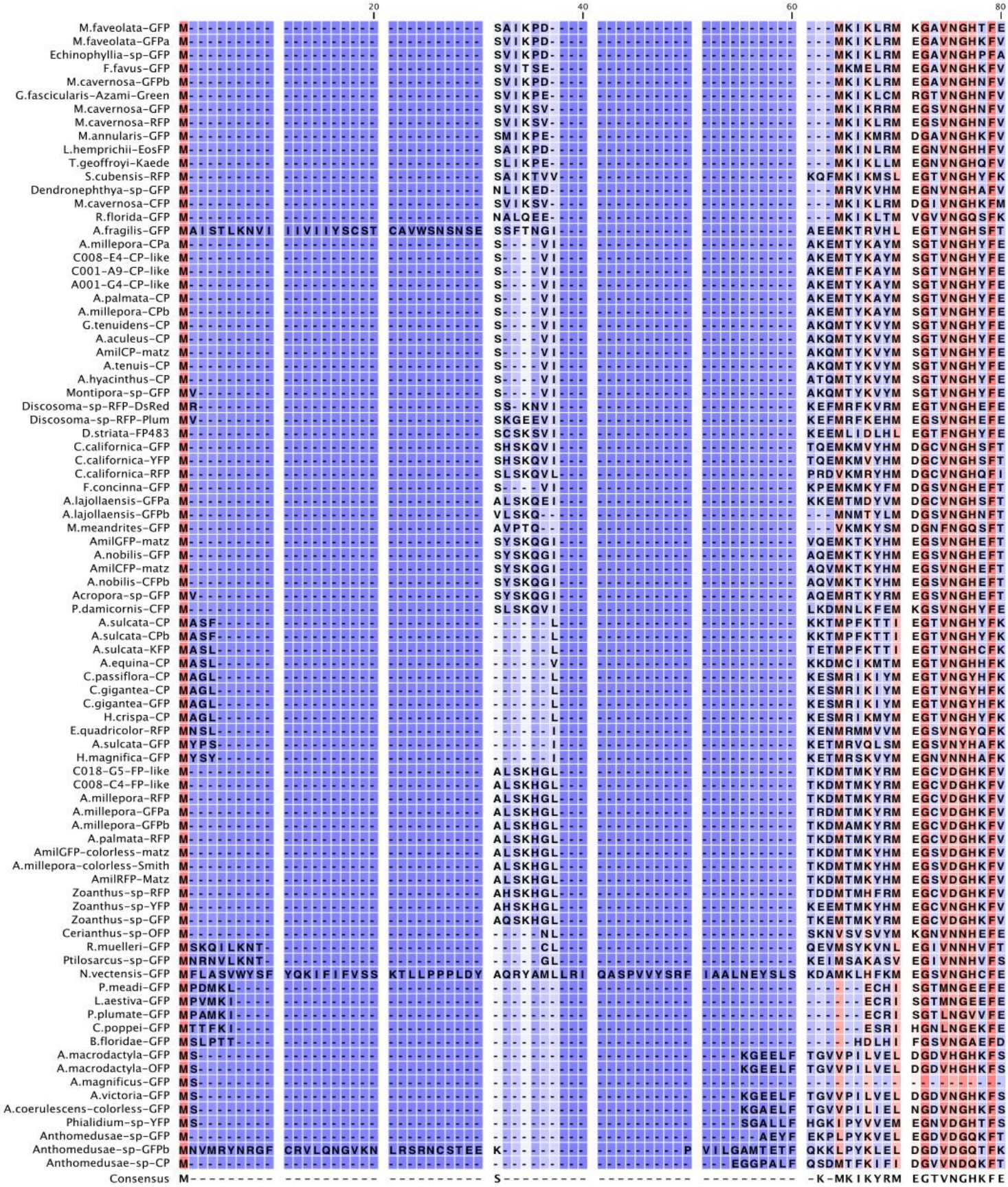

Figure 4.10 Amino acid alignment from representative GFP homologues across Metazoa.

The background shading pattern goes from red to blue for the more similar to dissimilar residues, respectively. Along the alignment, several *Acropora* variable positions are depicted. Information related to the NCBI accession number and reference of each sequence is summarised in Table 4**.**

```
                                                  100                                 120                                 140                                 160
                                                   |                                   |                                   |                                   |
M.faveolata-GFP      IEGEGKGKPF EGKQNMDLKV V--DGGPLPF AYDILTTVFH YGNRVFAKYP GDIA--DYFK QS---FPEGY SWERSMTYED
M.faveolata-GFPa     IEGDGKGKPF EGKQTMDLTV I--EGAPLPF AYDILTTVFD YGNRVFAKYP KDIP--DYFK QT---FPEGY SWERSMTYED
Echinophyllia-sp-GFP IEGVGLGKPF EGKQSMDLKV K--EGGPLPF AYDILTTVFC YGNRVFAKYP ENIV--DYFK QS---FPEGY SWERSMNYED
F.favus-GFP          ITGKGSGQPF EGIQNMDLTV I--EGGPLPF AFDILTTVFD YGNRVFVKYP EEIV--DYFK QS---FPEGY SWERSMSYED
M.cavernosa-GFPb     IEGEGKGKPF EGTQTINLTV K--EGGPLPF AYDILTAAFQ YGNRAFTKYP RDIA--DYFK QS---FPEGY SWERSMTYED
G.fascicularis-Azami-Green IEGEGKGNPY EGTQILDLNV T--EGAPLPF AYDILTTVFQ YGNRAFTKYP ADIQ--DYFK QT---FPEGY HWERSMTYED
M.cavernosa-GFP      IVGEGEGKPY EGTQSMDLTV K--EGAPLPF AYDIMTTVFH YGNRVFAKYP KHIP--DYFK QM---FPEGY SWERSMNFED
M.cavernosa-RFP      IVGEGEGKPY EGTQSMDLTV K--EGAPLPF AYDIMTTVFH YGNRVFAKYP KHIP--DYFK QM---FPEGY SWERSMNFED
M.annularis-GFP      ITGEGSGEPF EGKQTMNLTV I--DGGPLPF AFDILTTAFD YGXRVFAKYP EDIP--DYFK QS---FPEGF SWERSMTYED
L.hemprichii-EosFP   IDGDGTGKPF EGKQSMDLEV K--EGGPLPF AFDILTTAFH YGNRVFAEYP DHIQ--DYFK QS---FPKGY SWERSLTFED
T.geoffroyi-Kaede    IEGDGKGHPF EGKQSMDLVV K--EGAPLPF AYDILTTAFH YGNRVFAKYP DHIP--DYFK QS---FPKGF SWERSLMFED
S.cubensis-RFP       IVGEGDGTPF EGKQTLHLKV K--EGAPLPF AYDILTTALH YGNRVFVEYP ENIP--DYFK QS---FPKGY SWERSLTFED
Dendronephthya-sp-GFP IEGEGKGKPY EGTQTLNLTV K--EGAPLPF SYDILTTALH YGNRVFTKYP EDIP--DYFK QS---FPEGY SWERTMTYED
M.cavernosa-CFP      ITGEGEGKPF EGTHTIILKV K--EGGPLPF AYDILTTAFQ YGNRVFTKYP KDIP--DYFK QS---FPEGY SWERSMTFED
R.florida-GFP        IDGKGKGKPY EGSQELTLKV V--EGGPLLF SYDILTTIFQ YGNRAFVNYP KDIP--DIFK QTCSGLDGGY SWQRTMTYED
A.fragilis-GFP       IKGEGRGYPY KGEQFMSLEV V--NGAPLPF SFDILTPAFM YGNRVFTKYP PNIP--DYFK QT---FPEGY HWERNIPFED
A.millepora-CPa      VEGDGKGKPY EGEQTEKLTV T--KGGPLPF AWDILSPQCQ YGSIPFTKYP EDIP--DYVK QS---FPEGF TWERIMNFED
C008-E4-CP-like      VEGDGKGKPY EGVQTEKLTV T--KGGPLPF AWDILSPQCQ YGSIPFTKYP EDIP--DYVK QS---FPEGF TWERIMNFED
C001-A9-CP-like      VEGDGKGKPY EGEQTEKLTV T--KGGPLPF AWDILSPQCQ YGSIPFTKYP EDIP--DYVK QS---FPEGF TWERIMNFED
A001-G4-CP-like      VEGDGKGKPY EGEQTEKLTV T--KGGPLPF AWDILSPQCQ YGSIPFTKYP EDIP--DYVK QS---FPEGF TWERIMNFED
A.palmata-CP         VEGDGKGKPY EGEQTEKLTV T--KGGPLPF AWDILSPQCQ YGSIPFTKYP EDIP--DYVK QS---FPEGF TWERIMNFED
A.millepora-CPb      VVGDGKGKPY EGEQTEKLTV T--KGGPLPF SWDILSPQCQ YGSIPFTKYP EDIP--DYVK QS---FPEGF TWERIMNFED
G.tenuidens-CP       VQGDGKGKPY EGEQTVKLTV T--KGGPLPF AWDILSPQSQ YGSIPFTKYP EDIP--DYVK QS---FPEGY TWERIMNFED
A.aculeus-CP         VEGDGKGKPY EGEQTVKLTV T--KGGPLPF AWDILSPQSQ YGSIPFTKYP DDIP--DYVK QS---FPEGY TWERIMNFED
AmilCP-matz          VEGDGKGKPY EGEQTVKLTV T--KGGPLPF AWDILSPQCQ YGSIPFTKYP EDIP--DYVK QS---FPEGY TWERIMNFED
A.tenuis-CP          VEGDGKGKPY EGEQTVKLTV T--KGGPLPF AWDILSPQFQ YGSIPFTKYP EDIP--DYVK QS---FPEGY TWERIMNFED
A.hyacinthus-CP      VEGDGKGKPY EGEQTVRLTV T--KGGPLPF AWDILSPQSQ YGSIPFTKYP EDIP--DYVK QS---FPEGY TWERIMNFED
Montipora-sp-GFP     VEGDGKGKPY EGEQTVKLTV T--KGGPLPF AWDILSPLFQ YGSIPFTKYP EDIP--DYVK QS---FPEGY TWERTMNFED
Discosoma-sp-RFP-DsRed IEGEGEGRPY EGHNTVKLKV T--KGGPLPF AWDILSPQFQ YGSKVYVKHP ADIP--DYKK LS---FPEGF KWERVMNFED
Discosoma-sp-RFP-Plum IEGEGEGRPY EGTQTARLKV T--KGGPLPF AWDILSPQIM YGSKAYVKHP ADIP--DYLK LS---FPEGF KWERVMNFED
D.striata-FP483      IKGKGKGQPN EGTNTVTLEV T--KGGPLPF GWHILCPQFQ YGNKAFVHHP DNIH--DYLK LS---FPEGY TWERSMHFED
C.californica-GFP    IEGEGTGKPY EGNQTLKLRV T--KGGPLPF AFDILTATFC YGNRCFCEYP EDMP--DYYK QS---FPEGY SFERTMMFED
C.californica-YFP    IEGEGTGKPY EGNQTLKLRV T--KGGPLPF AFDILTATFC YGNRCFCEYP EDMP--DYYK QS---FPEGY SFERTMMFED
C.californica-RFP    IEGEGTGKPY EGKKILELRV T--KGGPLPF AFDILSSVFT YGNRCFCEYP EDMP--DYFK QS---LPEGH SWERTLMFED
F.concinna-GFP       VEGEGTGKPY EGHQEMTLRV TMAKGGPMPF SFDLVSHTFC YGHRPFTKYP EEIP--DYFK QA---FPEGL SWERSLQFED
A.lajollaensis-GFPa  VKGDGAGKPY EGHQRLSLHV T--GGQPLPF AFDILSAAFA YGNRCFTKYP KEIP--DMFK QA---FPGGM SWERTMTFED
A.lajollaensis-GFPb  VEGEGDGKPY EGHQCLKLRI T--KGEPLPF AFDILTAAFC YGNRCFVNYP AEIA--DYFK QA---FPEGL TWERSMTFED
M.meandrites-GFP     VVGEGTGNPY EGHQSLKLTV ---KGEPLPF AFDILSATFT YGNRVFTKYP EGKT--DYFK EA---FPGGL TWERTMTFED
AmilGFP-matz         IEGVGTGYPY EGKQMSELVI IKPAGKPLPF SFDILSSVFQ YGNRCFTKYP ADMP--DYFK QA---FPDGM SYERSFLFED
A.nobilis-GFP        IEGVGTGYPY EGKQMSELVI IKPAGKPLPF SFDILSSVFQ YGNRCFTKYP ADMP--DYFK QA---FPDGM SYERSFLFED
AmilCFP-matz         IEGVGTGNPY EGTQMSELVI TKPAGKPLPF SFDILSTVFQ YGNRCFTKYP EGMT--DYFK QA---FPDGM SFERSFLYED
A.nobilis-CFPb       IEGVGTGNPY EGTQMSELVI TKPAGKPLPF SFDILSTVFQ YGNRCFTKYP EGMT--DYFK QA---FPDGM SYERSFLYED
Acropora-sp-GFP      IEGVGTGNPY EGKQMSELVI IKSKGKPLPF SFDILSTAFQ YGNRCFTKYP ADMP--DYFK QA---FPDGM SYERSFLFED
P.damicornis-CFP     IEGEGRGKPY EGVQKSTFWV TK--GGPLPF SFDILSSAFK YGNRCFTKYP ADMP--DYFK EA---FPAGM SFERTFTFED
A.sulcata-CP         CTGKGEGNPF EGTQEMKIEV I--EGGPLPF AFHILSTSCM YGSKTFIKYV SGIP--DYFK QS---FPEGF TWERTTTYED
A.sulcata-CPb        CTGKGEGNPF EGTQEMKIEV I--EGGPLPF AFHILSTSCM YGSKTFIKYV SGIP--DYFK QS---FPEGF TWERTTTYED
A.sulcata-KFP        CIGKGEGNPF EGTQEMKIEV I--EGGPLPF AFHILSTSCM YGSKTFIKYV SGIP--DYFK QS---FPEGF TWERTTTYED
A.equina-CP          CVGEGEGKPF EGTQVEKIRI T--EGGPLPF AYDILAPCCM YGSKTFIKHV SGIP--DYFK ES---FPEGF TWERTQIFED
C.passiflora-CP      CEGEGDGNPY EGTQNMRIRV T--EGAPLPF AFDILSPCCA YGSKTFIKHT SGIP--DYFK QS---FPEGF TWERTTIYED
C.gigantea-CP        CEGEGDGNPF EGTQNMRIRV T--EGAPLPF AFDILSPCCA YGSKTFIKHT SGIP--DYFK QS---FPEGF TWERTTIYED
C.gigantea-GFP       CEGEGDGNPF EGTQNMRIRV T--EGAPLPF AFDILSPCCA YGSKTFIKHT SGIP--DYFK RS---FPEGF TWEGTTIYED
H.crispa-CP          CEGEGDGNPF TGTQSMRIHV T--EGAPLPF AFDILAPCCE YGSRTFVHHT AEIP--DFFK QS---FPEGF TWERTTTYED
E.quadricolor-RFP    CTGEGDGNPY MGTQTMRIKV V--EGGPLPF AFDILATSFM YGSKTFIKHT KGIP--DFFK QS---FPEGF TWERVTRYED
A.sulcata-GFP        CTGKGEGKPY EGTQSLNITI T--EGGPLPF AFDILSHAFQ YGIKVFAKYP KEIP--DFFK QS---LPGGF SWERVSTYED
H.magnifica-GFP      CTAEGEGKPY KGSQKLTITV T--EGGPLPF AFDILSHAFQ YGNKVFTKYP DDIP--DFFK QS---LSGGF TWKRVSNYED
C018-G5-FP-like      ITGHGNGSPF EGKQTINLCV V--EGGPLPF SEDILSAVFD YGNRVFTEYP QGMV--DFFK NSC---PAGY TWQRSLLFED
C008-C4-FP-like      ITGHGNGSPF EGKQTINLCV V--EGGPLPF SEDILSAVFD YGNRVFTEYP QGMV--DFFK NSC---PAGY TWQRSLLFED
A.millepora-RFP      ITGHGNGSPF EGKQTINLCV V--EGGPLPF SEDILSAVFD YGNRVFTEYP QGMV--DFFK NSC---PAGY TWQRSLLFED
A.millepora-GFPa     ITGHGNGSPF EGKQTINLCV V--EGGPLPF SEDILSAVFT YGNRVFTEYP QGMV--DFFK NSC---PAGY TWQRSLLFED
A.millepora-GFPb     ITGHGNGSPF EGKQTINLCV A--EGGPLPF SEDILSAVFT YGNRVFTEYP QGMV--DFFK NSC---PAGY TWQRSLLFED
A.palmata-RFP        ITGHGNGSPF EGKQTINLCV V--EGGPLPF SEDILSSVFV YGNRVFTEYP QSMV--DFFK NSC---PAGY TWQRSLLFED
AmilGFP-colorless-matz ITGHGDGNPF EGKQTINLCV A--EGGPLPF SEDILSAAFD YGNRVFTEYP QGMV--DFFK NSC---PAGY TWHRSLLFED
A.millepora-colorless-Smith ITGHGDGNPF EGKQTINLCV A--EGGPLPF SEDILSAAFD YGNRVFTEYP QGMV--DFFK NSC---PAGY TWHRSLLFED
AmilRFP-Matz         ITGHGNGNPF EGKQTMNLCV V--EGGPLPF SEDILSAAFD YGNRVFTEYP QGMV--DFFK NSC---PAGY TWHRSLLFED
Zoanthus-sp-RFP      IEGNGNGNPF KGKQFINLCV I--EGGPLPF SEDILSAAFD YGNRLFTEYP EGIV--DYFK NSC---PAGY TWHRSFRFED
Zoanthus-sp-YFP      ITGEGIGYPF KGKQTINLCV I--EGGPLPF SEDILSAGFK YGDRIFTEYP QDIV--DYFK NSC---PAGY TWGRSFLFED
Zoanthus-sp-GFP      ITGEGIGYPF KGKQAINLCV V--EGGPLPF AEDILSAAFN YGNRVFTEYP QDIV--DYFK NSC---PAGY TWDRSFLFED
Cerianthus-sp-OFP    YDGEGGGDPY TGKYSMKMTL RGQNC--LPF SYDIITTAFQ YGFRVFTKYP EGIV--DYFK DS---LPDAF QWNRRIVFED
R.muelleri-GFP       MEGCGKGNIL FGNQLVQIRV T--KGAPLPF AFDIVSPAFQ YGNRTFTKYP NDIS--DYFI QS---FPAGF MYERTLRYED
Ptilosarcus-sp-GFP   MEGFGKGNVL FGNQLMQIRV T--KGGPLPF AFDIVSIAFQ YGNRTFTKYP DDIA--DYFV QS---FPAGF FYERNLRFED
N.vectensis-GFP      IQGVGEGKAF DGEHWSKLCV V--KGKHLPF SFDILMPSMS YGTKQFAKYP AGMT--DFFK AAVEN--GGL SWERTMTFED
P.meadi-GFP          LIGSGDGNTD QGRMTNNMKS IK---GPLSF SPYLLSHILG YGYYHFATFP AGYE--NIYL HAMKN--GGY SNVRTERYED
L.aestiva-GFP        LVGAGDGNTD EGRMTNKMKS TK---GPLSF SPYLLSHIMG YGFYHYATFP AGYE--NVYL HAAKN--GGY TNTRTERYED
P.plumate-GFP        LVGGGEGIPE QGRMTNKMKS TK---GALTF SPYLLSHVMG YGFYHFGTYP SGYE--NPFL HAANN--GGY TNTRIEKYED
C.poppei-GFP         LVGGGVG--E EGRLEIEMKT KD---KPLAF SPFLLSHCMG YGFYHFASFP KGTK--NIYL HAATN--GGY TNTRKEIYED
B.floridae-GFP       LVGGGKGNPN DGTLETSVKS TR---GALPC SPLLIGPNLG YGFYQYLPFP GG-A--SPFQ TAITD--GGY QVHRVFKFED
A.macrodactyla-GFP   VRGEGEGDAD YGKLEIKFIC TT---GKLPV PWPTLVTTLG YGILCFARYP EHMKMNDFFK SA---MPEGY IQERTIFFQD
A.macrodactyla-OFP   VRGEGEGDAD YGKLEIKFIC TT---GKLPV PWPTLVTTLG YGILCFARYP EHMKMNDFFK SA---MPEGY IQERTIFFQD
A.magnificus-GFP     VKGEGEGDAT YGKLEIKFVC TT---GKLPV PWPTLVTTFS YGVQCFARYP EHMKMHDFFK SA---MPEGY IQERTIFFQD
A.victoria-GFP       VSGEGEGDAT YGKLTLKFIC TT---GKLPV PWPTLVTTFS YGVQCFSRYP DHMKQHDFFK SA---MPEGY VQERTIFFKD
A.coerulescens-colorless-GFP VSGEGEGDAT YGKLTLKFIC TT---GKLPV PWPTLVTTFS YGVQCFSRYP DHMKQHDFFK SA---MPEGY IQERTIFFKD
Phialidium-sp-YFP    IRGKGYGDAS VGKVDAQFIC TT---GDVPV PWSTLVTTLT YGAQCFAKYG PELK--DFYK SC---MPEGY VQERTITFEG
Anthomedusae-sp-GFP  VIGEGQGDAS TGRVEGKYVC TK---GEVPI SWVSLITSLS YGGKCFVRYP NVIK--DFFK ST---FPTGY HQERKITYED
Anthomedusae-sp-GFPb VIGEGVGDAT TGVIEGKYVC TE---GEVPI SWVSLITSLS YGAKCFVRYP NEIN--DFFK ST---FPSGY HQERKITYEN
Anthomedusae-sp-CP   IIADGSSKFP HGDFNVHAVC ---ETGKLPM SWKPICHLIQ YGEPFFAKYP NGI---SHFA QEC--FPEGL TIDRTVRFEN
Consensus            IEGEGKGKPY EGKQTMKLTV T--EGGPLPF AFDILSTAFQ YGNRVFTKYP EDIP--DYFK QS---FPEGY TWERSMTFED
```

