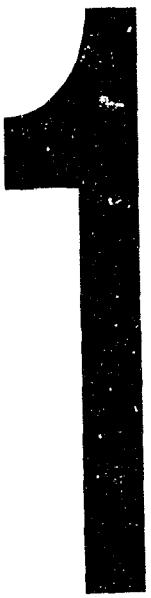

National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

Isolation and Characterization of Phase-specific cDNAs from Sporophytes and Gametophytes of *Porphyra purpurea* (Rhodophyta)

by

Qing Yan Liu

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
July, 1994

ISBN 0-612-05164-1

Canada

Name Qing Yan Liu

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Molecular Biology
SUBJECT TERM

0307
SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences

Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences

Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

DALHOUSIE UNIVERSITY

FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Isolation and Characterization of Phase-specific cDNAs from Sporophytes and Gametophytes of Porphyra purpurea (Rhodophyta)"

by QING YAN LIU

in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated 6/29/94

External Examiner [redacted]

Research Supervisor

Examining Committee

DALHOUSIE UNIVERSITY

DATE: July 5, 1994

AUTHOR: Qing Yan Liu

TITLE: Isolation and Characterization of Phase-specific cDNAs from Sporophytes and Gametsphytes of Porphyra purpurea (Rhodophyta)

DEPARTMENT OR SCHOOL: Biology

DEGREE: phD CONVOCATION: October YEAR: 1994

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

[redacted]

Signature of Author

Table of Contents

List of Figures

List of Tables

Statement of Retention of Copyright

Some of the research described in this thesis has been published (or submitted) in the scientific literature. I include references for these papers below. I retain copyright to these materials.

The research described in chapter 2 is also described in:

Liu, Q. Y., van der Meer, J. P. & Reith, M. E. (1994) Isolation and characterization of phase-specific cDNAs from sporophytes and gametophytes of *Porphyra purpurea* (Rhodophyta) using subtracted cDNA libraries. *J. Phycol.* 30:513-520.

The research described in chapter 3 is also described in:

Liu, Q. Y., Baldauf, S. L. & Reitn, M. E. (1994) Two different EF-1α genes from the red alga *Porphyra purpurea* show dissimilar developmental patterns of expression. *EMBO J.* submitted.

The research described in chapter 4 is also described in:

Liu, Q. L. & Reith, M. E. (1994) Isolation and sequence analysis of a sporophyte-specific cDNA encoding a serine protease-like protein from the red alga *Porphyra purpurea. Biochem. Biophys. Acta* submitted.

The research described in chapter 5 is also described in:

Liu, Q. L. & Reith, M. E. (1994) Isolation of a gametophyte-specific cDNA encoding a lipoxygenase from the red alga *Porphyra purpurea. Mol. Marine Biol.* submitted.

The research described in chapter 6 is also described in:

Liu, Q. Y., van der Meer, J. P. & Reith, M. E. (1994) A gametophyte-specific cDNA from the red alga *Porphyra purpurea* (Rhodophyta) encodes a protein with four apparent polysaccharide-binding domains. *J. Phycol.* submitted.

Abstract

The red alga *Porphyra purpurea* has a life cycle that alternates between shell-boring, filamentous sporophytes and free-living, foliose gametophytes. The significant differences between these two phases suggest that many genes should be developmentally regulated and expressed in a phase-specific manner. In this thesis, I prepared and screened subtracted cDNA libraries specific for the sporophyte and gametophyte of *P. purpurea*. This involved the construction of cDNA libraries, followed by the removal of common clones through subtractive hybridization. Of twenty putative phase-specific cDNAs selected from each subtracted library for further study, eight unique clones were obtained for the sporophyte and seven for the gametophyte. After confirming their phase-specificities by northern hybridizations, these 15 phase-specific cDNAs were sequenced and the deduced amino acid sequences were used to search protein data banks. The proteins encoded by two cDNAs from each phase were identified. One sporophyte-specific cDNA encodes an elongation factor 1α (EF-1α) normally involved in translation. Using this cDNA as a probe, the constitutively expressed EF-1α encoding gene (*tef*) was also isolated and sequenced. My results indicated that these are the only *tefs* in *P. purpurea*. The constitutively expressed gene encodes an EF-1α very similar to those of most eukaryotes; however, the sporophyte-specific EF-1α is one of the most divergent yet described. The other cDNA from the sporophyte encodes a serine protease-like protein that is structurally similar to mammalian tryptases that bind to sulphated glycosaminoglycans, suggesting that this protein may bind to sulphated galactans in the cell wall. One gametophyte-specific cDNA encodes a lipoxygenase for fatty acid metabolism and the other one encodes a protein containing repeated regions with strong similarity to fungal cellulose-binding domains. The results of *in vitro* translation studies of the latter cDNA suggest that, *in vivo*, the encoded protein enters the secretory system of the cell and may be a cell wall constituent that binds specific polysaccharides.

List of Abbreviations

bp	base pairs
cDNA	complementary DNA
DNA	deoxyribonucleic acid
EDTA	ethylenediamine tetraacetic acid
ER	endoplasmic reticulum
IPTG	isopropyl-thio-ß-D-galactoside
HEPES	N-2-hydroxyethylpiperacine-N'-2-ethanesulphonic acid
kb	kilobases
LB	Luria-Bertani
mRNA	messenger ribonucleic acid
mU	milliunits
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
polyA$^+$	polyadenylated
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
SM	sodium chloride and magnesium suspension medium
SSC	sodium chloride, sodium citrate solution
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactoside
YT	yeast tryptone

Acknowledgements

First and foremost, I would like to express my sincere appreciation to my supervisors Drs. John P. van der Meer and Michael E. Reith for their guidance, patience, encouragement and friendship over the past four years. I am also grateful to the members of my supervisory committee, Drs. Jonathan Wright and Tom MacRae for their valuable advice and for reviewing my thesis.

I appreciate the generosity of the National Research Council (Institute for Marine Biosciences) for providing financial support and laboratory facilities. Many people at NRC have given their help and support. I am thankful to the technical officers, Janet Munholland, Colleen Murphy, Jeff Gallant and Rama Singh for their assistance and for maintaining excellent working conditions in the laboratory. As well, I am grateful to Dr. Sandra Baldauf for helping me with the phylogenetic analysis, to Dr. Ron MacKay for his suggestions and technical guidance, and to Drs. Mark Ragan and Maurice Laycock for valuable discussions.

My special thanks go to my husband Wei Ding who has been very supportive and understanding during my study.

Chapter 1: General Introduction

The development of multicellular organisms involves a complex series of events that is generally assumed to be controlled at the level of gene expression. Thus, an understanding of how development is controlled requires information on the regulation of gene expression. With the growing accumulation of information on DNA structure and function, it is becoming evident that the molecular mechanisms controlling gene activity are broadly similar in all organisms but that they can differ considerably in detail from species to species and from gene to gene.

While vascular and nonvascular plants share a number of developmental features (Goldberg 1988), some unique characteristics of nonvascular plants offer experimental advantages over vascular plants for the study of common processes. For example, all plants undergo an alternation of a diploid spore-producing generation (sporophyte) with a haploid gamete-producing generation (gametophyte) during their life cycle. In vascular plants, the gametophyte is substantially reduced and is dependent on the sporophyte for nutrition and support. In nonvascular plants, the gametophyte is usually free-living and occupies a prominent position in the life cycle. In some algae, the gametophyte is the dominant phase of the life cycle with only a short period spent in the sporophytic phase. Thus, nonvascular plants, and particularly algae, are ideal for the study of processes that are dominant in the gametophyte.

Although morphologically simple multicellular organisms, macrophytic algae undergo a number of intriguing developmental processes. Among the red algae, members of the genus *Porphyra* (subclass Bangiophycideae) have a particularly interesting life cycle (Bold and Wynne 1985) that alternates between two dissimilar phases: the filamentous sporophyte, usually referred to as the "conchocelis" (Drew 1949), and the foliose gametophyte. The transition between these two phases is usually, but not always, associated with changes in chromosome ploidy level.

In nature, the conchocelis filaments bore into calcareous mollusc shells and emerge only to sporulate. Conchocelis filaments can also be cultured free of shells, using appropriate seawater medium. The cells of the conchocelis have peripheral chloroplasts (Pueschel and Cole 1985) and cell walls composed primarily of cellulose and galactans (Mukai et al. 1981). Between adjacent cells, proteinaceous "pit connections" (lens-shaped plugs) are present (Bourne et al. 1970). In contrast, the large leafy thalli of the gametophytes are conspicuous members of the benthic seaweed community, usually found attached to rocks, pebbles or other seaweeds. Cells of the gametophyte have a single, central, stellate chloroplast. The cell walls of the fronds contain mostly xylan and mannan, but no cellulose. Pit connections are absent in this phase. The striking differences between the two *Porphyra* life history phases offer an excellent opportunity to study developmentally regulated genes.

The *Porphyra* species I chose to work with is *Porphyra purpurea* (Roth)

C. Agardh. This species grows on the marine mudflats at Avonport, Nova Scotia, and its genetic and reproductive characteristics (Mitman and van der Meer 1994), as well as its chloroplast genome (Reith and Munholland 1993), have been described. In addition to the common characteristics of *Porphyra*, *P. purpurea* (Fig. 1.1) has a monoecious thallus that becomes longitudinally divided into a male part on one side and a more greenish female part on the other. Meiosis takes place in the germinating conchospore and sexual determination results from genetic segregation. The sporophyte of *P. purpurea* naturally grows in calcareous mud, but can be maintained as stock in culture (little or no growth) under very low light intensity (less than 5 $\mu E \cdot m^{-2} \cdot s^{-1}$). When the conchocelis is exposed to more intense light (30-40 $\mu E \cdot m^{-2} \cdot s^{-1}$), conchosporangial structures develop. This conchosporangial conchocelis can grow vigorously in culture by vegetative propagation and has therefore been used as material representing the sporophyte. Conchospore release and germination can be induced by shifting the growth temperature from 15°C to 13°C and the photoperiod from 16L:8D to 8L:16D (Mitman and van der Meer 1994).

As an initial approach to understanding the developmental regulation of gene expression in *P. purpurea*, the objective of my study was to isolate and characterize cDNAs for genes that are expressed in a phase-specific manner. This study could then provide a foundation for further study of the molecular aspects of development in *P. purpurea.*

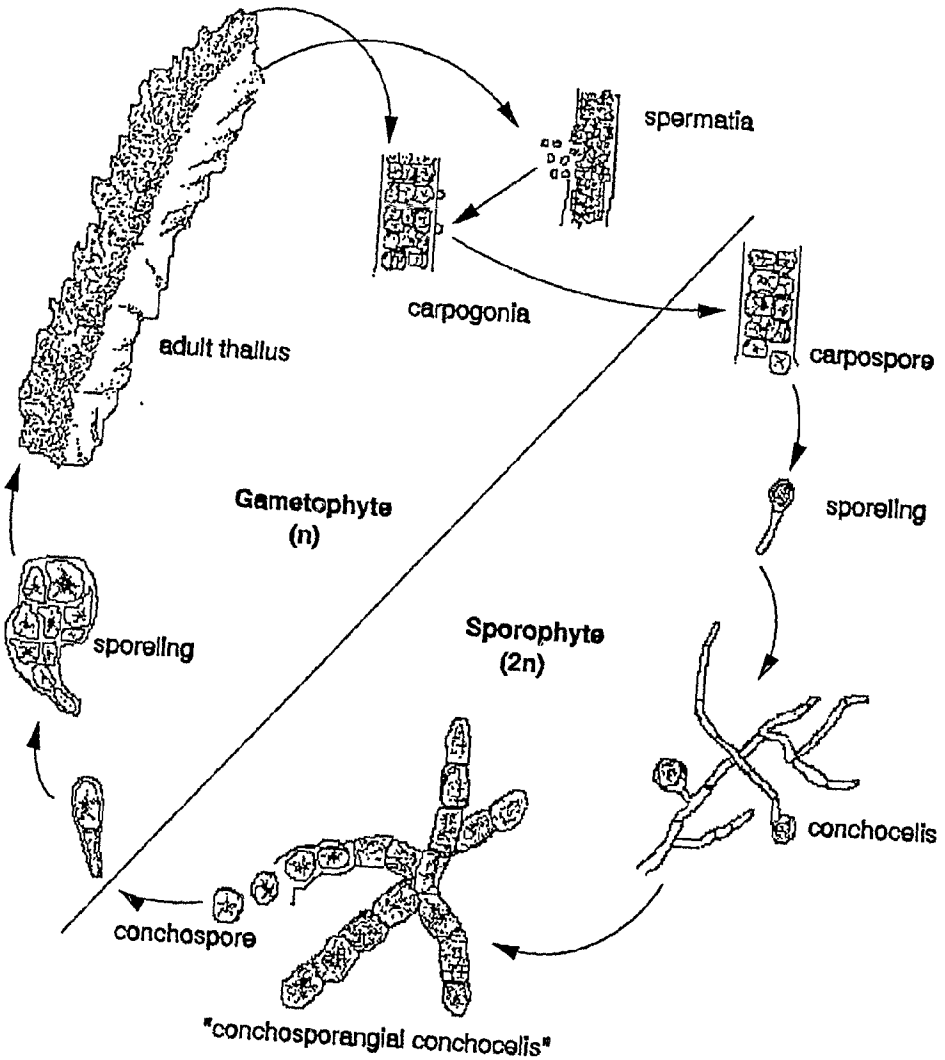

Figure 1.1 Life history of *Porphyra purpurea*.

Chapter 2: Isolation and Characterization of Phase-specific cDNAs Using Subtracted cDNA Libraries

INTRODUCTION

Early attempts to study phase-specific gene expression in algae were focused on the differential detection of proteins. Using tissue from gametophytes and sporophytes of *Ulva mutabilis*, Hoxmark (1976) found a protein that appeared to be present only in the sporophyte; however this difference was not apparent in a reexamination of this species (Hushovd et al. 1982). The techniques available at the time were simply too crude to obtain satisfactory results. This situation has changed dramatically with the advent of modern RNA and DNA technology.

Molecular biology studies of development now use two general approaches for the isolation of tissue- and developmental stage-specific mRNA sequences. These methods do not require any prior knowledge of the target genes or their products. The first method is the differential screening of a cDNA library constructed from a tissue of interest, by probing with two sets of labelled first-strand cDNAs. One probe is prepared from the same mRNA used to construct the library, whereas the second probe is derived from mRNA extracted from one or more control tissues. Clones from the library that do not hybridize with the control probe are specific to the target tissue (Conkling et al. 1990, John and Crow 1992). In a more sophisticated approach, "subtracted

libraries" are produced from two cDNA libraries by the removal of common sequences to enrich for sequences specific to the tissue of interest. For this approach, single-stranded cDNAs derived from the target tissue are mixed with single-stranded cDNAs or mRNAs from another tissue. Sequences that are common to both tissues hybridize, and the resultant double-stranded DNAs are removed from the mixture. The putative tissue-specific cDNAs remaining in solution are then used to produce the subtracted cDNA library or, in some cases, a probe. The advantage of libraries formed by subtractive hybridizations is that they are enriched for differentially expressed mRNA sequences, and thus are much easier to screen than a complete cDNA library (Duguid et al. 1988, 1989, Schweinfest et al. 1990, Kelly et al. 1990).

I have initiated a study of developmental regulation of gene expression in the red algae by constructing subtracted cDNA libraries for both the sporophyte and the gametophyte of *Porphyra purpurea*. In this chapter, I describe improvements to existing protocols for the construction and screening of subtracted cDNA libraries. I also describe a series of library screening procedures that led to the isolation of 15 phase-specific cDNAs and report the preliminary results obtained in their characterization.

MATERIALS AND METHODS

Algal Materials and Standard Molecular Biology Techniques

P. purpurea gametophytes and sporophytes were cultured in modified

D-11 medium at 15°C, under a 16:8 h light:dark cycle (Mitman and van der Meer 1994). Cultures were harvested two hours after initiation of a light period so that both developmental phases were in a comparable physiological condition. Total RNA was isolated according to MacKay and Gallant (1991) and polyA$^+$ RNA was purified by multiple passage through oligo(dT) cellulose. Standard methods for phage and bacterial growth, plasmid DNA isolation, DNA and RNA gel electrophoresis, Southern and northern blotting and labelling of first strand cDNAs were according to Sambrook et al. (1989).

cDNA Library Construction

Double-stranded cDNA was synthesized from polyA$^+$ RNA according to the method of Gubler and Hoffman (1983) (cDNA Synthesis System, Invitrogen). cDNA size selection, *Not*I/*Eco*RI linker-adapter addition and T4 polynucleotide kinase treatment were carried out according to the manufacturer's instructions. De-phosphorylated, *Eco*RI-cleaved λ ZAP II vector and Gigapack II Plus packaging extract (Stratagene) were used according to the supplier's recommendations. These procedures yield a cDNA library with the inserts oriented in both directions. Four-fifths of the original cDNA library were amplified using Stratagene's procedure.

Rescue and Purification of Single-stranded cDNAs from λZAPII Libraries

To isolate single-stranded DNA from the cDNA libraries, five ml of

Escherichia coli XL1-Blue cells (Stratagene) in SM buffer at OD_{600} = 1.0, 2 ml of the amplified λ ZAP II library containing 10^9 plaque forming units (pfu), and an appropriate amount of VCSM13 helper phage (Stratagene, see below) were combined in a 50 ml conical tube. The final volume was adjusted to 10 ml with SM buffer. After phage adsorption at 37 °C for 15 min, this phage-bacteria mixture was added to a 250 ml flask containing 100 ml 2 x YT medium and grown at 37 °C for 3 h, with vigorous agitation. The culture was then centrifuged for 30 min at 12,700 x *g* to pellet cellular debris. Single-stranded recombinant phage and helper phage were precipitated from the supernatant by adding one quarter volume of 20% (w/v) polyethylene glycol (8,000 Daltons), 3.5 M ammonium acetate. After 2-4 h at 4 °C, the precipitate was pelleted by centrifugation at 15,700 x *g* for 25 min. Contaminating bacterial nucleic acids were removed by dissolving the phage pellet in 4 ml SM (minus gelatin) containing 10 µg DNase I and 10 µg RNase A and incubating the phage suspension for 30 min at room temperature. The phage particles were lysed by the addition of SDS to 0.5% (w/v) and successive extraction with phenol, phenol:chloroform (1:1) and chloroform. Finally, the single-stranded phagemid DNA was precipitated with one tenth of a volume of 3 M sodium acetate and 2 volumes of ethanol.

Different helper phages and individual preparations of a given helper phage were found to differ in their abilities to rescue recombinant phagemid genomes. Helper phage strains R408, VCSM13, MMP (all from Stratagene)

and M13K07 (Pharmacia) were tested. The best ratio of recombinant λ ZAP II bacteriophage and helper phage had to be empirically determined. Typically, a series of quantities (10^5-10^7 pfu) of helper phage were used. For each attempt, 1/20 of the single-stranded DNA product was subjected to electrophoresis on a 1% agarose gel. Upon ethidium bromide staining, a faint band of double-stranded λ ZAP II DNA was visible at the top of the gel, a major band corresponding to the helper phage DNA occurred below and a long smear of phagemid single-stranded DNA ran below the helper phage band. The intensity and length of this smear were used to judge the effectiveness of the single-stranded DNA rescue.

Residual double-stranded DNA was removed by two rounds of digestion with restriction endonucleases (*Pvu* II and *Bgl* I) and magnesium-phenol extraction (Rubenstein et al. 1990). Siliconized microcentrifuge tubes were used for all experiments involving single-stranded DNA.

Construction of Subtracted cDNA Libraries

For subtractive hybridization, 50 µg of driver phase single-stranded cDNA was labelled with 75 µg of long-arm photobiotin (Clontech) (Rubenstein et al. 1990) and precipitated with 5 µg of single-stranded cDNA from the target phase. These DNAs were dissolved in 20 µl hybridization buffer containing 0.5 M NaCl, 50 mM HEPES pH 7.6, 2 mM EDTA and 40% (v/v) deionized formamide (Rubenstein et al. 1990). The sample was incubated at 68-70 °C for

5 min, sealed in a hybridization bag and immersed in a 52 °C water bath for 21 h.

To remove biotinylated DNA, as well as DNA hybridized to the biotinylated DNA, the hybridization mixture was diluted to 600 µl with binding buffer (100 mM Tris pH 7.5, 150 mM NaCl) and incubated with 250 mg Vectrex-avidin (Vector Laboratories), by gently rotating at 22 °C for 1-2 h (Swaroop et al. 1991). The beads with most of the biotinylated DNA were removed by centrifugation, leaving the target single-stranded cDNA in solution. The bead pellet was washed once with binding buffer and the wash supernatant was combined with the first supernatant. To determine whether or not the biotinylated DNA had been totally removed, 5 µl of supernatant was applied to a nylon membrane, incubated with streptavidin-alkaline phosphatase conjugate (Clontech) and stained for alkaline phosphatase activity according to the manufacturer's instructions. Known amounts of biotinylated DNA were applied to the same membrane as standards. After one round of Vectrex-avidin extraction, there was usually a small amount of biotinylated DNA left in the supernatant. To remove the remaining biotinylated DNA, the supernatant was first reduced to 200 µl by centrifugal ultrafiltration in a Centricon 100 device (Amicon) . The concentrate was then added to 24 mg of streptavidin-coated MagneSphere particles (Promega, prewashed twice with 0.5 x SSC containing 50 µg ml^{-1} sheared and denatured salmon sperm DNA) and incubated at 22 °C for 10 min. The particles and bound biotinylated DNA were trapped to the wall

of the tube with a magnetic stand. The supernatant was removed and single-stranded DNA was precipitated with ethanol in the presence of 5 µg of *E. coli* tRNA.