CRO

V26E-amilCP S53A-amilCP avicFP pos 96

Figure 4.10 Continuation

```
                                        180                                 200                                 220                                 240
                                         |                                   |                                   |                                   |
M.faveolata-GFP          GGVCIATNDI TL---VGDVF SRSEIRFEGV NFPANGPVMQ KKT-VKWEPS TEKMYV--RD GVLKGDVNMA LLL-EGGGHY
M.faveolata-GFPa         QGICIATNDI TMMKGVDDCF VY-KIRFDGV NFPANGPVMQ RKT-LKWEPS TEKMYV--RD GVLKGDVNMA LLL-EGGGHY
Echinophyllia-sp-GFP     GGICNATNDI TLD---GDCY IY-EIRFDGV NFPANGPVMQ KRT-VKWEPS TEKLYV--RD GVLKGDVNMA LSL-EGGGHY
F.favus-GFP              GGICLATNNI TMKKDGSNCF VY-EIRFDGV NFPANGPVMQ RKT-VKWEPS TEKMYV--RD GVLKGDVNMA LLL-QGGGHY
M.cavernosa-GFPb         QGICIIKSDI RME---GDCF IY-EIRYDGV NFPPSGPVMQ KKT-LKWEPS TEKMYV--RD GVLKGEVNMA LLL-EGGGHY
G.fascicularis-Azami-Green QGICTATSNI SMR---GDCF FY-DIRFDGV NFPPNGPVMQ KKT-LKWEPS TEKMYV--RD GVLKGDVNMA LLL-EGGGHY
M.cavernosa-GFP          GGICIARNEI TME---GDCF -FNKVRFDGV NFPPNGPVMQ KKT-LKWEPS TEKMYV--RD GVLTGDINMA LLL-EGGGHY
M.cavernosa-RFP          GGICTARNEI TME---GDCF -FNKVRFDGV NFPPNGPVMQ KKT-LKWEPS TEKMYV--RD GVLTGDINMA LLL-EGGGHY
M.annularis-GFP          GGICIATNDI KME---GDCF SY-EIRFDGV NFPANSPVMQ KKT-VKWEPC TXEMYV--RD GVLKGGLNMA LLL-EGGGHF
L.hemprichii-EosFP       GGICIARNDI TME---GDTF -YNKVRFHGV NFPANGPVMQ KKT-LKWEPS TEKMYV--RD GVLTGDITMA LLL-EGNAHY
T.geoffroyi-Kaede        GGVCIATNDI TLK---GDTF -FNKVRFDGV NFPPNGPVMQ KKT-LKWEAS TEKMYL--RD GVLTGDITMA LLL-KGDVHY
S.cubensis-RFP           GGICIARSDI KMV---GDTF -HNEVQFYGV NFPANGPVMQ RHT-VKWEPS TEKIYV--RD GVLTGDITMA LLL-KGGTHY
Dendronephthya-sp-GFP    KGICTIRSDI SLE---GDCF -FQNVRFNGM NFPPNGPVMQ KKT-LKWEPS TEKLHV--RD GLLVGNINMA LLL-EGGGHY
M.cavernosa-CFP          QGVCTVTSDI KLE---GDCF FY-EIRFYGV NFPSSGPVMQ KKT-LKWEPS TENMYV--RD GVLLGDVSRT LLL-EGDKHH
R.florida-GFP            GGVCTATSNV SV---VGDTF NY-EIHFMGA NFPPNGPVMQ KRT-VKWEPS TEIXFE--RD GLLRGDVPMS LLL-KGGDHY
A.fragilis-GFP           QAACTVTSHI RLEEE--ERR FVNNVRFHCV NFPPNGPVMQ RRI-LKWEPS TENIYP--RD GFLEGHVDMT LRV-EGGGYY
A.millepora-CPa          GAVCTVTNDS SIQ---GNCF IY-HVKFSGL NFPPNGPVMQ KKT-QGWEPN TERLFA--RD GMLIGNNFMA LKL-EGGGHY
C008-E4-CP-like          GAVRTVTNDS SIQ---GNCF IY-HVKFSGL NFPPNGPVMQ KKT-QGWEPH SERLFA--RD RMLIGNNFMA LKL-EGGGHY
C001-A9-CP-like          GAVCTVTNDS SIQ---GNCF IY-HVKFSGL NFPPNGPVMQ KKT-QGWEPH SERLFA--RD GMLIGNNFMA LKL-EGGGHY
A001-G4-CP-like          GAVCTVSNDS SIQ---GNCF IY-HVKFSGL NFPPNGPVMQ KKT-QGWEPH SERLFA--RD GMLIGNNFMA LKL-EGGGHY
A.palmata-CP             GAVCTVSNDS SIQ---GNCF IY-HVKFSGL NFPPNGPVMQ KKT-QGWEPH TERLFA--RD GMLIGNNFMA LKL-EGGGHY
A.millepora-CPb          GAVCTVTNDS SIQ---GNCF IY-HVKFSGL NFPPNGPVMQ KKT-QGWEPH SERLFA--RD GMLIGNNFMA LKL-EGGGHY
G.tenuidens-CP           GAVCTVSNDS SIQ---GNCF IY-NVKFSGL NFPPNGPVMQ KKT-QGWEPN TERLFA--RD GMLIGNNFMA LKL-EGGGHY
A.aculeus-CP             GAVCTVSNDS SIQ---GNCF IY-NVKFSGL NFPPNGPVMQ KKT-QGWEPN TERLFA--RD GMLIGNNFMA LKL-EGGGHY
AmilCP-matz              GAVCTVSNDS SIQ---GNCF IY-HVKFSGL NFPPNGPVMQ KKT-QGWEPN TERLFA--RD GMLLGNNFMA LKL-EGGGHY
A.tenuis-CP              DAVCTVSNDS SIQ---GNCF IY-HVKFSGV NFPPNGPVMQ KKT-QGWEPN TERLFA--RD GMLIGNNFMA LKL-EGGGHY
A.hyacinthus-CP          GAVCTVSNDS SIQ---GNCF IY-HVKFSGL NFPPNGPVMQ KKT-QGWEPN TERLFA--RD GVLIGNNFMA LKL-EGGGHY
Montipora-sp-GFP         GAVCTVSNDS SIQ---GNCF IY-NVKISGT NFPPNGPVMQ KKT-QGWEPS TERLFA--RD GMLIGNDYMA LKL-EGGGHY
Discosoma-sp-RFP-DsRed   GGVVTVTQDS SLQ---DGCF IY-KVKFIGV NFPSDGPVMQ KKT-MGWEAS TERLYP--RD GVLKGEIHKA LKL-KGGGHY
Discosoma-sp-RFP-Plum    GGVVTVTQDS SLQ---DGEF IY-KVKVRGT NFPSDGPVMQ KKT-MGWEAS SERMYP--ED GALKGEMKMR LRL-KDGGHY
D.striata-FP483          GGLCCITNDI SLT---GNCF YY-DIKFTGL NFPPNGPVVQ KKT-TGWEPS TERLYP--RD GVLIGDIHHA LTV-EGGGHY
C.californica-GFP        GACCTTSVHL SL-T--KNCF VHNST-FHGV NFPANGPVMQ KKT-LNWEPS SEKITP--FE GNLKGDVTMF LKL-EGGQQH
C.californica-YFP        GACCTTSVHL SL-T--KNCF VHNST-FHGV NFPANGPVMQ KKT-LNWEPS SEKITP--FE GNLKGDVTMF LKL-EGGQQH
C.californica-RFP        GGCGTASAHI SL-D--KNCF VHKST-FHGV NFPANGPVMQ KKT-LNWEPS SELITA--GD GILKGDVTMF LML-EGGHRL
F.concinna-GFP           GGFAAVSAHI SL-R--GNCF EHKSK-FVGV NFPADGPVMQ NQS-SDWEPS TEKITT--CD GVLKGDVTMF LKL-AGGGNH
A.lajollaensis-GFPa      GGVASVSADI SIEK--DGCF KHRSK-FVGV NFPVNGPVMQ NKT-RGWEPS IEKMTV--RD GTLKGHVTMF LKL-EGGGNY
A.lajollaensis-GFPb      GGFAAVSADL SLKN--GNWF VHQSK-FVGV NFPADGPVMQ KKT-VGWEPS TEKMIV--RD GTLEGDVTMY LKL-EGGGNY
M.meandrites-GFP         GGICTVAAEI SL-T--GSVF EHKSK-FVGV NFPANGPVIQ KKT-LGWETS TEKMAA--ND GSVQGYDTMF LKL-EGGGRH
AmilGFP-matz             GAVATASWNI RL-E--GNCF IHKSI-FHGV NFPADGPVMK KKT-IDWDKS FEKMTV--SK EVLRGDVTMF LML-EGGGSH
A.nobilis-GFP            GAVATASWKI RL-E--GNCF IHNSI-FNGV NFPADGPVME KKT-IGWDKS FEKMTV--SK EVLRGDVTMF LML-EGGGSH
AmilCFP-matz             GGVATASWNI RL-E--RDCF IHKSI-YHGV NFPADGPVMK KKT-IGWDKA FEKMTV--SK DVLRGDVTEF LML-EGGGYH
A.nobilis-CFPb           GGVATASWNI RL-E--RDCF IHKSI-YHGV NFPADGPVMK KKT-IGWDKA FEKMTV--SK DVLRGDVTEF LML-EGGGYH
Acropora-sp-GFP          GGVATASWSI RL-E--GNCF IHNSI-YHGV NFPADGPVMK KQT-IGWDKS FEKMSV--AK EVLRGDVTQF LLL-EGGGYQ
P.damicornis-CFP         GGVATASGHI CL-E--GNWF KHTSM-FHGV NFPANGPIMQ KRT-IGWDPS FEKMTV--SN NILRGDVTMF LQL-KGGGYH
A.sulcata-CP             GGFLTAHQDT SLD---GDCL VY-KVKILGN NFPADGPVMQ NKA-ERWEPA TEILYE--VD GVLRGQSLMA LKC-PGGRHL
A.sulcata-CPb            GGFLTAHQDT SLD---GDCL VY-KVKILGN NFPADGPVMQ NKA-GRWEPA TEIVYE--VD GVLRGQSLMA LKC-PGGRHL
A.sulcata-KFP            GGFLTAHQDT SLD---GDCL VY-KVKILGN NFPADGPVMQ NKV-GRWEPG TEIVYE--VD GVLRGQSLMA LKC-PGGRHL
A.equina-CP              GGYLTIHQDT SLQ---GNNF IF-KVNVIGA NFPANGPVMQ KKT-AGWEPC VEMLYP--RD GVLCGQSLMA LKC-TDGNHL
C.passiflora-CP          GGVLTAHQDT SLE---GNCL NY-KVKVLGT NFPADGPVMK NIS-GGWEPC TEIVYQ--DN GVLRGRNVMA LKV-SGRPPL
C.gigantea-CP            GGVLTAHQDT SLE---GNCL IY-KVKVLGT NFPADGPVMK KIS-GGWEPC TEIVYQ--DN GVLRGRNVMA LKV-SGRPPL
C.gigantea-GFP           GRVLTAHQDT SLE---GNCP IY-KVKVLGT NFPAASPVMK KVS-GGWEPS TEIVYQ--DN GVLRGRNVMA LKV-SGRPPL
H.crispa-CP              GGILTAHQDT SLE---GNCL IY-KVKVLGT NFPADGPVMK NKS-GGWEPC TEVVYP--EN GVLCGRNVMA LKV-GDRR-L
E.quadricolor-RFP        GGVFTVMQDT SLE---DGCL VY-HAKVTGV NFPSNGAVMQ KKT-KGWEPN TEMLYP--AD GGLRGYSQMA LNV-DGGGYL
A.sulcata-GFP            GGVLSATQET SLQ---GDCI IC-KVKVLGT NFPANGPVMQ KKT-CGWEPS TETVIP--RD GGLLLRDTPA LMLADG-GHL
H.magnifica-GFP          GGVLTVDQKT SLE---GDCI IC-NIKVHGT NFPADGPVMQ KQT-NGWEPS TETVIP--RG EGILLRDVPA LKLRNNKGHL
C018-G5-FP-like          GAVCTASADI TVS-VEENCF -YHESKFHGV NFPADGPVMK KMT-TNWEPC CEKIIPVPRQ GILKGDVAMY LLLKDGG-RY
C008-C4-FP-like          GAVCTASADI TVS-VEENCF -YHESKFQGV NFPADGPVMK KMT-TNWEPC CEKIIPVPRQ GILKGDVAMY LLLKDGG-RY
A.millepora-RFP          GAVCTASADI TVS-VEENCF -YHESKFHGV NFPADGPVMK KMT-TNWEPC CERIIPVPRQ GILKGDVAMY LLLRDGG-RY
A.millepora-GFPa         GAVCTASADI TVS-VEENCF -YHESKFHGV NFPADGPVMK KMT-TNWEPC CEKIIPVPRQ GILKGDVPMY LLLKDGG-RY
A.millepora-GFPb         GAVCTASADI TVS-VEENCF -YHESKFHGV NFPADGPVMK KMT-TNWEPC CEKIIPVPRQ GILKGDVPMY LLLKDGG-RY
A.palmata-RFP            GAVCTASADI TVS-VEENCF -YHESKFHGV NFPADGPVMK KMT-TNWEPC CEKIIPVSRQ GILKGDVPMY LLLKDGG-RY
AmilGFP-colorless-matz   GAVCTTSADI TVS-VEENCF -YHNSKFHGV NFPADGPVMK KMT-TNWEPS CEKIIPVPRQ GILKGDIAMY LLLKDGG-RY
A.millepora-colorless-Smith GAVCTTSADI TVS-VEENCF -YHNSKFHGV NFPADGPVMK KMT-TNWEPS CEKIIPVPRQ GILKGDIAMY LLLKDGG-RY
AmilRFP-Matz             GAVCTTSADI TVS-VEENCF -YHNSKFHGV NFPADGPVMK KMT-TNWEPS CEKIIPVPRQ GILKGDIAMY LLLKDGG-RY
Zoanthus-sp-RFP          GAVCICSADI TVN-VRENCI -YHESTFYGV NFPADGPVMK KMT-TNWEPS CEKIIPINSQ KILKGDVSMY LLLKDGG-RY
Zoanthus-sp-YFP          GAVCICNVDI TVS-VKENCI -YHKSIFNGM NFPADGPVMK KMT-TNWEAS CEKIMPVPKQ GILKGDVSMY LLLKDGG-RY
Zoanthus-sp-GFP          GAVCICNADI TVS-VEENCM -YHESKFYGV NFPADGPVMK KMT-DNWEPS CEKIIPVPKQ GILKGDVSMY LLLKDGG-RL
Ceriantus-sp-OFP         GGVLNMSSDI TYK---DNV LHGDVWAVGV NFPPNGPVMK NEI-VMEEPT EETFTP--KN GVLVGFCPKA YLLKDGSYYY
R.muelleri-GFP           GGLVEIRSDI NLI---EDKF VY-RVEYKGS NFPDDGPVMQ KTI-LGIEPS FEAMYM--NN GVLVGEVILV YKL-NSGKYY
Ptilosarcus-sp-GFP       GAIVDIRSDI SLE---DDKF HY-KVEYRGN GFPSNGPVMQ KAI-LGMEPS FEVVYM--NS GVLVGEVDLV YKL-ESGNYY
N.vectensis-GFP          GGYCTIVNTS ELK---DGS LHYHTNFHGI NLKPDGPVMQ KRT-MGWLPS VETNIP--RR DTLLGDINML LKVNDGS-FL
P.meadi-GFP              GGIISITFNY RYEGSK---- IIGDFKVIGT GFPTDSLIFT DKI-IKSNPT CENMFPKA-D NILVNAYTRT YLLKDGG-YY
L.aestiva-GFP            GGIISVNFTY RYEGNK---- VIGDFKVVGS GFPANSVIFT DKI-IKSNPT CEHIYPKG-D NILVNAYTRT WMLRDGG-YY
P.plumate-GFP            GGVLHVSFSY RYEAGR---- VIGDFKVVGT GFPEDSVIFT DKI-IRSNAT VEHLHPMG-D NVLVGSFART FSLRDGG-YY
C.poppei-GFP             GGILEVNFRY TYEFNK---- IIGDVECIGH GFPSQSPIFK DTI-VKSCPT VDLMLPMS-G NIIASSYARA FQLKDGS-FY
B.floridae-GFP           GGVLSCNFRY TYEGGK---- IKGEFQLIGS GFPAGGPVMS GGL-TTLDRS VAKLQCSD-D CTITGTNNWS FCTTDGK-RH
A.macrodactyla-GFP       DGKYKTRGEV KFE---GDT LVNRIELKGM DFKEDGNILG HKLEYNFNSH NVYIMPDKAN NGLKVNFKIR HNIEGGGVQ-
A.macrodactyla-OFP       DGKYKTRGEV KFE---GDT LVNRIELKGM DFKEDGNILG HKLEYNFNSH NVYIMPDKAN NGLKVNFKIR HNIEGGGVQ-
A.magnificus-GFP         DGKYKTRAEV KFE---GDT LVNRIELKGM DFKEDGNILG HKLEYSYNSH NVYVMADKPN NGLKVNFKIR HNIEGGGVQ-
A.victoria-GFP           DGNYKTRAEV KFE---GDT LVNRIELKGI DFKEDGNILG HKLEYNYNSH NVYIMADKQK NGIKVNFKIR HNIEDGSVQ-
A.coerulescens-colorless-GFP DGNYKSRAEV KFE---GDT LVNRIELTGT DFKEDGNILG NKMEYNYNAH NVYIMTDKAK NGIKVNFKIR HNIEDGSVQ-
Phialidium-sp-YFP        DGVFKTRAEV TFE---NGS VYNRVKLNGQ GFKKDGHVLG KNLEFNFTPH CLYIWGDQAN HGLKSAFKIM HEITGSKEDF
Anthomedusae-sp-GFP      DGVLETAAKV TLE---SGA IYNRISVKGV GFKKDGNVCK KRLHSS-PPQ VSYVVPY--G EGIRVLYSNI YPTKDGG--Y
Anthomedusae-sp-GFPb     DGVLETAAKI TME---SGA IVNRINVKGT GFDKDGHVCQ KNLESS-PPS TTYVVPE--G EGIRIIYRNI YPTKDGH--Y
Anthomedusae-sp-CP       DGTMTSHHTY ELD---GTCV I-SRITVNCD GFQPDGPIMK DQL-VDILPT ETHMFPHGSN AV-RQLCYIG FTTADGG---
Consensus                GGVCTASNDI SLE---GNCF IY-KVKFHGV NFPADGPVMQ KKT-LGWEPS TEKMYP--RD GVLKGDVTMA LKL-EGGGHY
```

N142H-amilCP
avicFP 148

P166A-amilFP

Figure 4.10 Continuation

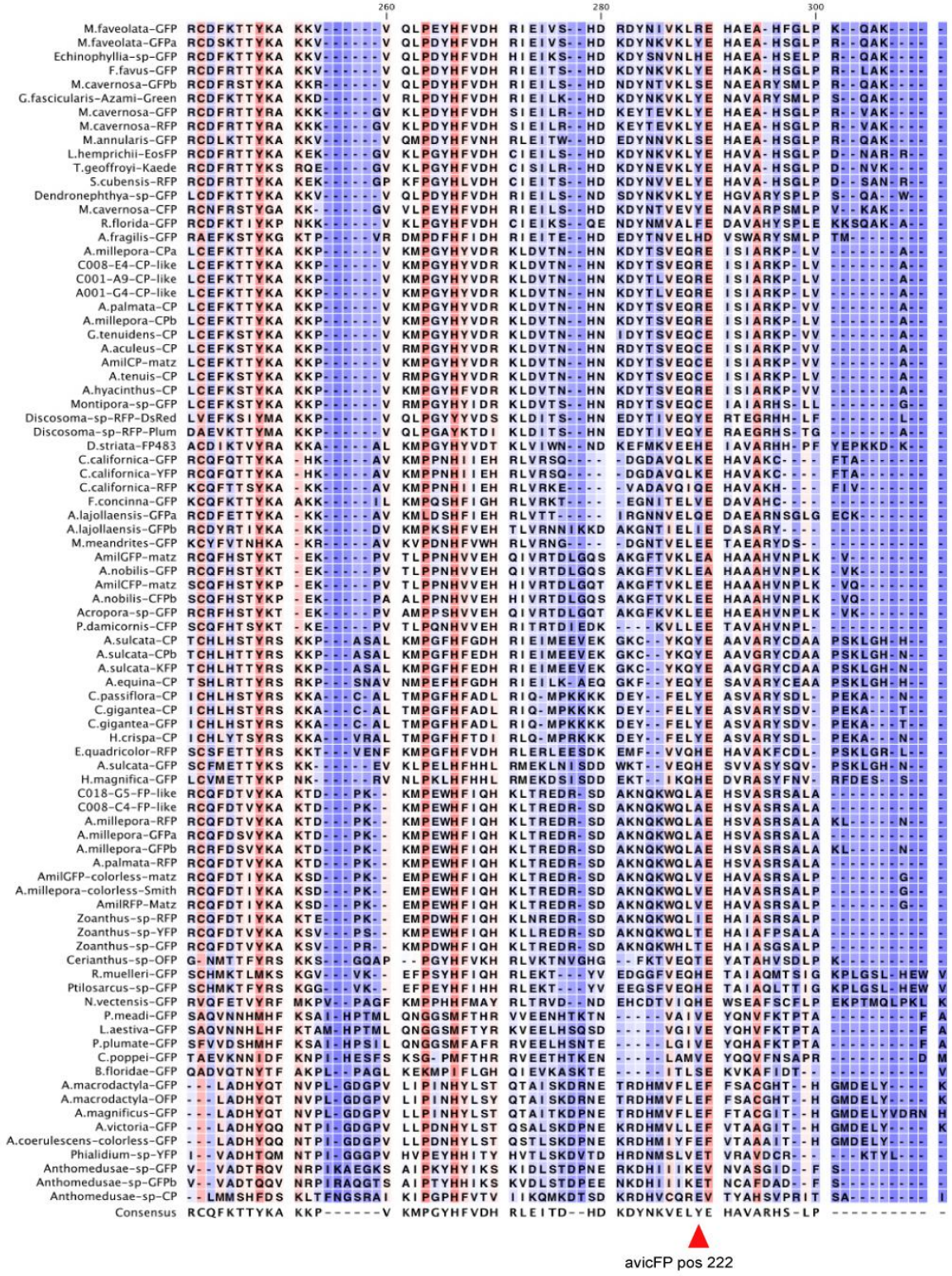

Figure 4.10 Continuation

The difference between the two Acropora CPs at this position may account for the observed red absorbance peak shift from 597 to 601 nm. Scleractineans (Figure 4.10) mainly have serine in this position, and only *A. palmata* shares the same residue as *A. millepora,* suggesting a deep genera-related substitution. Similar comparisons of cloned GFP-like molecules in other metazoans have shown no anthozoans that share similar amino acids, with the exception of a fluorescent hydrozoan homologue (Prasher et al. 1992, Shagin et al. 2004, Luo et al. 2006). It is tempting to speculate that this class-related amino acid residue is a key ancestral position retained in the majority of hydrozoans and is derived from anthozoans.

Aspargine 142 in the amilCP gene is widespread in Acroporids and is also a shared residue in CPs of *Goniopora tenuidens*, which belong to a suborder of Acropora. This implies that, similar to in its anemone-like homologues, the position is characteristic of the scleractinean CP gene. However, all fluorescent-like proteins have Cys148, a substitution observed in one favid and several anemone fluorescent CPs (Gurskaya et al. 2001, Shkrob et al. 2005).

The E26V substitution in the amilCP protein is located near the 142 or 143 site; however, the substitution falls just below the equatorial region (Figure 4.3), and the 142,143 sites are above the equator region. The substitution appears to have electrostatic properties in that protein surface region, as suggested by the isoelectric points of E26 (3.2) and V26 (5.96). The proportion of proteins with the V26 residue was 0.12 compared to 0.88 for those with the E26 amino acid (Figure 4.1). Compared to the rest of the scleractinians (Figure 4.10, general alignment), the consensus amino acid is glutamate (Ando et al. 2002, Karasawa et al. 2003, Carter et al. 2004, Karasawa et al. 2004, Sun et al. 2004, Kogure et al. 2006). However, there are fewer FP-like sequences from the *Montastraea* group, and *Scolymia cubensis* and *Meandrina meandrites* also show a similar V/E substitution (Kelmanson & Matz 2003, Carter et al. 2004, Sun et al. 2004). Comparisons of CPs and FPs from Acropora show that threonine occupies this position in all FPs (NCBI submitted sequences, Table 4). A similar position was found in sequences from distantly related *M. cavernosa* and *Favia favus* corals, indicating an ancient origin/fixation (Sun et al. 2004, Tsutsui et al. 2005).

With regard to the functional consequences of the E26V substitution, no connection with the genesis of the chromophore is expected. This is because the side chain faces the exterior of the protein; however, it is well known that certain residues

located in the surface of anthozoan GFP-like proteins are responsible for oligomerization (Baird et al. 2000, Wall et al. 2000, Zacharias 2002, Pakhomov et al. 2004, Verkhusha & Lukyanov 2004). The substitution, when compared with data from native and mutated proteins, is not located in the “interface” zone and therefore is not likely to be involved in the formation of the tetrameric complex. Alternatively, this amino acid might function in the binding/interaction with other proteins. It is possible that this protein region may promote its targeting to specific subcellular locales.

S53A was another substitution observed in the two CP genes. The dominant position was found to be alanine, with a relative proportion of 0.88 (Figure 4.1). The substitution was located at the base of the α-helix; hence, a low electronic interaction with the chromophore is expected to occur (Figure 4.3). The isoelectric point change between substitutions is small (6 to 5.68). Compared with the rest of the scleractinians (Figure 4.10), this substitution was also commonly alanine, and serine was the only other amino acid found in the coral-derived proteins. However, serine was the consensus in the FP contig (Figure 4.2), indicating a possible origin of the substitution in the CP gene and a fixation in the FPs. Further comparisons with the rest of the metazoans revealed that nearly all anthozoans as well as bilateria-derived molecules possess this A/S variation; the hydrozoan jellyfish was the only group to have a proline at this position (see Figure 4.10 and Table 4 for references). It is likely that this position plays a relatively important structural or functional role because of its restricted variability. As mentioned previously, the amino acid may not be entirely related to the acquisition of the colour/fluorescence phenotype.

From non-expressed CP variants, the S206L substitution was also abundant in the contig (Figure 4.1). What makes this interesting is that this position is relatively close to E26V; it also faces the outside of the barrel. Furthermore, according to the oligomer interaction/formation map, position 206 is not involved in tetramer formation, suggesting that it provides a larger area for facilitating protein–protein interactions.

In relation to the variation in the FP contig, the R10K substitution observed facing the exterior of the protein was highly abundant (Figure 4.2), with lysine being more abundant in the contig. Compared with other scleractinians (Figure 4.10 general alignment), lysine is the predominant amino acid at this position. Arginine at this position was also abundant in the FP contig; however, no coral GFP-like protein was

observed to have this amino acid, making the substitution especially unique to the AmilGFPb protein. The isoelectric points of the amino acids differ by 1.56 units, and they are not involved in tetrameric protein formation. It is possible that these amino acids are targets for selective recognition by similar proteins. An analogous situation may occur with the A13T substitution; however, structural changes in the surface of the protein are expected with this type of substitution.

A more important substitution with respect to the colour/fluorescence phenotype in the FP contig is D66T (Figure 4.2). This position is in the tripeptide motif involved in the generation of the chromophore. The change in green and red fluorescence observed in the two expressed variants (Figures 4.6C and 4.7C) is explained by this substitution. Threonine at this position is very rare in scleractineans and is found only in *M. meandrites* GFP (Figure 4.10). It is also rare in the rest of the metazoans; Phalidium GFP is the only non-anthozoan FP bearing this amino acid (Shagin et al. 2004, Sun et al. 2004). Aspartic acid, in contrast, was poorly represented in scleractinian GFP-like proteins and was observed only in distantly related *Favia* and *Montastraea* corals (Carter et al. 2004, Tsutsui et al. 2005).

P166A was a widespread surface variation in the contig (Figure 4.2). Two expressed green fluorescent variants had proline (Figures 4.4A and 4.4B), whereas the red fluorescent variant had alanine (Figure 4.4C). Overlapping the coralimorpharian DsRed map with the Acropora FP map shows that this substitution falls in the polar monomer–monomer interface (Wall et al. 2000). It is clear by comparing the substitution among the scleractinians (Figure 4.10) that no other coral bears this substitution; however, more extensive comparisons with sea anemone-, hydrozoan- and bilaterian-derived GFPs have shown that proline is present in the OFP of the octocoral *Cerianthus*, in the GFPs of the sea anemones *Anemonia sulcata* and *Heteractis magnifica*, and in the GFP of the coralimorpharian *Ricordea florida* (Wiedenmann et al. 2000, Tu et al. 2003). It is interesting that alanine was found only in non-cnidarian GFPs of the copepods *Pontellina plumate* and *Chiridius poppei* (Shagin et al. 2004, Masuda et al. 2006). There are two possible explanations for the similarities of this position with distantly related taxa. First, the presence of such substitutions in Acropora FPs represents an ancestral variation that may be related to the state/formation of the tetramer complex. Second, these similarities could be interpreted as *de novo* creations rather than inherited information from early ancestors because the Acropora genus is relatively recent in origin.

Two consecutive substitutions, R180Q and S183T, occur externally in the polar interface and internally in proximity to the chromophore, respectively, of the Beta-can (Figure 4.4). The difference in isoelectric point for the external substitution is 5.85 units, whereas that for the internal substitution is 0.02 units. R180 was scarce among the scleractineans and was found only in the GFP from *Acropora* sp (Figure 4.10), whereas glutamine at this position is dominant in *Acropora* genus GFPs. Glutamine at this position is also present in the closely related *Pocillopora damicornis* CFP and in distantly related proteins from *Fungia concinna* (Carter et al. 2004, Karasawa et al. 2004).