Single-stranded phagemid DNA was made double-stranded prior to transformation of *E. coli.* Single-stranded DNA was resuspended in 10 µl of annealing mixture containing 6.75 µl H_2O, 1.25 µl M13 reverse primer (3.75 ng) and 2 µl of Sequenase reaction buffer (US Biochemicals), incubated at 68 °C for 10 min and cooled over approximately 2 h to 30 °C. The annealing mixture was diluted to 20 µl by the addition of dithiothreitol to 5 mM, all four deoxynucleotides to 1.25 mM, and 2 units of Sequenase (US Biochemicals), and incubated at 37 °C for 1 h. Reactions with other DNA polymerases used the same concentrations of primer and deoxynucleotides in appropriate buffers (Sambrook et al. 1989). Five µl were used to transform 100 µl of *E. coli* XL1-Blue supercompetent cells. The bacteria were plated on LB-agar plates, supplemented with ampicillin, X-gal and IPTG (Sambrook et al. 1989). White colonies were picked into 96-well microtitre plates and stored at -80 °C as described by Schweinfest et al. (1990).

Screening of Subtracted cDNA Libraries

To test the phase-specificity of the subtracted cDNA libraries, colonies from each library were hybridized with labelled first-strand cDNAs from both the sporophyte and the gametophyte. Recombinant colonies from subtracted

libraries were grown in 96-well microtitre plates. Six 11.5 cm × 7.5 cm autoclaved Zeta Probe membranes (Bio-Rad) were placed on a 24.5 cm × 24.5 cm × 2.5 cm LB-ampicillin agar plate (Nunc Bio-Assay Dish) and three sets of 96 colonies from the subtracted library were imprinted in duplicate (one set of 96 colonies per membrane). The plates were incubated overnight at 37 °C. Membranes bearing 2-3 mm colonies were treated according to Sambrook et al. (1989). The two replica membranes of each 96-well plate were then hybridized separately with ^{32}P-labelled first-strand cDNAs synthesized from polyA$^+$ RNA of either the sporophyte or the gametophyte. Hybridizations were carried out at 68 °C in a hybridization buffer used by MacKay and Gallant (1991). Membranes were washed twice in 6 × SSC, 0.1% SDS, four times in 2 × SSC, 0.1% SDS and twice in 0.5 × SSC, 0.1% SDS; each wash was at 68 °C for 15 min. Colonies that hybridized with only one of the gametophyte or sporophyte first-strand cDNA probes were selected for further studies.

cDNA Characterization

Twenty positive colonies from each subtracted library were chosen for further investigation. Since more than one colony may have been selected for each expressed gene, these sets of colonies were checked for genetic relatedness by Southern hybridization. Initially, the cDNA insert from one colony was hybridized to the other 19 cDNA inserts. After removing duplicates, a second insert was hybridized to the remaining cDNAs, and so on, until all of

the duplications had been identified. cDNA probes were prepared by *EcoR*I digestion of clones and purification from agarose gels using GeneClean (Bio 101). All probe labelling and Southern hybridizations were done using the ECL kit (Amersham) according to the manufacturer's instructions.

Northern hybridizations were performed under the same conditions as used above for testing the specificity of clones. A ß-tubulin cDNA of a mRNA species produced in both gametophyte and sporophyte (MacKay and Gallant unpublished results) was used as a control probe.

Double-stranded plasmid templates were sequenced in both directions using synthetic oligonucleotide primers and a sequencing kit (US Biochemicals). Sequences were analyzed with the IBI/Pustell programs (Pustell and Kafatos 1984). Data bank searches and similarity analyses were carried out with the BLITZ (Smith and Waterman 1981), FASTA (Pearson and Lipman 1988) and BLAST (Altschul et al. 1990) programs.

RESULTS

By modifying established methods (Rubenstein et al. 1990, Schweinfest et al. 1990, Swaroop et al. 1991), subtracted cDNA libraries for gametophytic and sporophytic tissues of *Porphyra purpurea* were constructed. A general outline of our procedure is illustrated in Figure 2.1. In the first steps, two cDNA libraries were generated in a bacteriophage λ-based cloning vector, λ ZAP II, that allows the production of single-stranded cDNAs. A key feature of λ ZAP II

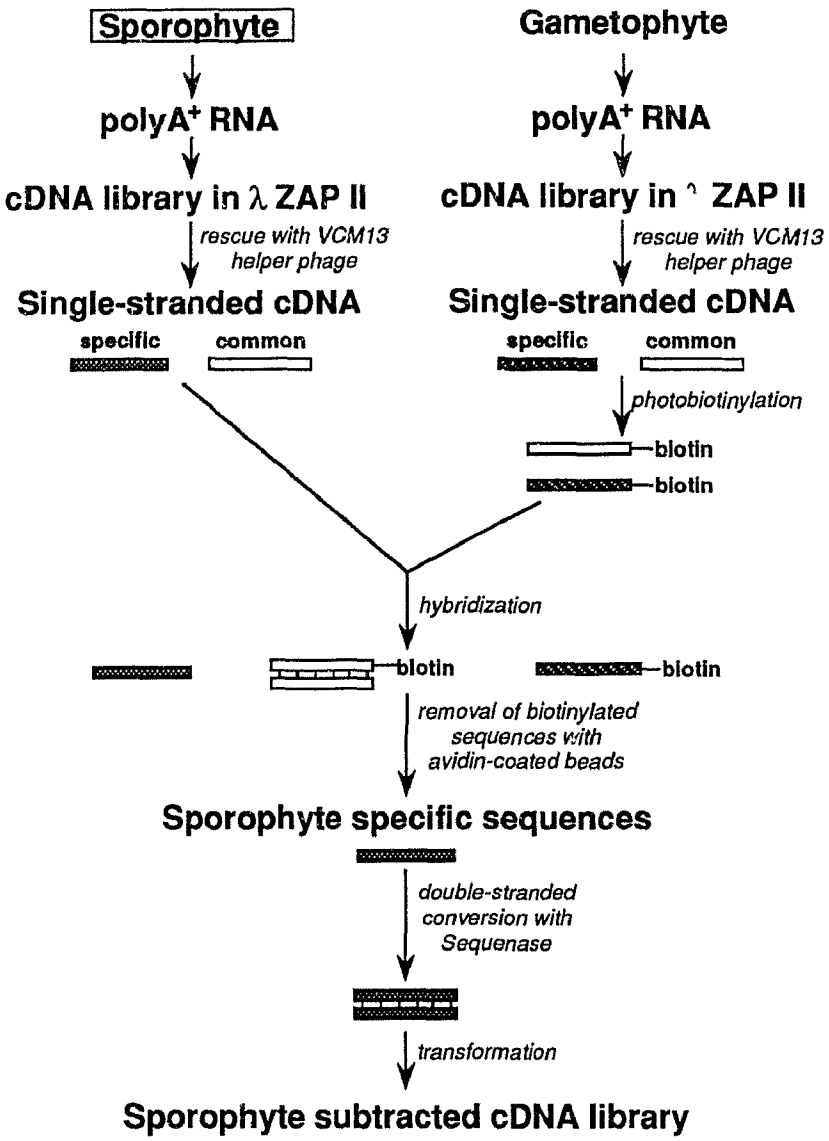

Figure 2.1 Flow chart for the construction of a subtracted cDNA library that is specific for the sporophytic phase. In practice, subtracted libraries were constructed for both phases. Details are presented in Materials and Methods and in Results.

is that a linearized phagemid, that is, a plasmid with a phage f1 origin of replication, has been incorporated into the λ genome. Co-infection of *E. coli* with phage from the λ ZAP II library and f1 helper phage results in single-stranded replication of the phagemid portion of λ ZAP II, which contains the cloned cDNA sequences, and the packaging of the single-stranded DNA into f1 phage particles. The resultant single-stranded DNAs generated for each phase were purified and aliquots of each were labelled with biotin. For the subtractive hybridizations, unlabelled single-stranded cDNA of the target phase was mixed with excess biotinylated cDNA from the driver phase. Biotinylated DNA, and any hybridizing sequences, were removed with streptavidin beads and the remaining single-stranded cDNA was made double-stranded with Sequenase. *E. coli* was transformed with the product to give the final subtracted cDNA libraries.

During the construction of the subtracted cDNA libraries, I found that the standard procedures for the isolation of single-stranded phagemid DNA, while suitable for individual phagemid clones, did not give satisfactory results for the rescue of a population of recombinant phagemid genomes from a λ ZAP II library. Both the quality and quantity of single-stranded cDNAs obtained by this technique appeared to be affected by the strain of helper phage that was used, the ratio of λ ZAP II recombinant bacteriophage to helper phage and the duration of infection. Various strains of helper phage (R408, VCSM13, M13K07 and MMP) were tested by analyzing the size distribution of the rescued single-

stranded DNA on agarose gels (data not shown). Only the use of R408 and VCSM13 gave a good smear of single-stranded DNA. Compared to R408, VCSM13 appeared to produce a higher yield and wider size range of single-stranded DNA. A 3 h infection period was sufficient to produce approximately 30 µg of single-stranded cDNA. A growth period shorter than 2 h reduced the yield significantly, while an infection period longer than 5 h appeared to give preferential rescue of certain individual clones, which were apparent as distinct bands by gel electrophoresis. There was also more helper phage and λ ZAP II DNA produced after prolonged growth of an infected culture.

After the subtractive hybridization, Vectrex-avidin and MagneSphere particles were used to remove the biotinylated DNAs. 250 mg of Vectrex-avidin removed approximately 80% of the biotinylated cDNA as judged by our biotin assay (see Materials and Methods). Because the Vectrex-avidin beads occupy a relatively large volume (approximately 1 ml), some single-stranded target DNA is unavoidably trapped between the beads. Repeating the Vectrex-avidin treatment to remove the remaining biotinylated DNA would have resulted in unacceptable losses of single-stranded target DNA. MagneSphere particles are much smaller and were thus used to remove residual biotinylated DNA. The lower biotin-binding capacity of the MagneSphere particles precluded their exclusive use.

To ensure efficient bacterial transformation, the phase-specific, single-stranded cDNAs were converted into double-stranded cDNAs. As previously

reported (Rubenstein et al. 1990), the transformation efficiency of double-stranded phagemid DNA was approximately 1000 times greater than single-stranded DNA. T4 DNA polymerase, Klenow fragment of DNA polymerase I, T7 DNA polymerase and Sequenase were tested for performing the double-stranded conversion. Double-stranded DNA prepared with Sequenase yielded twice as many colonies as that obtained from T7 polymerase, and several hundred-fold more than the product from T4 DNA polymerase or Klenow fragment.

Transformation of *E. coli* with one quarter of each of the available double-stranded recombinant phagemid DNAs resulted in two phase-specific subtracted cDNA libraries. Approximately 3000 recombinant (white) colonies were obtained for the sporophyte and 4,600 for the gametophyte. Thus the total number of white colonies that could be obtained from the subtracted cDNAs were 1.2×10^4 and 1.8×10^4 for sporophyte and gametophyte libraries, respectively.

Individual clones of the subtracted cDNA libraries were screened for phase-specificity by hybridizing with labelled first-strand cDNAs from both phases. Of approximately 2700 colonies screened from the sporophyte library, 266 showed good hybridization to sporophyte cDNA and no apparent hybridization to gametophyte cDNA (Fig. 2.2). Similarly, 180 of approximately 2300 colonies from the gametophyte library were apparently gametophyte-specific. Thus, 8 to 10% of the recombinant colonies in the subtracted cDNA

Figure 2.2 Specificity test of the subtracted cDNA libraries. Two replica membranes with 96 colonies were hybridized with ^{32}P-labelled first-strand cDNAs prepared from polyA$^+$ RNA of either the sporophyte or the gametophyte. **A**) Hybridization of sporophyte first-strand cDNA to colonies from the sporophyte-specific subtracted cDNA library. **B**) A duplicate membrane containing the same colonies as in "A)" hybridized to the gametophyte first-strand cDNA. **C**) Hybridization of sporophyte first-strand cDNA to colonies from the gametophyte-specific subtracted cDNA library. **D**) A duplicate membrane containing the same colonies as in "C)" hybridized to the gametophyte first-strand cDNA.

libraries were evidently phase-specific. Among the negative colonies from both subtracted libraries, most produced signals on both membranes but gave stronger signals with the cDNA probe from the expected phase. These appear to represent mRNAs that are differentially synthesized, but are not phase-specific. In some cases, different members of a gene family might be expressed in the two phases and have different affinities for the cDNAs in the recombinant colonies. There was also a small number of colonies that didn't hybridize to either labelled cDNA. These clones may encode cDNAs that are expressed at very low levels and could be phase-specific. A very low percentage of colonies showed equal hybridization signals to both labelled cDNAs.

Twenty putative, phase-specific colonies from each of the subtracted gametophyte and sporophyte libraries were chosen for further investigation. Since the intensity of the hybridization signal for each colony reflects the relative abundance of the corresponding mRNA species (Fig. 2.2), I was able to select colonies that represented a spectrum of mRNA levels in the source tissue. Because each cDNA in the subtracted library may be represented by more than one colony, the selected colonies were first checked for relatedness by Southern hybridization. It was assumed that any cDNAs that hybridized with each other most likely originated from the same gene or from different members of a gene family. Where cross-hybridizing cDNAs were detected, clones with the longest inserts were selected for further characterization. Eleven unique

clones were obtained from the sporophyte-specific library and seven from the gametophyte-specific library.

To confirm that these clones truly represented mRNAs expressed only in the sporophyte or only in the gametophyte, their cDNA inserts were hybridized to mRNA from both phases in northern hybridization experiments (Fig. 2.3). Eight cDNAs from the subtracted sporophyte library clones hybridized to the sporophyte mRNA but not to that of the gametophyte, confirming that their respective genes were only transcribed in the sporophyte phase. cDNAs from three of the selected clones hybridized to mRNAs from both phases. Two of these showed substantially stronger hybridization to sporophyte mRNA, suggesting that the respective genes are subject to differential but not phase-specific regulation (data not shown). Since the objective of this study was to isolate unambiguously phase-specific cDNAs, these three were not further characterized. The same protocol was repeated to confirm the phase-specificity of the seven gametophytic cDNAs. In this case, all seven clones hybridized only with the gametophyte mRNA and not at all with that of the sporophyte.

The 15 phase-specific cDNAs identified above were sequenced as the first step toward the identification of the corresponding genes. It was discovered that none of the cDNAs had a complete 5′ terminal sequence containing the protein synthesis initiation codon; however, all 15 contained one or more open reading frames (ORFs). The amino acid sequences encoded by

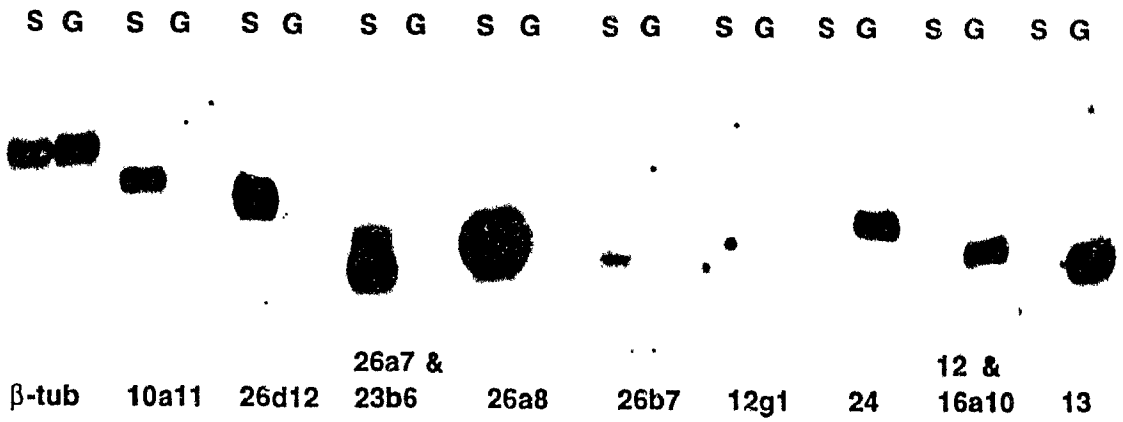

Figure 2.3 Northern analysis of the 11 unidentified phase-specific cDNA clones. One μg polyA+ RNA from the sporophyte (S lane in each pair of lanes) and the gametophyte (G lane) were electrophoresed on a formaldehyde/agarose gel, blotted to a nylon membrane and hybridized to a ^{32}P-labelled cDNA insert. Clone names of the labelled probes are indicated under each pair of lanes. Hybridization of a β-tubulin cDNA probe derived from an mRNA occuring in both phases is shown as a control. Clones 10a11, 26d12, 26a7, 23b6, 26a8 and 26b7 are from the sporophyte-specific library while clones 12g1, 24, 12, 16a10 and 13 are from the gametophyte-specific library. Exposure times varied from 1-72 h, depending on the intensity of the signal.

these ORFs were used to search data banks for similar proteins. The proteins encoded by two cDNAs from each phase were identified in this manner. For the sporophyte, one cDNA encodes a member of the elongation factor 1α (EF-1α) gene family involved in translation. The other cDNA encodes a serine protease-like protein. One of the identified gametophyte cDNAs encodes a protein containing multiple, non-identical regions with sequence similarity to cellulose-binding domains found in fungal cellulases. As there is little or no cellulose in *Porphyra* gametophytes (Mukai et al. 1981), these repeated motifs might be binding domains for other polysaccharides with similar structure. The other gametophyte cDNA encodes a lipoxygenase involved in fatty acid metabolism. The detailed characterization of these four genes will be described in subsequent chapters.

The putative proteins encoded by the remaining 11 phase-specific cDNAs did not show significant similarity to any proteins in the data banks, suggesting that their counterparts in other organisms have not yet been sequenced or have diverged beyond easy recognition. The established characteristics of these unidentified cDNAs from the sporophyte and the gametophyte are summarized in Table 2.1. Although these sequences for the most part are unremarkable, there were a few interesting observations. In the sporophyte phase, the cDNAs from clones 26a7 and 23b6 represent two members of a gene family. Both cDNAs hybridize to mRNAs of 0.8 and 1.0 kb (Fig. 2.3), but because cDNA 23b6 is 1008 bp long, it could only be derived

Table 2.1 Unidentified phase-specific cDNAs.

Clone #	Insert Size (bp)	mRNA Size (kb)	Abundance[a]	GenBank Accession #
sporophyte-specific				
10a11	510	1.5	++	U04722
26d12[b]	599	1.4	++	U04732
26a7[c]	746	0.8, 1.0	++	U04729
23b6[c]	1008	1.0	++	U04727
26a8	732	1.0	++++	U04731
26b7	632	0.9	+	U04730
gametophyte-specific				
12g1	558	2.6	+	U04724
24	335	1.2	++	U04728
12[d]	466	1.0	++++	U04723
16a10[d]	717	1.0	++++	U04726
13	553	0.9	++	U04725

[a] abundance scores are based on northern hybridization results
[b] contains repeated "ACG" motifs
[c] related cDNAs
[d] related cDNAs

from the longer mRNA. These cDNAs encode two very similar proteins with identical carboxy-terminal ends. The main difference between the two cDNAs is the presence of a longer 3′ untranslated region in 23b6. The two cDNAs are 88% identical, while their encoded amino acid sequences are 83% identical. The DNA sequence differences are due to single base substitutions, rather than insertions or deletions. Clone 26d12 contains "..ACGACGACGACG.." strings of various lengths in its coding region. The nature of these tandemly repeated codons will depend upon the identification of the correct reading frame for the cDNA. The longest ORF is in the "ACG" (threonine) reading frame but the "GAC" (aspartic acid) frame cannot be completely ruled out. Clones 12 and 16a10 from the gametophytic phase again represent different members of a gene family, but in this case the two mRNAs are the same size (Fig. 2.3). These two clones represent highly abundant mRNAs and have 77% DNA sequence identity. The sequences of these 11 unidentified cDNAs will appear in the GenBank DNA Database (Table 2.1).

DISCUSSION

In this chapter I describe the initial results from a long-term project designed to examine the molecular mechanisms regulating phase-specific gene expression in *Porphyra*. My first goal was to isolate genes that are unambiguously phase-specific in their expression. To initiate this work, I developed a cloning strategy based on the subtractive hybridization of cDNA

libraries. Although various techniques for the construction of such libraries have been described in the literature (Rubenstein et al. 1990, Schweinfest et al. 1991, Swaroop et al. 1991, Rodriguez and Chader 1992), I was unable to obtain subtracted libraries containing cDNAs that adequately represented an entire population of phase-specific mRNAs by following any of these protocols. By combining and modifying useful aspects of the methods reported by others, I have succeeded in constructing two subtracted cDNA libraries enriched for phase-specific genes expressed in the gametophytes and sporophytes of *Porphyra purpurea*.