The internal S183T substitution appears to be widely distributed among scleractinians. The distal hydroxyl residue of the serine side chain may have some effect in the electronic environment of the chromophore and may account for the observed red shift in fluorescence between the expressed proteins. Two interesting variants (EST clones c002-d8 and a011-h9; Figure 4.2) in the FP contig showed the typical green TYG chromophore; however, threonine at this position instead of serine resulted in a red shift similar to that observed in the RFP with the DYG chromophore and Thr183. Based on the literature, the AmillFP variation at 183 does not fall strictly into the amino acids that interact with the chromophore. Despite being internally buried, this amino acid is far from the supposed interacting position (Wiedenmann et al. 2004).

4.3.2 Acropora GFP-like protein electrophoretic signature

The theoretical molecular mass of the FPs and CPs were 27.6 and 26.3 kDa, respectively. The apparent masses were in agreement with these values when the proteins were heat-denatured in the presence of a reducing agent (Beta-mercaptopethanol; Figure 4.5A). Values found in the literature for several GFPs, including the former avGFP, *Acropora formosa* “pocilloporin” (abs max 580 nm), *Pocilliopora damicorns* GFP-like protein (abs max 560 nm), and many other anthozoa- and bilateria-based proteins, similarly were in the ~28 kDa range (Ward & Cormier 1979, Dove et al. 1995, Niwa et al. 1996, Dove et al. 2001, Wiedenmann et al. 2002, Pakhomov et al. 2004, Masuda et al. 2006). There were, however, a few extra bands in similar contig-related proteins. The CPs (Figure 4.5A, lanes 4 and 5) had extra bands that appeared to be tetramers and dimers. It is possible that the

oligomeric interaction was stronger in the CPs than the FPs and that heat denaturation did not fully dissociate the protein complexes.

The FP proteins (Figure 4.5A, lanes 1–3) did not show high molecular mass bands after being subjected to heat denaturation treatment. This is consistent with the idea that FPs form less stabile oligomers compared to CPs. The amilRFP protein displayed a band at 10 kDa, which probably represents a protein fragment; backbone fragmentation occurs near the chromophore as a consequence of its maturation (Wiedenmann et al. 2000, Martynov et al. 2003). Although these assumptions could be tested by examining the crystal structures of each protein, this is outside the scope of this study.

4.3.3 The AmilRFP602 spectral signature: similarities and hypotheses

The amilRFP602 protein in PBS possesses dual component excitation spectra with a peak and shoulder at 584 and 540 nm, respectively (Figure 4.7B). However, the incipient eluant containing this protein had twice as much NaCl than was recommended by the manufacturer. Although the double component excitation spectra were observed, the shoulder of the PBS-eluted protein was the main peak, and there was also a component that absorbed at 580 nm (Figure 6.1, Appendix). This double component signature has been observed in the *A. victoria* GFP. In hydrozoan proteins, these two components correspond to the protonated and deprotonated states of the chromophore (Miyawaki 2004). The different states of the chromophore possess different emission characteristics when the protein is illuminated with UV and blue lights; however, these states have also been linked with proton pumping behaviour (Agmon 2005). It is uncertain whether the amilRFP protein has the capability to pump protons under the ionic conditions used in these experiments. For this study, the comparison/description of the GFP-like proteins that eluted in PBS at pH 7.2 remained the subject of focus.

The larval excitation signal (measured *in situ* and from the cloned product) has differences in peak excitation maxima compared with its counterparts from adult tissue (Cox et al. 2007, Alieva et al. 2008); the larval protein exhibits a red shift of ~24 nm. The absorbance characteristics of the amilRFP602 protein were very similar to those observed in some non-fluorescent GFP-like proteins from *Anemonia sulcata* (Wiedenmann et al. 2000, Bulina et al. 2002), i.e., a double component curve and an

absorbance maximum of 560 nm. It is interesting that the adult version of the protein (amilRFP593) described by Cox et al. (2007) and Alieva et al. (2008) had an emission signal that was blue-shifted by 9 nm compared with the larval protein described here. Whether this spectral difference confers special attributes to adult metabolism or coral development remains unclear.

Lastly, similar to the excitation signal, the emission signal of the larval protein under stronger ionic conditions was shifted by 8 nm (Figure 6.1, Appendix). Regardless of the ionic conditions at the time of measurement, it is clear that the emission spectrum of this RFP shows the largest red shift of its class (Figure 4.7C). The actinarian-derived RFP from *E. quadicolor* showed a similar far-red emission at 611 nm, and a mutant product from one hydrozoan CP called KillerRed also had a similar emission maximum (610 nm; Wiedenmann et al. 2002, Bulina et al. 2006).

4.3.4 Larval AmilGFP517a, b: The most common emission for a coral GFP-like protein

It is interesting that an emission maxima of ~517 nm (Figure 4.6C, F) appears to be one of the most abundant spectral signatures detected in caribbean corals (Fux & Mazel 1999). However, spectral data from adult *A. millepora* colonies showed the presence of two green variants with emission peaks at 504 and 512 nm (Cox et al. 2007). These small spectral differences between the larval and adult proteins appear to be based in the amino acid composition (Alieva et al. 2008; Figure 4.10). In a phylogenetic analyses, the adult GFP protein clustered with the cyan FP rather than the larval GFP (Figure 3.2a, chapter 3). These data suggest that some spectral channels are under selective pressure.

The excitation peaks of the adult green FPs were blue-shifted to ~504 nm, whereas the peaks of the larval FPs were at ~517 nm. Using the fourth derivate of the spectral data, it was possible to detect components of the excitation spectra that were similar to those of the green-yellow mutant zFP538-D68N (Ianushevich Iu et al. 2002) cloned from a sea anemone (*Zoanthus* sp.). Before publication of the adult-derived *A. millepora* FPs in the NCBI gene bank, the anemone FP genes were closest at the amino acid level to the Acropora protein, which explains the similarities in their spectral signatures.

4.3.5 The amilCP products: similarities with other cnidarian proteins and potential FPs

Spectrally, AmilCP597 and AmilCP601 (Figure 4.8A, B) appear very similar to the aeCP597 protein (Shkrob et al. 2005) cloned from the blue margins of the pedal disc of the sea anemone *Actinia equina*. In *A. millepora* adult colonies, this colouration is distributed principally in the growing tips of the coral colony. In planula larvae, the blue colouration due to the presence of these proteins is distributed across the endodermal tissue. In a recent attempt to generate new biotechnological probes, a mutant gene of aeCP597 containing a few amino acid substitutions exhibited far-red fluorescence, which may be useful for multi-labelling purposes because of its non-overlapping emission characteristics. Similarly, the *A. millepora* CP genes could be manipulated by mutagenesis to express a red-shifted FP.

The creation of a far-red mutant from Acropora could be advantageous if a stable coral cell line becomes available. During spawning events, it may be possible to deliver the far-red mutant gene into the early dividing embryo cells. Difficulties in translation would be avoided because the probe is part of the same transcriptome and would use the same codons.

The photoinduction of fluorescence was investigated using recombinant CPs. As shown in Figure 4.5D, panel 1, similar amounts of protein were blotted onto 3M paper and then left under a 100 W Hg UV, blue, or green light power beam for 5 min. Proteins not exposed to light served as controls. No light effects in the red or yellow-green fluorescent state were observed (Figure 4.5D, panels 2 and 3). Once adsorbed by the paper, the solvent evaporated, leaving the proteins in a dry state. The dried proteins acquired fluorescence, which suggests that water aided in quenching the fluorescence of the solvated proteins.

Consistent with this, in *A. victoria* GFP, a complex network of water-mediated hydrogen bonding occurs between the chromophore and its surroundings; in particular, the hydroxyl groups of Try66 and Thr203 participate in hydrogen bonding (Ormo et al. 1996, Tsien 1998, Follenius-Wund et al. 2003). The elimination of the water may affect the internal charge environment of the protein. This level of dehydration is difficult to achieve in an aqueous cell environment such as the cytoplasm. However, if the protein is secreted to a hydrophobic environment (such as lipid vesicles), it is possible that these CPs may then fluoresce. *A. victoria* GFP is

expressed mainly in the endodermal planulae, and this tissue generally has a high lipid content that is used for energy. The endodermal fluorescence could be explained in part by the presence of CP-like proteins that were dehydrated and subsequently acquire a fluorescent state. It is important to note that FPs in the green state undergo one dehydrogenation reaction and that two dehydrogenation reactions occur with proteins in the red state during chromophore coordination. An extended dehydrogenation reaction may have occurred with the blotted CPs; this might have been aided by the physical removal of water and/or oxygen-mediated catalysis, which are unlikely to occur in nature.

4.3.6 Chemical nature of the GFP-like chromophore: hypsochromic pH-dependent shift

The three main GFP-like recombinant variants were subjected to one-step pH changes. The results showed similar (473–446 nm) blue shifts that occurred under alkaline conditions in all proteins; however, under acidic conditions only the RFP variant showed a greater shift toward the UV region (Figure 4.9). These extreme pH conditions cause denaturation of the protein, which exposes the chromophore directly to the solvent. Through its spectral characteristics, the type of chromophore can then be determined (Martynov et al. 2003, Pakhomov et al. 2004). The absorbing behaviour that occurred with pH change for the AmilGFP517 protein appears to be very similar to data from a Pectinidae coral protein called Dronpa (Ando et al. 2002, Ando et al. 2004). Furthermore, the normal absorbance at pH 7.4 appears to overlap with a peak at 503 nm. In acidic conditions, Dronpa displays a peak at 390 nm with a proportional decrease in the major (503 nm) peak, and this is also observed in the Acropora protein (Figure 4.9B). The 390 and 503 nm peaks correspond to the neutral and ionised states of the phenolic hydroxyl group of the chromophore (Tsien 1998).

The Acropora CP and RFP show a very similar alkaline-related peak (Figures 9A and 9C). A comparable behaviour has been recorded for Favid corals and implies that the chromophores share very similar chemical structures (Oswald et al. 2007). Detailed analyses of chromophore biogenesis have confirmed that some red FPs and non-fluorescent CPs share similar pathways in the three-step reaction that is responsible for the formation of the red-shifted chromophore (Gurskaya et al. 2000, Kelmanson & Matz 2003, Martynov et al. 2003, Shkrob et al. 2005).

4.3.7 The 3D structure of GFP and its nearest structural protein neighbour

The topological variability of GFP-like proteins is important for supporting the nervous systems of many distantly related organisms. Hopf et al. (2001) explored the topological similarity of perlecan binding of nidogen 1 and the GFP Beta-barrel shape. Despite the low amino acid similarity (10%) in the normal alignment, the overall shapes of both proteins appeared to overlap. Nidogens are a component of the basement membrane, which is a thin sheet made of extracellular matrix that separates tissues such as smooth muscle, skeletal muscle, parineural cells, lymph nodes, and fibroblasts. The building of such membranes involves other proteins like collagen IV, laminins, and proteoglycans. Collagen IV and laminins self-assemble in a network-like style, whereas the nidogens connect these networks. It would be interesting to compare the expression patterns of Acropora nidogen-like proteins and GFP-like proteins. The similarities in the β-barrel domain may also explain other mutations not related to the colour/fluorescence phenotype. The spatial distribution of some amino acids (Figures 4.3 and 4.4) on the surface of Acropora GFP-like proteins suggests the presence of an electrically charged “patch” environment. The surface anomalies of the clones and the binding-station phenotypes are similar to the nidogen proteins, and it may be possible that they can bind functionally related proteins.

4.4 Conclusions

A total of five recombinant proteins—two green FPs, one red FP, and two CPs—were characterised and compared with data from other related proteins. Analyses of the GFP-like protein diversity from the Acropora EST database showed important and abundant variants that likely contribute to the colour diversity that occurs in the embryo stage of *A. millepora*. The molecular masses of FPs were slightly larger than those of CPs but were in the range of other cloned GFPs from a variety of taxa. The spectral characteristics of the recombinant proteins showed a typical double excitation peak for the RFP and a widespread taxa-related green emission with unusual excitation wavelengths. Absorbance of the CPs was similar to other red-shifted cnidarian proteins, and, when blotted onto 3M paper, fluorescence occurred.

CHAPTER 5

TEMPORAL AND SPATIAL MRNA EXPRESSION AND FLUORESCENCE ANALYSES OF GFP-LIKE GENES DURING *ACROPORA MILLEPORA* EMBRYOGENESIS

5.1 Introduction

During metazoan embryogenesis, several genes that control morphogenesis appear to share a common spatial pattern of expression. It is during this time that organs are formed, but apart from organisms that undergo metamorphosis, morphogenesis occurs only once during the lifetime of an organism. For organisms that do undergo metamorphosis, a secondary morphogenesis event occurs. During this time, in some hydrozoans, the cellular fate of the planula ectoderm is changed by apoptotic mechanisms, creating new tissue characteristics for the newly created “polyp” (Seipp et al. 2001). This polyp differs from its planula stage in terms of its sessile character and, for some reef-dwelling cnidarians, the polyp stage is the starting point of exoskeleton formation.

Due to advances in mRNA manipulation, now it is possible to visualise the *in situ* expression of a specific gene in model organisms. This technology, when applied during embryological development, allows direct comparison of dynamic changes in gene expression across different animal phyla. With such information, the interpretation of possible evolutionary pathways becomes possible. In the coral *Acropora millepora,* for example, the expression pattern of important developmental genes, such *Snail* (involved in metazoan early gastrulation) and *Dpp* (involved in dorsoventral specification), are similar to those observed in the bilaterians *Drosophila* and mice (Ball et al. 2004; Hayward et al. 2004). Although cnidarians have typically been categorised as radial rather than bilateral animals, these expression patterns suggest that the common cnidarian-bilaterian ancestor was already a bilateral organism.

Genes involved in mesodermal tissue specification have also been found in diploblastic cnidarian genomes. The existence of such genes in a diploblastic

organism, which in higher metazoans specify the third tissue layer, has been explained in terms of potential initial/ancestral cellular (proliferation) activity that was later recruited/exploited in a triploblastic bauplan to create a new type of tissue (Technau 2001; Technau & Scholz 2003).

Symbiotic corals have several reproductive strategies, products of the combined effects of life history traits (Hall & Hughes 1996). They can reproduce asexually (via fragmentation) or sexually, in which hermaphroditic or gonochoric species fertilise gametes and development occurs internally (brooders) or in the water column (broad spawners) during synchronous mass spawning (Harrison et al. 1985). During gamete release into the water column, potential hybridisation between different species may have given rise to the diversity observed today, especially within the genus *Acropora* (van Oppen et al. 2000, 2001, 2002).

A. millepora is one of several coral broad spawners in the Great Barrier Reef. This species, along with others, usually spawns during the second or third night after the spring full moon, at around 9–10 p.m. Fertilisation occurs at the water surface and the first cell divisions begin 1–2 h later. After 12 h, the embryo remains at the surface and by then its morphology resembles that of an oriental prawn chip, an irregular milky brown-coloured bi-layered array of cells about 200 µm in diameter. Occasionally, it is possible to observe a red-pink colouration at this stage in the embryos (Figure 5.1; 12-h prawn chip embryo).

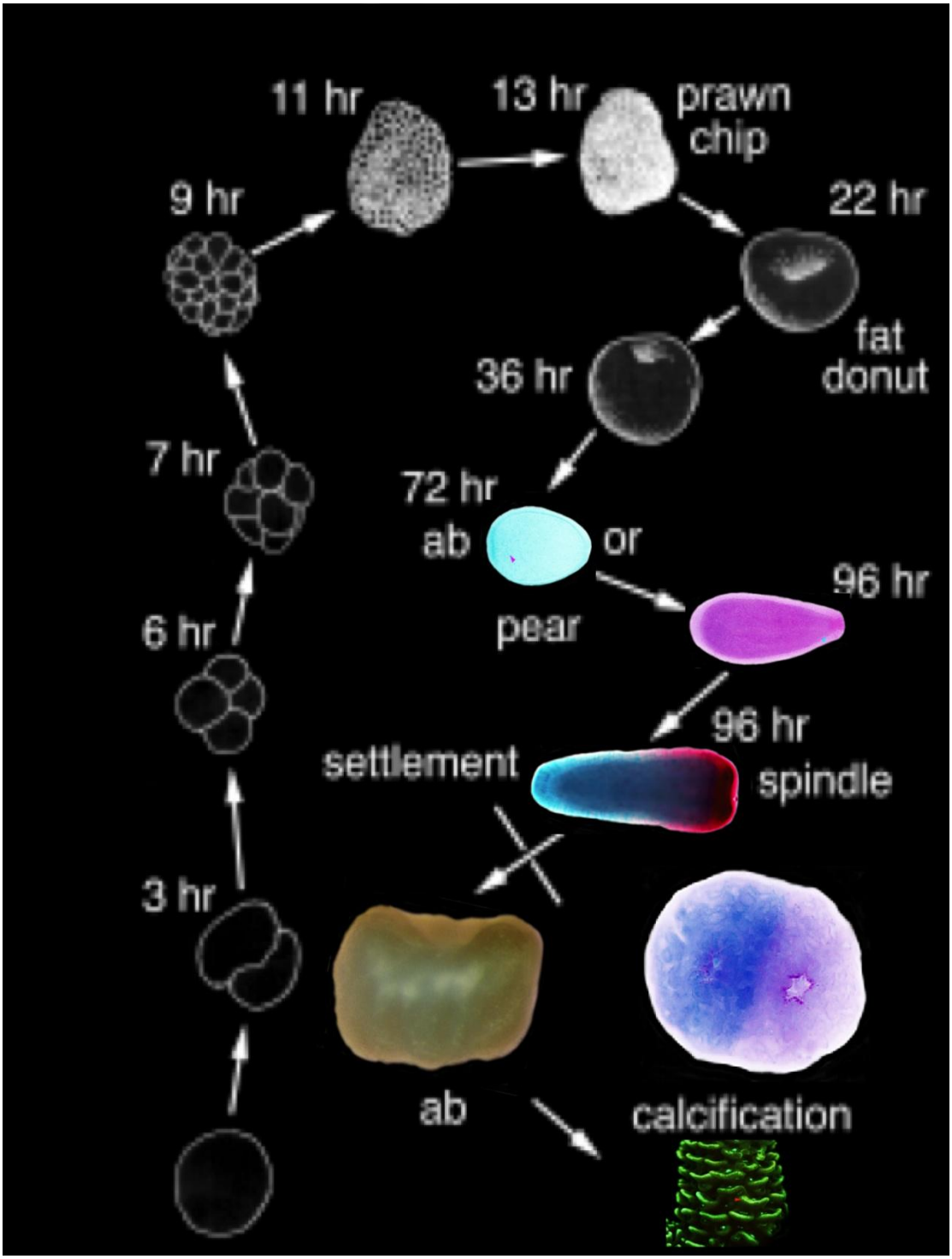

Figure 5.1 Embryological development of *Acropora millepora*.
Modified from Ball et al. (2002).

After approximately 24h post-fertilisation, the embryo usually remains non-motile, and its general morphology is similar to that of a fat donut, a product of the gastrulation-like process that forms the endodermal tissue and the oral pore (Figure 5.1, 22-h “fat donut” embryo). It is during the second night that gastrulation usually ends and the embryo acquires a spherical shape and rotation behaviour, driven by synchronous ciliary beating, begins.

From approximately 36 h post-fertilisation to the next day, no dramatic morphological changes are observed, but there is a slight elongation along the axis, forming a pear-like shape (Figure 5.1, 72 h post-fertilisation). Swimming behaviour is now directional (spinning on its own axis), with the aboral end first; however, the embryos are still close to the water surface, because of their high lipid content.

From this point to approximately 88–96 h post-fertilisation, the embryo experiences a dramatic elongation along its oral–aboral axis, acquiring the classic planuloid shape, with the appearance of diverse cellular types, including neurons, nematoblasts (nematocysts), and specialised secretory cells. In addition, thickening of the aboral-most ectodermal tissue is visible (Figure 5.1, 96-h planula larvae; Ball et al. 2001, 2002). At this point, the coral larvae generally leave the water column, actively following chemical cues leading them to the bottom of the reef. This bottom-searching period can be variable, but usually lasts approximately 2 days (Dr. Andrew Baird, JCU *comp pers*). After the planula finds the right spot to settle, it will attach via its aboral end, initiate metamorphosis to form a polyp, establish symbiosis, and begin exoskeleton formation (Figure 5.1, top view of two polyps). From then on, the coral polyp divides asexually until it reaches maturity (approximately 3–5 years). Then, it waits for the proper environmental cues to trigger the expulsion of its gametes into the water column, to continue the survival of the species.

This chapter describes the behaviour of green fluorescent protein (GFP)-like proteins *in situ* through the analyses of mRNA and protein expression at various stages during *A. millepora* development. Due to the ubiquitous fluorescence characteristics of the protein products, this combined method offers an excellent opportunity to observe the dynamics of gene and protein expression during the formation of a new individual. The information derived is discussed in terms of differences and similarities (observed in the pattern of expression and the cell type involved) with other gene expression patterns observed in this species and other cnidarians,

principally during the planula stage. Despite the great morphological/ecological plasticity of adults in this phylum (*e.g.*, jellyfishes *vs.* corals), the planula bauplan is similar across the phylum, allowing proper comparison. Using this temporal and spatial information, in conjunction with other evidence, the potential roles of specific genes and proteins will be explored at the organism and cellular levels.

5.1.1 Statement of Goals

Expression patterns in space and time can provide critical information about potential functions of genes during embryonic development, and studies of this kind are particularly important in non-model organisms where traditional genetics approaches cannot be applied. Based on expression data it is possible in principle to assign the relevance (timing and level of transcription) and putative domains of functionality (e.g. if the transcript or protein is localized at the aboral end and in a distinctive neuron-like cells, it is possible that the domains of functionality could be related with signal transduction) of the gene or protein. Gene expression analyses are commonly performed using *in situ* hybridization techniques, however, thanks to the inherent fluorescent properties of the gene products, GFP-like protein expression can be analysed *in vivo* by the use of fluorescence microscopy. In this thesis, the combination of these two approaches was extremely useful due to the potential for cross hybridization between mRNAs encoding spectroscopically distinct proteins.

5.2 Results

5.2.1 The major GFP gene types are differentially expressed in early coral development

The expression of the *fluorescent protein* (*FP*) and *chromoprotein* (*CP*) loci was investigated during embryonic and larval development in *Acropora* by Northern blot analyses and whole-mount *in situ* hybridisation. High levels of nucleotide sequence identity within the CP and FP clades effectively precluded resolution between specific spectral variants (red fluorescent protein [RFP] *vs.* GFP). Thus, the mRNA expression patterns reported here reflect general characteristics of the *CP* and *FP* classes.

Northern blot analyses showed major differences in the timing of *CP vs. FP* expression (Figure 5.2H). While *CP* mRNA expression was detected from the

gastrulation (35 h) stage, mRNAs corresponding to the FP types were only detectable much later in development (the pear stage at 52 h; Figure 5.3H). *In situ* hybridisation experiments indicated that *CP* expression was predominantly, but perhaps not exclusively, endodermal during larval development. Expression was first detected during gastrulation, in developing endodermal tissue and scattered ectodermal cells (Figure 5.2, panels A and B, respectively). All positive cells in the endoderm and the scattered ectodermal pattern remained from this stage to the planula stage (Figure 5.2, panels C–E). After metamorphosis, strong *CP* expression was observed throughout the endoderm, but was most prominently associated with the developing mesentery walls (Figure 5.2G). In older polyps, *CP* expression was specifically associated with the endoderm in the region of the developing tentacles (Figure 5.2F).

FP probes, on the other hand, showed a highly specific pattern of expression in the ectoderm during larval development. At the pear stage, staining was associated with a subset of transectodermal cells that appeared to be denser in the aboral half of the larva (Figure 5.2, panel I). In early planulae (~52 h post-fertilisation), the ectoderm tissue that formed the mouth started to stain (Figure 5.2, panel J, arrowheads); slightly later, there were two obvious domains of expression, the extreme oral ectoderm inside the mouth and the dense transectodermal staining at the aboral end of the planula (Figure 5.2, panel K). After metamorphosis, the aboral staining disappeared, whereas the oral staining remained; strong endodermal staining was also observed in the majority of surveyed organisms (Figure 5.2, panel L). In older polyps, the endodermal signal was primarily associated with the forming mesenteries (Figure 5.2M).