A number of points regarding the more important methodological adaptations deserve additional comment. In my protocol, the strain and quantity of helper phage were first carefully selected, and an appropriate infection period was determined experimentally because variations in these parameters had a substantial influence on the quality and quantity of the single-stranded phagemid cDNA pools obtained for the subtractive hybridizations. There are several DNA synthesis processes that take place simultaneously and even compete with each other in *E. coli* cells after co-infection with λZAP II and helper phage. Conditions that favour production of single-stranded cDNA while minimizing unwanted side reactions have to be defined empirically. Even so, side reactions such as the conversion of single-stranded into double-stranded phagemid cDNA will still be present to some extent, and thus the additional restriction endonuclease clean-up steps suggested by Rubenstein et

al. (1990) were retained in my protocol to help ensure a high specificity of the subtracted libraries. Double-stranded phagemid cDNA would not be removed by the hybridization step and would be retained as non-specific clones in the subtracted library. In another adaptation, I used both Vectrex-avidin beads and streptavidin-coated MagneSphere particles to remove the biotinylated DNA after the subtractive hybridization. I found that the consecutive use of both types of beads allowed efficient and economical removal of biotinylated DNA with minimal loss of target DNA. The conversion from single-stranded to double-stranded cDNA prior to bacterial transformation was another crucial step for ensuring success. In contrast to the use of the Klenow fragment of DNA polymerase I (Schweinfest et al. 1990), I found that double-stranded DNA prepared with Sequenase produced several hundred-fold more colonies after transformation. This improvement is likely due to the absence of both exonuclease and strand displacement activities in Sequenase (Sambrook et al. 1989).

The subtracted cDNA libraries obtained using my cloning procedure contained a large number of phase-specific clones. This outcome indicates that there are many transcriptionally regulated genes making contributions to the differences between the two phases of *P. purpurea*. The 15 phase-specific cDNAs I isolated are derived from mRNAs that are produced at various levels, indicating that my subtraction techniques have the sensitivity necessary to identify low-abundance, phase-specific mRNAs. There were some recombinant

colonies from the subtracted libraries that did not show hybridization to either sporophyte or gametophyte first-strand cDNA probes. Such colonies might contain cDNAs derived from very rare mRNAs that encode regulatory proteins. This possibility needs further examination. I also obtained several clones derived from abundant mRNA species. These may encode structural proteins needed at high levels such as those presumably found in cell walls and cuticles (both of which have different compositions in the two phases of *Porphyra*). It is interesting that the gene encoding the putative cellulose-binding protein was one of those expressed at a high level.

One of the surprising results of genomic sequencing efforts is the large number of genes being discovered that encode proteins with no known homologues. As many as half the sequenced genes of simple organisms have no identified function (Cole and Hamilton 1993). I have encountered this same situation in *Porphyra*. It is not yet possible to predict the functions of the proteins encoded by the 11 unidentified phase-specific cDNAs I have isolated. Nevertheless, these cDNAs can be used to retrieve genomic clones for the isolation of red algal promoters sensitive to phase-specific regulation. Such promoters would be useful tools for studies of red algal gene regulation and for construction of genetically engineered strains for biotechnological applications.

I have demonstrated the power of subtracted cDNA libraries in the identification of phase-specific algal genes. However, careful monitoring of each step of the protocol was required to obtain a large number and variety of

clones. The construction of a subtracted cDNA library is more challenging than the construction of a cDNA library, but the screening procedures for a subtracted library are efficient due to the enrichment of specific cDNAs. The disadvantage of selecting cDNAs solely on the basis of their expression pattern is that they often show no sequence similarity to known proteins in the data banks and thus cannot be identified. To characterize the encoded proteins, extensive further study is required, for example, *in situ* hybridization or protein purification and analysis (Ursin et al. 1989, Koltunow et al. 1990, Theerakulpisut et al. 1991). Thus, in cases where one is interested in a tissue- or phase-specific gene that encodes a previously characterized protein, traditional approaches using oligonucleotide probes based on a partial amino acid sequence or using antibodies to screen clones producing recombinant proteins might be more efficient (Young and Davis 1983). The method chosen will depend on the objectives of the study and the available information. With such a wealth of molecular tools available, questions concerning the regulation of gene expression during algal life cycles have now become tractable.

Chapter 3: Two Different EF-1α Genes Show Dissimilar Developmental Patterns of Expression

INTRODUCTION

Elongation factor 1α (EF-1α) is a well-studied housekeeping protein found in all eukaryotic cells, where it plays a central role in protein synthesis. It is part of the elongation factor 1 (EF1) complex, which also includes the EF-1ß and EF-1γ proteins. EF-1α facilitates GTP-dependent binding of aminoacyl-tRNA to the A site of the ribosome during the elongation phase of translation, while EF-1ß/γ promotes the exchange of GDP for GTP on EF-1α (Riis et al. 1990). In this process, EF-1α must bind to GTP/GDP, aminoacyl-tRNAs, the 80S ribosome and the EF-1ß/γ protein complex. These multiple interactions of EF-1α probably limit its evolution, and it is therefore not surprising that the EF-1αs characterized to date are highly conserved in their amino acid sequences.

EF-1α is essential for the production of all cellular proteins and at least one active copy of the EF-1α-encoding gene (*tef*) is necessary for cell viability (Cottrelle et al. 1985). Both EF-1α and its prokaryotic homologue EF-Tu are required in large quantities; EF-Tu is possibly the most abundant protein in *E. coli* (Pedersen et al. 1978). Many organisms contain multiple copies of *tef*. Even organisms with markedly reduced genomes such as *Arabidopsis thaliana* (Liboz et al. 1990) and *Saccharomyces cerevisae* contain four and two active *tef* genes, respectively (Cottrelle et al. 1985). Some bacteria also have more

than one EF-Tu-encoding gene (Yokota et al.1980, Sela et al. 1989). However, in almost all cases members of each gene family encode nearly identical proteins.

Divergent EF-1αs have, however, been found in a few animals. The genes encoding these divergent EF-1αs invariably exhibit tissue or developmental stage-specific patterns of expression. *Drosophila* has two EF-1αs, F1 and F2; F1 is expressed throughout the life cycle, while F2 is mainly expressed in the pupal stage (Walldorf et al. 1985, Hovemann et al. 1988). Of the three EF-1αs characterized from *Xenopus laevis* (EF-1αS, EF-1αO, 42Sp50), only the somatic elongation factor, EF-1αS, shows constitutive expression. The oocyte-specific copy, EF-1αO, is closely related to EF-1αS in sequence, while the RNA storage particle protein 42Sp50, which is also expressed only in the oocyte, is highly divergent (Dje et al. 1990). A major function of the 42Sp50 protein appears to be the storage of tRNAs for later use in oogenesis and early embryogenesis. Purified 42Sp50 can function as an EF-1α, but is unique in that it exchanges GDP for GTP without the assistance of the EF-1ß/γ and binds to uncharged tRNA more tightly than charged tRNA (Viel et al. 1991). Rats and humans also have an extra, slightly divergent *tef* gene that is expressed only in terminally differentiated tissues such as heart, brain and muscle (Lee et al. 1993, Knudsen et al. 1993). No studies on the developmental regulation of *tef* genes have been reported in fungi, plants or protists.

Among the cDNAs for genes that are only expressed in the sporophyte of *P. purpurea*, one was found to encode an EF-1α. Following this identification, a constitutively expressed *tef* gene was also isolated. Here, I describe the isolation and characterization of these two *tef* genes from *P. purpurea*. I present evidence that these are the only members of the *P. purpurea tef* gene family and that one is expressed in both phases while the other is specific to the sporophyte. I also report the phylogenetic relationship of these two EF-1αs to those from other eukaryotes.

MATERIALS AND METHODS

Genomic DNA manipulation and analysis

P. purpurea sporophytic cultures grown in modified D-11 medium (Mitman and van der Meer 1994) were used for DNA isolation. Nuclear DNA was extracted and purified according to Rice and Bird (1990), except that the proteinase K treatment was omitted.

For Southern hybridization, 12 µg of nuclear DNA were digested with 120 units of restriction enzyme (*Bam*HI, *Kpn*I, *Pst*I or *Sst*I) at 37 °C overnight. The restricted DNA samples were electrophoresed on a 0.6% agarose gel and transferred to Zeta Probe GT nylon membranes (BioRad). Blotted membranes were hybridized to probe DNAs and washed as described in chapter two, except that the final wash was in 2XSSC, 0.1%SDS, at 68 °C.

Bacteriophage clones harbouring *tef* genomic sequences were obtained

by screening a *P. purpurea* genomic library kindly provided by Dr. R. MacKay. Plaque lifts were prepared according to standard procedures (Sambrook et al. 1989). Membranes were hybridized and washed under the same conditions as the genomic Southern blotting. Recombinant bacteriophage DNA was prepared from positive clones by the method of Sambrook et al. (1989). The insert DNA was mapped by Southern hybridization using the ECL kit (Amersham). The DNA fragment containing *tef* was subcloned into the plasmid vector, pUC18, and sequenced in both directions from double-stranded plasmid template using an ABI 370A sequencer.

cDNA characterization

PolyA$^+$ RNA extraction, subtracted cDNA library construction and screening, northern hybridization and DNA sequencing were performed as described in chapter two. The missing 5' end of the cDNA was obtained by PCR from single-stranded cDNA extracted from the unsubtracted sporophyte cDNA library, which presumably contains longer cDNA inserts. The 5' PCR primer (with its 3'-terminus directed towards the cDNA insert) was complementary to the multiple cloning site sequence adjacent to the *Eco*RI cDNA insertion site on the T7 side of pBluescript II SK$^-$ (Stratagene). The 3' PCR primer (with its 3' terminus directed towards the 5' PCR primer site) was a gene-specific sequence 120 bp from the 5' end of the originally isolated cDNA. The PCR product of expected size was purified from an agarose gel by

Geneclean (BIO 101), cloned into a TA cloning vector (Invitrogen) and sequenced as described in Chapter 2. Protein secondary structure prediction was carried out with the PHD program (Rost and Sander 1993).

Phylogenetic analysis

Sequences were aligned on a SUN workstation using the GCG program Pileup with default gap penalties (Devereux et al. 1984). Minor modifications were made by eye to minimize insertion/deletion events. For phylogentic analyses, regions not alignable with confidence among all taxa were deleted. Deleted regions include the amino-terminus, one small internal region, the carboxyl-terminus, and all insertions unique to *Porphyra* sequences. These correspond to positions 1-4, 157-165, 191-216, 240-243, 417-435 and 485-524 of the *Porphyra* alignment (Fig. 3.1). All parsimony analyses were performed on amino acid sequences using PAUP 3.0r (Swofford 1991) with 50 replicates of random addition and branch-swapping by tree bisection reconnection. Bootstrap analyses used 100 replicates of simple addition with one tree held at each step. Distance matrices were calculated using the PHYLIP 3.5c program PROTDIST with the George-Hunt-Barker and Dayhoff amino acid substitution matrices. Phylogenetic trees were constructed by the NEIGHBOR or FITCH programs (Felsenstein 1991). The eukaryotic EF-1α consensus sequence shown in Figure 3.1 was compiled from all known EF-1αs (Baldauf and Palmer 1993).

RESULTS

Isolation and characterization of a sporophyte-specific EF-1α cDNA

Among the cDNAs from the sporophyte-specific library (Chapter 2) was one whose deduced amino acid sequence showed significant (>60% identity) similarity to EF-1αs. As this 1.2 kb cDNA sequence contained only the C-terminal half of EF-1α, PCR was used to isolate the 5' half of the gene from the unsubtracted sporophyte cDNA library. The combined sequence from these two clones is 1882 bp in length and contains a single open reading frame (ORF) encoding an entire EF-1α of 515 amino acids (Fig. 3.1). This ORF is preceded by 53 bp of 5' leader sequence and followed by 281 bp of 3' untranslated region without a poly A tail at its extreme 3' end (databank accession number U08841).

Northern hybridization of the 1.2 kb cDNA insert with poly A^+ RNAs from both the sporophyte and the gametophyte detected a single 2.1 kb mRNA that is present only in the sporophyte (Fig. 3.2A). This sporophyte-specific EF-1α has been designated Porphyra-s (P-s), while its gene is referred to as *tef-s*.

Isolation and characterization of a second, constitutively expressed, EF-1α

Since *tef-s* was apparently only expressed in the sporophyte and EF-1α is essential for translation, it was expected that there must be at least one more *tef* that is expressed in either the gametophyte or in both phases. Using the P-s cDNA as a probe to screen an unamplified *P. purpurea* genomic library,

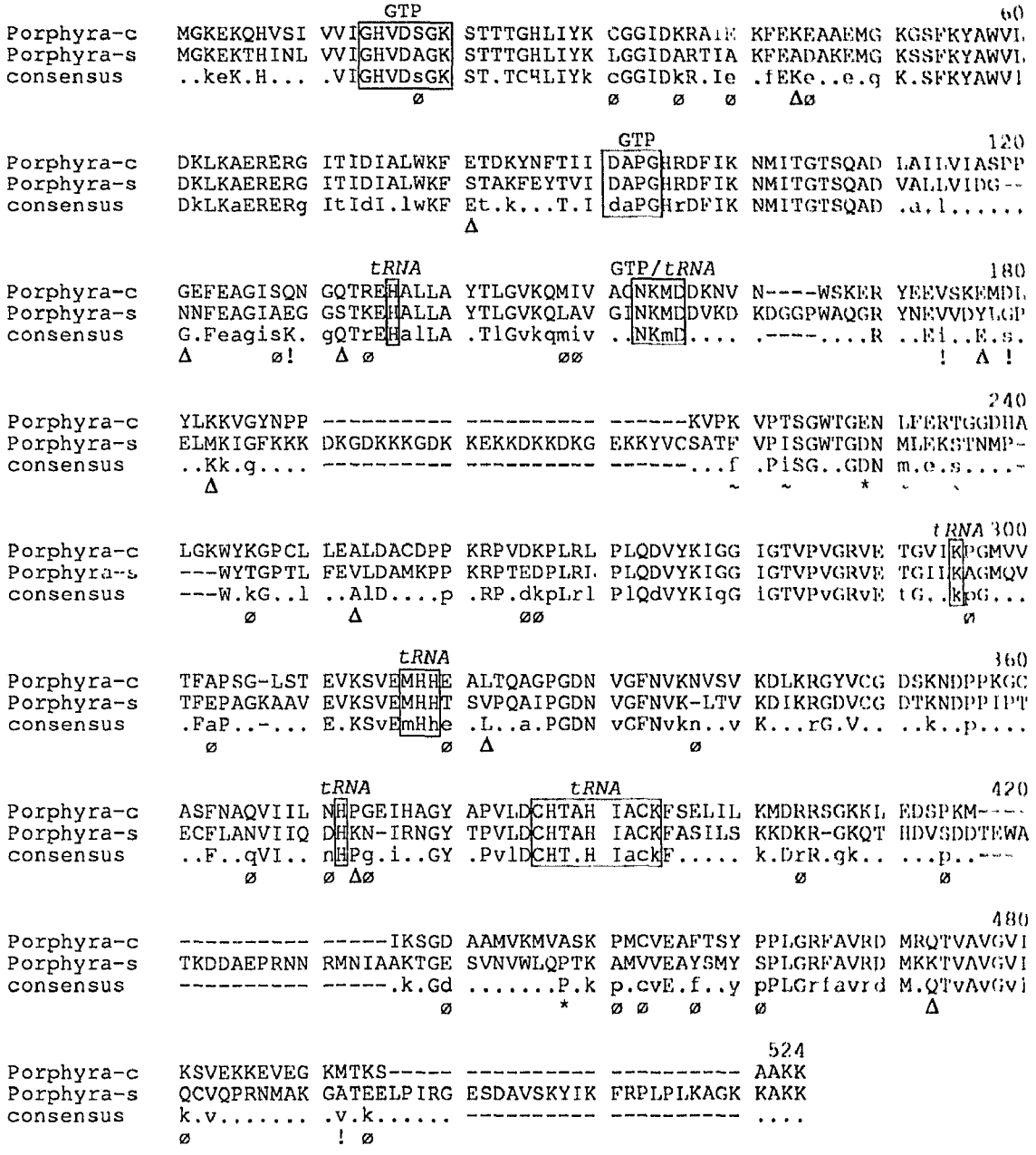

Figure 3.1 Alignment of *Porphyra* EF-1αs with a eukaryote EF-1α consensus sequence. Amino acids involved in GTP and tRNA binding are boxed. In the consensus sequence, upper case letters indicate absolutely conserved amino acids while lower case letters denote either 1-3 conservative substitutions, or 1-2 non-conservative substitutions. Dashes indicate insertions/deletions. Symbols under the consensus sequence are as follows: Δ - "universally" conserved sites, not conserved in Porphyra-s; ø - highly conserved sites, not conserved in Porphyra-s; * - "universally" conserved sites, not conserved in Porphyra-c; ~ - highly conserved sites, not conserved in Porphyra-c; ! - universally or highly conserved sites, not conserved in either *Porphyra* EF-1α.

Figure 3.2 Northern analysis of the two *tef* genes from *P. purpurea*. One µg polyA+ RNA from the sporophyte (S lanes) and the gametophyte (G lanes) were electrophoresed on a formaldehyde agarose gel, blotted to a nylon membrane and hybridized to a ^{32}P-labelled DNA fragment ($2x10^6$dpm ml^{-1}). Final washing was in 0.5 x SSC, 0.1%SDS at 68 °C. Signals were detected by autoradiography with an intensifying screen overnight. **A**) Hybridization with the 3'-specific probe of *tef-s* (from the XhoI site to the end of the cDNA including the 3' untranslated region, see Fig. 3A). **B**) Hybridization with the 3'-specific probe of *tef-c* (from KpnI site to SstI site, see Fig. 3A).

four positive clones were obtained from approximately 200,000 plaques. The cloned fragments were mapped, subcloned and sequenced. Three of these genomic clones contained a complete *tef* whose sequence differed from *tef-s*, and the fourth contained a portion of the same gene. No genomic clone carrying *tef-s* was isolated during this screen. The coding region of the second *tef* sequence revealed a single ORF of 449 amino acids (Fig. 3.1) with no interruptions by any introns (databank accession number U08844).

To test the expression of this new *tef*, northern hybridization was performed under high stringency using the 3' coding region of this sequence as a gene-specific probe. The results indicate that the transcript of this new *tef* is present at equal levels in both the sporophyte and the gametophyte (Fig. 3.2B) suggesting that it is constitutively expressed. I designate this gene as *tef-c* and its encoded EF-1α as Porphyra-c (P-c). The mRNA size of this *tef* is approximately 1.5-1.6 kb, a size consistent with the coding region determined by sequencing. The larger size of the P-s mRNA is apparently due to several insertions in the coding region (see below).

Estimation of gene copy number of *tef* in *P. purpurea*

The number of copies of *tef* in the *P. purpurea* genome was determined by Southern hybridizations with the 5' or the 3' coding region of *tef-c* (Fig. 3.3). The 5' probe of *tef-c* gave rise to two hybridization signals in the *Bam*HI, *Pst*I and *Sst*I lanes (Fig. 3.3B). Based on similar hybridizations with both 5' and 3'

Figure 3.3 Physical maps of *P. purpurea tef* genes and Southern analysis of genomic DNA. **A**) Physical maps of *tef-c* genomic DNA and *tef-s* cDNA. Positions of relevant restriction enzyme digestion sites are those predicted from the nucleotide sequencing data. Hatched boxes represent coding regions. DNA fragments used as probes for Southern hybridization are indicated by arrows. **B**) and **C**) Hybridizations with the 5' and 3' coding regions of *tef-c* respectively. B, BamHI; K, KpnI; P, PstI; S, SstI. Size markers are in kb.