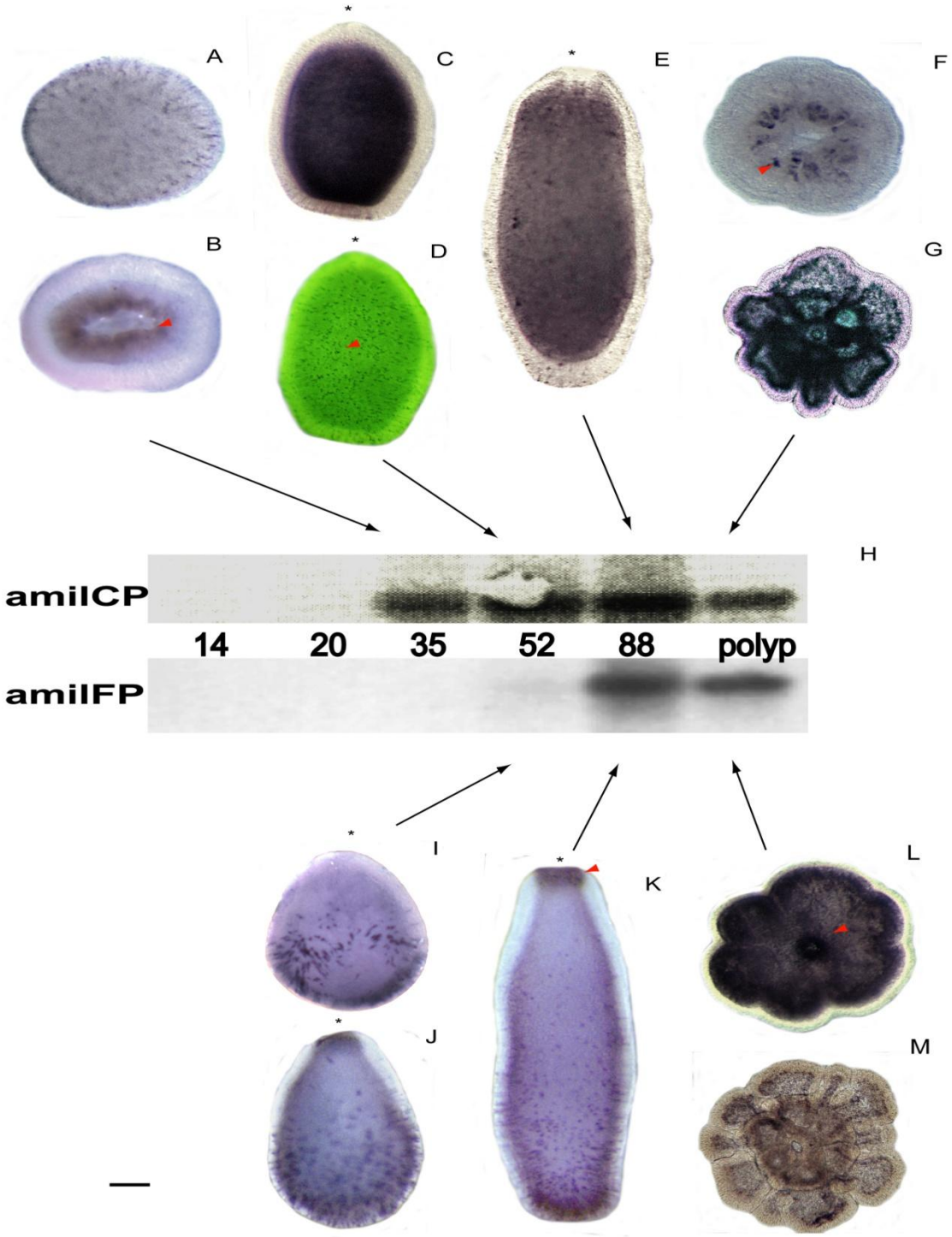

Figure 5.2 *In situ* and *in vitro* time-dependent detection of the GFP-like mRNAs in *Acropora* development.
A, B) *A. millepora* chromoprotein (*amilCP*) mRNA detected in a cross section and the whole 35-h embryo, respectively. C, D) Pear-stage *amilCP*-associated detection observed under normal light and fluorescent enhanced view of the ectodermal surface, respectively. E) *AmilCP* was detected in the planula (approx. 88 h). F, G) Early and later expression of *amilCP* in metamorphosed polyps, respectively. H) Northern detection of the *amilCP* and fluorescent protein (*amilFP*) genes in *Acropora* development. I, J) Early and late detection of the *amilFP* mRNA in pear-like embryos. K) Detection of the *amilFP* in planula larvae. L, M) Early and later staining pattern of *amilFP* in young polyps. Scale bar equals 200 μm (Beltran-Ramirez *et al.* 2009)

5.2.2 Fluorescence microscopy of GFP variants

Although *in situ* hybridisation was unable to resolve *GFP* and *RFP* mRNA expression, the strong and specific spectroscopic characteristics of the corresponding proteins enabled their distribution patterns to be determined via fluorescence microscopy. To supplement the *in situ* hybridisation data, we examined the fluorescence characteristics of coral larvae during development.

After 52 h of development, at the pear stage, the majority of the embryos started to show a red colouration at the aboral end (Figure 5.3, panel A). Green fluorescence was detected throughout the endodermal cavity and was restricted ectodermally to the mouth area (Figure 5.3, panel D). At this stage, embryos observed under a CY5 filter showed variable low endodermal red fluorescence; ectodermally, all embryos showed strong cell-specific red signals towards the aboral end (Figure 5.3, panel G). Using a combined filter, embryos showed yellow endodermal fluorescence, because of the overlapping green and red signals, whereas tissue forming the oral and aboral ends appeared green and red, respectively (Figure 5.3, panel J).

The strongest GFP-like mRNA signal was detected in the planula stage via Northern analyses; *in vivo,* larvae displayed a blue endodermal colouration, due perhaps to CP expression, whereas the red ectodermal aboral colouration observed previously was more evident (Figure 5.3, panel B). Planulae observed under GFP2/FITC filter showed strong green fluorescence throughout the endoderm, as well as within the ectodermal mouth and the surrounding area (Figure 5.3, panel E). Detailed inspection of the oral and mid-ectodermal areas showed tightly packed columnar cells forming the mouth sphincter, and isolated neuron-like cells running along the ectodermal basement, both with associated green fluorescence (Figure 5.3, panels M and N, respectively). A red fluorescent signal was detected in specific ectodermal cells with higher abundance towards the aboral end, forming a gradient along the planula body (Figure 5.3, panel H). Closer observation of the mid-ectodermal section under the CY5 filter showed columnar-type cells, with a broad basement layer, with observable basal prolongation and small spherical processes facing the larvae surface (Figure 5.3, panel O).

Planulae observed using the combined filter typically showed yellow endodermal fluorescence, a product of overlapping green and red signals (Figure 5.3, panel K). At greater magnification, the ectodermal mid-body section showed a greater abundance of RFP-expressing cells and isolated GFP-expressing ones. These cells also exhibited centrally positioned nuclei and columnar morphology with spherical processes facing the surface of the planula, and they appeared to extend along the ectodermal basement (Figure 5.3, panel P).

After metamorphosis, the polyps appeared translucent, with zooxanthella in their endodermal tissue (Figure 5.3, panel C). In young polyps observed under the GFP2 filter, green fluorescence appeared to be predominantly associated with the developing mesenteries, whereas it was localised to the mouth area in the ectoderm (Figure 5.3, panel F). Similarly, newly metamorphosed polyps observed under the CY5 filter showed primarily red fluorescence associated with the mesenteries, but no cell-specific ectodermal signal was observed (Figure 5.3, panel I). Later, after the polyps started to form the exoskeleton, green/red endodermal fluorescence was observed; however, the green mouth ectodermal signal remained in the majority of the polyps (Figure 5.3, panel L)

.

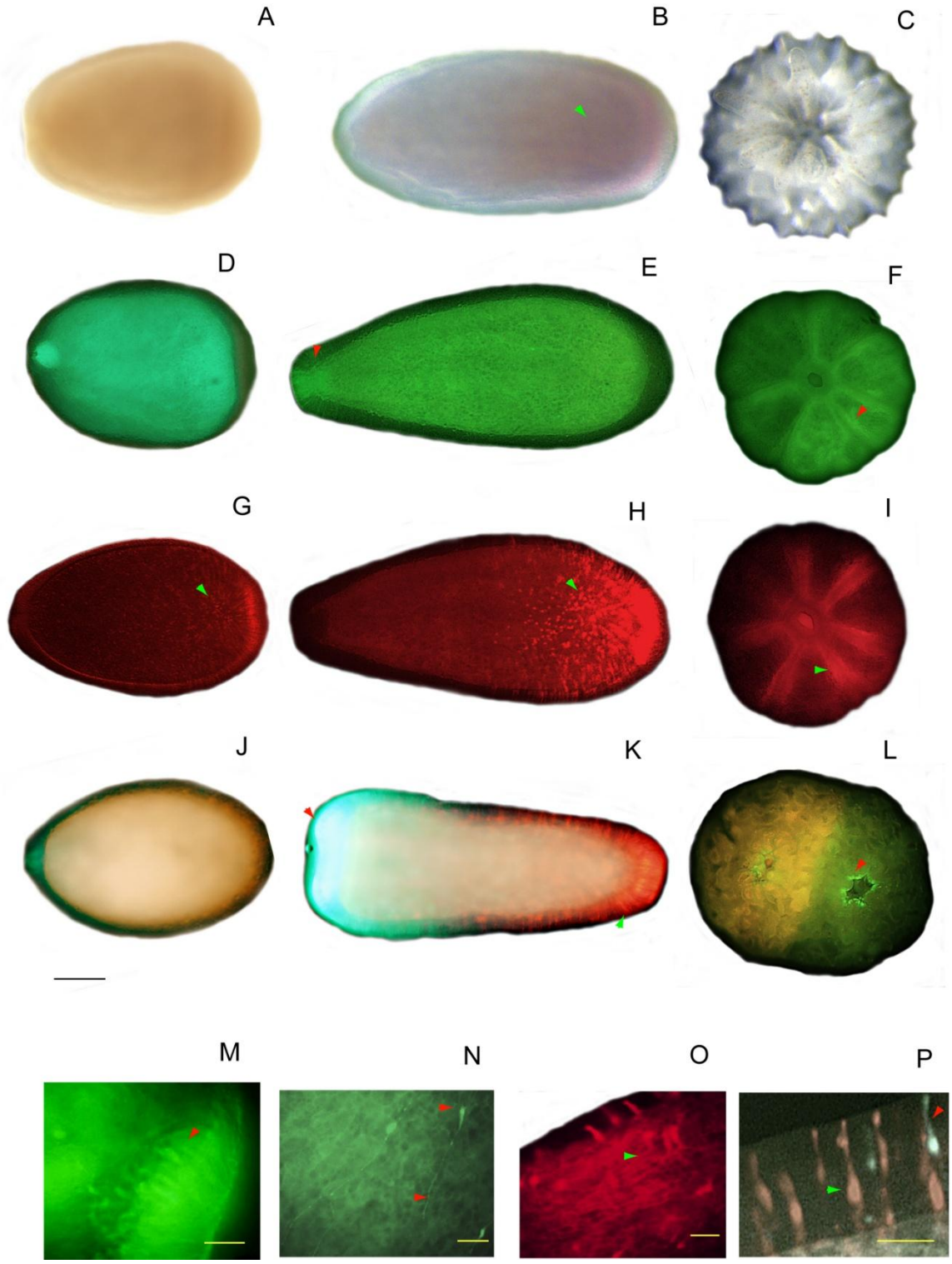

Figure 5.3 *In vivo* observation of the principal morphological and fluorescence characteristics of *Acropora millepora* embryogenesis.
A–C) Pear, planula, and polyp observed under normal lighting conditions, respectively. D–F) Pear, planula, and polyp observed under the GFP2 filter, respectively. G–I) Pear, planula, and polyp observed under the Cy5 filter, respectively. J–L) Pear, late planula, and polyps observed under the combined (WIB) filter, respectively. M) Detail of planula's mouth, observed under the GFP2 filter. N–P) Mid-section of planula ectoderm, observed under the GFP2, Cy5, and WIV filters, respectively. Oral opening is oriented to the left in the larval pictures. Arrowheads indicate important features, such as specific cells or domains with associated fluorescence. Scale bars at the organism (A–L) and cellular (M–P) levels indicate 200 and 20 μm, respectively (Beltran-Ramirez *et al.* 2009).

5.4 Discussion

5.4.1 GFP-like message temporal regulation

The “virtual Northern” has become an important comparative tool for relative gene expression comparisons during coral development (Anke et al. 2001; Hayward et al. 2001, 2002; Hislop 2003). The assay results demonstrate significant up-regulation of both *Acropora* GFP homologues during the planula stage (Figure 5.2H). In this planula stage, the organism is active and planktonic.

Comparing the levels of expression between polyp and planula (Figure 5.2, panel H), it is clear that GFP down-regulation occurs from the planula to polyp stage. Interestingly, *in situ* hybridisation revealed a considerable reduction in endodermal staining as the polyp became older and started to form mesenteries (Figs. 2 and 3, panels K and O, respectively), whereas the ectoderm seemed to lose its strong aboral cell-specific signal (Figure 5.2, panel L). In relation to the observed down-regulation of GFP-like proteins after metamorphosis, work using hydra as a model (Seipp et al. 2001) showed that almost all of the ectoderm of the metamorphosis-induced planulae underwent apoptosis, during which cell death and phagocytosis appeared to occur from the posterior to the anterior part of the planulae. These results indicated that the anterior cells that once functioned as a sensory organ on the active planulae had disappeared, to become the calicoblastic ectoderm, responsible for the deposition of exoskeleton. Similar to the case in *Hydra,* the observed changes in GFP-like transcript levels after *Acropora* metamorphosis may also be related to tissue characteristic modifications, from columnar cells in the planula to squamous tissue in the early calcifying polyp (Vandermeulen 1974, 1975).

5.4.2 Nature of the observed colouration in the early stages

According to the mRNA norther analyses, the GFP-like message appears after gastrulation (Figure 5.2, panel H).. There were however, some cases when a colouration was observed in the proliferating (timing before the GFP mRNA detection) tissue of some embryos (Figure 6.3, Appendix). This coloration was attributed, instead of GFP-like based ones, to the presence of another type of pigment of maternal origin. The orange-pink pigment is likely a carotenoid compound. Carotenes are small, soluble, organic, non-fluorescent molecules (Fingerman, 1965). The distribution of the pigmentation in the prawn chip embryo

was apparently on the surface of the cell, indicating its hydrophobic character. Also, during the 2006 spawning season, some of the eggs showed purple and orange pigments, which turned out to be insoluble in water, but soluble in ethanol (Wiedenmann Ulm Univ., personal communication).

5.4.3 Green-red signal distribution in planulae

Atoda (1951) reported on pigmentation in planulae from *Acropora bruggemanni,* in which the distribution of colouration appeared to follow a pattern similar to that in *A. millepora*: green at the oral end and red at the aboral end. More recently, a study by Hollingsworth et al. (2006) on the distribution and appearance of fluorescence in *Fungia scutaria* embryos described similar patterns to those observed in the planulae of *A. millepora*. In this previous study, the authors performed fluorescence microscopy and immunohistochemical analyses of protein extracts at specific time intervals (using an anti-hydrozoan GFP antibody). In general, they found up-regulation of these proteins occurring later in development. These results, compared with those of the virtual Northern (Figure 5.2, panel H) and fluorescence microscopy at later stages (Figure 3, panels J and K), are consistent with the purposed similarities in the up-regulation of mRNA expression and protein fluorescence relative to the age of the *A. millepora* embryo.

The mushroom coral planula, in comparison with *A. millepora*, does not display red transectodermal cells towards the aboral end (Figure 5.3, panels G and H), but it does display green fluorescence in the oral region (Figure 5.3, panels D and E). In *Fungia* embryos, the up-regulation of this oral green fluorescence was stronger as the planulae aged. *Acropora* green fluorescence in the mouth region was also similar, with increasing fluorescence at later stages. It has been hypothesised that green fluorescence in the oral tissue of the planula may function as an attractor for the dinoflagellate symbiont (Hollingsworth et al. 2005). This suggestion is based on indirect evidence of green phototaxis in free living coral symbionts. This phototactic behaviour has been also observed in other dinoflagellate species, each with its own spectral preference (Horiguchi et al. 1999). If such light preferences also occur in *Symbiodinium* clades, this behaviour may be correlated with slightly different mouth fluorescence signals. This hypothesis, however, may only be applicable in open systems where the symbionts are acquired from the environment; in closed systems

(*i.e.,* maternally inherited), the presence of this signal may have other roles than the attraction of suitable algal strains.

Hollingsworth et al., (2006) also reported red fluorescence in *F. scutaria*, detected only in the endodermal cavity, and the up-regulation of this signal was less dramatic relative to its green counterpart. In *A. millepora,* the red fluorescent signal was also distributed in endodermal tissue, and this signal was up-regulated more quickly than in *F. scutaria* planulae. However, in both species, green fluorescence appeared first, and red fluorescence appeared later.

5.4.4 Non-endodermal *in situ* staining; no fluorescent product was observed

During our survey of fluorescence characteristics, some embryos exhibited variable endodermal signals, even though mRNA levels showed very little variation in the endoderm. Despite the correlation between mRNA and protein signals, these cases of under-detection of both signals could be explained by a combination of colourless protein production and undetectable fluorescence, and, perhaps, the down-regulation of the gene product (mRNA) with the acquisition of normal levels of protein per cell. To address the question, it will be necessary to systematically separate the variation by controlled coral crosses, and increase the sampling rate (fluorescent visual inspections on the developing embryos), particularly at later stages when the “normal” up-regulation would be expected to occur. To preferentially avoid interference by the ectodermal signal, small endodermal sections could be used for selective RNA detection/quantitation, relative to the actual state of tissue fluorescence. By doing so, it may be possible to determine whether the observed changes are actually a product of up-regulation of a colourless GFP-like protein, or due to the down-regulation of a specific fluorescent gene.

5.4.5 GFP-like protein-expressing cell types in *Acropora*

The first cell-specific signal occurred just after the end of gastrulation (early pear embryo, Figure 6.4, Appendix), but are more conspicuous at the pear stage (Figure 5.3 panel D, G). After this stage, diverse cell types became apparent (Ball et al. 2001, 2002, 2004). It appeared that the relatively low abundance of the newly appearing green transectodermal cells remained constant, based on closer examination of middle planula ectodermal tissue (Figure 5.3, panel P), whereas the

proliferation of red fluorescent cells began along with the elongation process (Figure 5.3, panel G) and became more pronounced towards the aboral end as the embryo became more elongated.

Closer examination of the green fluorescent cells distributed along the basement of the ectodermal tissue (Figure 5.3, panel N) showed strong nuclear signals and neuron-like cell morphology, based on long bipolar projections. Red fluorescent cells, on the other hand, were apparently absent from the basement of the ectoderm; instead, there were columnar-type cells, with small spherical processes at the edge of the planula surface, with a broad basement showing slender projections running along the ectoderm/mesoglea boundary (Figure 5.3, panels O and P). Both types seemed to be sensory neurons that reacted against RFamide antisera (Martin 1992).

In cnidarians, two general types of neurons can be found in the ectoderm: sensory cells distributed from the free planular surface to the mesoglea, characterised by the presence of an apical cilium and dense core vesicles, and ectoderm-basal ganglionic cells, distributed along the entire anterior–posterior axis of the planula (Kolberg & Martin 1988). These neurosensory cells are known to arise via the transdifferentiation of epithelia/muscle cells, whereas the ganglionic cells arise from migratory undifferentiated interstitial cells (Martin 1992).

GFP expression has been also linked with bioluminescence. In relation with this, photocyte characterisation in pennatulids (*Venterillum, Renilla, Stylatula,* and *Acanthoptilum)* and hydroids (*Obelia, Campanularia,* and *Clytia)* has revealed the presence of one or more cytoplasmic projections larger than about 20 μm. These ectodermal photocytes, according to the description, were also similar in shape to the *A.millepora* planula green fluorescent transectodermal cells (Figure 5, panel D). Based on this similarity and the relatively low cell-type diversity within the phylum (Miller & Ball 2000; Ball et al. 2001, 2002, 2004; Technau et al. 2005; Kamm et al. 2006), metabolic and functional similarities may also be expected. In the case of bioluminescent cnidarians, sensory neurons represent the cell type most capable of combining reception and control of a response via calcium homeostasis. Among the Scleractinian corals and their planula larva, however, no bioluminescence is observed. However, the presence of GFP-expressing cells with similarities to those observed in bioluminescent organisms suggests that although aequorin-like activity

was lost, GFP-like protein expression persisted throughout coral-specific evolutionary history.

Early detailed anatomical descriptions of *Acropora* planulae (Ball et al. 2001) showed that the mouth tissue, in contrast to the rest of the ectoderm, consisted of only one cell type. These distinct cells were of a slender, secretory type, with many vesicles along the cell body and ciliary processes at their ends, which presumably function in the creation of water currents to draw in small food particles. In the present study, most of the planulae observed under the GFP2 filter and stained with the amilFP probe *in situ*, showed strong signals associated with this tissue (Figs 5.2 and 5.3, panels K and M, respectively). A recent study in eukaryotic cells pointed out the specific compartmentalisation of GFP-like recombinant proteins within lysosomes (Katayama et al. 2008). Lysosomes are acidic compartments primarily dedicated to molecular degradation. That these secretory-like cells were fluorescent, combined with eukaryotic evidence of lysosome GFP-like protein compartmentalisation, suggests a role for amilGFP517 in digestive functions.

Closer observation of the red fluorescent cells (Figure 5.3, panel O) shows that, in addition to having neuron-like morphology, the cells contain several intracellular granules (Figure 6.5, Appendix) that were apparently highly fluorescent. Unfortunately, no data were available for comparison, at least at the planular level. However, several studies in adult corals, including some in *A. millepora,* have also reported these aggregations in actively GFP-expressing cells, and attributed this observation to the increased tendency of this protein to form oligomers (Salih et al. 1998, 2000, 2004; Cox & Salih 2005; Cox et al. 2007).

The endoderm also showed some fluorescence (Figure 5.3, panel N). The cells, however, were not easily recognisable as independent units; instead, a field of evenly distributed round, non-fluorescent bodies embedded within a relatively homogenous fluorescent region were observed, with occasional highly fluorescent, but small, spherical granulations. Combining the fluorescence and *in situ* data (Figure 6.6, Appendix) it is probable that the small granulations observed via fluorescence microscopy are nuclei, as the *in situ* hybridisation data identified similar small stained units. Few data are available regarding cell characterisation in the endoderm of the planula; instead, the cells are typically categorised as yolk cells, interstitial cells, nematoblasts, or gastrodermal cells capable of acquiring intracellular dinoflagellates that digest/distribute the food content (Spurlock & Cormier 1975; Ball

et al. 2002; Dandar-Roh et al. 2004). However, characterisation of G protein-coupled receptors (GPCRs) showed that these genes were expressed exclusively in the endodermal cavity, in a subset of cells described as round or ovoid-shaped and clearly different from yolk cells (Anctil et al. 2007). Thus, the GFP-like mRNA-expressing cells observed in this study may be the same cells expressing GPCRs. It would be interesting to perform a double *in situ* study to clarify this and to better characterise endodermal cell diversity.

Finally, regarding the green fluorescence associated with the ectodermal neuron-like cells, it is possible that this signal is not only due to GFP-like proteins. Considerable evidence has indicated that some neurotransmitters have fluorescent properties, including catecholamines and dopamine (Reynolds et al. 1981; Mabuchi Michiaki et al. 2001). The pure compound can readily be excited by ultraviolet–blue light; however, its detection in live material requires concomitant glyoxylic acid treatment (Battenberg & Bloom 1975; Kolberg & Martin 1988). Although the planulae were not treated in this way, the natural fluorescent properties of the neurotransmitters may contribute to the whole cellular signal. However, this study also confirmed the presence of true GFP-derived fluorescence, based on the similarities between the protein and mRNA signals detected in living and fixed cells.

5.4.6 Possible interaction of position-governing genes and differential GFP-like protein expression along the anteroposterior axis in *Acropora* embryos

Despite the great similarity observed in the phylogenetic comparison (Chapter 4), green and red fluorescent proteins were differentially expressed along the planula body (Figure 5.3, panels J and K). *In situ amilFP* detection was insufficient to resolve the mouth- or aboral end-specific expression; instead, both patterns were observed at the same time regardless of the probe used (GFP or RFP). However, detection of both showed that the signal was distributed at the extremities of the principal axis. Using this spatial information and previously published work related to the *in situ* detection of important developmental genes in other cnidarian planulae, it is possible that the GFP-like genes are being regulated by such genes.

For example, expression of *Snail*-Am in early gastrulation (Hayward et al. 2004) resembles both GFP-like expression patterns, in the sense that both appeared in the proliferating tissue that subsequently formed the endoderm. In other organisms, *Snail*

plays important roles in mesoderm specification during embryogenesis (Technau & Scholz 2003). In *Nematostella vectensis* and *Hydra* embryogenesis, *Wnt* and *Brachyury* expression patterns show some similarities with *Acropora amilFP* expression, specifically in the oral pore domain (Technau 2001; Kusserow et al. 2005). These transcription factors are involved in mesoderm specification and the blastoporal signalling centre (Technau & Scholz 2003). Hydrozoan *Hox*-like gene expression overlaps with aboral RFP expression in *Acropora* (Masuda-Nakagawa et al. 2000). *Hox* genes are primarily responsible for proper anteroposterior patterning, at least in bilateral animals (Finnerty & Martindale 1999; Finnerty et al. 2003, 2004; Wellik & Capecchi 2003). Additionally, cell-specific staining for the *Emx-Am* and *Pax C* genes showed a neuron-like ectodermal population that clearly resembled the cells identified when staining for GFP-like genes. Both genes play important roles in early neural tissue specification, as well as in eye morphogenesis (Miller et al. 2000; Ball et al. 2002; de Jong et al. 2006).

Considering the correlative spatial expression of key regulatory genes (see above) in a variety of cnidarians and the expression of GFP-like genes in *Acropora*, it may be that despite the apparently ambiguous physiological properties of these widespread genes/proteins (at least in the Scleractinia), their early and localised expression may indicate their involvement in tissue/cell characterisation, a potential role in chemo- and photoreception (in the case of aboral expression; Pang et al. 2004; Chia & Koss 2005), and perhaps also symbiosis-related functions.

5.4.7 Cell surface array

The endodermal expression of both GFP-like genes made it impossible to observe ectodermal cell surface expression (*e.g.*, Figure 5.2, panel E). Fortunately, using the natural autofluorescence induced by the fixation process, this endodermal signal disappeared when the organisms were observed under the GFP2 filter, allowing the observation of the spatial pattern formed by ectodermal GFP-expressing cells (Figure 5.2, panel D). This aggregation characteristic has also been observed in the *Hydra* body column, using genes like *Hyzic* or *Nowa,* which are apparently expressed in proliferating stages by neuronal cell types, the nematocytes (Engel et al. 2002; Lindgens et al. 2004). In *Acropora,* however, GFP-like expression was not detected in that cell lineage, but was apparent in another neuronal cell type.

5.5 Conclusions

In *Acropora* embryogenesis, the GFP-like genes are maximally up-regulated during the planula stage, prior to metamorphosis. *In situ* and live fluorescent microscopy confirmed the expression of red and green fluorescent proteins at the aboral and oral ends of the planula, respectively. *AmilCP* was primarily distributed in the endodermal cavity.

In the ectoderm, two neuron-like cell types and one glandular-type cell showed associated mRNA and fluorescence signals. In the endoderm, GFP-expressing cells did not appear to be yolk cells, but rather may represent an as yet undescribed cell type.

In vitro assays of the temporal regulation of genes showed that *amilCP* was expressed before the *amilFP* gene during development.