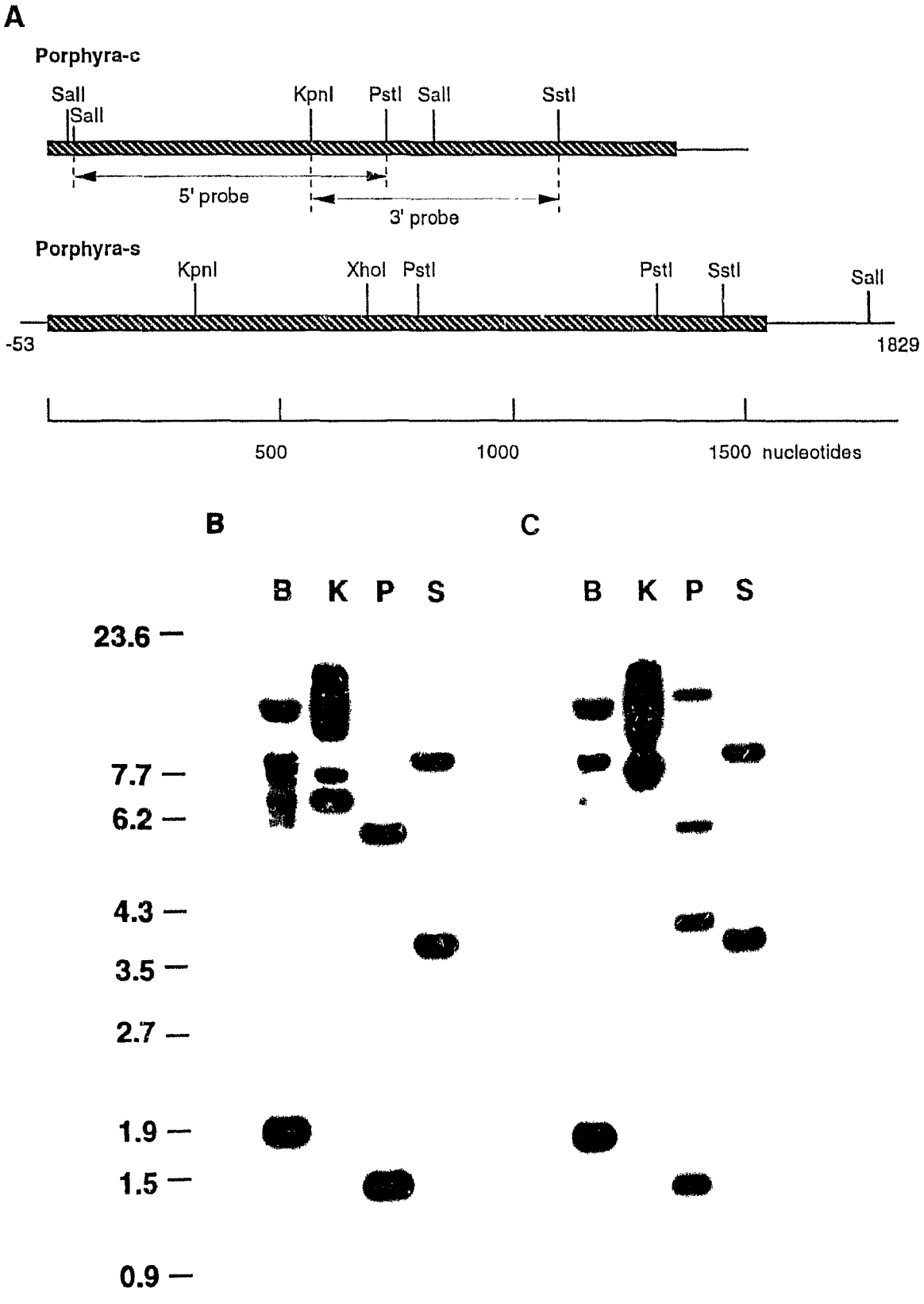
A
Porphyra-c
SalI
SalI
KpnI
PstI
SalI
SstI
5' probe
3' probe
Porphyra-s
KpnI
XhoI
PstI
PstI
SstI
SalI
-53
1829
500
1000
1500 nucleotides
B
C
B K P S
B K P S
23.6
7.7
6.2
4.3
3.5
2.7
1.9
1.5
0.9

Figure 3.3

end probes of *tef-s*, sequencing data and the restriction mapping results of the genomic clones (data not shown), one fragment in each digestion can be assigned to *tef-c* (15 kb *Bam*HI, 6kb *Pst*I and 8kb *Sst*I fragments), while the other signal in each lane is produced by *tef-s* (1.9 kb *Bam*HI, 1.5 kb *Pst*I and 3.9 kb *Sst*I fragments). The probe from the 3' coding region of *tef-c* (Fig. 3.3C), which overlaps the 5' *tef-c* probe, hybridizes to the same *Bam*HI, *Pst*I and *Sst*I fragments as the 5' end probe and to two additional *Pst*I fragments. The two additional *Pst*I fragments are due to the existence of a *Pst*I site in the 3' region of both genes. According to the sequence data, there is a second *Pst*I site in the 3' end of *tef-s*. However, a 513 bp *Pst*I fragment was not detected in the 3' probe hybridization, suggesting that one of two *P-s Pst*I sites might be protected by DNA methylation. The hybridization signals obtained with *Kpn*I digests also appear to be consistent with those of the other enzymes except that some of the hybridizing fragments are large and either partially digested, or possibly degraded during isolation, making it difficult to determine their exact sizes. Taken together, these data indicate that there are two *tefs* in *P. purpurea* and each gene is present as a single copy per haploid genome.

Sequence comparison and phylogenetic analysis of EF-1αs

Amino acid sequence comparison of the two *Porphyra* EF-1αs with each other reveals 63% sequence identity and 17% conserved residue substitutions. P-s shows a relatively low sequence identity (60-67%) with other eukaryotic

EF-1αs, while the identity between P-c and other eukaryotic EF-1αs is 70-78%. In addition, comparison of the P-s sequence with other eukaryotic EF-1αs revealed four unique insertions of 4, 19, 20-25 and 26 amino acids (aa) each (Fig. 3.1). Among known eukaryotic EF-1αs even small insertions are very rare and only one insertion longer than 10 aa is known (Baldauf and Palmer 1993). The amino acid composition of the 26 aa insertion is also striking in that it contains 50% lysine residues.

Secondary structure predictions of the amino acid sequences of both *Porphyra* EF-1αs reveal a striking conservation of α-helix and ß-sheet elements with the bacterial EF-Tu (Kjeldgaard and Nyborg 1992, Berchtold et al. 1993). The important elements of EF-1α predicted to be involved in GTP-binding and hydrolysis based on tertiary structure studies of EF-Tu, as well as residues demonstrated by cross-linking studies to interact with the aminoacyl tRNA, are all completely conserved in both *Porphyra* EF-1αs (Fig. 3.1) (Woolley and Clark 1989, Metz-Boutigue et al. 1989, Kinzy et al. 1992). A comparison of the secondary structure predicted by the PHD program (Rost and Sander 1993) of P-s with bacterial EF-Tu showed that the 4 aa and the 26 aa insertions are located within loops in the GTP-binding domain. The 26 aa insertion, in particular, appears to extend the existing loop that protrudes from the surface of this domain. The 19 aa insertion extends a loop that occurs in domain III. Finally, the 20-25 aa insertion (depending on the alignment) occurs four residues from the extreme 3' end of EF-1α, and part of this insertion may form

an extra α-helix. Only a single insertion was found in the P-c sequence relative to other eukaryotic EF-1αs. This 4 aa insertion is located within a loop of the GTP binding domain in a position different from any of the P-s insertions.

The relationship between the two *Porphyra* EF-1α sequences and other known EF-1α and EF-1α-like proteins was investigated by phylogenetic analysis. All trees derived by either parsimony (Fig. 3.4) or distance methods (data not shown) show that both EF-1αs from *Porphyra* are of eukaryotic origin, although only very distantly related to each other (Fig. 3.4). In all trees P-c occupies a similar and relatively derived position, close to plants, animals and fungi, though its exact position relative to the slime mold, *Dictyostelium*, and green plants is unresolved. Four equal length trees were found, two of which differed from the tree shown in placing P-c as a separate branch immediately below that of *Dictyostelium*, while the single distance tree placed P-c and *Dictyostelium* together on a branch with plants (data not shown). All methods placed P-s as the deepest branch among eukaryotes after *Giardia* EF-1α, identical to the tree in Figure 3.4. However, it should be noted that P-s has a similarly low level of identity to all known EF-1αs suggesting that it may, instead, be evolving at a much faster rate than the other proteins in the tree. This is also reflected in the relatively long terminal branch of P-s in the tree (Fig. 3.4). Thus the position of P-s may be an artefact, and it may, in fact, share a recent common ancestry with P-c.

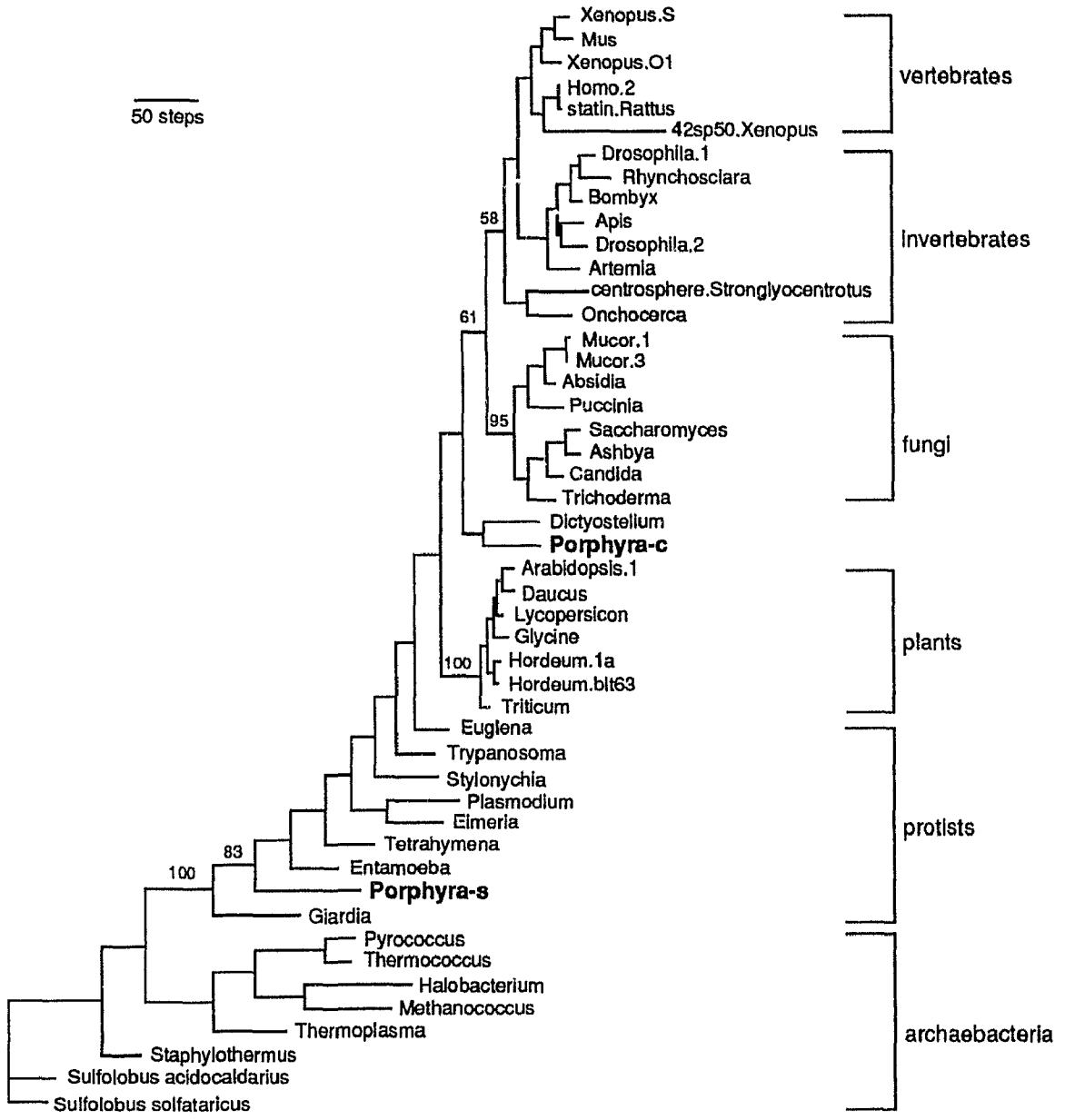

Figure 3.4 The relationship among *Porphyra* and other eukaryotic EF-1αs as predicted by phylogenetic analysis. The phylogenetic tree shown was derived by parsimony analysis of amino acid sequences with all amino acid changes weighted equally. Branches are drawn proportional to the number of inferred amino acid substitutions as indicated by the scale bar. The tree is rooted with the archaebacteria (Iwabe et al. 1989). Bootstrap values above 50% are indicated above the nodes defining major groups only. This tree is one of four shortest trees found and has a length of 2309 steps, a consistency index, excluding uninformative characters, of 0.496, and a retention index of 0.615. The three additional trees found at this length differ from the tree shown in placing P-c as a separate branch immediately below that of *Dictyostelium* and in the branching patterns within plants and within animals.

DISCUSSION

In this chapter, I describe the isolation and characterization of two *P. purpurea tef* genes that are strikingly different from each other with respect to sequence and expression pattern. According to the results of genomic Southern analysis, only one copy of each *tef* exists in the *Porphyra* genome. My data demonstrate that *tef-c* encodes a typical EF-1α that is expressed in both the sporophyte and the gametophyte, while *tef-s* encodes an EF-1α with an unusual primary structure that is expressed only in the sporophyte. While it has been shown that *Drosophila* and *Xenopus* contain *f* genes that are developmentally regulated (Walldorf et al. 1985, Hovemann et al. 1988, Dje et al. 1990), this is the first example of a developmental stage-specific *tef* outside of the animal kingdom. My results provide direct evidence that the evolution of specialized, stage-specific *tef* genes is more widespread than previously thought and has occurred independently more than once in the evolution of multicellular eukaryotes.

The sporophyte and gametophyte of *Porphyra* differ in both morphology and subcellular details and there are many transcriptionally regulated genes contributing to the differences between the two phases (Chapter 2). Under the growth conditions used in this study, the *P. purpurea* sporophyte rapidly develops conchosporangial branches, in which conchospores develop. The conchospores are eventually released and germinate to form the gametophyte,

which then produces male and female gametes. One possible role for the sporophyte-specific EF-1α in *P. purpurea* may be that P-s functions specifically in the conchosporangia. This would parallel the situation in *Xenopus* where EF-1αO and 42Sp50 are found only in the developing oocyte.

Secondary structure predictions for both *Porphyra* EF-1αs show that the overall arrangement of structural elements in the GTP and the aminoacyl-tRNA binding sites is very similar to that of EF-Tu. P-c, in particular, is very highly conserved in terms of primary sequence and predicted secondary structure, all of which suggest that P-c is a typical eukaryotic EF-1α representing the "housekeeping" EF-1α required in all living cells. Although P-s has a highly divergent primary structure, all the elements involved in GTP and tRNA binding are absolutely conserved. Amino acid substitutions are mostly in the non-conserved regions. Studies in *Xenopus* have shown that such amino acid differences have not led to an important modification in function (Dje et al. 1990, Viel et al. 1991). Although three of the insertions in P-s are quite large, they all appear to be restricted to pre-existing loops with the possible exception that part of the last insertion may form an extra helix located at the very end of the protein. Given the locations and structural properties of these insertions, it is unlikely that they would affect the proper folding or function of the protein. Thus, P-s may still be a functional EF-1α.

Alternatively, it is possible that P-s has an altered or additional function relative to other known EF-1αs. Since both *tef-c* and *tef-s* are actively

transcribed in the sporophytic-phase of *P. purpurea*, one cannot simply infer that P-s participates directly in translation. The high degree of conservation of the important elements in its primary and secondary structure suggests that P-s retained most EF-1α-related functions which may still require interacting with GTP, tRNAs, ribosomes or EF-1ß/γ. However, the unique insertions and numerous sequence substitutions may reflect the evolution of function. For example, the lysine-rich 26 aa insertion in the GTP-binding domain probably forms a large positively charged loop located on the surface of the protein, which would allow an interaction with negatively charged molecules.

Probably the largest and most highly divergent member of the EF-1α/EF-Tu family known is the *E. coli* selenocysteyl-tRNA^{sec}-specific elongation factor. This protein is 614 amino acids long and is homologous to, but extremely divergent from, the eubacterial EF-Tu (Forchhammer et al. 1990). Although selenocysteyl-tRNAs that decode UGA have been found in representatives of all major groups of organisms (Lee et al. 1990, Hatfield et al. 1992), a eukaryotic selenocysteyl-tRNA^{sec}-specific EF-1α has not yet been identified. Comparison of the amino acid sequence of P-s with the *E. coli* selenocysteyl-tRNA^{sec} elongation factor reveals no sequence similarity in the regions where the insertions are located and the overall sequence identity is 29%. Since two functionally related proteins, P-c of *P. purpurea* and EF-Tu of *E. coli*, only have 30% sequence identity, the low degree of sequence similarity between P-s and the *E. coli* selenocystyl-tRNA^{sec} elongation factor may not mean that they are

functionally unrelated. More information is required to determine the relationship between P-s and the selenocysteyl-tRNA^{sec}-specific EF-1α.

There are a number of reports that relate EF-1α or EF-1α-like proteins to functions other than peptide elongation. The EF-1α in *Dictyostelium* reversibly binds to the actin cytoskeleton. This may affect the activity or availability of EF-1α in protein synthesis and may therefore contribute to temporal and spatial regulation of protein synthesis (Yang et al. 1990). However, the proposed actin binding site in *Dictyostelium* EF-1α does not appear to have any sequence similarity to the corresponding region in P-s. In *E. coli*, EF-Tu also regulates rRNA synthesis by stabilizing a particular conformation of the RNA polymerase (Travers 1973) and is a subunit of RNA dependent RNA polymerase of the bacteriophage Qß (Blumenthal et al. 1972). In carrot, an EF-1α-like protein functions as a phosphatidyl inositol 4-kinase activator, binds actin and facilitates actin polymerization, while still retaining EF-1α function in an *in vitro* assay (Yang et al. 1993). A mitotic-apparatus-associated protein from sea urchin eggs is also structurally and functionally related to EF-1α (Kuriyama et al. 1990, Ohta et al. 1990) and the suf12 suppressor protein of yeast has a high degree of sequence identity with EF-1α (Wilson and Culbertson 1988). Most of these observations concerning the alternative functions of EF-1αs and EF-1α-like proteins in eukaryotes are quite recent and it is as yet difficult to relate any of these functions to P-s. Nevertheless, the maintenance of *tef-s* in *Porphyra* as an open reading frame

despite the numerous substitutions and large insertions suggests that there is selective pressure to maintain this gene.

Although phylogenetic analysis suggests an ancient origin for P-s, the low sequence similarity of this protein to P-c and all other known EF-1αs leads to an alternative explanation: P-s is evolving more rapidly than other eukaryotic EF-1αs. This rapid rate of evolution interferes with accurate phylogenetic analysis (ie.,long branch effects; Swofford and Olsen 1990). This is especially a problem when there is a lack of taxonomic representation. A similar obstacle was encountered in early phylogenetic analyses of the 42Sp50 protein of *Xenopus* (Viel et al. 1991). These placed 42Sp50 very distantly from the *Xenopus* EF-1αO and EF-1αS, presumably due to the rapid evolutionary rate of 42Sp50 and the lack of representation of higher eukaryotic EF-1α sequences. In all of my trees, which contain a much broader taxonomic representation, all three *Xenopus* sequences group together with other vertebrates, with 42Sp50 having the expected long terminal branch. I would predict that with the addition of more protistan sequences, P-s would show a closer relationship with P-c. This would, in turn, suggest a relatively recent origin for P-s perhaps reflecting the evolution of a new function specific for a unique developmental stage.

Chapter 4: Isolation and Sequence Analysis of a Sporophyte-specific cDNA Encoding a Serine Protease-like Protein

INTRODUCTION

As already demonstrated in the first two chapters of this thesis, specific phase differences between sporophytes and gametophytes of *Porphyra* can be detected at the level of gene expression. While screening and characterizing the subtracted cDNA libraries constructed for this study (Chapter 2), I found one sporophyte-specific clone, 10c3, that contained an interesting insert of 842 bp. Preliminary sequence analysis revealed a single open reading frame (ORF) of 233 amino acids that was followed by a 140 bp 3' untranslated sequence. This ORF did not contain an initiation codon near the N-terminus indicating that it likely was missing the 5' end of the protein coding region. Protein data bank searches (Altschul et al. 1990, Pearson and Lipman 1988, Smith and Waterman 1981) with the deduced amino acid sequence (Pustell and Kafatos 1984) indicated that the protein encoded by the ORF was similar to serine proteases.

Not all members of the serine protease family have enzyme activity. One serine protease-like protein, the α-subunit of 7S nerve growth factor, contains all the active-site residues, but lacks catalytic activity due to an amino acid replacement near one of the active-site residues (Isackson and Bradshaw 1984). Other serine protease-like proteins have substituted two of the three active-site residues, histidine and serine, with other amino acids and

subsequently lost their proteolytic activity (Højrup et al. 1985, Kurosky et al. 1980, Nakamura et al. 1989). Most of these serine protease-like proteins have adopted an alternative function as a subunit of a protein complex, such as the ß-chain of human hepatocyte growth factor, the α-subunit of the 7S nerve growth factor and the ß-chain of human haptoglobin, which binds to haemoglobin. The precise function of these serine protease homologs in their protein complexes is still obscure. The amino acid sequences of these homologs are not very similar to each other and some of the proteins have replaced some of the cysteine residues that are involved in disulphide bridge formation in active serine proteases. They appear to have evolved independently and are thus unlikely to have maintained a similar function.

In view of the diversity of serine proteases and related proteins, I decided to obtain a full length copy of the ORF in clone 10c3 so that a proper comparison could be made with other members of this family of proteins, and a possible function for the *Porphyra* protein more accurately deduced.