CHAPTER 6

GENERAL DISCUSSION

6.1 Differential exon evolution, functionality, origins and regulators of the GFP-like genes.

During the process of sequencing the two GFP-like genes, the significance of protein structure for understanding the evolution of the corresponding genes became clear. The *Acropora* loci encoding both the GFP-like chromoproteins and their fluorescent counterparts shared a common intron/exon structure. When the Acropora's GFP-like structure was compared with other nuclear homologues across cnidarians and bilaterians, the first two splicing cites do match extraordinarily well along the ORF in all organisms. In view of these similarities, it would be interesting to explore whether regulatory elements upstream and in introns are also conserved. The ability of putative regulatory elements to drive expression of reporter genes could be explored in well characterised mammalian systems, or in the emerging *Hydra* transgenic system (Bosch et al. 2002, Wittlieb et al. 2006, Khalturin et al. 2007).

To better understand the evolution of the coding region, amino acid sequences of GFP homologs were aligned and used to infer the phylogeny of the *Acropora* GFP-like proteins. The general topology of the resulting trees did not differ significantly from published trees, however, the most significant novel finding was that the larval proteins are largely resolved from their adult counterparts. This unexpected finding suggests that the adult and larval GFP-like genes may have different functions. It would be interesting to test this idea by gene silencing technology, which has been successfully applied in symbiotic anemones (Dunn et al. 2007), promising future applicability for coral embryos.

Finally, because the GFP-like proteins in general are capable of forming its chromophore, and some of them may be capable of light induced proton-pumping, it is possible to think of these proteins as enzyme like molecules. In terms of functional categorization, it was interesting to see that the complete protein sequence, and fragments (exons) hit the category of lipid and energy metabolism. However, the presence of several substitutions facing the outside of the B-barrel structure suggest

functions in protein-protein interaction, which could perhaps be explored using the yeast two-hybrid assay technology.

6.2 Protein world

It was fascinating to observe how bacterial colonies turn blue, or red after being transformed with expression constructs containing the coral GFP-like genes. However, the “cut and paste” did not always work exactly as expected. On occasion, the expressed proteins (particularly those in the RFP class) appeared to be toxic to the bacterial cell, as evidenced by poor growth and tendency to delete the plasmid under selective conditions. There are at least two possible explanations for this toxic behaviour. The first is that the self-aggregation and extraordinary resistant nature of the protein could lead to its physical disruption of sensitive structures inside the cell. The second potential mechanism of toxicity is more subtle, and is related to the secondary properties of the fluorescence, as has been documented in the case of killerRed (Bulina et al. 2006), which is derived from a hydrozoan cnidarian, highly reactive oxygen species may be generated as secondary products upon illumination. Further biophysical characterization of the *Acropora* RFP (amilRFP) promise interesting insights into biological functions of these proteins and their physical properties.

The AmilGFP517a (clone 114 c8m4) on the other hand, appeared to be only poorly soluble, but its spectral characteristics were almost indistinguishable (although slightly lower) from those of its paralog Amil517b (clone c012-c9). Similar low solubility phenomena were observed in the case of one chromoprotein. However by adding Triton X100 to 2%, the protein was fully solubilized. This behaviour was attributed to electrostatic changes on the surface of the protein. Whether such solubility behaviour applies in the coral cell remains unclear, however the majority of observed substitutions fall outside the barrel indicating that mutations in this latter protein domain may be strongly selected against.

The process of the acquisition of fluorescence by GFPs belonging to the non-fluorescent chromoprotein class has previously been reported (Gurskaya et al. 2001, Chudakov et al. 2003, Martynov et al. 2003), and occurs under strong light excitation. In the case of the *Acropora* chromoproteins, however, fluorescence is not acquired through light induction, rather dehydration causes the appearance of fluorescence.

This water-quenched fluorescence represents an important biotechnological characteristic of these proteins. Further studies on the crystal structure may reveal the mechanisms behind the conversion of non-fluorescent protein to the fluorescent state.

Finally, the spectral overlap between the GFP-like proteins, suggests that FRET may occur in living planulae, as has previously been demonstrated in adult coral tissue. (Salih et al. 2004, Cox & Salih 2005). The biological significance of these energy transfer processes is as yet unclear and, in the case of the planula may be restricted to the endoderm, since the GFP-like proteins have largely discrete expression domains in the ectoderm.

6.3 Too much attention to the core

Most structural studies on GFPs have focussed on the chromophore and the amino acids involved in its maturation (Chalfie et al. 1994, Cubitt et al. 1995, Gross et al. 2000, Gurskaya et al. 2001, Bulina et al. 2002, Campbell et al. 2002, Barondeau et al. 2003, Chudakov et al. 2003, Andresen et al. 2005). However, in terms of overall structure, the GFP proteins closely resemble the Nidogen (G2F) proteins. Whereas GFPs are classified as bioluminescent proteins in gene ontology, G2F proteins are categorized as constituents of the basal membranes (Ekblom et al. 1994, Kadoya et al. 1997). The significance of the overall structural similarity has not yet been explored, but lead to questions about the possible function of the ancestor of these very different proteins. The question is not easy to address; no GFP-like proteins have been identified yet in sponges or other “lower” animal phyla. The priority for future work should be to determine the expression pattern of the *Acropora* Nidogen-like gene (contig 1301). It is hypothesized that the Nidogen *in situ* pattern should overlap with those of the GFP-like transcripts, the logic being that it may be more difficult for an organism to create a new cell type which express the “derived” gene, than retain both messages in the same cell.

6.4 Environmental influences on early embryogenesis: does light controls GFP expression?

Light constitutes a form of energy that may be transduced by living organisms in a series of biological processes. The most obvious example is photosynthesis, the fixation of carbon dioxide driven by light energy. Another example is the circadian

clock, which is responsible for ensuring the correct metabolic state of organisms during active (growing, hunting) or inactive (sleeping) periods in terms of production/suppression of hormones, proteins and other secondary metabolites (McGinnis et al. 1994, Delaunay et al. 2000, Krishnan et al. 2001, Delaunay & Laudet 2002, DeBruyne et al. 2004, Kaneko & Cahill 2005). Light is also responsible of the evolution of complex mechanisms of photodetection and transduction in neural impulses, commonly called vision.

Each of these complex biological processes (photosynthesis, circadian clocks and vision) has two common characteristics, the first of which is that all are based on the absorption of light by a “chromophore” followed by processes of excitation and decoupling of the electrons excited or captured by this chromophore, generating an electrical potential that can then be used to produce “work”. The second characteristic common to each of these processes is the constancy of the input over evolutionary time, allowing organisms to adapt to the input.

Considering the evolutionary age and reactivity of the GFP chromophore, these proteins belong to a family with ancient origins, dating from before the split between the cnidarians and the bilaterians. A high proportion of the organisms which contain GFP-like proteins are regularly exposed to sunlight, as presumably were their evolutionary ancestors. On this basis, it is reasonable to postulate that some of the variation seen in coral GFP-like proteins reflects selective pressures for specific roles in light attenuation. However, this scenario may well not apply in the case of bioluminescent cnidarians, where different selective pressures may apply. If sunlight shapes in some degree the evolution of GFP-like proteins, then one might expect light to directly or indirectly regulate the expression of GFP-like genes. The AvicGFP for example, is capable of pumping protons upon light radiation, these protons could be sensed by other neighbour proteins and a network of activation or repression of GFP mRNA could be achieved. However, experimental support for this idea is lacking. In preliminary experiments, coral embryos raised in darkness showed relatively normal levels and distribution of mRNAs encoding GFP-like proteins (data not shown). Nevertheless, the notion of a link between light exposure and expression of GFP-like genes merits further exploration.

6.5 Future directions

Fluorescence Activated Cell Sorting could be used to isolate GFP/RFP expressing cells, allowing microarray profiling of the cell types. Electron microscopy suggest that treatment with a cocktail of digestive enzymes might be a more appropriate strategy to disaggregate larvae rather than mechanical disruption (Gates & Muscatine 1992). Cells released in this way could be cell sorted based on fluorescence characteristics, and then either be used directly in array experiments or used to found cell cultures.

Another important approach is that of examining the phenotypic consequences of knocking down the transcription of specific GFP-like genes during development. Although methods for knocking down coral genes are not yet available, morpholino oligonucleotide technology has been successfully applied to developmental genes in the case of the sea anemone *Nematostella vectensis* (Magie et al. 2007), and the hydrozoan *Clytia hemisphaerica* and *Eleutheria dichotoma* (Jakob & Schierwater 2007, Momose et al. 2008). Also RNA interference downregulation/knockdown has been applied successfully in the symbiotic anemone *Aiptasia pallida* (Dunn et al. 2007). These precedents offer great promise in terms of functional analyses of coral genes in general, and GFP genes specifically.

REFERENCES

Agmon N (2005) Proton pathways in green fluorescent protein. Biophys J 88:2452-2461

Alieva NO, Konsen KA, Field SF, A. ME, Hunt ME, Beltran-Ramirez VH, Miller DJ, Salih A, Matz M (2008) Diversity and evolution of coral fluorescent proteins. PLoS one 3:1-12

Anctil M, Hayward DC, Miller DJ, Ball EE (2007) Sequence and expression of four coral G protein-coupled receptors distinct from all classifiable members of the rhodopsin family. Gene 392:14-21

Ando R, Hama H, Yamamoto-Hino M, Mizuno H, Miyawaki A (2002) An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. PNAS 99:12651-12656

Ando R, Mizuno H, Miyawaki A (2004) Regulated fast nucleoplasmic shuttling observed by reversible protein highlighting. Science 306:1370-1373

Andresen M, Wahl MC, Stiel AC, Grater F, Schafer LV, Trowitzsch S, Weber G, Eggeling C, Grubmuller H, Hell SW, Jakobs S (2005) Structure and mechanism of the reversible photoswitch of a fluorescent protein. Proc Natl Acad Sci U S A 102:13070-13074

Anke HE, Kerkx M, Boer Ed, Loon AEv (2001) Spatio-temporal expression of a gene encoding a putative RNA-binding protein during the early larval development of the mollusc *Patella vulgata*. Dev Genes Evol 211:423-427

Atoda K (1951) The larva and postlarval development of the reef-building corals. Journal of Morphology 89:1-15

Ausubel FM, Brent R, Kingston RE, Moore DD, Siedman JG, Smith J, Struhl K (1996) Current protocols in molecular biology. New York, USA: John Wiley & Sons, Inc

Baird AH, Salih A, Trevor-Jones A (2006) Fluorescence census techniques for the early detection of coral recruits. Coral Reefs 25:73-76

Baird GS, Zacharias DA, Tsien RY (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. PNAS 97:11984-11989

Ball EE, Hayward DC, Catmull J, Reece-Hoyes JS, Hislop NR, Harrison PL, Miller DJ (2001) Molecular control of development in the reef coral *Acropora millepora*. Proceedings 9th Intl Coral Reef Symposium:1-7

Ball EE, Hayward DC, Reece-Hoyes JS, Hislop NR, Samuel G, Saint R, Harrison P, Miller DJ (2002) Coral development: from clasical embryology to the molecular control. Int J Dev Biol 46:671-678

Ball EE, Hayward DC, Saint R, Miller DJ (2004) A simple plan- cnidarians and the origins of developmental mechanisms. Nature Reviews Genetics 5:567-577

Barondeau DP, Putnam CD, Kassmann CJ, Tainer JA, Getzoff ED (2003) Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. PNAS 100:12111-12116

Battenberg EL, Bloom FE (1975) A rapid, simple and more sensitive method for the demonstration of central catecholamine-containing neurons and axons by glyoxylic acid induced fluorescence: I. Specificity. Psychopharmacol Commun 1:3-13

Baumann D, Cook M, Ma L, Mushegian A, Sanders E, Schwartz J, Ron Yu C (2008) A family of GFP-like proteins with different spectral properties in lancelet *Brachiostoma floridae*. Biology Direct 3:1-12

Beltran-Ramirez VH, Tomljenovic L, Hislop N, Hayward D, Ball EE, Matz MV, Miller DJ (2010) Differential expression of GFP-related proteins during early coral development. Biology Direct *in prep*

Benton W, Davis R (1977) Screening lambda-gt recombinant clones by hybridisation to single plaques *in situ*. Science 196:180-182

Bosch TCG, Augustin R, Gellner K, Khalturin K, Lohmann JU (2002) In vivo electroporation for genetic manipulations of whole Hydra polyps. Differentiation 70:140-147

Bradford M (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein dye binding. Analytical Biochemistry 74:248-254

Bridge D, Cunningham CW, DeSalle R, Buss LW (1995) Class-level relationships in the phylum Cnidaria: molecular and morphological evidence. Mol Biol Evol 12:679-689

Bridge D, Cunningham CW, Schierwater B, DeSalle R, Buss LW (1992) Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. Proc Natl Acad Sci U S A 89:8750-8753

Bulina ME, Chudakov DM, Britanova OV, Yanushevich YG, Staroverov DB, Chepurnykh TV, Merzlyak EM, Shkrob MA, Lukyanov S, Lukyanov KA (2006) A genetically encoded photosensitizer. Nat Biotechnol 24:95-99

Bulina ME, Chudakov DM, Mudrik NN, Lukyanov KA (2002) Interconversion of Anthozoa GFP-like fluorescent and nonfluorescent proteins by mutagenesis. BMC Biochemistry 3:1-8

Butler WL, Hopkins DW (1970) An analyses of fourth derivative spectra. Photochem Photobiol 12:451-456

Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY (2002) A monomeric red fluorescent protein. PNAS 99:7877-7882

Carter RW, Schmale MC, Gibbs PD (2004) Cloning of anthozoan fluorescent protein genes. Comp Biochem Physiol C Toxicol Pharmacol 138:259-270

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. Science 263:802-805

Chan MCY, Karasawa S, Mizuno H, Bosanac I, Ho D, Prive BG, Miyawaki A, Ikura M (2006) Structural characterization of a blue chromoprotein and its yellow mutant from the sea anemone *Cnidiopus japonicus*. JBC http://www.jbc.org/cgi/doi/10.1074/jbc.M606921200

Chen JY, Oliveri P, Gao F, Dornbos SQ, Li CW, Bottjer DJ, Davidson EH (2002) Precambrian animal life: probable developmental and adult cnidarian forms from Southwest China. Dev Biol 248:182-196

Chia F-S, Koss R (2005) Fine structural studies of the nervous system and the apical organ in the planula larva of the sea anemone *Anthopleura elegantissima*. Journal of Morphology 160:275-297

Chudakov DM, Belousov VV, Zaraisky AG, Novoselov VV, Staroverov DB, Zorov DB, Lukyanov S, Lukyanov KA (2003) Kindling fluorescent proteins for precise in vivo photolabeling. Nat Biotechnol 21:191-194

Collins AG (2002) Phylogeny of medusozoa and the evolution of cnidarian life cycles. J Evol Biol 15:418-432

Collins AG, Schuchert P, Marques AC, Jankowski T, Medina M, Schierwater B (2006) Medusozoan phylogeny and character evolution clarified by new large and small subunit rDNA data and an assessment of the utility of phylogenetic mixture models. Syst Biol 55:97-115

Cormier MJ, Totter JR (1964) Bioluminescence. Annu Rev Biochem 34:431-458

Cox G, Matz M, Salih A (2007) Fluorescence lifetime imaging of coral fluorescent proteins. Microsc Res Tech 70:243-251

Cox G, Salih A (2005) Fluorescence lifetime imaging of symbionts and fluorescent proteins in reef corals. Proceedings of SPIE 5700

Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY (1995) Understanding, improving and using green fluorescent proteins. Trends Biochem Sci 20:448-455

Daly M, Fautin DG, Cappola VA (2003) Systematics of the Hexacoralia (Cnidaria:Anthozoa). Zoological Journal of the Linnean Society:419-437

Dandar-Roh AM, Rogers-Lowery CL, Zellmann E, Thomas MB (2004) Ultrastructure of the calcium-sequestering gastrodermal cell in the hydroid Hydractinia symbiolongicarpus (Cnidaria, Hydrozoa). J Morphol 260:255-270

de Jong DM, Hislop NR, Hayward DC, Reece-Hoyes JS, Pontynen PC, Ball EE, Miller DJ (2006) Components of both major axial patterning systems of the Bilateria are differentially expressed along the primary axis of a 'radiate' animal, the anthozoan cnidarian *Acropora millepora*. Dev Biol

DeBruyne J, Hurd MW, Gutierrez L, Kaneko M, Tan Y, Wells DE, Cahill GM (2004) Isolation and phenogenetics of a novel circadian rhythm mutant in zebrafish. J Neurogenet 18:403-428

Deheyn DD, Kubokawa K, McCarthy JK, Murakami A, Porrachia M, Rouse GW, Holland ND (2007) Endogenous green fluorescent protein (GFP) in *Amphioxus*. Biol Bull 213:95-100

Delaunay F, Laudet V (2002) Circadian clock and microarrays:mammalian genome gets rhythm. Trends in Genetics

Delaunay F, Thisse C, Marchand O, Laudet V, Thisse B (2000) An Inhereted functional circadian clock in zebrafish embryos. Science 289:297-300

Dove SG, Hoegh-Guldberg O, Ranganathan R (2001) Major colour patterns of reef-building corals are due to a family of GFP-like proteins. Coral Reefs 19:197-204

Dove SG, Takabayachi M, Hoegh-Guldberg O (1995) Isolation and partial characterization of the pink and blue pigments of Pocilloporid and Acroporid corals. Biol Bull 189:288-297

Dunn SR, Phillips W, Douglas R, Weis V (2007) Knockdown of actin and caspase gene expression by RNA interference in the symbiotic anemone *Aiptasia pallida.* Biol Bull 212:250-258

Ekblom P, Ekblom M, Fecker L, Klein G, Zhang HY, Kadoya Y, Chu ML, Mayer U, Timpl R (1994) Role of mesenchymal nidogen for epithelial morphogenesis in vitro. Development 120:2003-2014

Engel U, Ozbek S, Streitwolf-Engel R, Petri B, Lottspeich F, Holstein TW (2002) Nowa, a novel protein with minicollagen Cys-rich domains, is involved in nematocyst formation in Hydra. J Cell Sci 115:3923-3934

Fingerman M (1965) Chromatophores. in: The control of Chromatophores, Pergamon press 119:296-339

Finnerty JR, Martindale MQ (1999) Ancient origins of axial patterning genes: Hox genes and *ParaHox* genes in the Cnidaria. Evol Dev 1:16-23

Finnerty JR, Pang K, Burton P, Paulson D, Martindale MQ (2004) Origins of bilateral symmetry: *Hox* and *dpp* expression in a sea anemone. Science 304:1335-1337

Finnerty JR, Paulson D, Burton P, Pang K, Martindale MQ (2003) Early evolution of a homeobox gene: the parahox gene Gsx in the Cnidaria and the Bilateria. Evol Dev 5:331-345

Follenius-Wund A, Bourotte M, Schmitt M, Iyice F, Lami H, Bourguignon JJ, Haiech J, Pigault C (2003) Fluorescent derivatives of the GFP chromophore give a new insight into the GFP fluorescence process. Biophys J 85:1839-1850

Fradkov AF, Chen Y, Ding L, Barsova EV, Matz MV, Lukyanova SA (2000) Novel fluorescent protein from Discosoma coral and its mutants possesses a unique far-red fluorescence. FEBS letters 479:127-130

Fux E, Mazel C (1999) Unmixing coral fluorescence emission spectra and predicting new spectra under different excitation conditions. Applied Optics 38:466-494

Gates RD, Muscatine L (1992) Three methods for isolating viable anthozoan endoderm cells with their itracellular symbiotic dinoflagellates. Coral Reefs:143-145

Graham GJ (1995) Tandem genes and clustered genes. J Theor Biol 175:71-87

Grasso LC, Hayward D, Trueman JW, Hardie KM, A. JP, Ball EE (2001) The evolution of nuclear receptors: evidence from the coral Acropora. Mol Phylogenet Evol:93-102

Gross LA, Baird GS, Hoffman RC, Baldridge KK, Tsien RY (2000) The structure of the chromophore within DsRed, a red fluorescent protein from coral. PNAS 97:11990-11995

Gurskaya NG, Fradkov AF, Pounkova NI, Staroverov DB, Bulina ME, Yanushevich YG, Labas YA, Lukyanov S, Lukyanov KA (2003) A colourless green fluorescent protein homologue from the non-fluorescent hydromedusa *Aequorea coerulescens* and its fluorescent mutants. Biochem J 373:403-408

Gurskaya NG, Fradkov AF, Terskikh A, Matz MV, Labas YA, Martynov VI, Yanushevich YG, Lukyanov KA, Lukyanov SA (2001) GFP-like chromoproteins as a source of far-red fluorescent proteins. FEBS Lett 507:16-20

Gurskaya NG, Savitsky AP, Yanushevich YG, Lukyanov SA, Lukyanov KA (2000) Color transitions in coral's fluorescent proteins by site-directed mutagenesis. BMC Biochemistry 2:1471-2091

Gurskaya NG, Verkhusha VV, Sheglov AS, Staroverov DB, Chepurnykh TV, Fradkov AF, Lulyanov S, Lukyanov KA (2006) Engineering of a monomeric green-to-red photoactivatable fluorescent protein by blue light. Nature Biotechnology advance online publication:1-5

Hall VR, Hughes TP (1996) Reproductive strategies of modular organisms: comparative studies of reef building corals. Ecology 77:950-963

Hayward DC, Catmull J, Reece-Hoyes JS, Berghammer H, Dodd h, Hann SJ, Miller DJ, Ball EE (2001) Gene structure an larval expression of cnox-2Am from the coral *Acropora millepora*. Dev Genes Evol 211:10-19

Hayward DC, Miller DJ, Ball EE (2004) snail expression during embryonic development of the coral Acropora: blurring the diploblast/triploblast divide? Dev Genes Evol 214:257-260

Hayward DC, Samuel G, Pontynen PC, Catmull J, Saint R, Miller DJ, Ball EE (2002) Localized expression of a dpp/BMP2/ ortholog in a coral embryo. PNAS 99:8106-8111

Hislop NR (2003) Characterisation of anterior patterning genes in the staghorn coral, *Acropora millepora* (Cnidaria; Anthozoa; Scleractinia). PhD Thesis, James Cook University:199pp

Hislop NR, de Jong D, Hayward DC, Ball EE, Miller DJ (2005) Tandem organization of independently duplicated homeobox genes in the basal cnidarian *Acropora millepora*. Dev Genes Evol 215:268-273

Hollingsworth LL, III RAK, Lewis TD, Krupp DA, Leong J-AC (2005) Phototaxis of motile zooxanthellae to green light may facilitate symbiont capture by coral larvae. Coral Reefs 24:523

Hollinsworth LL, Lewis TD, Krupp DA, Leong J-AC (2006) Early onset and expression of fluorescent proteins in the larvae of the mushroom coral *Fungia scutaria* Lamark 1801. Proc of 10th int Coral Reef Symp 1:99-105

Hopf M, Gšhring W, Ries A, Timpl R, Hohenester E (2001) Cristal structure and mutation analyses of a perlecan-binding fragment of nidogen-1. Nature structural biology 8:634-640

Horiguchi T, Kawai H, Kubota M, Takahashi T, Watanabe M (1999) Phototactic responses of four marine dinoflagellates with different types of eyespot and chloroplast. Phycologycal Research 47:101-107

Ianushevich Iu G, Bulina ME, Gurskaia NG, Savitskii AP, Lukianov KA (2002) Key amino acid residues responsible for the color of green and yellow fluorescent proteins from the coral polyp Zoanthus. Bioorg Khim 28:303-307

Iwama H, Gojobori T (2004) Highly conserved upstream sequences for transcription factor genes and implications for the regulatory network. Proc Natl Acad Sci U S A 101:17156-17161

Jakob W, Schierwater B (2007) Changing hydrozoan bauplans by silencing Hox-like genes. PLoS ONE 2:e694

Janknecht R, De Martynoff G, Lou J, Hipskind R, Nordheim A, Stunnenberg H (1991) Rapid and efficient purification of native histidene-tagged protein expressed by recombinent vaccinia virus. PNAS 88:8972-8976

Kadoya Y, Salmivirta K, Talts JF, Kadoya K, Mayer U, Timpl R, Ekblom P (1997) Importance of nidogen binding to laminin gamma1 for branching epithelial morphogenesis of the submandibular gland. Development 124:683-691

Kamm K, Schierwater B, Jakob W, Dellaporta SL, Miller DJ (2006) Axial patterning and diversification in the Cnidaria predate the Hox system. Curr Biol 16:920-926

Kaneko M, Cahill GM (2005) Light-dependent development of circadian gene expression in transgenic zebrafish. PLoS Biol 3:e34

Kao HT, Sturgis S, Desalle R, Tsai J, Davis D, Gruber DF, Pieribone VA (2007) Dynamic Regulation of Fluorescent Proteins from a Single Species of Coral. Mar Biotechnol (NY)

Karasawa S, Araki T, Nagai T, Mizuno H, Miyawaki A (2004) Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer. Biochem J 381:307-312

Karasawa S, Araki T, Yamamoto-Hino M, Miyawaki A (2003) A green-emitting fluorescent protein from Galaxeidae coral and Its monomeric version for use in fluorescent labelling. The Journal of Biological Chemistry 278:34167-34171

Katayama H, Yamamoto A, Mizushima N, Yoshimori T, Miyawaki A (2008) GFP-like proteins stably accumulate in lysosomes. Cell Struct Funct 33:1-12

Kawaguti S (1969) Effect of the green fluorescent pigment on the productivity of the reef corals. Micronesica 5:313

Kelmanson IV, Matz MV (2003) Molecular basis and evolutionary origins of color diversity in great star coral *Montastraea cavernosa* (Scleractinia: Faviida). Mol Biol Evol 20:1125-1133

Khalturin K, Anton-Erxleben F, Milde S, Plotz C, Wittlieb J, Hemmrich G, Bosch TCG (2007) Transgenic stem cells in Hydra reveal an early evolutionary origin for key elements controlling self-renewal and differentiation. Developmental Biology 309:32-44

Kim S, Wadsworth WG (2000) Positioning of longitudinal nerves in *C. elegans* by nidogen. Science 288:150-154

Kogure T, Karasawa S, Araki T, Saito K, Kinjo M, Miyawaki A (2006) A fluorescent variant of a protein from the stony coral *Montipora* facilitates dual-color single-laser fluorescence cross-correlation spectroscopy. Nat Biotechnol 24:577-581

Kolberg KJ, Martin VJ (1988) Morphological, cytochemical and neuropharmacological evidence for the presence of catecholamines in hydrozoan planulae. Development 103:249-258

Krishnan B, Levine JD, Kathlea M, Lynch S, Dowse HB, Funes P, Hall JC, Hardin PE, Dryer SE (2001) A new role for cryptochrome in a *Drosophila* circadian oscillator. Nature 411

Kummer A (2000) Fluorescence kinetics on the green fluorescent protein (GFP). PhD Thesis, Technische Universitat Munchen:52pp

Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, Technau U, von Haeseler A, Hobmayer B, Martindale MQ, Holstein TW (2005) Unexpected complexity of the Wnt gene family in a sea anemone. Nature 433:156-160

Labas YA, Gurskaya NG, Yanushevich YG, Fradkov AF, Lukyanov KA, Lukyanov SA, Matz MV (2002) Diversity and evolution of the green fluorescent protein family. PNAS 99:4256-4261

Laemmli U (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685

Lindgens D, Holstein TW, Technau U (2004) Hyzic, the Hydra homolog of the zic/odd-paired gene, is involved in the early specification of the sensory nematocytes. Development 131:191-201

Lockett TJ (1990) A bacteriophage lambda DNA purification procedure suitable for the analyses of DNA from either large or multiple small lysates. Analytical Biochemistry 185:230-234

Lukyanov KA, Chudakov DM, Lukyanov S, Verkhusha VV (2005) Innovation: Photoactivatable fluorescent proteins. Nat Rev Mol Cell Biol 6:885-891

Lukyanov KA, Fradkov AF, Gurskaya NG, Matz MV, Labas YA, Savitskyi AP, Markelov ML, Zaraisky AG, Zhao X, Fang Y, Tan W, Lukyanov SA (2000) Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. The journal of biological chemistry 275:25879–25882

Luo WX, Cheng T, Guan BQ, Li SW, Miao J, Zhang J, Xia NS (2006) Variants of Green Fluorescent Protein GFPxm. Mar Biotechnol (NY)

Mabuchi Michiaki, Shimada J-i, Okamoto K, Kawakami Y, Fujita S, Matsushige K (2001) Time-resolved fluorescence spectroscopy of dopamine in the single cells. Proceedings of BiOS 4252:1-9

Magie CR, Daly M, Martindale MQ (2007) Gastrulation in the cnidarian *Nematostella vectensis* occurs via invagination not ingression. Developmental Biology 305:483-497

Martin VJ (1992) Characterization of a RFamide-positive subset of ganglionic cells in the hydrozoan planular nerve net. Cell Tissue Res 269:431-438

Martindale MQ, Pang K, Finnerty JR (2004) Investigating the origins of triploblasty: " mesodermal" gene expression in diploblastic animal, the sea anemone *Nematostella vectensis* (phylum Cnidaria, Class Anthozoa). Development:2463-2474

Martynov VI, Maksimov BI, Martynova NY, Pakhomov AA, Gurskaya NG, Lukyanov SA (2003) A purple-blue chromoprotein from *Goniopora tenuidens* belongs to the DsRed subfamily of GFP-like proteins. The Journal of Biological Chemistry 278:46288-46292

Masuda H, Takenaka Y, Yamaguchi A, Nishikawa S, Mizuno H (2006) A novel yellowish-green fluorescent protein from the marine copepod, *Chiridius poppei*, and its use as a reporter protein in HeLa cells. Gene 372:18-25

Masuda-Nakagawa LM, Groger H, Aerne BL, Schmid V (2000) The HOX-like gene Cnox2-Pc is expressed at the anterior region in all life cycle stages of the jellyfish Podocoryne carnea. Development Genes and Evolution 210:151-156

Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML, Lukyanov SA (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. Nature Biotechnology 17:969-973

Matz MV, Lukyanov KA, Lukyanov SA (2002) Family of the green fluorescent protein: journey to the end of the rainbow. BioEssays 24:953-959

Mazel CH, Fuchs E (2003) Contribution of fluorescence to the spectral signature and perceived color of corals. Limnol Oceanogr 48:390-401

Mazel CH, Lesser MP, Gorbunov MY, Barry TM, Farrell JH, Wyman KD, Falkowski PG (2003) Green-fluorescent proteins in Caribbean corals. Limnol Oceanogr 48:402-411

McGinnis JF, Austin BJ, Stepanik PL, Lerious V (1994) Light-dependent regulation of the transcriptional activity of the mammalian gene for arrestin. J Neurosci Res 38:479-482

Medina M, Collins AG, Takaoka TL, Kuehl JV, Boore JL (2006) Naked corals: skeleton loss in Scleractinia. Proc Natl Acad Sci U S A 103:9096-9100

Miller DJ, Ball EE (2000) The coral *Acropora*: what it can contribute to our knowledge of metazoan evolution and the evolution of developmental processes. Bioessays 22:291-296

Miller DJ, Hayward DC, Reece-Hoyes JS, Scholten I, Catmull J, Gehring WJ, Callaerts P, Larsen JE, Ball EE (2000) Pax gene diversity in the basal cnidarian *Acropora millepora* (Cnidaria, Anthozoa): implications for the evolution of the Pax gene family. Proc Natl Acad Sci U S A 97:4475-4480

Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, Agata K, Bosch TCG (2007) The innate immune repertoire in Cnidaria - ancestral complexity and stochastic gene loss. Genome Biology 8:45-57

Miyawaki A (2004) Fluorescent proteins in a new light. Nat Biotechnol 22:1374-1376

Mocz G (2007) Fluorescent Proteins and Their Use in Marine Biosciences, Biotechnology, and Proteomics. Mar Biotechnol (NY)

Momose T, Derelle R, Houliston E (2008) A maternally localised Wnt ligand required for axial patterning in the cnidarian *Clytia hemisphaerica*. Development 135:2105-2113

Morin JG (1974) Coelenterate bioluminescence. In: Coelenteare Biology - Reviews and new perspectives Ed Leonard Muscatine and Howard M Lenhoff, Academic Press London:397-438

Niwa H, Inouye S, Hirano T, Matsuno T, Kojima S, Kubota M, Ohashi M, Tsuji FI (1996) Chemical nature of the light emitter of the *Aequorea* green fluorescent protein. Proc Natl Acad Sci U S A 93:13617-13622

Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein. Science 273:1392-1395

Oswald F, Schmitt F, Leutenegger A, Ivanchenko S, D'Angelo C, Salih A, Maslakova S, Bulina M, Schirmbeck R, Nienhaus GU, Matz MV, Wiedenmann J (2007) Contributions of host and symbiont pigments to the coloration of reef corals. Febs J 274:1102-1109

Pakhomov AA, Martynova NY, Gurskaya NG, Balashova TA, Martynov VI (2004) Photoconversion of the chromophore of a fluorescent protein from Dendronephthya sp. Biochemistry (Mosc) 69:901-908

Pang K, Matus DQ, Martindale MQ (2004) The ancestral role of COE genes may have been in chemoreception: evidence from the development of the sea anemone, Nematostella vectensis (Phylum Cnidaria; Class Anthozoa). Dev Genes Evol 214:134-138

Prasher DC (1995) Using GFP to see the light. TIG 11:320-323

Prasher DC, Eckenrode VK, Wardc WW, Prendergast FG, Cormier MJ (1992) Primary structure of the *Aequorea victoria* green-fluorescent protein. Gene:229-233

Prendergast FG (2000) Bioluminescence illuminated. Nature 405:291-293

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317:86-94

Remington SJ, Wachter RM, Yarbrough DK, Branchaud B, Anderson DC, Kallio K, Lukyanov KA (2005) zFP538, a yellow-fluorescent protein from *Zoanthus*, contains a novel three-ring chromophore. Biochemistry 44:202-212

Reynolds CP, German DC, Weinberg AG, Smith RG (1981) Catecholamine fluorescence and tissue culture morphology. Techniques in the diagnosis of neuroblastoma. Am J Clin Pathol 75:275-282

Rinkevich B, Loya Y (1979) Reproduction of the Red-Sea Coral *Stylophora pistillata*.1. Gonads and Planulae. Marine Ecology-Progress Series 1:133-144

Rinkevich B, Loya Y (1985) Intraspecific competition in a reef coral: effects on growth and reproduction. Oecologia 66:100-105

Rogozin IB, Sverdlov AV, Babenko VN, Koonin EV (2005) Analyses of evolution of exon-intron structure of eukaryotic genes. Brief Bioinform 6:118-134

Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV (2003) Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. Curr Biol 13:1512-1517

Romano SL, Palumbi SR (1997) Molecular evolution of a portion of the mitochondrial 16S ribosomal gene region in scleractinian corals. J Mol Evol 45:397-411

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science:487-491

Salih A, Hoegh-Guldberg O, Cox G (1998) Photoprotection of symbiotic dinoflagellates by fluorescent pigments in reef corals. Proc Aust Coral Reef Society 75 Anniv Conference, ed Greenwood & Hall:217-230

Salih A, Larkum A, Cox G, Kuhl M, Hoegh-Guldberg O (2000) Fluorescent pigments in corals are photoprotective. Nature 408:850-853

Salih A, Larkumb A, Croninc T, Wiedenmannd J, Szymczake R, Cox G (2004) Biological properties of coral GFP-type proteins provide clues for engineering novel optical probes and biosensors. Proceedings of SPIE 5329:61-72

Sambrook J, Russell D (2001) Molecular Cloning: A laboratory Manual. Cold Spring Harbor Laboratory Press

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. PNAS 74:5463-5467

Seipp S, Schmich J, Leitz T (2001) Apoptosis--a death-inducing mechanism tightly linked with morphogenesis in *Hydractina echinata* (Cnidaria, Hydrozoa). Development 128:4891-4898

Shagin DA, Barsova EV, Yanushevich YG, Fradkov AF, Lukyanov KA, Labas YA, Semenova TN, Ugalde JA, Meyers A, Nunez JM, Widder EA, Lukyanov SA, Matz MV (2004) GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. Mol Biol Evol 21:841-850

Shkrob MA, Yanushevich YG, Chudakov DM, Gurskaya NG, Labas YA, Poponov SY, Midrik NN, Lukyanov S, Lukyanov KA (2005) Far-red fluorescence proteins evolved from a blue chromoprotein from *Actinia equina*. Biochem J 392:649-654

Siebert S, Thomsen S, Reimer MM, Bosch TCG (2005) Control of foot differentiation in Hydra: Phylogenetic footprinting indicates interaction of head, bud and foot patterning systems. Mechanisms of Development 122:998-1007

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503-517

Spurlock BO, Cormier MJ (1975) A fine structure study of the anthocodium in *Renilla mulleri.* Evidence for the existence of a bioluminescent organelle, the luminelle. J Cell Biol 64:15-28

Sun Y, Castner EW, Jr., Lawson CL, Falkowski PG (2004) Biophysical characterization of natural and mutant fluorescent proteins cloned from zooxanthellate corals. FEBS Lett 570:175-183

Sverdlov AV, Babenko VN, Rogozin IB, Koonin EV (2004) Preferential loss and gain of introns in 3' portions of genes suggests a reverse-transcription mechanism of intron insertion. Gene 338:85-91

Technau U (2001) Brachyury, the blastopore and the evolution of the mesoderm. Bioessays 23:788-794

Technau U, Rudd S, Maxwell P, Gordon PMK, Saina M, Grasso LC, Hayward DC, Sensen CW, Saint R, Holstein TW, Ball EE, Miller DJ (2005) Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. Trends in Genetics 21:632-639

Technau U, Scholz CB (2003) Origin and evolution of endoderm and mesoderm. Int J Dev Biol 47:531-539

Terao T, Yamashita A, Katoh S (1985) Chlorophyll b-deficient mutants of rice I. Absorption and fluorescence spectra and chlorophyll a/b ratios. Plant Cell Physiol 26:1361-1367

Terskikh A, Fradkov A, Ermakova G, Zaraisky A, Tan P, Kajava AV, Zhao X, Lukyanov S, Matz M, Kim S, Weissman I, Siebert P (2000) "Fluorescent timer": Protein that changes color with time. Science 290:1585-1588

Thomsen S, Till A, Wittlieb J, Beetz C, Khalturin K, Bosch TC (2004) Control of foot differentiation in Hydra: in vitro evidence that the NK-2 homeobox factor CnNK-2 autoregulates its own expression and uses pedibin as target gene. Mech Dev 121:195-204

Tsien RY (1998) The green fluorescent protein. Annu Rev Biochem 67:509-544

Tsutsui H, Karasawa S, Shimizu H, Nukina N, Miyawaki A (2005) Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. EMBO Rep 6:233-238

Tu H, Xiong Q, Zhen S, Zhong X, Peng L, Chen H, Jiang X, Liu W, Yang W, Wei J, Dong M, Wu W, Xu A (2003) A naturally enhanced green fluorescent protein from magnificent sea anemone (*Heteractis magnifica*) and its functional analyses. Biochem Biophys Res Commun:879-885

van Oppen MJ, Willis BL, Vugt HWJAV, Miller DJ (2000) Examination of species boundaries in the *Acropora cervicornis* group (Scleractinia, Cnidaria) using nuclear DNA sequence analyses. Molecular Ecology 9:1363-1373

van Oppen MJH, McDonald BJ, Willis B, Miller DJ (2001) The evolutionary history of the coral genus *Acropora* (Scleractinia, Cnidaria) based on a mitochondrial and a nuclear marker: reticulation, incomplete lineage sorting, or morphological convergence? Mol Biol Evol 18:1315-1329

van Oppen MJH, Willis BL, Rheede Tv, Miller DJ (2002) Spawning times, reproductive compatibilities and genetic structuring in the *Acropora aspera* group: evidence for natural hybridization and semi-permeable species boundaries in corals. Molecular Ecology 11:1363-1367

van Roessel P, Brand AH (2002) Imaging into the future: visualizing gene expression and protein interactions with fluorescent proteins. Nature Cell Biology 4:E15-20

Vandermeulen JH (1974) Studies on reef corals. II. Fine structure of planktonic planula of *Pocillopora damicornis*, with emphasis on the aboral epidermis. Marine Biology 27:239-249

Vandermeulen JH (1975) Studies on reef corals. III. Fine structural changes of calicoblast cells in *Pocillopora damicornis* settling and calcification. Marine Biology 31:69-77

Verkhusha VV, Chudakov DM, Gurskaya NG, Lukyanov S, Lukyanov KA (2004) Common pathway for the red chromophore formation in fluorescent proteins and chromoproteins. Chem Biol 11:845-854

Verkhusha VV, Lukyanov KA (2004) The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins. Nature Biotechnology 22:289-296

Wall MA, Socolich M, Ranganathan R (2000) The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. Nature structural biology 7:1133-1138

Ward WW, Cormier MJ (1975) Extraction of Renilla-type luciferin from the calcium-activated photoproteins aequorin, mnemiopsin, and berovin. Proc Natl Acad Sci U S A 72:2530-2534

Ward WW, Cormier MJ (1979) An energy transfer protein in coelenterate bioluminescence. Characterization of the *Renilla* green-fluorescent protein. J Biol Chem 254:781-788

Wellik DM, Capecchi MR (2003) Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. Science 301:363-367

Wiedenmann J (2000) The identification of new proteins homologous to GFP from *Aequorea victoria* as coloring compounds in the morphs of *Anemonia sulcata* and their biological function. PhD Thesis, Universitat Ulm, Germany:67

Wiedenmann J, Elke C, Spindler K-D, Funke W (2000) Cracks in the B-can: Fluorescent proteins from *Anemonia sulcata* (Anthozoa, Actinaria). PNAS 97:14091-14096

Wiedenmann J, Ivanchenko S, Oswald F, Nienhaus GU (2004) Identification of GFP-like proteins in nonbioluminescent, azooxanthellate anthozoa opens new perspectives for bioprospecting. Mar Biotechnol (NY) 6:270-277

Wiedenmann J, Schenk A, Rocker C, Girod A, Spindler KD, Nienhaus GU (2002) A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from *Entacmaea quadricolor* (Anthozoa, Actinaria). Proc Natl Acad Sci U S A 99:11646-11651

Wilson T, Hastings JW (1998) Bioluminescence. Annu Rev Cell Dev Biol 14:197-230

Wittlieb J, Khalturin K, Lohmann JU, Anton-Erxleben F, Bosch TCG (2006) Transgenic Hydra allow in vivo tracking of individual stem cells during morphogenesis. Proceedings of the National Academy of Sciences of the United States of America 103:6208-6211

Zacharias DA (2002) Sticky caveats in an otherwise glowing report: oligomerizing fluorescent proteins and their use in cell biology. Sci STKE 2002:PE23

APPENDIX

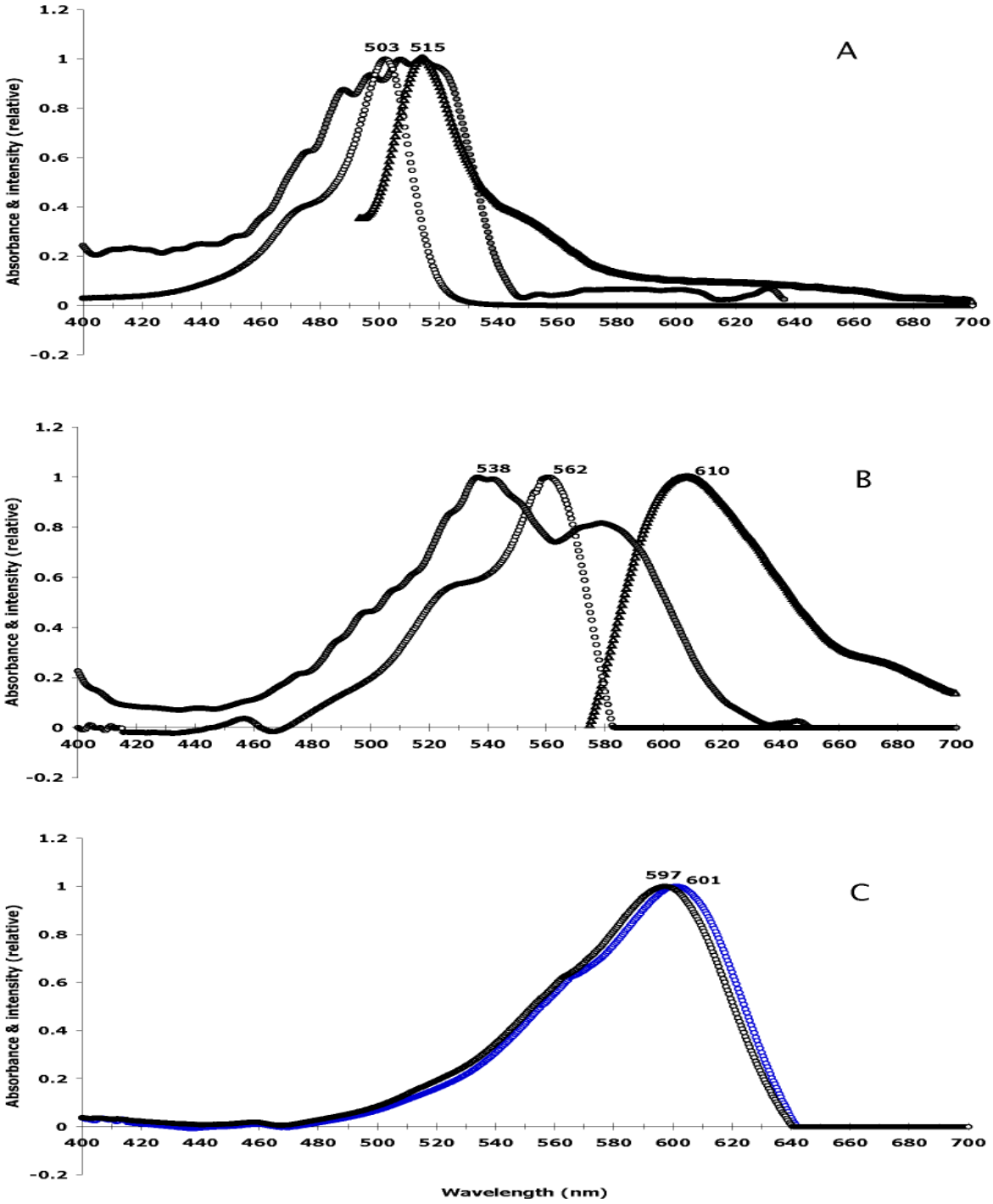

Figure 6. 1 Spectral characteristics of the recombinant proteins from *Acropora millepora* presettlement library.

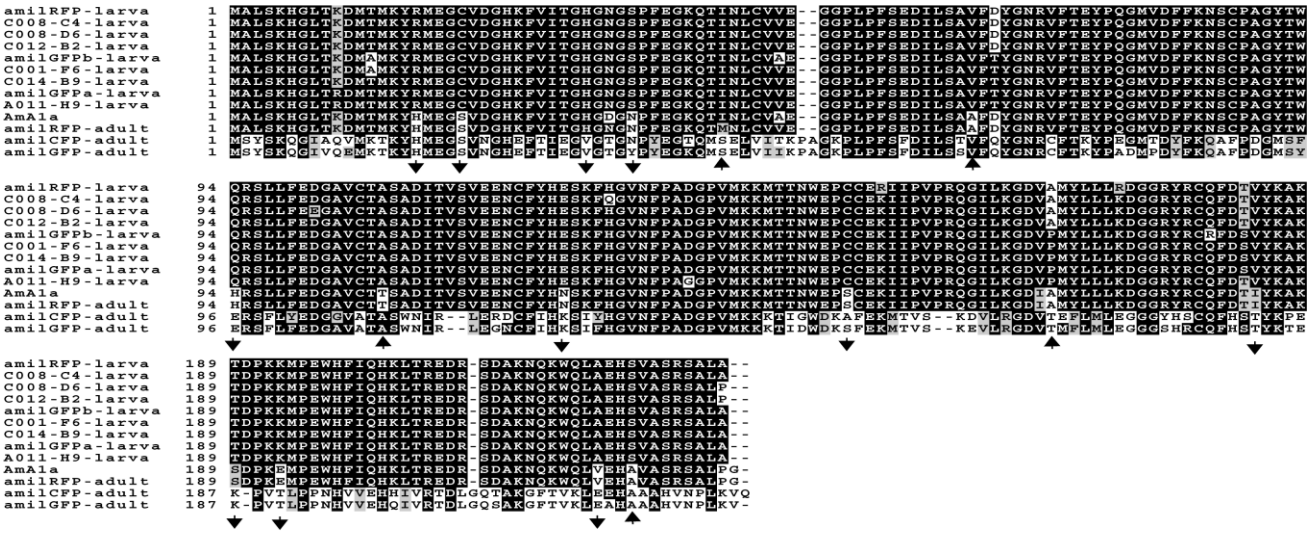

Figure 6. 2 Box shade alignment between larval and adult *Acropora* fluorescent proteins.
Arrowheads points to the orientation of the residue (in or outside) the GFP-like ß barrel (Beltran-Ramirez *et al* 2010)

Figure 6. 3 Prawn chip coloured embryos.

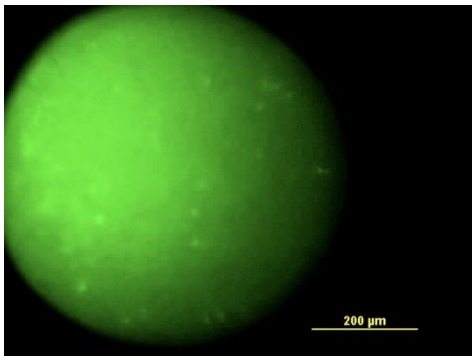

Figure 6. 4 38h-FITC coral embryo

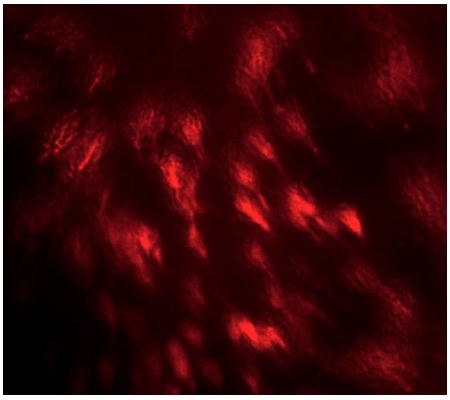

Figure 6. 5 Red granulations in planula cells.

Figure 6. 6 Planula section stained against amilFP

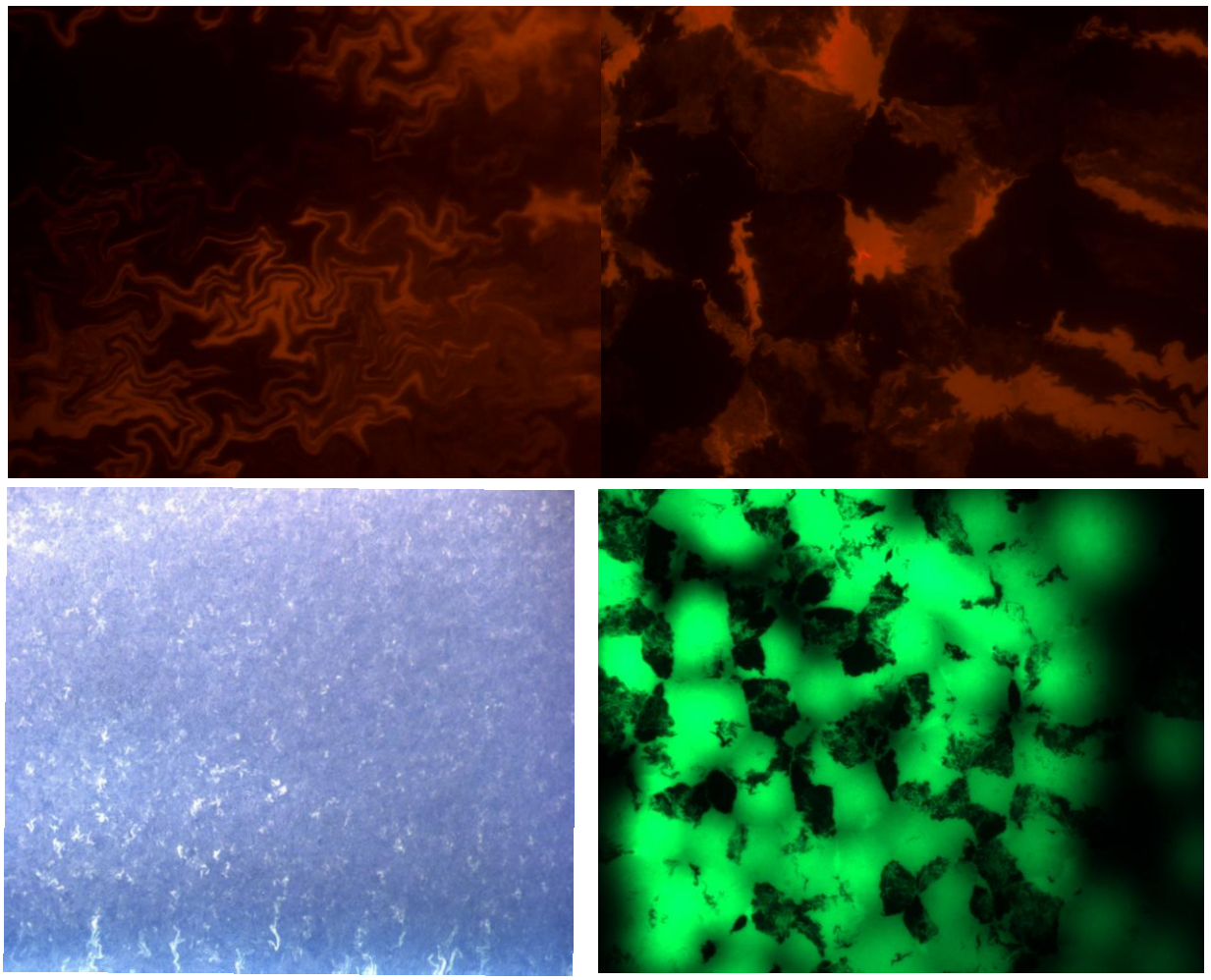

Figure 6.8 Lawn *of E. coli* expressing the amilCP and FPs produced during this research.