MATERIALS AND METHODS

General methods have been described in Chapter 2. To obtain a full length cDNA of the ORF in clone, 10c3, first-strand cDNA was synthesized from sporophyte mRNA using a biotinylated poly(T) primer that was attached to streptavidin magnetic beads (Lee and Vacuier 1992). An "anchor" oligonucleotide containing an *Eco*RI sequence was then ligated to the 5' end of

the cDNA (Troutt et al. 1992). PCR was performed using the bead-attached cDNA as the template, a 5' end primer that was complementary to the anchor oligonucleotide sequence and a 3' end gene-specific primer that was complementary to the sequence 49 bp downstream of the 10c3 cDNA stop codon. A single 1.1 kb PCR product was obtained and cloned into *Eco*RI - digested pUC18.

RESULTS AND DISCUSSION

As confirmed by northern hybridization (Fig. 4.1), the 10c3 cDNA represents a gene that is expressed only in the sporophyte . The intensity of the hybridization signal clearly indicates that the mRNA for this cDNA is present in high abundance. The size of the mRNA estimated from the northern hybridization is 1.2 kb, which is in keeping with the conclusion that the 10c3 cDNA is not full length.

The new clone obtained by PCR contained a cDNA fragment encoding a protein of 303 amino acids that is preceded by a 147 bp 5' non-coding region. Combining the sequences from both cDNAs gives a total length of 1199 bp (Fig. 4.2).

Protein data bank searches showed that the *Porphyra* protein had highest similarity to the tryptase family of serine proteases. Tryptases are a class of serine proteases in mammalian mast cells that have trypsin-like activity (Miller et al. 1990). Comparison of this *Porphyra* tryptase-like protein (TLP)

Figure 4.1 Northern analysis of the 10c3 cDNA. One μg polyA$^+$ RNA from the sporophyte (S lanes) and the gametophyte (G lanes) were electrophoresed on a formaldehyde agarose gel, blotted to a nylon membrane and hybridized to a ^{32}P-labelled cDNA fragment ($2x10^6$ dpm ml^{-1}). Final washing was in 0.5 x SSC, 0.1% SDS, at 68 °C. The first pair of lanes is a control hybridization with a β-tubulin cDNA probe derived from an mRNA occurring in both phases. The second pair of lanes is the hybridization with the 10c3 probe. The signals in the control panel were detected by autoradiography with an intensifying screen for 48 h. The signal for the 10c3 hybridization was detected by autoradiography with an intensifying screen for 1 h.

```
-147          AGCAAGGTGTCACACATCGTCCACCCTCAACGTCCGCCGCCGACTCGGACCCCCAACCACTCCTCTTG

-79  GCTCGCTTTTCTCTACCTCCCCCTCCTTCACGCCACCCCCTCCCTCCCCCGCGTCCTCGGCAATCTGTCCTTGCCCGCC

1    ATG GCC CGC TTG ACG AGC ACT ACT ACG CTC ATC GCC GCG CTG CTG CTT GTC GCC GTC TCC
     M   A   R   L   T   S   T   T   T   L   I   A   A   L   L   L   V   A   V   S   20

61   TTC ACC GCC GTG GCC GCT GTC AAC ACG AAG GAG AGC AGC ACT CGC AAA AAG GCG CTG AAG
     F   T   A   V   A   A   V   N   T   K   E   S   S   T   R   K   K   A   L   K   40

121  GCA AAG GCC CCA CGT GGG CGC AGT GGG CGC ATC GTC GGC GGC CGC GAA GTC GAC GAC TAT
     A   K   A   P   R   G   R   S   G   R   I   V   G   G   R   E   V   D   D   Y   60

181  GAC GAG GAC ACG GGC GTT CAC TTC ATT GCC AAG CTC TTC CTC CCC GAC GGC AAC GGC TTT
     D   E   D   T   G   V   H   F   I   A   K   L   F   L   P   D   G   N   G   F   80

241  TAC TGC AGT GGG TCT GTC ATC TCA AAG AGC GGG CAT GTG CTG ACG CGC GCC GGC TGC GAG
     Y   C   S   G   S   V   I   S   K   S   G   H   V   L   T   R   A   G   C   E   100

301  CCG CGC GTC AAT GAC GTT GTG CGT CTC GGC GGA TCG CGC CTG TAC AAC GGG GTC GTC GCG
     P   R   V   N   D   V   V   R   L   G   G   S   R   L   Y   N   G   V   V   A   120

361  CGG GTG GCC AAG GTG TCG ATC CAC CCC AAG TAC GAC CCT GCG GGC GAA GTG GCC GAC GTG
     R   V   A   K   V   S   I   H   P   K   Y   D   P   A   G   E   V   A   D   V   140

421  GCG GTG CTC AAA CTC AAG GGC GTC TCT GAG AGC CGA CTC CTC CGC GCC GGC GTC GTG CCC
     A   V   L   K   L   K   G   V   S   E   S   R   L   L   R   A   G   V   V   P   160

481  GTC TTC CTC AAC CGC GTG TGG GAC AAC CCG CAC GGC ATG TAC TTT ACG GGG TAC GGC GCA
     V   F   L   N   R   V   W   D   N   P   H   G   M   Y   F   T   G   Y   G   A   180

541  ACG GAT AAG GCG GCG CAG TCG GCG GGT AGC CTG ACG CTC AAA CGC GCC TAC CTC CCG GTC
     T   D   K   A   A   Q   S   A   G   S   L   T   L   K   R   A   Y   L   P   V   200

601  GCG CCG TGG TGG AAC TGC CGC CGC ATT ACC GAC ACG GTC GTG GTG CCG GGG CTG TCA CGC
     A   P   W   W   N   C   R   R   I   T   D   T   V   V   V   P   G   L   S   R   220

661  CCC GGC CTG CCC ATC TCC CCT GCC GCG CAG GTG TGT CTG CGG GGC GGC CGC GGG GCG GGT
     P   G   L   P   I   S   P   A   A   Q   V   C   L   R   G   G   R   G   A   G   240

721  GCT CTT TGC GAG CGT GAC CCT GGG GGC CCG ATG TAT CGG GTC AGC ACG CAC CGC GGC GTC
     A   L   C   E   R   D   P   G   G   P   M   Y   R   V   S   T   H   R   G   V   260

781  AAG ATC TAC ACC CTG TAC GCG GTC AGC AGC TAC TGG ATT GGC TTG GGG GCG GAC AAC CGC
     K   I   Y   T   L   Y   A   V   S   S   Y   W   I   G   L   G   A   D   N   R   280

841  TGC CCG CGC GCG ATG CCA AAC GTC GGC TCC AAG GTC GCC TTC TAC TAC AGC TGG ATC CAG
     C   P   R   A   M   P   N   V   G   S   K   V   A   F   Y   Y   S   W   I   Q   300

901  AAC CAG GTG TAG AGAGGTGTGGAGAGGGACTCGCCGTGCAGCGAGCGGCGCGTCTGGGCCGATGTGTGACACC TT
     N   Q   V   -

976  GCTGCGCTTGCGCGGCGGCTCACAGATGTTGTCATGACTTGATTGTGTGGCTTGTGTGGGCTGCTCTAGCATCTCAA
```

Figure 4.2 cDNA nucleotide sequence and deduced primary structure of the encoded protein. Numbers in the left margin refer to nucleotide positions. Numbers in the right margin refer to amino acid residues. The signal peptide is underlined. The two amino acids replacing the usual histidine and serine in the charge relay triad of serine proteases are indicated by thin boxes. The conserved aspartic acid residue in the charge relay triad is indicated by a thick box.

with representative tryptases revealed 34%, 31% and 29% sequence identity plus 28%, 27% and 28% conserved residue substitutions for dog tryptase precursor (Vanderslice et al. 1989), mouse mast cell protease 6 (Reynolds et al. 1991) and human α-tryptase precursor (Miller et al. 1989), respectively. Alignment of the *Porphyra* TLP with these tryptase precursors shows that the amino acid sequence of the TLP contains all regions that are highly conserved in tryptases (Miller et al. 1989) (Fig. 4.3). These include three regions surrounding the three active site amino acids and the IVGG activation site region, where cleavage of the propeptide occurs to produce the mature protein. In addition, the six cysteine residues that are characteristic of serine proteases and important for determination of structure (De Haën et al. 1975, Yun and Davis 1989) are all conserved in the TLP (Fig. 4.3). Although serine proteases of different species have different numbers of cysteine residues, the ancestral active enzyme appears to have contained three disulphide bridges that occur in analogous positions in the *Porphyra* TLP. Some vertebrate serine proteases have more than three disulphide bonds, such as the mammalian mast cell tryptases (Fig. 4.3), but three disulphide bridges are common to all known invertebrate serine proteases (Yun and Davis 1989, Jany et al. 1983). Further, the *Porphyra* TLP also has a typical signal peptide as one would expect for a secretory protein (Fig. 4.2). The signal peptidase cleavage site determination was based on the rule of von Heijne (1984).

The His/Asp/Ser charge relay triad, the active site for all serine

```
PORPHYRA        MARLTSTTTLIAAL-LLVAVSFTAVAAAVNTKESSTRKKALKAKAPRGRSGR
TRYT-CANFA      M-----PSPLVLALALLGSLVPVSPAPGQ---------ALQ------RVG-
MCP6-MOUSE      M--LKRRLLLLWALSLLASLVYSAPRPAN----------Q------RVG-
TRYA-HUMAN      M-----LSLLLLALPVLASRAYAAPAPVQ---------ALQ------QAG-
                *        *. ** .* .   . . .          .      . *

PORPHYRA        IVGGREVDD--YDEDTGVHFIAKLFLPDGNGFYCSGSVISKSGHVLTRAGC
TRYT-CANFA      IVGGREAPGSKWPWQVSLRLKGQYWRH-----ICGGSLI-HPQWVLTAAHC
MCP6-MOUSE      IVGGHEASESKWPWQVSLRFKLNYWIH-----FCGGSLI-HPQWVLTAAHC
TRYA-HUMAN      IVGGQEAPRSKWPWQVSLRVRDRYWMH-----FCGGSLI-HPQWVLTAAHC
                ****.*         . ...              *.**.*  .  *** * *

PORPHYRA        -EPRVNDV----VRLGGSRLY-NGVVARVAKVSIHPKYDPAGEVADVAVLK
TRYT-CANFA      VGPNVVCPEEIRVQLREQHLYYQDHLLPVNRIVMHPNYYTPENGADIALLE
MCP6-MOUSE      VGPHIKSPQLFRVQLREQYLYYGDQLLSLNRIVVHPHYYTAEGGADVALLE
TRYA-HUMAN      LGPDVKDLATLRVN-SGTHLYYQDQLLPVSRIMVHPQFYIIQTGADIALLE
                  * .       *      **  . .  . .. .** .       **.*.*

PORPHYRA        LKGVSESRLLRAGVVPVFLNRVWDN-PHGM--YFTGYGATDKAAQSAGSLT
TRYT-CANFA      LEDPVN---VSAHVQPVTLPPALQTFPTGTPCWVTGWGDVHSGTPLPPPFP
MCP6-MOUSE      LEVPVN---VSTHIHPISLPPASETFPPGTSCWVTGWGDIDNDEPLPPPYP
TRYA-HUMAN      LEEPVN---ISSRVHTVMLPPASETFPPGMPCWVTGWGDVDNDEPLPPPFP
                *    .   . . . . *    .  * *  ^. **.*  .     . .

PORPHYRA        LKRAYLPVAPWWNCRRITDTVVVPGLSRPGLPISPAAQVCLRGGRGAGALC
TRYT-CANFA      LKQVKVPIVENSMCDVQYHLGLSTG---DGVRIVREDMLC--AGNSKSDSC
MCP6-MOUSE      LKQVKVPIVENSLCDRKYHTGLYTG---DDFPIVHDGMLC--AGNTRRDSC
TRYA-HUMAN      LKQVKVPIMENHICDAKYHLGAYTG---DDVRIIRDDMLC -AGNSQRDSC
                **.  .*.     *    .     *    .  *     .*  .*      *

PORPHYRA        ERDPGGPMYRVSTHRGVKIYTLYAVSSYWIGLGADNR---CPRA-MPNVGS
TRYT-CANFA      QGDSGGPL-------------VCRVRGVWLQAGVVSWGEGCAQPNRPGIYT
MCP6-MOUSE      QGDSGGPL-------------VCKVKGTWLQAGVVSWGEGCAQPNKPGIYT
TRYA-HUMAN      KGDSGGPL-------------VCKVNGTWLQAGVVSWDEGCAQPNRPGIYT
                  *.***.              .^ * . *.  *  ..   *...  * . .

PORPHYRA        KVAFYYSWIQNQV----
TRYT-CANFA      RVAYYLDWIHQYVPKEP
MCP6-MOUSE      RVTYYLDWIHRYVPEHS
TRYA-HUMAN      RVTYYLDWIHHYVPKKP
                .*..*  **.  *
```

Figure 4.3 Comparison of the *Porphyra* TLP with mammalian mast cell tryptases. The alignment was done using the CLUSTAL V program (Higgins et al. 1992) with minor adjustments made by eye. "*" - conserved residues. "." - conserved substitutions. Highly conserved regions characteristic of serine proteases are boxed. Cysteines that are involved in disulphide-bridge formation common to all sequences are bold. Two cysteines forming an extra disulphide bond in sequences other than the *Porphyra* TLP are indicated by "^". TRYT-CANFA: dog tryptase precursor, MCP6-MOUSE: mouse mast cell protease 6, TRYA-HUMAN: human α tryptase precursor.

proteases, does not occur in the *Porphyra* TLP because two of the three active residues, histidine and serine, are replaced by glycine-98 and proline-247, respectively (Fig. 4.2). It is clear that the TLP cannot show any proteolytic activity due to the replacement of these essential amino acids. However, the conservation of the activation site sequence and the amino acids surrounding the active-site residues, as well as the disulphide bridges strongly suggest that the TLP is evolutionarily related to the tryptases.

The amino acid sequence of the TLP does not show substantial similarity to any of the other known serine protease-like proteins that lack a catalytic function. Although the *Porphyra* TLP lacks the active-site histidine and serine residues, it has retained all the important cysteines required to maintain the structural features of tryptases. Thus it appears quite possible that the *Porphyra* TLP has a function that remains related to specific characteristics of tryptases other than their proteolytic activity.

Functional characteristics unique to tryptases are their tetrameric structure and their binding to and stabilization by highly sulphated glycosaminoglycans, such as heparin and chondroitin sulphate (Vanderslice et al. 1989, Schwartz and Bradford 1986). Sulphated polysaccharides stabilize tryptases through a direct interaction with the enzymes rather than by an indirect, ionic association (Schwartz and Bradford 1986). Both the sugar chain and the sulphate groups of the polysaccharide contribute to this interaction and the basic residues of the enzymes are assumed to be important for binding

(Vanderslice et al. 1989, Evans et al. 1992, Petersen 1983). The *Porphyra* TLP, excluding the signal peptide, contains 39 basic residues (15 lysines and 24 arginines) and has a net cumulative charge of +19. Although this protein does not have proteolytic activity, the similarity of its primary structure to the tryptases suggests that it might have maintained the capacity to bind sulphated polysaccharides.

Some of the striking differences between the two life cycle phases of *Porphyra* exist in the composition and structure of the cell walls (Mukai et al. 1981, Gretz et al. 1983, Gretz et al. 1986). The cell walls of the sporophyte are composed mainly of galactans, which frequently contain 6-sulphated residues, whereas the gametophyte cell wall consists mostly of mannans with galactans as only minor components. Northern hybridization demonstrates that the gene encoding the TLP is expressed only in the sporophyte which is the phase with abundant sulphated residues. The deduced amino acid sequence for the TLP begins with a typical signal peptide suggesting that it enters the cell's secretory system and thus may be transported to the cell wall. The mRNA species encoding the TLP is highly abundant indicating that the TLP exists in large amounts *in vivo* as would be required for a major wall component. Taken together, these characteristics suggest a possible role of the TLP as a cell wall protein that interacts with the sulphated galactans in the sporophytic phase.

Chapter 5: Isolation and Characterization of a Gametophyte-specific cDNA Encoding a Lipoxygenase

INTRODUCTION

Lipoxygenases are nonheme iron-containing enzymes that catalyze the oxygenation of fatty acids containing a cis,cis-1,4-pentadiene moiety. The oxidized lipid products, hydroperoxides, are important precursors in a variety of metabolic pathways (Vick and Zimmerman 1987). Lipoxygenases have been isolated from plants, animals, fungi and algae, indicating that these enzymes are widely distributed in nature (Hildebrand et al. 1988). The primary structures of some plant and animal lipoxygenases have been determined by molecular cloning and sequencing of genomic DNA or mRNA species (as cDNAs). Animal and plant lipoxygenases do not show a high degree of overall amino acid sequence identity; however, they are more similar in the C-terminal half of the proteins where five conserved residues acting as ligands for the enzymatically active iron are located (Minor et al. 1993). A region of 38 amino acids containing five conserved histidine residues (Minor et al. 1993, Steczko et al. 1992) is found in all known lipoxygenases.

Some mammals and higher plants contain isozymes for lipoxygenase that are developmentally regulated (Funk et al. 1992, Hildebrand et al. 1988). Most lipoxygenases found in mammals are tissue-specific (Boado et al. 1992, Funk et al. 1992, Thiele et al. 1987). In plants, the level of lipoxygenase

activity can vary not only among different organs of a given plant, but also between developmental stages of a tissue (Siedow 1991). Lipoxygenase activity has been reported in algae (Beneytout et al. 1989, Moghaddam and Gerwick 1990, Zimmerman and Vick 1973), but the amino acid sequences of these enzymes have not been determined. In this chapter, I report the isolation and characterization of a lipoxygenase cDNA from the red alga *Porphyra purpurea* and demonstrate that its gene is only expressed in the gametophytic phase of the life cycle.

MATERIALS AND METHODS

General materials and methods have been described in Chapter 2. The missing 5' end of the cDNA was obtained by a PCR approach described in Chapter 3 except that the 3' PCR primer was complementary to a sequence that occurs 149 bp from the truncated end of the 12g5 ORF.

RESULTS AND DISCUSSION

Following construction of a subtracted cDNA library for the gametophyte of *P. purpurea* (Chapter 2), several gametophyte-specific clones were randomly selected for further characterization. One of these, 12g5, contained an insert of 2037 bp with a single open reading frame (ORF) that encodes 582 amino acid residues. Protein data bank searches (Altschul et al. 1990, Pearson and Lipman 1988, Smith and Waterman 1981) with the deduced amino acid

sequence (Pustell and Kafatos 1984) indicated similarity to plant and animal lipoxygenases. As this cDNA did not encode the N-terminus of the lipoxygenase, a PCR approach was used to isolate the N-terminal coding region. The second cDNA clone obtained by this method contained an 891 bp fragment encoding the N-terminal region of the protein. The combined sequences of these two clones (2779 bp) define a single ORF of 830 codons (including the termination codon) and a 3' untranslated region of 289 bp (databank accession number U08842). The first ATG codon occurs at position 78 and there is no stop codon upstream of this ATG. These observations suggest that this ATG is not the initiation codon and that these cDNAs do not encode the N-terminus of the protein.

Alignment of the *Porphyra* sequence with rice, soybean and human lipoxygenases (Fig. 5.1) (Ohta et al. 1992, Shibata et al. 1987, Matsumoto et al. 1988) shows that the N-terminal portion of the proteins are divergent, while regions in the middle and C-terminal portion exhibit substantial sequence similarity. This result is in agreement with other alignments of various animal and plant lipoxygenase amino acid sequences (Ohta et al. 1992, Siedow 1991, Sigal et al. 1988). For example, the N-termini of the three soybean lipoxygenases are not at all similar, while the C-terminal nine amino acids of all known plant lipoxygenases are identical. All plant lipoxygenases are about 25-30% longer than the animal proteins due to the presence of several large insertions (relative to animal proteins) in the N-terminal half of the plant

Figure 5.1 Sequence comparison of different lipoxygenases. *Porphyra* lipoxygenase was aligned with rice lipoxygenase-2, soybean lipoxygenase-1 and human lipoxygenase-5 using MULTALIN (Corpet 1988) with minor adjustments made by eye. "*" indicates conserved residues. ":" indicates conserved substitutions. The conserved 38 amino acid region is underlined. The five conserved residues known to bind the enzymatically active iron atom in soybean lipoxygenase-1 are boxed.