Nucleotide sequence of the nuclear GFP-like genes from *Nematostella vectensis* and *Acropora millepora*. Exonic sequences are marked in bold.

>nvecFP

AAAATAATTTCCCTAAAAAACAAAAAAGGAAAAAGCTTTTAATTCCTGGATAAGTGGCCGGAAGGACCCCATGCGCC
AATTAACCGTTTGTCGAAAAGATAAGAATTGTGAAATTATGTTATCCAGTAACCCCACCCACCAACCCCAATTTGTTC
TCCCTCCATGGGGGTTCCGCTAATGACCCGTTCAGAGTTAATTTTTAACAAAAAACAAATCAATCGCCTTTATTGTTT
TAATAGAATGGGTTATATTTTTATTGACTTAAAAAAACATCATGTTTTTCGTGAAAGAACGGCAGAGAGATCTTTGGA
AGCTCTTGACTTTATAAAATGGCTAAAATCAAAAACACACTTTGTTGATTGAGACTAATTTTGATATATAACTTTAGTA
TACATGGCATCTTATAAATAGTTGATACCCTTATTCATCATTAAGGGGGGAAAGTTGGAATTGCATAACAAATATCCT
ATTATTCACACATCAGCCAACAAAACTGCAACGAAGCAAGACACAACGGAACGAACCCCAAACTTTGATAAAAAAAC
ACGGTCAAACAACAACAAGAACCCCTACCGATTGAGATTTAGCATAGTCAGACAGCTGCTCCGAGGCCTTGCATTT
GGACGCCATGTTCAAAGCAAACTCAGATGTTTTCCACACAAAAAATGGATTCGGCCGACTTTCTCCTTGCACGACG
ATCAGAGAAGGAAGAGGTCAACGATCACGTGACTTAGGCGAAACGGCGTGGTGGGGGTCCTGTGGTAAGACTGTG
TCTGCTTGCTTCAATTTCTCTCTTTTTTTTCATTTATACACATTTAACCCCCCATCTAAAACATCTGTTTCATAAATCCT
ACAGCACTGATCAAGAATATAAAAACCCTCCTCTTTAGCGATCTTGTATGGGGAGAGGCCAAAAAAGGATTGTACTT
CTTAGCCCTGGGCAGGGAATCGTCGGCAAAATCAATTTTAACCCCAGGGGATAGCCCTTTGGGCTTGTTCAACAGA
AATTCCTACCGTTAAGAGAAAGGAATTCTTAACCCTAAGAGTTGCAGAATCGGAACAAAATGTTTAACACCCCCTTA
GGAATAGCAATTGGTCGTGAATTTTATACCCTAAGAGATAGCAAAATCGGAAAAAAACCTCCAACCTCCCCCCCCCC
CCCCCCCATAAGGGATGAAAGTTGGACGAAATTTTATACCCTAAAAGATGGGACGATCACCCCCGTCTACTTTTGAA
TCCGTGTTAGTGTGGACGACCCCTTTGTCGCGTAGACAAGAGAGCTAGCTACAGGCTTGTCAATAAAATCTATCCA
CCTCCATCTATCGCAGCCACGATAACGATAGCGTGCTTAGAAAGAGAAAAAAGGGAAGAAGGAGCAAAGTGGGCC
TTTACTCGTATTTGACCCATCGCGCAAATAAATGGACGTGGGACAGGAACTATAAATGTATGAAATGAATATAACAA
TTCCACATGAGTCAATTGCACATAAGCGAATTTAAATGAGTATATTACTTGTGAGTAGCATAAGACTACTCCAAAGG
GTAAAACGTATATATGATAAGACAATACACACTCCATCTATATGGACTGGTACCGGATGCATGCTACGTGACGTGAG
TCTAATTTTCTAGCCTGGGTAGCCTATGGTATTATTATAAGGCTTCCTCTACATAAGAGGACTAATATAATCACGTGC
CTATCACGTGGTCTACTTCAGCTTGGGCAGCTGCATGGTGGGTTTCTCCGGCAGGAAGCAGGAGAACGCCTCGGA
CCACTCGTGCTGAATCACAGTGTCACAGTGCTCGTCGTTGTCAACCCTTGTCAATCTGTAGGCCATGAAGTGGTGC
GGTGGCATCTTGAATCCCGCAGGCACGGGCTTCATGAACCTGAAGCAACACAGGGAACCCGATTATAATTACACAC
AGGGAACCCACCAAGATACACAGGAAACCCGATAACAATTACACAAAGCGAACCTATCAAGATACACAGGTATCCC
GTTAACAATTACACACAGGGAGCCCATTAAGATGCATAGGTATCCCTATAACAATTACACACAAGGAACCCATCAAG
ATACACAGGTATCCCGTTAACAATCACACACAGGGAACCCATTAAGATGCATAGGTATCCCGATAACAATTACACAC
AAGGAGCCCATCAAGATACACAGGTATCCCGCTTACAATTACACAAGGCGAACCCATTTAGATGCATAGGTATCCC
GTTAACAATTACACACAGAGAACCCATTAAGATGCATAGGTATACCTTTAACAATTCCACACAGAGAACCATTAAGAT
GCATAGGTATACTATCCTAGCACGAAAGCTAGCTTTTCCATATACTGTATCCTAGCACGAAAGCTAGCTTTTCCATAT
ACTGTATCCTAGCACGAAAGCTAGCTTTTCCATATACTGTATCCTAGCACAAAAGCTAGCTTTTCCTATACTGTATCC
TAGCCCTAGCACGAAAGCTAGCTTTTCCTATAGTGTATCCTAGCCCTAGCACGAAAGCTAGCTTTCCCATTCTGTAT
CCTAGCACAAAAGCTAGACTCTGGCAGGTCGGAGTTTGTCTACAGGCGTCTTGCAAGGATATTAAACTTTCTGCTT
GTTGTACAATTCTAAAAAAACATAAGCACTTGGTTTAATGGGGTAAAGATGCGAGTTTGGTATTAAAACAAATCTGTT
TTACGATAATGATGAAGGAATTCTTATCAAACGTCCAATACAAACTTCCCAGCGATTTCCACGCCTTTTCGAGATATT
TAGAATATAATTACATGCGTGCTTTGCGGTTTGGAGACGGTTAAACCATGCAAGCTCGTAATCGTTTAATATAATTAA
AACCAGGTGGAAATATGTGCTCAGAGTTCACTCAGATTAAGTAAAATAACAGTTTTAACTATGCTGTGAGTCTCGGC
TTTTAACTTTGCTTATGACTGGATGTACCAAAAATGTGAAATTTAGTGAGTTAGAACTTTTCTCATGTGATTCATACTC
TTCAAAATATCGCCAAAATCGAATTATTTTAAGAATTCTTAGGAAGAAGTCTTAAAATATGAGATAAGTTGACCTTTAT
ACTTTATTCATATGGTGTTGTAATAATTAATAGTCATGCCGTGTTGTTAAATAGATAAATGATGATAAATAGTGCACAG
TTCCAACGCGTCCTCGCGACTGTTGTTAACAGCGGAAATTATTCCTTTCATTAGCCGAAATAGCATCAATTCCTTAG
CCGAAATAGCATCTTCTGGTTCCTACACAAAACAGGCATCACAATTTCACAATTGATGGAAGGAAGCTTTTTGTAGC
TTACTGAATTAGAAAATTAAAATAAATCATTTACGGTGATTCCTGAGTCGAACAAATAATATTTCCTCATCATATAAGC
CATCTTGGATTATACTGGGATTATTCATTAGCAAATTTAAAGCATGTTAATGACATGACTCAGCAAAAAACAGTGTGC
GCTTTTGTTAACATTCTGCAGCGCACCTATGCGCACATTGCTTGCACTTTGCACTTAAAACAACCATATTTTGATCTA
TGAGAAAAATATAAAAGTTTTTAAGACGGCACGGTCTTTTGCGCGCAATCCATTTATTATATGGTCTGGTCGAATAAG
CATTAATCTTAACCATGATATTTCTACTTGGCGTCACTGGAGAGTCAGGAGAGAGATGTAGAATATTCTAAATATAGA
TGGCCTTTATTTTCAATTGATTTCTTCTAGAAAAACAGACAGCCTGCCGCAGATTAGCGGTCCTTTTTGGTCTCATGA
CCTTGATAAACAGATTTATTCCAGTGAGTATGTATCATATTAAGCTGTCTAAAATCTAATTAGTTTGAAAGCCCAAGT
GTTTGTCTCTTGCATTTTTTATTGTCTTCGGAGTATAATAAAAAGCGCCCTCTCAGTAACTTTTCAGAAGTGTCTACT
AGCGCGCCAAGCACTATTGCACCAAGTCCTGATTGAGAGGTTGTCTTCAAAGGCCAGTAACAAAAACACAATGGTA
ATAGTATTTGTCCTTATTTGTGTGCATGACAATTGACTGAATTCGGT**ATGTTCCTGGCTAGTGTGTGGTATAGTTTTT**
ACCAAAAGATTTTCATTTTTGTGAGCAGTAAGACATTGCTTCCTCCACCTCTGGATTACGCTCAACGTTATGCTAT
GCTACTTAGAATCCAAGCGAGCCCTGTGGTGTACTCTAGGTTTATTGCGGCCTTAAACGAATATGTAAGAAGAAT
AACCAACAAATCTTTCTCTTTCTTATGGTCAACTCCCTGCTAGCGTAGGCCACTTTTCTTTAGATTTTGAATGCAACC
CTTAGTGTGGCCAAAGTGGAGCTCAAAACTTTTTTTAACAAGTCAACAAATCTCTAGTCAACATCCGGTGAAAGTTT
CATTTTAATGTCTCTTAGTGTATGTCTGTCATTTTTCTCCCTTAG**AGTCTTAGTAAAGATGCTATGAAGCTTCACTTT**
AAAATGGAGGGGTCCGTCAACGGACATTGCTTCGAGATCCAAGGAGTAGGAGAAGGGAAAGCATTCGAGTGAG
TTGAATGTATTTTGTTTGACTTTTTTTTTCGGTTTTAACAGCATTTAAAAACACATTGTACGCAGTGTAAAAAGATATT
AAGTAAAAGTATATAAACGGTTTAGCGGGCAAATTGTTTATCTTTAATTGATAATGCTTGTTTTCGGTCTTACAGCTT
CAGTTTTGCAGCTTGTTTATCTGTTCTATAAATTGCTTGGTATTTCACAGGATAAGTTGACAACAACTCTAAATTCAAA
CTAAAATTAATGTTATATATCTTCTCCTGTTGGCTTGCATAAGGATTATCTAGTTCCCGTATCCCCAATACCAGTCCT
TTTTGTTTTCTATTTTCAG**CGGGGAGCATTGGTCTAAGCTTTGTGTTGTTAAGGGGAAGCATCTACCGTTTTCCTTT**
GACATCCTCATGCCAAGCATGTCTTACGGAACCAAGCAGTTCGCCAAGTACCCAGCTGGCATGACTGACTTCTTC
AAAGCTGCCGTGGAAAATGGCGGGCTGAGCTGGGAACGAACCATGACGTTCGAGGATGGAGGCTACTGTACGA
TCGTCAACACATCAGAGTAAGCGATGAAATTAATTAAATGGTCTGCAGATCTAAGAAAAAGCAAGGAGGGGCAAAA
GACAAGGAGGATCGATCCTCCTTGTGTTTTGCCCCACCTTGCTTTTTTTTTTACTATGTATAAAATCTATTTCGTATGG
TTTAAAG**GCTCAAAGACGGCAGCCTTCATTATCACACCAACTTCCACGGAATCAACCTCAAACCGGACGGACCC**
GTAATGCAGAAGAGGACGATGGGGTGGTTACCAAGCGTGGAGACAAACATCCCCCGACGAGACACTCTCCTCG
GTGACATCAACATGCTCCTTAAAGTCAACGACGGCAGTTTTTTGCGCGTTCAGTTTGAGACGGTCTACAGGTATG

TGCACTCCCTGATTTATGGGCTACCTGTGTATTGAATAATACCGGATTTCCTTTAGATTTTCTAATCTCTTTTTT
GGAGGTACGTCGGAAATATATATGTTTTGGATTCATTCAGAAACGTCATAAAAGATGGCAGACTGTCAATATTTTCA
GGGTTTCGTAAAGAATTACACGATAGATGGACTGAATACAGTCGTAAATGTAATAAATAACACAGGTACTCTTCTAA
GTCGACACAAAACCATGTCCCTAGTCCCCCATCAGACAATATTGACTGTTTCTTAAAAAAAATAAATTACATGTGACT
AACCTGTGTAGTCTGAAGCTGTTTACCCGCCCATCTTTGGTCTCATCAAAAGAGCGCGAATTCAATTATATACATAA
ATGACCTGTTTCATACTTTAAATCTTCTTTCCTTTATTCTTTTTGCAGAAGATACTAACATATTATTCTTCCGACATCAA
AATATACACAGTTTAATCTCCATTGTCAATAAGGAGCTCTCTCTTGTCTCTAAATGGTTCAATTTAAACAAACTGACAT
TACATCCTGATAAAACAAAATTTATTTTGTTTCACCCCCCAAGAAAGAAAATCATAGTTGATCATCCTGTCAAAAATAA
AAATAGTATCATATCTCGTGTTGACAGTACTAAGTTCTTAGGAGTTATTATCCATGGAAACCTCTCTTGGAAACAGCA
TATTCTAACAATTCGCTCCAAAGTCCCAAAAGTCATCAGTATCATCTCCAAAGCTAGACAATTTCTTTTTCCAACTAC
ACTCTTGACTCTATACAACTCTCTTTTCTTACCTTATCTTAATTATTGTACACTAATTTGGGGATCAACATTTGATTCTT
ATACAAGACCTCTATTTCTCTTACAAATAAAAAATAATCCGTATCATTACTTCCTCACCTTGGAGGGCCCATACAAAA
ACACTCTTTCAAATGTATAATCTCCTGACATATGCCATCTATATACAAATATCAAACATCTTGCTTTGTTTTCTCACAC
CTTTATAATTTACTACCCCTGTCTCTATCTTCACTCTTGTGTGTCAATTCCATAATTTACATAAAAAAAAACCAAATAT
CACTTTTCCATCTCCTATCAAGCCCCTACTATCTGGAATGATATTCCACCCTCTCTTAGAAAGTCCTTAAAGCCCCAC
AACTTCAGATCCAAACTCAAACACCACTTCCTGGATCCATAATCTTGTCTTTTTCACCACACCTGGAAGTCCGTTGA
CGACCTCTACTACAAGCCAACGATTTTAGTTGAGTTTAGTTATTGTTATTAATATTATATTGTTATTTTATTATTTATAT
ATATTTTTACTTTTTTCATTGTACCAAGCTATACATCGATCCTTCACTCCTGTACTCTTAATATATCTGACATTAAGCC
ACTTAAACTTCCGGATTTGGCTTTTCCCATGTATGTTATCACCAAGTTTTACCAATAAAATTGATTGATTGATTGATTG
ATTGATGGGCTCCCTAACACCCGGTCGTCTGTTTCTTGGTGAAATTGTTAGCGGGATACCTTTTTATTGTGATGGGT
TCCCTGTGTGTAAGTAAAATTGGCTGCCCTGTGTATCTAAAAGGGTTCCAGGTGTGTGTAATTCTTACTGGGATATC
TATGCATCTTAATGGGTTCTCTGTGTGTAATTGTTAACGGTATATCTATGCATCTTAATGGGTTCTCTGTGTGGAATT
GTTAAAGGTATACCTATGCATCTTAATGGGTTCCCTGTGTGTAATTGTTAACGGGATACCTATGCATCTAAATGGGTT
CGCCTTGTGTAATTGTAAGCGGGATACCTGTGTGTCTTAATGGGATTCCCTGTGTGTAATTGTTAGCGGGATACCTA
TGCATCTTAATAGTTCGCTTTGTGTAATTGTTAACGGGCTCCCTGTGTATCTTGATGGGCTCCTTGTGTGTAATTGTT
ATCGGGATACCTATGCATCTTAATGGGCTCCCTGTGTGTAATTGTTAATGGGATACCTGTGTATCTTGATGGGTTCG
CTTTGTGTAATTGTTATCGGGTTTCCTGTGTATCTTGGTGGGTTCCCTGTGTGTAATTATCATCGGGTTCCCTGTGTT
ACTTCAG**GTTCATGAAGCCCGTGCCTGCGGGATTCAAGATGCCACCGCACCACTTCATGGCCTACAGACTGACA**
AGGGTTGACAACGACGAGCACTGTGACACTGTGATTCAGCACGAGTGGTCCGAGGCGTTCTCCTGCTTCCTGCC
GGAGAAACCCACCATGCAGCTGCCCAAGCTGAAGTAGATGCACGACTTAGCACGTGATGAGCACGTGCTTATACT
AGTCCCCTTATGTAGAGGATCCTTATAATTATACCACAGGCTGCAAAAGCTAGAGAATTAGATCACATGATAAGCAC
AGAGTACCTATTTGTCTTATCATAACACATGTTTACTCAAGCTGTAGTTTACCCCCATCCTACCTATAATTGATATATC
CATTTATTAAAGACACTGAAATCGACTGATCGACAGAATTCATTTCGTACTTCGGAGTTGCTGTTTCGCGTCAAATAA
GTTGTCGCGATTTCAAATCTGAGTCAAGGCCACCTTAGCTCCTTTTTATCTATTTTCTTCTCTTAAAATTAATCCGGT
CATCCACAGTTATACAGATTGATGGGCCATACCCTGAAGTCTGTTGACCAAAACCCATACTTAGAAGTTGAATTATC
AGTTGCCTTCGACTGGAACCACTATATAGACAGTAATGTTTCTAAGGCTACCTACTAGGCAAGCTTGAATGGCCCTC
ACTCTAACAAAGAAGGAGAAATTCCCGCTTAACTCTGCCATGGTGACACTGCGCTTGGGGTCCCTCCATATACAGA
TAAATAAGTTGCCTTAACTCGCTCAAATTTGCCAATGGGAGTGACAATTACGGCCATTCTTTGTCAGATTCCATGTAA
TTTCATGTCTCATTTCACCAATCTGTCGGACATATTAGATTCGGTAAATACCCTTGTGTCACAGGTCTACCTGTGAAA
ATAAGCCAGCCTCCGCCAGGGTCTTGGTCAGCCAGACGACATATTGAAAATTACCCAGAAAACCCAGGGGACGAG
GTTGGTTATGTGCGCAGTGTGTGTGTGCGTGTCGCTATGTGCGCAGCGTGTGCGTGTCGTTATGTGCAGTGTGTAA
ATGTCGTCATGTGCGAAGTGTATTTGTACCGTTGTGAGTAGGTGTGTCGTTATGTGCGCATTTGTGTGCGGGTAGT
TATATGCGCTGTTTATTTATACCGATGTGCGCAGTGCGTTGTGATATGTGATCAGTGTGTGCGTGTCGTTATGTGCA
GTGTGTCCTTGTTGTTATGTGCACAGTGGTACGTGTCGTTATGTGCGCAGCTTGTGCATGTCGTTATATGCGCAGT
GCAAACGTGTCTTCATGTGTACAGTGTGTGCGTTTAGTAATGTGCAAAGTTCGTGCGTATCTTCAAGTGCACAGTTC
GTGCGTGTCGTCATCGTCATTTGCGCAGTGCGTGCGTTTAGTTATGCGCAAAATTCGTGCGTGTWGTTATGTGCAA
AGCTCGTGTGTTTCGTCATGTGCACAGTGTGTGCGTGTGGTTATATGCGCAGTGTGTGTGTCATGTGCGCAGTGTG
TGCGTGTGGTTATGAGCGCAGTGTGTCATGTTTGCAGTGTGTGTGTCGTTATGTGCGCAGTGTTGTGGGTCGTTAT
GTGCAGTGGGTGCATGTTGTCATGTGCGCAGTGTGTCCTTATGTTCTCTGTGATTTTGACATCGCTATTTTATTGTC
AAAGTGTGTGCGAGTCGTAATGTTCTCAGTGTGGGCGTGTCGTCCTGTGCGCAGGTGTGCGTGTGGTTATGTGCG
CAGTATGTGTGTTTTGTTATGTGCGCAGTGTAGACGTGTCGTCATGTGCGCGTGTGCGTGTCGCTATGTGCACAGT
GTGTGTGTCATGTGCGCGGTGTGTGCGTGTCGCTATGTGCACAGTGTGTGTGTGTCATGTGCGCGGTGTGTGCGT
GTCGCTATATGCGCAGTGTGTGTGTGTCATGTGCGCGGTGTGTGCGTGTCGCTATGTGCGTAGTGTGCGTGTGTC
ATTATGTGCGCGGTGTGTGCGTGTCGCTATGTGCGCAGAGTGTGCGTGTCATTATGTGCGCGGTGTGTGCGTGTC
GTTATGTGCACAGTTTGTGCCTGTCGTCATGTGCATAGTTTGTGCGTATCGCTGTTTGCGCAGTGTGTGCGTGTCG
TTATGAACACGGTGTGCGTGTTATTATGTGCGCAATGTATTTGTACCGCTTTGCGCAGTGCGTGCCGTTCTAATACG
CAGTGTGCCGTTATATACATTGAGTCGTTATGTGCACAGTGTGTGTGCGTTTCGTTATGTGCACAGTGAGTGCGTGT
CGTTATGTGCACAGTGTGTGCGTTTTGTTATGTGCGCAGTGTATTTGTACCGTTGTGCGCAGTGCGTAGTGTTGTG
CGCAGTGTGCGTGTCGTTATGAGTGTGTGCGTGCGGTTATGTGCGCAGTGTGTGTCGTTATGTCTGCAGTGTGTG
CGTGTCGTTATGTGCGCGGTGTGTGCGTGTCGTTACTTGCAAAGTGGTGGTGTCGTCATGTGCGCAGTGTGTGCG
TGTCGTTATGTGCGCAGTGTTGCATGTCGTTATGTGCAGTGTGTGCATGTCGTCATGTGCGCAGTGTGTTCTTGTG
TGCTTTGGGATTTTGACATCGCTATGTTATAGTCAGTGTGTGCGAGTCGTCATGTGTATAGTGTGAGTGTCGTTATG
TGCACAGTGTGTGTGTCGTTATGTGCGCATTGTGTACGTGTTGTTATGTGCAAAGTGCGTGTGTCATATGCGCAGT
GTGTGCGTGCCGTTATGTACAGTGTGTGTGTGTCGTTATGTGGGCAGTGCGTGTGGTTGTGTGCGCAGTGTGTGT
GTGTCATATGCGCAGTGTGTGCGTGCCGCTATGTACAGTGTGTGTGTGTCGTTATTTGCGCAGGTGTGCAGGGTTA
TGTACAGTGTGTGCAGGTCGTGATGTGTACACGGTGTGCGGGTGTCGTTATGTGCATGTTGTAATGTGCGCATTGT
GTGCGGGTCATTATGTGCGCTGTTTATTTATACCGTTGTGTGCAGTGCGTGGTGGTATGTGCGCAGGCGTGTTGTC
TTGTACGCAGTGTGTGTGTGTCGTTATGTGCGCAGTGTGTGCGTGTCGTTATGTGCGCAGTGTTGTGGATCATTAT
GTGCAGTGTTTGCGTGTCGTTATGTGGCAGTGTATGCGTGTCATTATTTGCGCAGGTGTGCAGGGTTATGTGCAGT
GTGTGCAGGTCATGATGTGTACATGGTGTGCGCGTGTCGTTAGGTGCGCATTGTGTGCGGGTCATTATGTGCGCT
GTTTATTTACACCGTTGTGTGCAGTGCGTGGTGGTATGTGCGCAGGTGTGTTGTCTTGTGCGCAGTGTTTGCGTGT
TGTCATGTGCGCAGTGTGTCCTTATGTGCTTTGTGATTTTGACATCGTAGTGTGTGCATGGTTGTGTGTGCGCAGTT
TATGGAGTACACCCCAATACCCACAAACCGAGAGGCAAATAATATATTATGGTAAATTTCATCAATGGAAAAACAAA
GCACATTCA