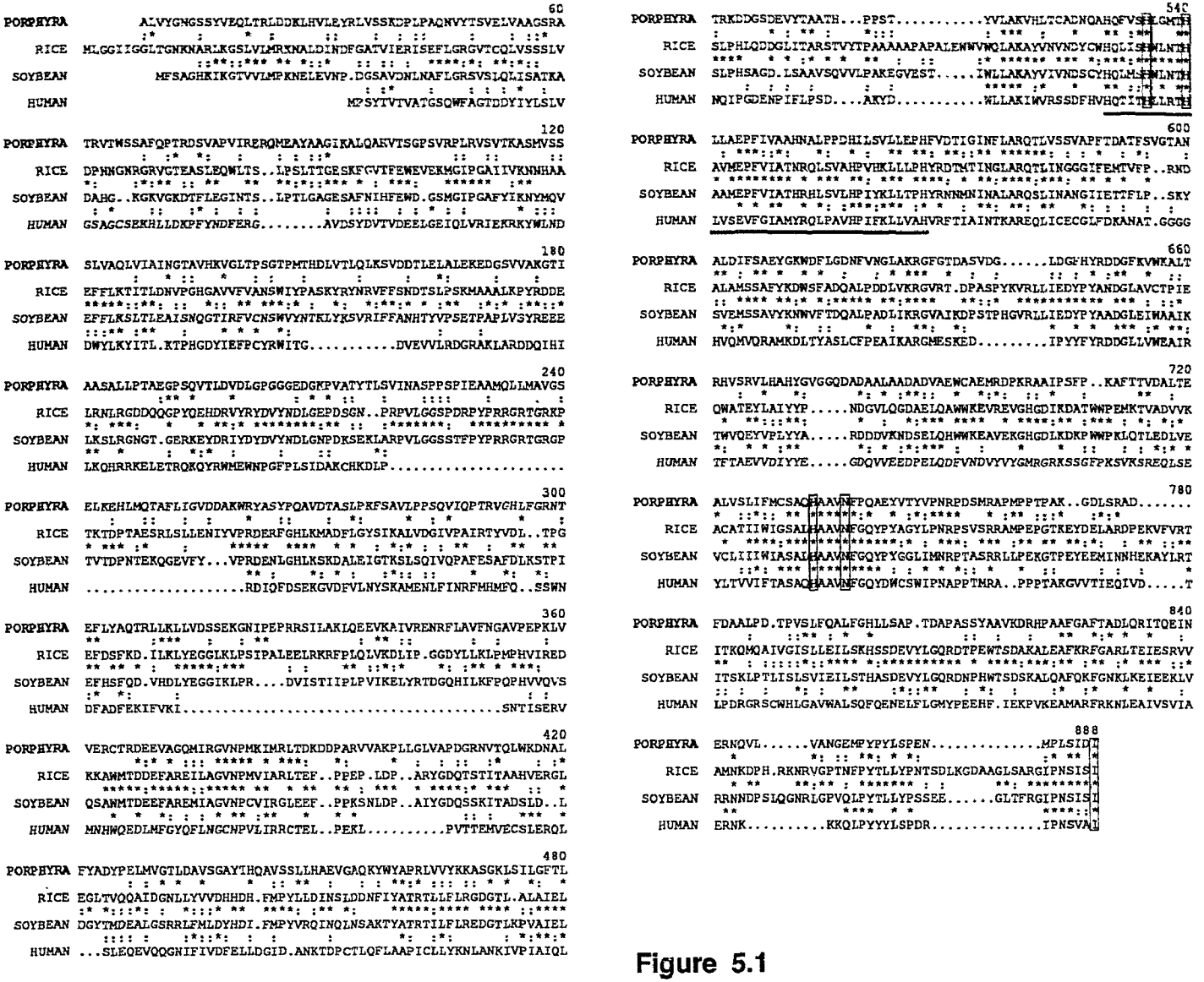

Figure 5.1

lipoxygenases (Fig. 5.1). A 38 residue region containing a cluster of five histidines in the form of His-X_4-His-X_4-His-X_{17}-His-X_8-His is found in all known lipoxygenase sequences (Steczko et al. 1992) and in the *Porphyra* lipoxygenase (Fig. 5.1). An additional conserved histidine residue, about 150-160 residues towards the C-terminus from this histidine-rich region, is also present in the *Porphyra* sequence. Further, the five residues that act as ligands for the enzymatically active iron atom in the soybean lipoxygenase-1 (Minor et al. 1993) are all found in the *Porphyra* lipoxygenase (Fig. 5.1). The evidence clearly shows that this cDNA encodes a lipoxygenase.

The original *Porphyra* lipoxygenase clone was isolated from a subtracted cDNA library enriched for gametophyte-specific cDNAs. To confirm the expression of the originating lipoxygenase gene, I performed northern hybridization with the C-terminus-encoding half of the cDNA insert as a probe of polyA$^+$ RNA from both the sporophyte and the gametophyte (Fig. 5.2). The lipoxygenase transcript (3.2 kb) is only detectable in the gametophyte and is present in high abundance relative to the major β-tubulin transcript. The *P. purpurea* lipoxygenase gene that generated this transcript is the first such gene shown to be developmentally regulated in algae. This phase-specific expression resembles the pattern of lipoxygenase gene expression in plant and animal tissues (Boado et al. 1992, Hildebrand et al. 1988, Thiele et al. 1987).

Plant lipoxygenases do not have a clearly defined developmental or physiological role. They function by providing fatty acid hydroperoxides, which

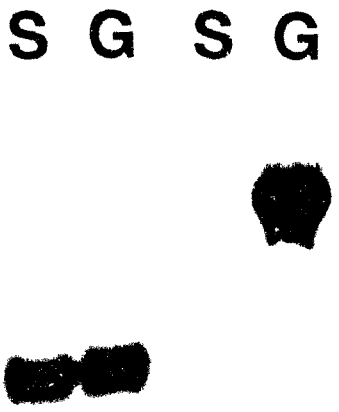

Figure 5.2 Northern analysis of the *Porphyra* lipoxygenase cDNA. One µg polyA$^+$ RNA from the sporophyte (S lanes) and the gametophyte (G lanes) were electrophoresed on a formaldehyde agarose gel, blotted to a nylon membrane and hybridized to a ^{32}P-labelled cDNA fragment ($2x10^6$ dpm ml^{-1}). Final washing was in 0.5 x SSC, 0.1% SDS, at 68 °C. The left panel is a control hybridization with a β-tubulin cDNA probe derived from an mRNA occurring in both phases. The right panel is the hybridization with the *Porphyra* lipoxygenase cDNA probe. The signals in the control panel were detected by autoradiography with an intensifying screen for 48 h. The signal for the *Porphyra* lipoxygenase cDNA was detected by autoradiography with an intensifying screen for 1.5 h.

act as precursors in a variety of metabolic pathways. Plant lipoxygenases are active in many different stages of normal growth and development, and in senescence, pest resistance and wound responses (Hildebrand et al. 1988, Siedow 1991). My results have shown that this *Porphyra* lipoxygenase gene is expressed only in the gametophyte. Although the sporophyte and gametophyte of *Porphyra* differ in morphology and subcellular structure (Pueschel and Cole 1985), and many genes are differentially transcribed in the two life cycle stages (Chapter 2), it is not possible to assign any of the general functions mentioned above exclusively to the gametophyte and thus relate lipoxygenase activity to gametophyte development.

Chapter 6: A Gametophyte-specific cDNA Encodes a Protein with Four Apparent Polysaccharide-binding Domains

INTRODUCTION

Red algal cell walls contain a variety of sulphated galactans, some of which (the agars and carrageenans) are commercially exploited and, consequently, have been chemically and physically well characterized (Percival and McDowell 1981, Craigie 1990). The sulphated galactans are flexible polysaccharides that constitute the bulk of the cell wall matrix. Embedded within this matrix, are more rigid, structural polysaccharides such as cellulose, mannans and xylans. Cellulose appears to be present in all red algae investigated, although in the gametophyte of bangiophytes it is replaced by mannan and xylan microfibrils (Turvey and Williams 1970, Mukai et al. 1981, Craigie 1990). Compared to the polysaccharides, very little is known about red algal cell wall proteins. The composition of proteins extracted from red algal cell walls varies among species and between life cycle stages of a single species; however, specific cell wall proteins have not been isolated or studied to any significant extent (Craigie 1990). Hydroxyproline, a distinctive component of the extensins, which are the major higher plant and green algal cell wall proteins (Siegel and Siegel 1973, Adair and Apt 1990, Schowalter 1993), is rarely detected in rhodophyte cell wall proteins. This observation suggests that, if the polysaccharides in rhodophyte cell walls are cross-linked

by proteins, a class of proteins different from the extensins must be involved.

The differences between the two life cycle phases of *Porphyra* are evident in the subcellular anatomy (Bourne et al. 1970, Mukai et al. 1981, Pueschel and Cole 1985) and the cell wall structure and composition (Mukai et al. 1981, Gretz et al. 1983, Gretz et al. 1986). The cell wall microfibrils of the sporophyte are primarily cellulose, while the gametophyte forms xylan microfibrils and contains little or no cellulose. Further, the gametophyte cell wall contains a large amount of a granular, insoluble mannan, whereas sporophyte walls have a high proportion of sulphated galactans, with mannan as only a minor component. The sporophyte galactans are chemically distinct from those of the gametophyte. In addition, the amino acid compositions of the total cell wall proteins of the two phases are different, with the sporophyte proteins distinguished by their high arginine, histidine and aspartic acid content and the gametophyte by high levels of alanine and glycine (Mukai et al. 1981, Gretz et al. 1986). Thus far, no cell wall proteins have been purified and characterized for either life cycle phase of *Porphyra*.

Among several cDNA clones isolated from a *P. purpurea* gametophyte-specific subtracted cDNA library was one that encodes a protein containing four apparent polysaccharide-binding domains (PBDs). *In vitro* translation studies demonstrate that this protein is glycosylated and is released to the lumen of the ER. The deduced characteristics of this protein suggest that it may be a secreted protein that binds cell wall polysaccharides.

MATERIALS AND METHODS

P. purpurea gametophytes and sporophytes were cultured as described by Mitman and van der Meer (1994). PolyA$^+$ RNA isolation, subtracted cDNA library construction and screening, northern hybridization and DNA sequencing were as described in Chapter 2. Since the isolated cDNA did not appear to encode an initiation codon, the amino terminus-encoding end of the mRNA sequence was obtained by a PCR approach described in Chapter 3, except that the 3' PCR primer was complementary to a sequence 60 bp downstream from the termination codon of the truncated cDNA and included an *Eco*RI recognition site at its 5' terminus. A PCR product of the approximate expected size (based on the mRNA length) was purified from an agarose gel using Geneclean (BIO 101) and was cloned into *Eco*RI digested pUC18 to produce clone 12a5L. This clone was sequenced as described previously (Chapter 2) and analyzed with the IBI/Pustell programs (Pustell and Kafatos 1984). Data bank searches and similarity analyses were conducted with the BLITZ (Smith and Waterman 1981), FASTA (Pearson and Lipman 1988) and BLAST (Altschul et al. 1990) programs. Sequence alignments were done with MULTALIN (Corpet 1988) and slightly modified after visual inspection.

For *in vitro* transcription and translation, the 12a5L cDNA was subcloned (Sambrook et al. 1989) downstream of the T7 promoter of the pTZ18R vector (Phamacia). The cloned gene was transcribed with the MegaScript kit (Ambion) and 3 µg of the product was translated in a reticulocyte lysate system (Ambion).

Two microliters of canine pancreatic microsomal membranes (Promega) were added to the *in vitro* translation reactions when testing signal peptide function. Translation products were digested (Andrews 1987) with either proteinase K (0.1 mg ml^{-1}, 0 °C, 45 min) or endoglycosidase H (0.5 mU ml^{-1}, 37 °C, 16 h), both from Boehringer Mannheim. Microsomal membranes were separated from the soluble proteins as described (Fujiki et al. 1982). Typically, 25 µl of translation mixture was diluted to 500 µl with 0.1 M sodium carbonate pH 11.5, incubated at 0 °C for 30 min and centrifuged at 4 °C for 1 h at 175,000 x *g* in a Beckman TL-100 rotor. The membrane pellet was washed with ice-cold distilled water and dissolved in 25 µl SDS gel buffer (Schagger and van Jagow 1987). Soluble proteins in the supernatant were precipitated with 10% trichloroacetic acid, washed three times with ice-cold ethyl ether and dissolved in 100 µl SDS gel buffer. One fifth of each of the soluble and membrane samples was loaded on an SDS-polyacrylamide gel. SDS-PAGE was according to Schagger and van Jagow (1987).

RESULTS

cDNA characterization

In the process of screening the *P. purpurea* subtracted cDNA libraries (Chapter 2), I found a cDNA clone, 12a5, that was apparently derived from a gene expressed only in the gametophytes. Northern hybridization experiments confirmed a gametophyte-specific expression pattern and showed that the 1 kb

transcript of the 12a5 gene was present in high abundance relative to the major β-tubulin transcript of *P. purpurea* (Fig. 6.1). The 12a5 cDNA was sequenced and found to contain a single open reading frame (ORF) of 193 codons (including a termination codon) and a 3' untranslated region of 170 bp. Protein data bank searches with the deduced amino acid sequence revealed that it contains four regions similar to the cellulose-binding domains (CBDs) of fungal cellulases. However, the 12a5-encoded amino acid sequence does not contain a methionine codon near its N-terminus, indicating that the ORF is incomplete. A cDNA (clone 12a5L) with the complete ORF (as defined by the presence of a complete signal peptide and initiation methionine codon at its N-terminus) was obtained by PCR amplification with the original, unsubtracted gametophyte cDNA library as the template. The complete ORF encodes a protein of 210 amino acids (Fig. 6.2A) that I have tentatively named the *P. purpurea* polysaccharide-binding protein or PBP.

Amino acid sequence analysis

Inspection of the amino acid sequence of the PBP indicates that the entire protein can be divided into discrete domains (Fig. 6.2B). The major feature of the protein is the occurrence of four very similar regions of 40 - 45 amino acids that resemble fungal CBDs. This protein also has a typical (von Heijne 1985) eukaryotic signal peptide of 22 amino acids that includes a charged N-terminus, a hydrophobic core and a probable cleavage site

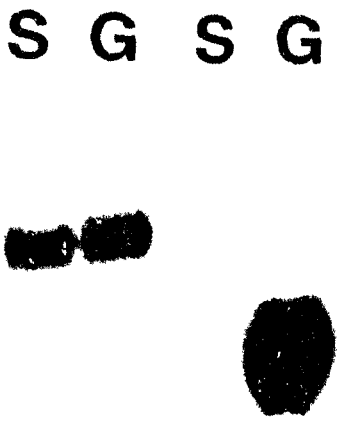

Figure 6.1 Northern analysis of the 12a5 cDNA. One µg polyA$^+$ RNA from the sporophyte (S lanes) and the gametophyte (G lanes) were electrophoresed on a formaldehyde agarose gel, blotted to a nylon membrane and hybridized to a ^{32}P-labelled cDNA fragment ($2x10^6$ dpm ml^{-1}). Final washing was in 0.5 x SSC, 0.1% SDS, at 68 °C. The first pair of lanes is a control hybridization with a β-tubulin cDNA probe derived from an mRNA occurring in both phases. The second pair of lanes shows hybridization with the 12a5 cDNA probe. The signals in the control panel were detected by autoradiography with an intensifying screen for 48 h. The signal for the 12a5 hybridization was detected by autoradiography with an intensifying screen for 1.5 h.

A

```
                                                                                GAATTCGCTGCCCTCTTCCCATCCCCCGGCGCTGCTGCTCCTTGCCCCGCCACACCACCCACAAG

    ATG GGC TTC CTC AAG GGC ACC GCC GCC GCG CTG ACG CTC CTG TCG GCC GCT GCC GCC GCA
1   M   G   F   L   K   G   T   A   A   A   L   T   L   L   S   A   A   A   A   A

    TCG GCG TGC GGT GTC CTC TAT GAG CAG TGC GGT GGC ATT GGC TTT GAC GGC GTC ACC TGC
21  S   A   C   G   V   L   Y   E   Q   C   G   G   I   G   F   D   G   V   T   C

    TGC TCC GAG GGG CTG ATG TGC ATG AAG ATG GGC CCC TAT TAC AGC CAG TGC CGC GCC ATG
41  C   S   E   G   L   M   C   M   K   M   G   P   Y   Y   S   Q   C   R   A   M

    CCC GGC ATG ATG GGC CAG GTT AAG CCC TAC GGC CAG TGC GGC GGC ATG AAC TAC TCT GGC
61  P   G   M   M   G   Q   V   K   P   Y   G   Q   C   G   G   M   N   Y   S   G
                                                                    *   *   *

    AAG ACC ATG TGC AGC CCC GGC TTC AAG TGC GTG GAG TTG AAC GAG TTT TTC TCG CAG TGC
81  K   T   M   C   S   P   G   F   K   C   V   E   L   N   E   F   F   S   Q   C

    GAC CTG GCC AAC AAG TCG CCG GTC GCG ACG CCC AAG GTG TCG CCG ACG TCC CCG CCG GGT
101 D   L   A   N   K   S   P   V   A   T   P   K   V   S   P   T   S   P   P   G
                *   *   *

    TCG GCG CAG GTG TGC GGC AAG GAG TAC GCA GCG TGC GGT GGC GAG ATG TTC ATG GGG GCG
121 P   A   Q   V   C   G   K   E   Y   A   A   C   G   G   E   M   F   M   G   A

    AAG TGC TGC AAG TTC GGC CTC GTC TGC TAC GAG ACG TCG GGC AAG TGG CAG TCG CAG TGC
141 K   C   C   K   F   G   L   V   C   Y   E   T   S   G   K   W   Q   S   Q   C

    CGC GCG CCG CCG CCC AAG ATG GGT GAG GTC GGG CGC TAT GCG CAG TGC GGT GGG ATG GGC
161 R   A   P   P   P   K   M   G   E   V   G   R   Y   A   Q   C   G   G   M   G

    TAC ATG GGC TCG ACC ATG TGC GTG GGC GGC TAC AAG TGT ATG GCG ATC TCG GAG GGC TCG
181 Y   M   G   S   T   M   C   V   G   G   Y   K   C   M   A   I   S   E   G   S

    ATG TAC AAG CAG TGC CTC CCC ATG CAC CCG TAG TGCGCCTCTTGACTGGACGGCGGGGGCCCTGTGGG
201 M   Y   K   Q   C   L   P   M   H   P   -

    GTCGCCATCATCGGCTAACATGATAGCGTAGCGCGCGGGGCCTTGCTGACAGGCTCTGGTTCGTCAAACACTTGTAGTC

    ACTTGCCTTCTGGCGCGCGTGGGGTAGCACATTATTCTATTTTTGTCAGACGGCGT
```

B

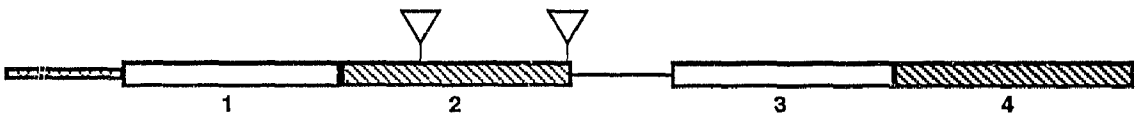

Figure 6.2 A) cDNA nucleotide sequence and deduced amino acid sequence of the encoded protein. Numbers in the left margin refer to amino acid residues. The signal peptide is underlined. Bold amino acids indicate the linker sequence. Asterisks under the amino acids indicate putative N-linked glycosylation signals. **B)** A schematic diagram showing the organization of the 12a5L-encoded protein. The thin hatched bar represents the signal peptide. The solid line refers to the linker. The four similar polysaccharide-binding domains are shown by shaded thick bars. The two glycosylation sites are indicated by inverted triangles.

"Ala-Ser-Ala" (Fig. 6.2A). The signal sequence is immediately followed by polysaccharide-binding domains (PBDs) I and II and a 19 amino acid linker sequence rich in proline and hydroxy amino acids connects domain II to domain III, with domain IV adjacent to III (Fig. 6.2A, 6.2B). The linker region is similar to the linkers joining the catalytic and cellulose-binding domains of cellulases from bacteria and fungi (Gilkes et al. 1991). There is no linker between PBD I and II, or PBD III and IV. Two putative N-linked glycosylation sites (Asn-Tyr-Ser and Asn-Lys-Ser) are located in the second PBD (Fig. 6.2A, 6.2B).

An alignment of the four PBDs of the *Porphyra* PBP with the CBDs of fungal endoglucanases and cellobiohydrolases is shown in Figure 6.3A. All the conserved residues in the various fungal CBDs, including the cysteines shown to be involved in disufide-bridge formation (Kraulis et al. 1989), are conserved in the PBDs. The important aromatic residues that presumably make contact with the polysaccharides (Poole et al. 1993, Beguin 1990) are present in the *P. purpurea* PBDs. There is a single amino acid insertion in domain III and a two amino acid insertion in domain IV, relative to domains I and II. These insertions appear to be located at the same site within the respective domains. An alignment of the two adjacent pairs of PBDs (PBDs I and III with PBDs II and IV) shows the high degree of sequence similarity between these two halves of the protein (Fig. 6.3B). In addition, PBDs I and III both have an extra potential disulphide-bridge, similar to that of the CBD of the *T. reesei* cellobiohydrolase II (Kraulis et al. 1989).