>amilCP

GCAAATGTGGATGTCGGAGCTGAAATGCCCTCTGGCCCAAGTGAACCGACCAAAAAGGTTCCCGCTGGCTGCTGA
ATAATAATATATCACTTTCACAACATTAGATTACGTAACGATAAGTATTGTGTCATACACGTCGTTGACCTTTTCCCAT
AAATGGTCCTATTGACGCCATTAATTTGGTCTGTTATCTGTTGGATATCTTTTAAAACACGACTAATAAAAGAGGTGTA
CTGGGTGAGGCGTGCATTCACATCAGCGTCTCTTTCGCG**GAAGGAGCTACTTAAGAAACGTTTGCGAATCGTTGTT**
CTGCGCTACTTAGTGTCAATATGGTAAGTTTTCTTCCAATTTAGAACTCACAGTCAGTTAAACCTTTAATAATAGCCA
CCTGCGTATCGTCTCAGTCCTTTTTACGAAGGCCACTGCTTTTTGTTCCGACTAACAGTCTTTAAATTCCATTTTATG
CTTCCTTCTCTTCAACACGGCCACCACCCTACAATGCAGTAAGGTCCTTGCCTCATTTACGAAAATCTTATCACAATG
GCCATTTGAAATGCAATTGATCTACTAACTAAACTTTAGTAGCTTAATGCTTTGTTAATGTTTTTTTCCTGACCACTTG
GGTTTCACTGTGTTATTTATGTTATGCTAAAATAATAGCTCACTACAAATTAAAGTTCAACCTAAATCATGTCAACTTG
GATTTTCGTGCAGTAGGCTTATTTTACGTACCCAGGAGAGTTTCTTAACTTCGCACAAAAAAAACGCAAAGACCTTGC
CTGTCGTTTGTTTTTTCAATATTTTTACCCTGAACATTTTACTTTGGGCTTAAATTTGTTACTTAATGATGTTTTAAAAT
GAACATTTTTAAAACTACAAAAAAATCTACTTTTTAACGCTCTTCTGTTTGGGGCTTTCTCTTCCTTTGCATGCCCTGG
AGAGTTTCTCGTAGTTTCGTTCTTTACTACGTAGAAGAAACGTTGGAAAGAAAAAGCTTTTCAGGAGAAAACCGTCC
CTTCTGCTTGATGGATGACAAATTTTACTAACCTCTTTCAAATGAAAACTTGCATATCTTGATAGTTTGCTGTATAAAT
AAAATTGAAACAAAATGCAGCTTATCACAGCCTATGATATAAAACTAAACAAAATTAAGATTTTCTTTTCAATATGTTTG
AATTTATTTACCTACTATGGTCACTGATATTTATAAAATCTAATAGATGTGTTGTTGAAGTTCAACAAGACCGATAAAG
TTGAAGAACTCGGCGAGAGATCGGCAAGTCAAGCACATGCGAATATAATCTTTGGATAAATTATTCACATTTTGCAG
CCTGTATAAACCAATTTTTACCTCTTGATTATTCGTAGTGGTCTCTAGAACCAATAAACATTTTGAGGAAGTCCAGAAT
TGGCTTCAAACAACTCCAAAACAGCGTAAAACGTGATGAAATCAGAAGTCGCTCGGTTCTTTGTTTACCAATTTAACG
TGAGCTTGTTTTGAAATGCCGTCTTCCTCACTTCGAACTTCATTGTCCTCCACTACACTGATCTTCGATTACAGAGAA
TATTAGTTTCAAAGCAGAATTCACTTTGAGTAGTTCGCCTGTCAAAGTAGGTTTTATGGACATTACGATAAAGCATTC
TTGAAAAATTCCGGTCGCTTCGTTTTCTAGCGGAGGAAAAATAATACAAAAAAAACCCAACGCATAGTATGACTCATA
CACACGGCCTACGGCCTGGGCACGAGTAAAAAGTCAACTTTGATACATTCCTGAGCGTCGCGAATCCTTTACTTCG
CGCTCCGCCGAAGTAAAAAGCGATTCAATGAATTGAGCAGTGATTGACTGGGTGTCATAAGTAAGTGGGCAAGGGC
ATCTTAGTTTTTAACTAGGCTGTGAGCATTTAAGTGAAATAGCGAAAAGGATATTATGTTCGGAGAAGCAAACATGGC
CTAGATTTTTCTATAGCTGGAGGACAGTTCAAAATCTATCCGATTATCGGAAGCATAGTTTTTCACGTTTCACTCTCC
ATGTCAGGCTATTTTTCGTCGAAAAAAATCAAAAATTTGAGAATTCTCTCTTTCGTAAATCTTTAAATTGCATCTAAAAT
TTTTGGGAATATAATACTGAAGCCCTTGTAATTTCGAAGAGACTCAAGTTACCCTAGGATGACCATGCAATCAGATAC
AAAATCTACATAGCTGCTTAGAATACGTTTCTAGAGATCTAAAAGTACTCTGGATAGCCTAAAAAGTGTTTTAAATGTA
TTCAGGGGGGACACTGTCGTTGGGTGCCGCTGTAAAATAGACAATTTGCAACTCTTTGTTAGTATAATGCATCCCAA
CGCATTCTTTTGCGAAGTTGCAGAACGTTCAGATCGGTGCTAGCACGTATTATTGTATAGACTAGGAAAACTCGGGC
ACATTATACTCCTGTATCTTGAAAGAACTTTACTGGCTACGAGGTAGAGCTGAGAATAGTTTATAGAATAAATTTCCTT
ACTTTTAAATATATTCATGGATTCACCCCCTTAGTATTTACGAGAATTATTTATTGGATTGTCTACAAGTCTAATACATC
CGACTACGCTGCTCTGATGCGATTAGACTAAACCTGTCTCGGTATAATCTGAATTCTTATGGACTGAGGTGCTGTTG
AAGGAACTTCCATCTGAATTACAGTATATTAAGTATCTTTAAGACTAAGCTGAACACACAACTTTTTAAAGGCATATTA
TAATTAATTTTTGTTTTAGTCAGCTCTATTTATTAATTTATTTATCTATTTCATCTTTTCTTTCCTCTTTTGCGATTTTTTT
GACACTAGAAACATAGTTAGAGAATTGTAAAGCTCTTGGAACAATTTATACAGGCCGTACATATATAAAAATGTTCGT
TACAACAATAATAATAATAATAATAATAATAATAACTATTATTATTTGTAATTTGTTATTATTTGTTATCACTGACTACCT
AACAAAAGGACCTAAGTAAGTAGGTGATGCGTTTGAAGTCTGACCACTGCATAGATGACTGTGTGTAATGACCAATT
ATTTAAAATGACCTGTGTAAGAACAGCCTGCTTTCTCGTGATGTACGATCTGCAAGAGTCTGATGATGCCTTTAAACC
TAAAGCGCATCACCACAATTTTGAAATAGTTCAATGCGTTTTATCAAAAGGATCCTTGAAAAGGAGCTTTATTGTATC
ATAAGTTATATATTTGTTACTTGTAGTCCGTTCAGTTAGGCTTTTTTTTTGGCCGTGAGCACATTTTCAAAGTTACAAC
TGACCCTTGTAATAAAAAGTTTCACGCTGTATTTTTTTCACTTTACAG**AGTGTGATCGCTAAAGAAATGACCTACAA**
GGCTTATATGTCAGGCACGGTCAATGGACACTACTTTGAGGTCGAAGGCGATGGAAAAGGAAAGCCTTACGAGT
ACGTTTGTGTTTAATGGATGTAATTTATATCTAATACATGTTCAATATGTGTTGGTGTCTCTTCAATGCATGCGTGAAG
CGTTTTCGTAGCGAACCCTTTTTCACTTCGTTCCATGATACTTTCGAGGATGTTTTTGGAAGAAGAGGAGGTAGAATT
ACGATAGAACTAGGGAGTTGGGAAGTGTTTTCGGCGTTCTCCAAAAACGCCGGTTATGCATTATCTTTGATGGACAG
AAATTATGTGTGACCTACATCGAAAGCTATTACACAAAATTAACCTCTTCATCAATTTAATAAAATAAATAGCAACCAA
CTGGCATCACATGCACAAAGTGGTTTCTTATCCACCATTCCAAAAAGAATTGATATTTCAAAATGTGTTTGTTTTTAAA
AGGGTGGTTTTGGTTTTTTAAGGAAAAAGGGGAGGGAAATCCAAGCAATATAAGAGAAAATCCTGTAATAGTAAGAA
CTAATAGCACCTTTTTAAGCAAGATCCATAGAGAATTTTTTAGTCAATTAGCCCCCATTTTATATAAATGTGCCTAATY
TGCAATGTCAATTCGTTCTTTTTAAG**GGGGGGGAGCAGACAGAGAAGCTCACTGTCACCAAGGGCGGACCTCTGC**
CATTTGCTTGGGATATTTTATCACCACAGTGTCAGTACGGAAGCATACCATTCACCAAGTACCCTGAAGACATCCC
TGACTATGTAAAGCAGTCATTCCCGGAGGGATTTACATGGGAGAGGATCATGAACTTTGAAGATGGTGCAGTGTG
TACTGTCACCAATGATTCCAGGTACATCGCGCAAAATAAGTTAATTGAATTTTAGTTTCCGCAAAAGCAGAATGTCA
GAACTGGAAAGCATTTTCCTTTGCATTAGAAAATGCTAAGAAATCTCAGCAGATCAGTCTCACTTTCTCTTCGTTAAA
AATGGCGAGTTGTAGAACCTGACATGCGCTGTAGAGTATCATTATGAAAAATATCCACAACCATTCCATTGCGCTTT
GTTCAAACATTTTTTCAACTCTTTTTCAAAAG**CATCCAAGGCAACTGTTTCATCTACCATGTCAAGTTCTCTGGTTTG**
AACTTTCCTCCCAATGGACCTGTGATGCAGAAGAAGACACAGGGCTGGGAACCCCACTCTGAGCGTCTCTTTGCA
CGAGATGGAATGCTGATAGGAAACAACTTTATGGCTCTGAAGTTAGAAGGAGGCGGTCACTATTTGTGTGAATTC
AAAACTACTTACAAGTAAGATATCACAATTGCCATCTAAACCATTAAGGAGGGACCACAAATGTAGGTTTTCCGTGA
TGTTGCCCGGTCGATAATCCTGAAGGCCACGATTCGGGATTTTAAGGAAAAACGGGCCAGAATCGGTTTAAGGTTA
AATAAACCTCTGAAAAAGAAAAGAGGTTTTCACAGTTTTTGCATTTGAGGTAAAGAATTGTTTCAAATGAAAAAGAGA
TTTAGAGAAAAGAATCAATTATGAGTAATACTATATATTGTGTGGCACTCATTATGCATGTGTAGATATCTAATTCGTC
ATTGTATATTTTTTTCACGTTTAAGAATTAGTTAACCGTTTTGGAGTAGATTTGCACGTCGACATCGAGTTATATGACA
TCCTCCACTTGTAGGACACTTTTTGTCGTAAATGACATGTGAAGCGGATTTTGCAAACCATCAGCAGCAGCCTGTTA
GGACGTACAAAGGGAAGCCTCATTTCAAGCCCCGTCCCCTTCGTGGATTTCAACCTGTTTGTTTGGGGCAAAACGA
TGATGGCTATTGTTCAAAAAGGGTATTTCGGTGAAGATTTAAGTTGATGCTTTGCCTTACAAGCGATCTTATTTCGTT
CTTCTTTTGTTAG**GGCAAAGAAGCCTGTGAAGATGCCAGGGTATCACTATGTTGACCGCAAACTGGATGTAACCA**
ATCACAACAAGGATTACACTTCCGTTGAGCAGCGTGAAATTTCCATTGCACGCAAACCTTTGGTCGCCTGACGTTT
TTTCAGAGTCAAATCAAGGCACAAATAAGCAGTGGCGTAAAAAACGTAGATTATGATTTTAGCTTAAAGAAGTAG
GAACGAAGAAGTGTAAACAACCTTCAATGATTAAACTTTTGAAAACAACGCTTAAGTGCGACTCGTCATTATATTAC
TGGCACCTGATTAGTTTTAGCTTGATCTGTTTTTATAATTGGCAATAACAGCTCTTTAAGCCCTAGCGAAATACAGCT
ATCCAATTTGCTTTGTCTGCGGACGTCTTAAGAGTGAATTTCACCTATTACCAAACTGACGACGGCAGAAAGTGAAA

AATCAATCTTCCTCTATTTTCGTCTAGGTATAGCCCTTAGGTAGATATGAAATGTAATGTTGTTTTTTTCCCTCTT
GCAACGCTTTAAATTCGAGAAAATCAAGAAAAACGAATTTCGTCCGACTATGGCCAAAATTGTTTCCCTTCTAGCAGA
GGCTCTTTTCCTGGTGTTCGCTGACGGGAGGAAAGAGACCTCTGCCATGGGTCGAAAATGGCTTTGATTTAACCGC
CGTCCACGATCTTGAACGGCTCTATTTGAAACGCATGCTCCTTGCGTTGTTTGCTGCCCGCAACCTTAGCTCGCTGC
GTCTCGCGCATTTGTTTTCATTAATTCCGCTGTCATATGAGCCCACAGTGTAAACATGACCAGCTCATCGGTGCGCT
CGAAAGAGGCAGTCGTAAGGAGTTCCTGAGTTTGCTTCCTTCCGTTTTGGAGAACCTTCACGAAGGAAAGATAAAAA
AAAATGACTCACGAGCGGTTTAAGGCAGTGTATAATTTTTTAAACGGGCGAGACACGTTTGCGTGCTTGCTGACTGG
TCACGGAAAAACTCTGATGTACCGTATAGCTGTTAAAATGGAGCAGGAGCTATTCGACAACGATGATAAACTGACAG
AGCTAGAAGAAGCCGAAGTTGTTTACGTGCTGCAAACATCACGTGATGGCTCGGTTGCTAAGCCACGGAGCATGCT
CAACAAAGCCATTTTCGACACATGGCAGAGGTCTCTTTTCTCCCGTCAGCGAACACCAGGAAAAGAGCCTCTGTTAG
TAGGGAACAAATTGGTCGAACGCAATGGCTGAAAATAAGAAATCTAAACTAAATTTGCGTTCGCGTGTCAAAAAGAA
CCAAAACTTCCCGATACTTTTGCTTGAAAGACAAAGCATTCAGCTATCAAAAACAGGCTCTAACTTGTCTTGAAAAAA
ACCACAAATGTTTAACAATTACCTTCGTTTACTCTCCTGATTATCCCAGTTTCAAAGTAAAATTCGGAGACGTTTTTCG
ACAAGTGCCATGACTCCAAATTCCCTCAAACTTGGCCAGCGTAAACATTTTATGGTAGTCTAAAACATGTATGAAAAT
CAGCTTGGGTTCTTCTTCGAGTGGTTTTTAAGGGGCCGCTGAAAAAGCTACATTTTAATGAAAACATCAGCAAAAAAT
AAACAATAGACTGCTGCATGTTGTTTGAGGACAGATGGGTATGAGCTTTTAACGACCTTGTATTATTAGTTATTTGTA
CAGATTAGCAATTTGCTTGCTTTCACTTATAGTATAAATTGTTTATTTATACGTACTATTGAGAAATATTTTCGAAATCG
ATGGAACTCACCACTCGTGAACCTTCTCCAGAAGTTTGCGAAAACTCGGTGTTTCGATGCGAAACCGCTATACGACC
TACTGGAGACCCGTAACAGTTTATCAGTCCTTCACCAACATTACTATGCTCGCTATCGTTCCTTACATCAACGGACGT
TGCTTCGCTTTGGTTCATCTTCTCGCTGAGATTTCTCTCACTTCTCACGCACGTGTGTATAACAAGATGACATGTGAG
AAATGGCCACGAAGTCCCCGGATAATAACAATCACAGAGGCGCGCAAAGCACACTGGTTGCGTGTGGTCCTTTAAT
CGAATTAAGTGTTTCGTGAGCAGTTCGCATTATTCGCCATTCAGTTGGAACGAGATGTCGATATCACGAGTAAGAAA
CCGGAAGTCAATAGTCTGTGCCTGCAAGTAATAAAATAACCCATCGCAAAAACATGTATCAAAATCAGCTTGAGTTTT
TGTTCGAGTGGTTTTTGAGGCGCCGCTGAAAGAGCTAAATTATGATGAAAACATTAGCGAAAAGAAAACCTTAGAAT
GTTGCTTAGTTGTTTTGAAGCCAAGAGAGTGAGTGTTGAGATATAAAGGCCCTTTGACTTCTCTCACCATTATATCTT
CGTTCTTCAACTCAGGATCGTTTTGTTCGTCCACCAGCAATTGTACATTGCAGCATTGTTATCTGGGTCCCTAGACAT
TGGTT

>amilFP

CACTTTTGTGGCATCAAGCGATCTGTTCCAAGCATGGCTCTGTCAAAGCACGTAAGTGAATTCTCAATTCGAACTA
GCGAGATGGTATTTAATAGCAAAGGTCTTTACCAAATTCGAGTGTTTAAATGTTGCGAGAATTTGTGTCGATTTTTGA
AGCCCTCACTTGGTTATTGAAATATCGGCTCCTGCAAGCTACTTCACGTAGAGAAACATTTAAACTCAGAACAGTAAA
CTTTTTTATTCCTGAATTAAACATGTCAAAGAAGTGTGTACCAGCCAAACACCTGGAAAACCAATATTTCCGGTTCCT
TTTATTTAAAGTGATAAGTTCTACCTATGGTGAAAATAGGTTGGTTCTTTGTTTTCCACGCAAACAAGCCGAAAATTAA
ACCTTGGATAACTTAAATACCAGAGATAAATAATCTTCAGTGTTCGGTCTTCGATCTTTTTTGCTTTCTCACAAATTTT
GTTGAGCGTCTTATGACTATTTTCAACCGCCCACTGCATTGGTCGAGGTGTGGAAGGCCTTGCCGGGCTCGAACCA
CGTTCGGACCAACACTCAGGATCTTACGATAACTACGGAAAAAGTGCTAGAACCTTTGCAATTACATCTGCACTGAT
TTCATAAGGCTCAAGTTTTCTGGAATAAGGACGGTACACCTTAGACGCCGTTTCATAACCCTGGCTTAACTTTATTGT
TTATAGACTTCGTAAGACGTTAAGAACTCGCACACTATTCGCAAAGAGCATTAAGGGCATAGAATTCCCGGTGTTCC
CGTGATTTGTCCTTGAATTGAATATTATTAATTGGAGCCTCTCTCTTTCGTATGATGCTGATCACCCTCGTGTAAGATT
AATCAAAGAAAAACAACTAACACTGAAAACCATAGTTAAACTTCAATCTTCTTTTATCTACTTGTAAGCCACGTTTATT
GTCATCTGGTGACAATTTTCTGCATTATGTGCATAG**GGTCTAACAAAGGACATGACGATGAAATACCGGATGGAAG**
GGTGTGTCGATGGGCATAAATTTGTGATCACGGGCCACGGCAATGGAAGTCCTTTCGAGTAAGTGGTTTAAAATTT
TTGTAAGATATAGGACAAGGACAAGGGCAAGTTTGACTAGTTAAAGTCGGAAAAGTTCGGACGTTCAGTAAAAACAG
CTTGGCCTAAGTTAAGCGCGTTAAAAATGCAAGCTGTCTTGCATATGAAAAGACAAGGGATTATCTCGCCTTAAGTTT
TTGACCGGAAATTGCTGCAGAGTTATACTACCTGGCGAGGATGGTATAACTTGCCCTGGTAAATGCCGCAGCCAAT
CACATTCACAAAGTATGAATGCGTATATGGAGACACTTAACATTTACCACAGTACAAGGTGAATGTTATAGTCGTTTA
ACAGTTCCTAAAGTTTGTTTTGAATAAACATTTTTTTTTCTTTTCATTTAGGCGCATCAAAGCTAGCCTTTAAGCTCCC
GAATAACACGAGTTGGACCCTGTAAATCGATTTGGACCCAGCTGTAAATCGATTTCATGTACACTGTTTTAGGTCATC
CTGCAATTTAGATACAATGATAATAGTTTAGTCCTCAACAGCTTTCCTTTTTGATATAAGGTTTTGAATTTGTATGCCC
CAATTTGTTAACACGATTTTTTTCGTATTTCTTAG**AGGGAAACAGACCATCAATCTGTGTGTGGTTGAAGGAGGACC**
CCTGCCATTCTCCGAAGACATTTTGTCTGCTGTGTTTACCTACGGAAACAGGGTCTTCACTGAATATCCTCAAGGC
ATGGTTGACTTTTTCAAGAATTCATGTCCAGCGGGATACACATGGCAAAGGTCTTTACTCTTTGAAGATGGAGCAG
TTTGCACAGCCAGTGCAGATATAACAGTGAGGTACATAAATTCACTCAGTCACTCTAGGATCTTAATAAAAAAAATG
GAAAGCTTTTGGTTTGCTCCATTTGAGACTTGAGGTTTTGCTCTTTGGTGGTAGTCAAAGCAAAGACAAATATTACCT
TGAGCCGCCAGGTTGCTGGTTGGGGAAGTCTCCTCTATGTTTTTGTTAGTCCCTGGTGCCCCAATATTAATTAACTT
TTTATCACTTTTAATTATCCAATGACCTTGATCTCCATTGCTATAGCCATTAGGTATTGAACCCTAAAATCAACCTGGA
TAGGGTTTTCTCCCTACGAATTTGAGACCTTTCCTCCCTTTAAACAACCAGTTAAAGCACAAATTTTGATTCCATGTTC
CATTTCGAATGTCGAAATTAAATACGGCTTGGACTGTGGTCACCTATTTGTATAAAGCCCTATGTATATAAGCACACC
CTTGCGCGAATGATTATTATTATTTTTTTTATAGTAGAGAATATATATGATATTACAGTATCATAAGAACGGTTGATCGT
CCCTTCAATTCTATTTGAAGTCTACAGAATTTGTTTCGTGAGGGGCCAGCCCTTGAGTTTACGATAATTCCTAATATC
TCATCGCGATCATGATCAATTACATCTGCATCACTCTTTGTGTATGTATACTCATCTACATTTCGCAAAAATATGATGC
GTGTTACAGCTACTCGCGAGTTCAGATCTTTTGTTAATGCCGCTTTGTCTCTACGTCTGCATAAATTGCATTGCTTCG
ATATAAGCCAGTAAGGATGAGCTTTGTCAATTATCGTCCAAGTGGTACGGTGCTGTATGTTCTTCTGTTTTAGCTCCC
ACACGACTTACTTAGCTCGGTTTGGGTGGAGTACATCTGCTTTTTAAAGGACTGTTTGTGGTTCCTAAAGTGGGACT
TAAACTCAGTGTCCGTCATGCCTATGTATGTTTCGCTTCCCTCATAGCCTGGTGAATGATACCCTTAAGTTGGGCATT
TGTTCTCGAATGGAAGACCAACCATTTCCCGTGATGAATCAAGATCCTATCGAGTGATTGCAGGGTACTGTTTGAGG
GTAACTGGAATACTGCATTCTATGATATTACGTATATTTCAACTGCGGAGTTGAGATGATTGATGGTAGAGATATCAT
CACAATGAAATGGGATGGTCTCTATCAGCAATCGAATTTTTTTTACTTTGATAG**TGTTGAGGAGAACTGCTTTTATCA**
CGAGTCCAAGTTTCATGGAGTGAACTTTCCTGCTGATGGACCTGTGATGAAAAAGATGACAACTAATTGGGAGCC
ATGCTGCGAGAAAATCATACCAGTACCTAGACAGGGGATACTGAAAGGGGATGTCCCCATGTACCTCCTTCTGAA
GGATGGTGGGCGTTACCGGTGCCAGTTCGACTCAGTTTACAAGTCATTTTGTTTCTACATAAATAGCTCTAAATTCT
TTTTTTGAGAGTGACAATCATTTGAAAGCACTTCTATTGATTTCATGATCACAACTTTGACATTGAGAATAGCTAGCTT
AGAATAGCAGAGCACAGGGAGCTCGTATCATGACAGAGACCATTTCTTTACAAAAAGTGAATTTTCACGTCTTCAGT
CGAGTCTTTCCTTGCATAATAAATAGTTCCATATTGAAATTCCACTAGCCGTAAACCACTTGAGCACCAAGGAAGAAA
TTAGATTAATTTCCCCAACCCGTTGAGTGTGCACAAAAATATTAAAAAATGTGTAAGAGACAATGGACCAGTCCTTGT

CTAAGCGCATGACAGGCGAGTTTCAATCAAGTTCACCGTTCTCCATTGCCCCAGGGACCATTGTCCACTTACC
AGATTTCCCTAGTGCCAAATGGCATTTAAAAAATGGCCTTCTTTCTTCCCTTCCTCTTACATTCCAGGATGGAATGCT
AAAGAATGATTCTTCTTTCTTCTCCACAG**AGCAAAGACTGACCCGAAAAAGATGCCGGAGTGGCACTTCATCCAAC**
ATAAGCTCACCCGGGAAGACCGCAGCGATGCTAAGAACCAGAAATGGCAACTGGCAGAACATTCTGTTGCTTCC
CGATCCGCATTGGCCTGATAAGAACATGATATAGTTCAAACATGTTGTTACATGCGCATGCTTATTTCTCTGATGA
CAATGTAGTYCSAACCAGGCCAGTAGAAATAAAKCACATTTGAAACATTCTACTTCGAGTTGTTTCGTTACATTAAT
TAATCCACTAAATAAATGCTCACAAGAAAATTAATCCATCAAACGTATTTATCAATAACCGCTTATTCCGTTACACTCG
CACCCCTCTCAAAAAGTCTCATCTGATTTCAAGCTGTTACCCATTCAAACAGCAGTTTGATCATCAAACTGCTGCTGA
ATTGTTCAATTTCTGAGAAAGCAAACTCGCTTTTTCTGAGGCAGTTATTTCTAATTTCTATAAGTCACAAACAACACAC
TTTAAATATTGGACAATCCATGTAAAAAAACCATATAATAACCTTGGAGTCCGAATAGATTTGTCCAAGACTTTCTCTT
TGGCAAATGATTTTTTTCTAATGGCCAGGTCGCGGTTCCTTTTTTCGTAGGAACTACCCCATACTGCTTATGCT