A

```
Porphyra #1      CGVLYE QCGG IGFD G VTC C SEGLM C MKM--GPYYS QC RAMPG
Porphyra #2   MMGQVKPYG QCGG MNYS G KTM C SPGFK C VEL--NEFFS QC DLANK
Porphyra #3      CGKEYA ACGG EMFM G AKC C KFGLV C YETS-GKWQS QC RAPPP
Porphyra #4   KMGEVGRYA QCGG MGYM G STM C VGGYK C MAISEGSMYK QC LPMHP
                                                                  Ref.
TR  EGLIII      *QQTVWG QCGG IGWS G PTN C APGSA C STL--NPYYA QC IPGAT  1
TR  CBHII     *QACSSVWG QCGG QNWS G PTC C ASGST C VYS--NDYYS QC LPGAA  2
TR  CBHI       GPTQSHYG QCGG IGYS G PTV C ASGTT C QVL--NPYYS QC L*     3
TR  EGLI       SCTQTHWG QCGG IGYS G CKT C TSGTT C QYS--NDYYS QC L*     4
PC  CBHI       GVTVPQWG QCGG IGYT G STT C ASPYT C HVL--NPYYS QC Y*     3
TV  CBHI       GPTQTHYG QCGG IGYI G PTV C ASGST C QVL--NPYYS QC L*     5
HG  CBHI       GPKAGRWQ QCGG IGFT G PTQ C EEPYI C TKL--NDWYS QC L*     6
                    ^         ^                     ^^
```

B

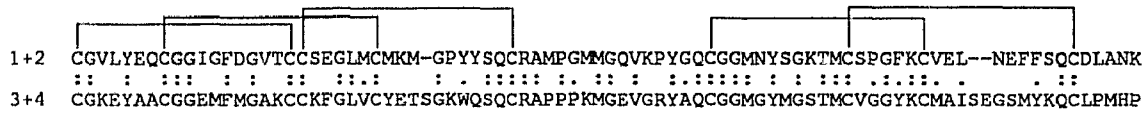

Figure 6.3 A) Alignment of the *Porphyra* PBDs with the CBDs from fungi. Highly conserved residues among all sequences are boxed. "^" indicates the conserved aromatic amino acids. TR - *Trichoderma reesei*, PC - *Phanerochaete chysosporium*, TV - *Trichoderma viride*, HG - *Humicola grisea*, EGL - endoglucanase, CBH - cellobiohydrolase, "*" - N or C terminus of the protein. Ref. 1 - Saloheimo et al. 1988, 2 - Teeri et al. 1987, 3 - Sims et al. 1988, 4 - Penttila et al. 1986, 5 - Cheng et al. 1990, 6 - Azevedo et al. 1990. **B)** Comparison of *Porphyra* paired PBDs. The possible disulfide-bridges equivalent to those in fungal CBDs are also indicated.

In vitro translocation and glycosylation of the PBP

To establish that the first 22 residues in the *P. purpurea* PBP do function as a signal peptide, 12a5L transcripts were translated in a rabbit reticulocyte *in vitro* translation system in the presence of microsomal membranes (Fig. 6.4A). Inclusion of microsomes in the translation mixture should yield a protein of lower molecular weight if the 22 amino acid signal peptide is cleaved after transport across the microsomal membrane. In contast to the β-lactamase control, the major protein detected by SDS-PAGE did not decrease in size with the inclusion of microsomes in the translation mixture (Fig. 6.4A, lanes 2 and 3), indicating that the signal peptide was not cleaved. However, two additional higher molecular weight proteins appeared on the gel. One explanation for the appearance of these proteins is that they result from the glycosylation of the PBP at either one or both of the putative glycosylation sites (Fig. 6.2A, 6.2B). Since glycosylation is another feature indicative of proper translocation across membranes, this result suggests that the signal peptide does function in directing the protein into the microsomes.

If translocation across the microsomal membrane does occur, the proteins in the lumen of the microsomes should be protected from proteolytic digestion. When the translation mixture was digested with proteinase K, both the non-glycoslylated and the apparently glycosylated forms of the protein were efficiently protected from digestion (Fig. 6.4A, lane 4). Protein molecules smaller than the expected product, probably premature termination products

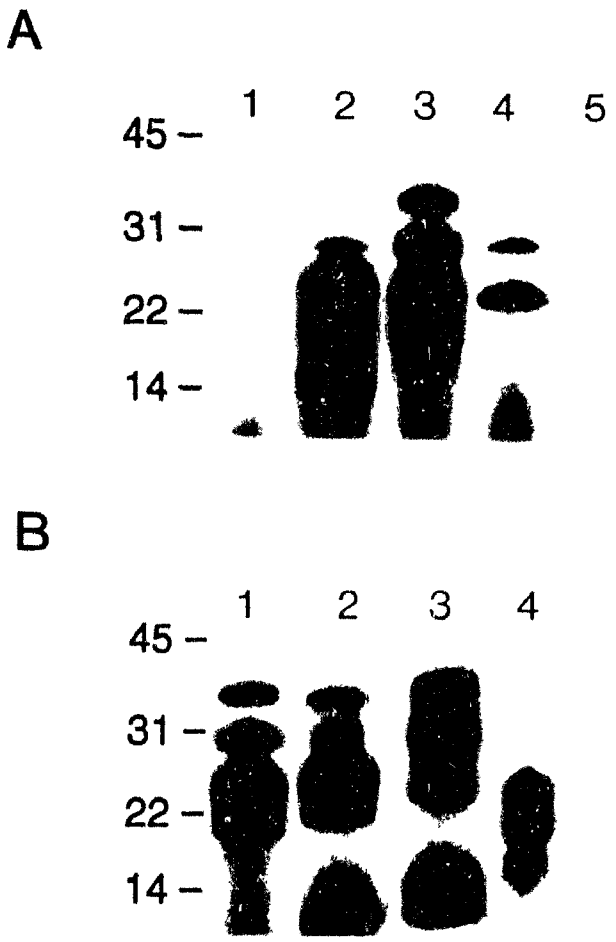

Figure 6.4 *In vitro* translocation and glycosylation of the *Porphyra* polysaccharide-binding protein (PBP). Signals were detected by autoradiograpy for 18 h. **A**) Translocation of the PBP: lane 1 - translation without 12a5L transcript; lane 2 - translation with 12a5L transcript but without microsomes; lane 3 - translation with 12a5L transcript and microsomes; lane 4 - translation as in lane 3 followed by proteinase K digestion; lane 5 - translation as in lane 3, followed by 0.1% Triton X-100 treatment and proteinase K digestion. **B**) Glycosylation and solubility of the PBP in the microsomes: lane 1 - translation with 12a5L transcript and microsomes; lane 2 - translation as in lane 1 followed by endoglycosidase H digestion; lane 3 - translation as in lane 1, soluble proteins in the supernatant of the microsomal lysate; lane 4 - translation as in lane 1, membrane proteins in the pellet of the microsomal lysate.

from in vitro synthesis, were completely digested during the proteinase K treatment. This experiment confirmed that the PBP is efficiently translocated into the microsomal lumen, even though the signal peptide is not cleaved (the bands in lane 4 are weaker than those in lane 3 due to the unavoidable rupture of some of the microsomes during handling). Solubilization of the microsomes with Triton X-100 allowed digestion of all bands by proteinase K (Fig. 6.4A, lane 5).

To test the hypothesis that the higher molecular weight species in Figure 6.4A (lanes 3 and 4) resulted from the addition of N-linked oligosaccharides, the translocated products were treated with endoglucosidase H, which cleaves between the two basal sugar residues that anchor the oligosaccharides to the protein. This enzyme treatment reduced the size of the two higher molecular weight proteins to near that of the non-glycosylated species (Fig. 6.4B, lanes 1 and 2), confirming that the PBP is glycosylated in the microsomes.

The observation that the signal peptide of the PBP is not cleaved after translocation suggested that the signal peptide might remain anchored in the membrane and keep the PBP associated with the membrane. Alternatively, the PBP might be released from the membrane by a mechanism other than the cleavage of the signal peptide. To test these hypotheses, I separated soluble and membrane-attached proteins by using high pH to lyse the microsomal membranes after translation. The soluble fraction (Fig. 6.4B, lane 3) of the lysate contained all of the glycosylated proteins and a portion of the

non-glycosylated species. Only the small, incomplete proteins and some of the non-glycosylated molecules remained attached to the membranes (Fig. 6.4B, lane 4), suggesting that the mature PBP is not membrane-bound. Thus, although the signal peptide is not cleaved, the PBP is released to the lumen of the microsomes.

DISCUSSION

In this study, I have characterized a cDNA derived from a gene that apparently encodes a polysaccharide-binding protein and is expressed specifically in the *P. purpurea* gametophyte. I have obtained circumstantial evidence that this protein is a structural element of the gametophyte cell wall: (1) the encoded protein consists almost entirely of four domains that each resemble the CBDs of fungal cellulases and it lacks any obvious catalytic site; (2) the cDNA is derived from an abundant mRNA species as would be expected for a gene encoding a major cell wall protein, given that cell walls make up a large portion of the gametophyte biomass, and; (3) the encoded protein has a typical signal peptide that allows it to be translocated into the microsomal lumen, suggesting that the protein is secreted *in vivo*.

The hypothesis that the repeated domains of the *P. purpurea* 12a5L-encoded protein bind polysaccharides is further supported by their similarity to the chitin-binding domains of plant chitinases and wheat germ agglutinins (WGAs). Chitin-binding domains are also 40 - 45 amino acids in length but

have four, rather than two or three disulphide bridges. Nearly all of the boxed residues of PBDs in Fig. 6.3 (except the first C and the last Q) can be aligned with their counterparts in chitin-binding domains by the insertion of a three amino acid gap near the beginning of the chitin-binding domains. The overall amino acid identity of the PBDs to chitin-binding domains is approximately 25%. In addition, the structural organization of the *P. purpurea* PBP is similar to that of WGAs, which also contain a signal peptide and four domains with similar amino acid sequences (Wright and Raikhel 1989). However, WGAs lack the linker region between PBDs II and III of the PBP. The similarity of the *P. purpurea* PBDs to both cellulose- and chitin-binding domains strongly suggests that PBP binds carbohydrate polymers.

While amino acid sequence similarities point toward a polysaccharide-binding function for the *Porphyra* PBP, the identity of these polysaccharides remains unclear. The *Porphyra* gametophyte lacks cellulose, instead having xylan microfibrils and a granular, insoluble mannan (Mukai et al 1981). The PBDs are similar to domains that interact with cellulose or chitin, both of which are β(1,4)-linked polymers of six carbon sugars (glucose or N-acetyl glucosamine, respectively). Since mannose differs from glucose only in the orientation of the hydroxyl group on carbon-2, it seems possible that a β(1,4)-linked mannan is the likely target for the *Porphyra* PBP. An additional uncertainty about the *P. purpurea* PBP is the number of polysaccharide polymers it might bind. Since a single CBD in fungal cellulases binds cellulose

efficiently, it is possible that the PBP might interact simultaneously with four individual polymers. However, the placement of the flexible linker domain between PBDs II and III suggests that only two polymers are bound simultaneously, with PBDs I and II or PBDs III and IV binding adjacent regions of the same polymer, and each pair of PBDs binding a different polymer. Such a mechanism would probably result in the tight crosslinking of the two polysaccharide chains.

The observation that the *Porphyra* PBP has a typical signal peptide but appears to undergo translocation across the microsomal membrane without cleavage of the signal peptide is unusual. An N-terminal signal peptide is necessary to direct the protein translation machinery to the ER and to promote entry of the nascent polypeptide chains into the secretory pathway (Chrispeels 1991). Usually, signal peptides of secretory proteins are then cleaved by a signal peptidase; however, exceptions have been reported (Lingappa et al. 1979, Chrispeels 1991), and thus the actual cleavage of the signal peptide appears to be dispensable in at least some systems (Gorlich et al. 1992, Rapoport 1992). The signal peptide in the *Porphyra* PBP contains a typical cleavage site (Ala-X-Ala) at the C-terminus, so recognition of the cleavage site by the signal peptidase should not be a problem. However, immediately adjacent to the cleavage site is a cysteine residue that, based on the known disulphide bridges in fungal CBDs, is likely to participate in a disulphide bridge (Fig. 6.3B). If this disulphide-bridge is formed prior to the complete

translocation of the protein through the ER membrane, the cleavage site might not be accessible to the signal peptidase and the PBP would have to be released from the membrane through a different mechanism.

The cellular destination of a signal peptide-containing protein can often be inferred from its amino acid sequence. The default pathway for any protein entering the ER is secretion through the Golgi apparatus to the cell surface. All diversions from this path to cellular compartments require additional information in the protein's amino acid sequence (Pryer et al. 1992). The amino acid sequence of the *P. purpurea* PBP does not contain any of the known signals for ER retention (Chrispeels 1991, Bednarek and Raikhel 1992), Golgi retention (Hurtley 1992) or vacuole targeting (Chrispeels and Raikhel 1992), suggesting that this protein is probably secreted to the cell wall. However, Golgi retention signals have not been specifically identified in plants or red algae and the biosynthesis of some red algal cell wall polysaccharides occurs in the Golgi (Hawkins 1974, Tsekos 1985). Thus, an alternative hypothesis is that the PBP has a role within the Golgi to aggregate polysaccharides destined for the cell wall.

In vitro studies have shown that the *P. purpurea* PBP is glycosylated at two sites, but it is not known whether the glycosylation events are critical to the function of the mature protein. N-linked oligosaccharides added to proteins in the ER promote the correct folding of the nascent polypeptide and prevent proteolytic breakdown in the Golgi (Chrispeels 1991). Glycosylation of WGAs

on the C-terminal propeptide is important for their transport from the Golgi bodies to the vacuoles (Wilkins et al. 1990). The propeptide together with the N-linked oligosaccharides are then removed in the vacuole to yield the mature protein. The glycosylation sites of the *Porphyra* PBP are located near the middle of the protein and are unlikely to be removed by the cleavage of a propeptide, suggesting that the mature protein is glycosylated. A glycosylation signal occurs in the *T. reesii* cellobiohydrolase II CBD at the same position as the first putative glycosylation site of PBP (Fig. 6.3A), but is absent on most other CBDs suggesting that the occurrence of glycosylation at this position does not influence substrate binding.

The isolation of a *P. purpurea* gametophyte-specific cDNA that apparently encodes a cell wall structural protein has demonstrated that subtracted cDNA libraries and database-searching methods can provide a novel approach to phycological questions. If the protein described in this work does bind cell wall polysaccharides, its isolation and characterization by traditional methods would probably prove difficult. With the present information, protein purification could be guided by the characteristics expected for the amino acid sequence, or antibodies could be made to part or all of the expressed protein. Further characterization of the PBP, and further screening of the *P. purpurea* subtracted libraries for additional developmentally regulated cDNAs encoding cell wall proteins could provide new insights into the structure of red algal cell walls.

Chapter 7: Conclusions

This study has demonstrated that subtracted cDNA libraries are powerful tools for isolating a large number and variety of developmentally regulated algal genes without requiring any prior knowledge about the genes or their products. Following the isolation and sequencing of such phase-specific cDNAs, the putative functions of those genes whose counterparts have been characterized from related organisms can be identified through protein data bank searches. These approaches have allowed me to isolate and identify phase-specific cDNAs encoding the first developmentally regulated EF-1α outside the animal kingdom, two putative red algal cell wall proteins that would be difficult to obtain by traditional methods, and a gametophyte-specific lipoxygenase. For those phase-specific genes that cannot be identified by data bank searches, the isolation and sequencing of their respective cDNAs is still an important step towards the identification of their functions. The cloned cDNAs could be expressed in *E. coli* and the purified proteins could then be used either to produce antibodies for immunological studies or other assays leading to the identification of these proteins. The phase-specific cDNAs could be used to produce RNA probes for *in situ* hybridization experiments to determine the spatial and temporal regulation of these genes. In addition, the labelled cDNA inserts could be used as probes to retrieve genomic clones for the isolation of red algal promoters sensitive to phase-specific regulation. Once transient

expression systems or the capability to produce transgenic plants are available in *Porphyra*, the cis-acting elements controlling phase-specific gene expression can be more precisely identified and characterized. The various cis-acting elements would be useful probes for the identification of specific regulatory proteins essential to the control of development. Furthermore, phase-specific promoters would be useful tools for construction of genetically engineered strains for biotechnological applications.

References

Adair, W. S. & Apt, K. E. (1990) Cell wall regeneration in *Chlamydomonas*: accumulation of mRNAs encoding cell wall hydroxyproline-rich glycoproteins. *Proc. Natl. Acad. Sci.* USA 87:7355-7359.

Andrews, D. (1987) Assaying protein translocation across the endoplasmic reticulum membrane. *Promega Notes* 11:1-4.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

Azevedo, M. de O., Felipe, M. S., Astolfi, S. Filho & Radford, A. (1990) Cloning, sequencing and homologies of the cbh-1 (exoglucanase) gene of *Humicola grisea var. thermoidea. J. Gen. Microbiol.* 136:2569-2576.

Baldauf, S. L. & Palmer, J. D. (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci.* USA 90:11558-11562.

Bednarek, S. Y. & Raikhel, N. V. (1992) Intracellular trafficking of secretory proteins. *Plant Mol. Biol.* 20:133-150.

Beguin, P. (1990) Molecular biology of cellulose degradation. *Annu. Rev. Microbiol.* 44:219-248.

Beneytout, J. L., Andrianarison, R. H., Rakotoarisoa, Z. & Tixier, M. (1989) Properties of a lipoxygenase in green algae (Oscillatoria sp.). *Plant Physiol.* 91:367-372.

Berchtold, H., Reshetnikova, L., Reiser, C. O. A., Schirmer, N. K., Sprinzl, M. & Hilgenfeld R. (1993) Crystal structure of active elongation factor Tu reveals major domain rearrangements. *Nature* 365, 126-132.

Blumenthal, T., Landers, T. A. & Weber, K. (1972) Bacteriophage Qß replicase contains the protein biosynthesis elongation factor EF Tu and EF Ts. *Proc. Natl. Acad. Sci.* 69, 1313-1317.

Boado, R. J., Pardridge, W. M., Vinters, H. V. & Black, K. (1992) Differential expression of arachidonate 5-lipoxygenase transcripts in human brain tumours: evidence for the expression of a multitranscript family. *Proc. Natl. Acad. Sci.* USA 89:9044-9048.

Bold, H. C. & Wynne, M. J. (1985) *Introduction to the algae*. 720 pp. Prentice-Hall, Inc., Englewood Cliffs, N.J. 07632

Bourne, V. L., Conway, E. & Cole, K. (1970) On the ultrastructure of pit connections in the conchocelis phase of the red alga *Porphyra perforata* J. Ag *Phycologia* 9:79-81.

Cheng, C., Tsukagoshi, N. & Udaka, S. (1990) Nucleotide sequence of the cellobiohydrolase gene from *Trichoderma viride*. *Nucl. Acids Res.* 18:5559.

Chrispeels, M. J. (1991) Sorting proteins in the secretory system. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 42:21-53.

Chrispeels, M. J. & Raikhel, N. V. (1992) Short peptide domains target proteins to plant vacuoles. *Cell* 68:613-616.

Cole, S. & Hamilton, A. (1993) First int'l. symposium on the mapping and sequencing of small genomes. *Genetic Engineering News* 13:34-35.

Conkling, M. A., Cheng, C., Yamamoto, Y. T. & Goodman, H. M. (1990) Isolation of transcriptionally regulated root-specific genes from tobacco. *Plant Physiol.* 93:1203-1211.

Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* 16:10881-10891.

Cottrelle, P., Cool, M., Thuriaux, P., Price, V. L., Thiele, D., Buhler, J. M. & Fromageot, P. (1985) Either one of the two yeast EF-1 alpha genes is required for cell viability. *Curr. Genet.* 9:693-697.

Craigie, J. S. (1990) Cell walls. In *Biology of the Red Algae*, Eds. K. M. Cole & R. G. Sheath, pp. 221-257 Cambridge University Press, Cambridge, UK.

De Haën, C., Neurath, H. & Teller, D. C. (1975) The phylogeny of trypsin-related serine proteases and their zymogens. New methods for the investigation of distant evolutionary relationships. *J. Mol. Biol.* 92:225-259.

Devereux, J., Haeberli, P., & Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* 12:387-395.

Dje, M. K., Mazabraud, A., Viel, A., Le-Maire, M., Denis, H., Crawford, E. & Brown, D. D. (1990) Three genes under different developmental control

encode elongation factor 1- alpha in Xenopus laevis. *Nucl. Acids Res.* 18:3489-3493.

Drew, K. M. (1949) Conchocelis-phase in the life history of *Porphyra umbilicalis* (L.) Kutz. *Nature* 164:748-749.

Duguid, J. R., Bohmont, C. W., Liu, N. & Tourtellotte, W. W. (1989) Changes in brain gene expression shared by scrapie and Alzheimer disease. *Proc. Natl. Acad. Sci.* USA 86:7260-7264.

Duguid, J. R., Rohwer, R. G. & Seed, B. (1988) Isolation of cDNAs of scrapie-modulated RNAs by subtractive hybridization of a cDNA library. *Proc. Natl. Acad. Sci.* USA 85:5738-5742.

Evans, D. L., Marshall, C. J., Christey, P. B. & Carrell, R. W. (1992) Heparin binding site, conformational change, and activation of antithrombin. *Biochemistry* 31:12629-12642.

Felsenstein, J. (1991) Phylogeny Inference Package [PHYLIP 3.5C] (University of Washington, Seattle).

Forchhammer, K., Rucknagel, K. P. & Bock, A. (1990) Purification and biochemical characterization of SELB, a translation factor involved in selenocysteine synthesis. *J. Biol. Chem.* 265:9346-9350.

Fujiki, Y., Hubbard, N. L., Fowler, S. & Lazarow, P. B. (1982) Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *J. Cell Biol.* 93:97-102.

Funk, C. D., Funk, L. B., Fitzgerald, G. A. & Samuelsson, B. (1992) Characterization of human 12-lipoxygenase genes. *Proc. Natl. Acad. Sci.* USA 89:3962-3966.

Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, JR R. C. & Warren, R. A. J. (1991) Domains in microbial ß-1,4-glycanases: sequence conservation, function, and enzyme families. *Microbiol. Rev.* 55:303-315.

Goldberg, R. B. (1988) Plants: novel developmental processes. *Science* 240:1460-1467.

Gorlich, D., Hartmann, E., Prehn, S. & Rapoport, T.A. (1992) A protein of the endoplasmic reticulum involved early in polypeptide translocation. *Nature* 357:47-52.

Gretz, M. R., Aronson, J. M. & Sommerfeld, M. R. (1986) Cell wall composition of the conchocelis phases of *Bangia atropurpurea* and *Porphyra leucosticta* (Rhodophyta). *Bot. Mar.* 29:91-96.

Gretz, M. R., McCardless, E. L., Aronson, J. M. & Sommerfeld, M. R. (1983) The galactan sulphates of the conchocelis phases of *Porphyra leucosticta* and *Bangia atropurpurea* (Rhodophyta). *J. Exp. Bot.* 34:705-711.

Gubler, U. & Hoffman, B. J. (1983) A simple and very efficient method for generating cDNA libraries. *Gene* 25:263-269.

Hatfield, D., Choi, I. S., Mischke, S. & Owens, L. D. (1992) Selenocysteyl-tRNAs recognize UGA in *Beta vulgaris*, a higher plant, and in *Gliocladium virens*, a filamentous fungus. *Biochem. Biophys. Res. Commun.* 184:254-259.

Hawkins, E. K. (1974) Golgi vesicles of uncommon morphology and wall formation in the red alga, *Polysiphonia. Protoplasma* 80:1-14.

Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992) CLUSTAL V: improved software for multiple sequence alignment. *CABIOS* 8:189-191.

Hildebrand, D. F., Hamilton-Hemp, T. R., Legg, C. S. & Bookjans, G. (1988) Plant lipoxygenases: occurrence, properties and possible functions. *Curr. Top. Plant Biochem. Physiol.* 7:201-219.

Højrup, P., Jensen, S. M. & Petersen, T. E. (1985) Amino acid sequence of bovine protein Z: a vitamin K-dependent serine protease homolog. *FEBS Lett.* 184:333-338.

Hovemann, B., Richter, S., Walldorf, U. & Cziepluch, C. (1988) Two genes encode related cytoplasmic elongation factors 1 alpha (EF-1 alpha) in Drosophila melanogaster with continuous and stage specific expression. *Nucl. Acids Res.* 16:3175-3194.

Hoxmark, R. C. (1976) Protein composition of different stages in the life cycle of *Ulva mutabilis*, Føyn. *Planta* 130:327-332.

Hurtley, S. M. (1992) Golgi localization signals. *Trends Biochem. Sci.* 17:2-3.

Hushovd, O. T., Gulliksen, O. M. & Nordby, Ø. (1982) Absence of major differences between soluble proteins from haploid gametophytes and

diploid sporophytes in the green alga *Ulva mutabilis* Føyn. *Planta* 156:89-91.

Isackson, P. J. & Bradshaw, R. A. (1984) The α-subunit of mouse 7 S nerve growth factor is an inactive serine protease. *J. Biol. Chem.* 259:5380-5383.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phyiogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. USA* 86:9355-9359.

Jany, K. D., Bekelar, K., Pfleiderer, G. & Ishay, J. (1983) Amino acid sequence of an insect chymotrypsin from the larvae of the hornet, *Vespa orientalis*. *Biochem. Biophys. Res. Com.* 110:1-7.

John, M. E. & Crow, L. J. (1992) Gene expression in cotton (*Gossypium hirsutum* L.) fiber: Cloning of the mRNAs. *Proc. Natl. Acad. Sci.* USA 89:5769-5773.

Kelly, A. J., Zagotta, M. T., White, R. A., Chang, C. & Meeks-Wagner, D. R. (1990) Identification of genes expressed in the tobacco shoot apex during the floral transition. *Plant Cell* 2:963-972.

Kinzy, T. G., Freeman, J. P., Johnson, A. E. & Merrick, W. C. (1992) A model for the aminoacyl-tRNA binding site of eukaryotic elongation factor 1 alpha. *J. Biol. Chem.* 267:1623-1632.

Kjeldgaard, M. & Nyborg, J. (1992) Refined structure of elongation factor EF-Tu from *Escherichia coli*. *J. Mol. Biol.* 223:721-742.

Knudsen, S. M., Frydenberg, J., Clark, B. F. & Leffers, H. (1993) Tissue-dependent variation in the expression of elongation factor-1 alpha isoforms: isolation and characterisation of a novel variant of human elongation-factor 1 alpha. *Eur. J. Biochem.* 215:549-554.

Koltunow, A. M., Truettner, J., Cox, K. H., Wallroth, M. & Goldberg, R. B. (1990) Different temporal and spatial gene expression patterns occur during anther development. *Plant Cell* 2:1201-1224.

Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J. & Gronenborn, A. M. (1989) Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and

hybrid distance geometry-dynamical simulated annealing. *Biochemistry* 28:7241-7257.

Kuriyama, R., Savereide, P., Lefebvre, P. & Dasgupta, S. (1990) The predicted amino acid sequence of a centrosphere protein in dividing sea urchin eggs is similar to elongation factor (EF-1 alpha). *J. Cell Sci.* 95:231-236.

Kurosky, A., Barnett, D. R., Lee, T-H., Touchstone, B. Hay, R. E., Arnott, M. S., Bowman, B. H. & Fitch, W. M. (1980) Covalent structure of human haptoglobin: a serine protease homolog. *Proc. Natl. Acad. Sci.* USA 77:3388-3392.

Lee, B. J., Rajagopalan, M., Kim, Y. S., You, K. H., Jacobson, K. B. & Hatfield, D. (1990) Selenocysteine $tRNA^{[Ser]Sec}$ gene is ubiquitous within the animal kingdom. *Mol. Cell. Biol.* 10:1940-1949.

Lee, Y. H. & Vacuier, V. D. (1992) Reusable cDNA libraries coupled to magnetic beads. *Anal. Biochem.* 206:206-207.

Lee, S., Wolfraim, L. A. & Wang, E. (1993) Differential expression of S1 and elongation factor-1α during rat development. *J. Biol. Chem.* 268:24453-24459.

Liboz, T., Bardet, C., Le-Van-Thai, A., Axelos, M. & Lescure, B. (1990) The four members of the gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1 alpha are actively transcribed. *Plant Mol. Biol.* 14:102-110.

Lingappa, V. R., Lingappa, J. R. & Bloble, G. (1979) Chicken ovalbumin contains an internal signal sequence. *Nature* 281:117-121.

MacKay, R. M. & Gallant, J. W. (1991) Beta-tubulins are encoded by at least four genes in the brown alga *Ectocarpus variabilis. Plant Mol. Biol.* 17:487-492.

Matsumoto, T, Funk, C. D, Radmark, O., Hoog, J. O., Jornvall, H. & Samuelsson, B. (1988) Molecular cloning and amino acid sequence of human 5-lipoxygenase. *Proc. Natl. Acad. Sci.* USA 85:26-30.

Metz-Boutigue, M. L., Reinbolt, J., Ebel, J. P., Ehresmann, C. & Ehresmann, B. (1989) Crossliking of elongation factor Tu to $tRNA^{Phe}$ by *trans*-diamminedichloroplatinum (II). *FEBS Letters* 245:194-200.

Miller, J. S., George, M. & Schwartz, L. B. (1990) Cloning and characterization

of a second complementary DNA for human tryptase. *J. Clin. Invest.* 86:864-870.

Miller, J. S., Westin, E. H. & Schwartz, L. B. (1989) Cloning and characterization of complementary DNA for human tryptase. *J. Clin. Invest.* 84:1188-1195.

Minor, W., Steczko, J., Bolin, J. T., Otwinowski, Z. & Axelrod, B. (1993) Crystallographic determination of the active site iron and its ligands in soybean lipoxygenase-1. *Biochemistry* 32:6320-6323.

Mitman, G. G. & van der Meer, J. P. (1994) Meiosis, blade development and sex determination in *Porphyra purpurea* (Rhodophyta). *J. Phycol.* 30:147-159.

Moghaddam, M. F. & Gerwick, W. H. (1990) 12-lipoxygenase activity in the red marine alga *Gracilariopsis lemaneiformis*. *Phytochemistry* 29:2457-2459.

Mukai, L. S., Craigie, J. S. & Brown, R. G. (1981) Chemical composition and structure of the cell walls of the conchocelis and thallus phases of *Porphyra tenera* (Rhodophyceae). *J. Phycol.* 17:192-198.

Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K. & Shimizu, S. (1989) Molecular cloning and expression of human hepatocyte growth factor. *Nature* 342:440-443.

Ohta, K., Toriyama, M., Miyazaki, M., Murofushi, H., Hosoda, S., Endo, S. & Sakai, H. (1990) The mitotic apparatus-associated 51-kDa protein from sea urchin eggs is a GTP-binding protein and is immunologically related to yeast polypeptide elongation factor 1 alpha. *J. Biol. Chem.* 265:3240-3247.

Ohta, H., Shirano, Y., Tanaka, K., Morita, Y. & Shibata, D. (1992) cDNA cloning of rice lipoxygenase L-2 and characterization using an active enzyme expressed from the cDNA in *Escherichia coli*. *Eur. J. Biochem.* 206:331-336.

Pearson, W. R. & Lipman, D. J. (1988) Improved tool for biological sequence analysis. *Proc. Natl. Acad. Sci.* USA 85:2444-2448.

Pedersen, S., Bloch, P. H., Reeh, S. & Neidhardt, F. C. (1978) Patterns of protein synthesis in *E. coli*: a catalog of the amount of 140 individual proteins at different growth rates. *Cell* 14:179-190.

Penttila, M., Lehtovaara, P., Nevalainen, H., Bhikhabhai, R. & Knowles, J. (1986) Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the endoglucanase I gene. *Gene* 45:253-263.

Percival, E. & McDowell, R. H. (1981) Algal walls - composition and biosynthesis. *Encyclopedia Plant Physiol.* N. S. 13B:277-316.

Petersen, L. C. (1983) modulatory effects on proteinase kinetics caused by association of both enzyme and substrate to heparin. *Eur. J. Biochem.* 137:531-535.

Poole, D. M., Hazlewood, G. P., Huskisson, N. S., Virdon, R. & Gilbert, H. J. (1993) The role of conserved tryptophan residues in the interaction of a bacterial cellulose binding domain with its ligand. *FEMS Microbiol. Lett.* 106:77-84.

Pryer, N. K., Wuestehube, L. J. & Schekman, R. (1992) Vesicle-mediated protein sorting. *Annu. Rev. Biochem.* 61:471-516.

Pueschel, C. M. & Cole, K. M. (1985) Ultrastructure of germinating carpospores of *Porphyra variegata* (Kjellm.) Hus (Bangiales, Rhodophyta). *J. Phycol.* 21:146-154.

Pustell, J. & Kafatos, F. C. (1984) A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determinations. *Nucl. Acids Res.* 12:643-655.

Rapoport, T. A. (1992) Transport of proteins across the endoplasmic reticulum membrane. *Science* 258:931-936.

Reith, M. & Munholland, J. (1993) A high-resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. *Plant Cell* 5:465-475.

Reynolds, D. S., Gurley, D. S., Austen, K. F. & Serafin, W. E. (1991) Cloning of cDNA and gene of mouse mast cell protease-6. *J. Biol. Chem.* 266:3847-3853.

Rice, E. L. & Bird, C. J. (1990) Relationships among geographically distant populations of *Gracilaria verrucosa* (Gracilariales, Rhodophyta) and related species. *Phycologia* 29:501-510.

Riis, B., Rattan, S. I., Clark, B. F. C. & Merrick, W. C. (1990) Eukaryotic protein

elongation factors. *Trends Biochem. Sci.* 15:420-424.

Rodriguez, I. R. & Chader, G. J. (1992) A novel method for the isolation of tissue-specific genes. *Nucl. Acids Res.* 20:3528.

Rost, B. & Sander, C. (1993) Improved prediction of protein seconda y structure by use of sequence profiles and neural networks. *Proc. Natl. A :ad. Sci. USA* 90:7558-7562.

Rubenstein, J. L. R., Brice, A. E. J., Ciaranello, R. D., Denney, D., Porteus, M. H. & Usdin, T. B. (1990) Subtractive hybridization system using single-stranded phagemids with directional inserts. *Nucl. Acids Res.* 18:4833-4842.

Saloheimo, M., Lehtovaara, P., Penttila, M., Teeri, T. T., Stahlberg, J., Johansson, G., Pettersson, G., Claeyssens, M., Tomme, P. & Knowles, J. K. C. (1988) EGIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. *Gene* 63:11-21.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Schagger, H. & von Jagow, G. (1987) Tricine-sodium dodecyl sulphate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biol. chem.* 136:368-379.

Schowalter, A. M. (1993) Structure and function of plant cell wall proteins. *Plant Cell* 5:9-23.

Schwartz, L. B. & Bradford, T. (1986) Regulation of tryptase from human lung mast cell by heparin. *J. Biol. Chem.* 261:7372-7379.

Schweinfest, C. W., Henderson, K. W., Gu, J. R., Kottaridis, S. D., Besebas, S., Panotopoulou, E. & Papas, T. S. (1990) Subtraction hybridization cDNA libraries from colon carcinoma and hepatic cancer. *Genet. Anal. Techn. Appl.* 7:64-70.

Sela, S., Yogev, D., Razin, S. & Bercovier, H. (1989) Duplication of the tuf gene: a new insight into the phylogeny of eubacteria. *J. Bacteriol.* 171:581-584.

Shibata, D., Steczko, J., Dixon, J. E., Hermodson, M., Yazdanparast, R. &

Axelrod, B. (1987) Primary structure of soybean lipoxygenase-1. *J. Biol. Chem.* 262:10080-10085.

Siedow, J. N. (1991) Plant lipoxygenases: structure and function. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 42:145-188.

Siegel, B. Z. & Siegel, S. M. (1973) The chemical composition of algal cell walls. *Crit. Rev. Microbiol.* 3:1-26.

Sigal, E., Craik, C. S., Highland, E., Grunberger, D., Costello, L. L., Dixon, R. A. F. & Nadel, J. A. (1988) Molecular cloning and primary structure of human 15-lipoxygenase. *Biochem. Biophys. Res. Com.* 157:457-464.

Sims, P., James, C. & Broda, P. (1988) The identification, molecular cloning and characterization of a gene from *Phanerochaete chrysosporium* that shows strong homology to the exo-cellobiohydrolase I gene from *Trichoderma reesei*. *Gene* 74:411-422.

Smith, T. F. & Waterman, M. S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.* 147:195-197.

Steczko, J., Donoho, G. P., Clemens, J. C., Dixon, J. E. & Axelrod, B. (1992) Conserved histidine residues in soybean lipoxygenase: functional consequences of their replacement. *Biochemistry* 31:4053-4057.

Swaroop, A., Xu, J., Agarwal, N. & Weissman, S. M. (1991) A simple and efficient cDNA library subtraction procedure: isolation of human retina-specific cDNA clones. *Nucl. Acids Res.* 19:1954.

Swofford, D.L. (1991) PAUP: Phylogenetic Analysis Using Parsimony version 3.0r (Illinois Natural History Survey, Champaign, Illinois).

Swofford, D. L. & Olsen, G. J. (1990) In *Molecular Systematics*, eds. Hillis, D.M. & Moritz, C. (Sinauer Assoc., Sunderland, Mass.), pp.411-501.

Teeri, Y. T., Lehtovaara P., Kauppinen S., Salovuori I. and Knowles J. (1987) Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. *Gene* 51:43-52.

Theerakulpisut, P., Xu, H., Singh, M. B., Pettitt, J. M. & Knox, R. B. (1991) Isolation and developmental expression of Bcp1, an anther-specific cDNA clone in *Brassica campestris*. *Plant Cell* 3: 1073-1084.

Thiele, B. J., Fleming, J., Kasturi, K., O'Prey, J., Black, E., Chester, J.,

Rapoport, S. M. & Harrison, P. R. (1987) Cloning of a rabbit erythroid-cell specific lipoxygenase mRNA. *Gene* 57:111-119.

Travers, A. (1973) Control of ribosomal RNA synthesis *in vitro*. *Nature* 244:15-17.

Troutt, A. B., McHeyzer-Williams, M. G., Pulendran, B. & Nossal, D. J. V. (1992) Ligation-anchored PCR: a simple amplification technique with single-sided specificity. *Proc. Natl. Acad. Sci.* USA 89:9823-9825.

Tsekos, I. (1985) The endomembrane system of differentiating carposporangia in the red alga *Chondria tenuissima*: occurrence and participation in secretion of polysaccharide and proteinaceous substances. *Protoplasma* 129:127-136.

Turvey, J. R. & Williams, E. L. (1970) The structure of some xylans from red algae. *Phytochemistry* 9:2383-2388.

Ursin, U. M., Yamaguchi, J. & McCormick, S. (1989) Gametophytic and sporophytic expression of anther-specific genes in developing tomato anthers. *Plant Cell* 1:727-736.

Vanderslice, P., Craik, C. S., Nadel, J. A. & Caughey, G. H. (1989) Molecular cloning of dog mast cell tryptase and a related protein: structural evidence of a unique mode of serine protease activation. *Biochemistry* 28:4148-4155.

Vick, B. A. & Zimmerman, D. C. (1987) Oxidative systems for modification of fatty acids: the lipoxygenase pathway. In *Lipids: structure and function* (The Biochemistry of Plants Vol. 9), Ed. P. K. Stumpf, pp. 53-91 Academic Press, Inc. Orlando.

Viel, A., Le-Maire, M., Philippe, H., Morales, J. & Mazabraud, A. (1991) Structural and functional properties of thesaurin a (42Sp50), the major protein of the 42 S particles present in *Xenopus laevis* previtellogenic oocytes. *J. Biol. Chem.* 266:10392-10399.

von Heijne, G. (1984) How signal sequences maintain cleavage specificity. *J. Mol. Biol.* 173:243-251.

von Heijne, G. (1985) Signal sequences. The limits of variation. *J. Mol. Biol.* 184:99-105.

Walldorf, U., Hovemann, B. & Bautz, E. K. F. (1985) F1 and F2: Two similar

genes regulated differentially during development of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 82:5795-5799.

Wilkins, T. A., Bednarek, S. Y. & Raikhel, N. V. (1990) Role of propeptide glycan in post-translational processing and transport of barley lectin to vacuoles in transgenic tobacco. *Plant Cell* 2:301-313.

Wilson, P. G. & Culbertson, M. R. (1988) SUF12 suppressor protein of yeast (A fusion protein related to the EF-1α family of elongation factors). *J. Mol. Biol.* 199:559-573.

Woolley, P. & Clark, B. F. C. (1989) Homologies in the structure of G-binding proteins: an analysis based on elongation factor EF-Tu. *Biotechnology* 7:913-920.

Wright, C. S. & Raikhek, N. (1989) Sequence variability in three wheat germ agglutinin isolectins: products of multiple genes in polyploid wheat. *J. Mol. Evol.* 28:327-336.

Yang, F., Demma, M., Warren, V., Dharmawardhane, S. & Condeelis, J. (1990) Identification of an actin-binding protein from *Dictyostelium* as elongation factor 1α. *Nature* 347:494-496.

Yang, W., Burkhart, W., Cavallius, J., Merrick, W. C. & Boss, W. F. (1993) Purification and characterization of a phosphatidylinositol 4-kinase activator in carrot cells. *J. Biol. Chem.* 268:392-398.

Yokota, T., Sugisaki, H., Takanami, M. & Kaziro, Y. (1980) The nucleotide sequence of the cloned *tufA* gene of *Escherichia coli*. *Gene* 12:25-31.

Young, R. A. & Davis, R. W. (1983) Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci.* USA 8:1194-1198.

Yun, Y. & Davis, R. L. (1989) Levels of RNA from a family of putative serine protease genes are reduced in *Drosophila melanogaster* dunce mutants and are regulated by cyclic AMP. *Mol. Cell. Biol.* 9:692-700.

Zimmerman, D. C. & Vick, B. A. (1973) Lipoxygenase in *Chlorella pyrenoidosa*. *Lipids* 8:264-266.