



I. Identification and phylogenetic analysis of plant TERT genes, II. Identification of Ribulosebisphosphate carboxylase/oxygenase activase in wheat  
by Ruschelle Ann Love

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Plant Sciences

Montana State University

© Copyright by Ruschelle Ann Love (2000)

Abstract:

Telomerase is a ribonucleoprotein complex that adds telomeres to the ends of eukaryotic chromosomes. Two subunits are associated with telomerase, the catalytic subunit, commonly referred to as TERT, and the RNA subunit. Telomerase has been extensively studied in mammalian and yeast systems, but little is known about the roles of telomeres and telomerase in plants or their effect on plant growth and development. This thesis reports the identification of two plant TERT genes, *Arabidopsis thaliana* and *Oryza sativa* (rice). From the multiple sequence alignments it was concluded that, as with previously identified TERTs, plant TERTs contain the conserved reverse transcriptase motifs 1, 2 and A-E as well as the TERT specific motif T. In addition, the alignment of all known TERTs showed seven additional conserved motifs, five upstream of motif T (TEL 1-5) and two downstream of the RT motif E (F and G). Phylogenetic analysis revealed that the TERT catalytic subunit in plants is conserved throughout evolution, and, that plant TERT proteins resemble those of higher eukaryotes (human, mouse and hamster) more closely than to those of lower eukaryotes (yeast or ciliates).

Ribulose bis-phosphate carboxylase activase (Rubisco activase) is an enzyme that catalyzes the activation of Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). Rubisco activase has been identified in a number of plant species, including barley, rice, maize, *Arabidopsis*, tobacco, tomato and spinach, but only within the monocot barley have two distinct Rubisco activase genes been identified. All other plants have one gene associated with Rubisco activase. In this thesis, we set out to identify the Rubisco activase gene(s) in an additional monocot, wheat. Screening of a cDNA library resulted in identification of two wheat Rubisco activase genes, WRcaA and WRcaB. The 432 amino acid WRcaB polypeptide was found to be 96% identical to the barley BRcaB polypeptide and the 363 amino acid WRcaA polypeptide was found to be 95% identical to the barley BRcaA polypeptide. Multiple sequence alignment also showed WRcaA and WRcaB contain the same structural features found through studies on other plant species.

I. IDENTIFICATION AND PHYLOGENETIC ANALYSIS OF PLANT TERT GENES

II. IDENTIFICATION OF RIBULOSEBISPHOSPHATE  
CARBOXYLASE/OXYGENASE ACTIVASE IN WHEAT

by

Ruschelle Ann Love

A thesis submitted in partial fulfillment  
of the requirements for the degree

of

Master of Science

in

Plant Sciences

MONTANA STATE UNIVERSITY  
Bozeman, Montana

July, 2000

N/378  
29425

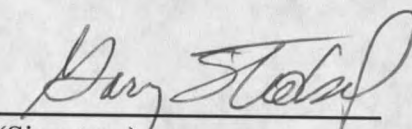
APPROVAL

of a thesis submitted by

Ruschelle Ann Love

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style and consistency, and is ready for submission to the College of Graduate Studies.

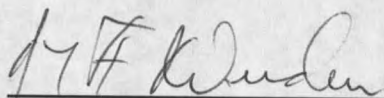
Dr. Gary A. Strobel

  
(Signature)

7/5/00  
Date

Approved for the Department of Plant Sciences

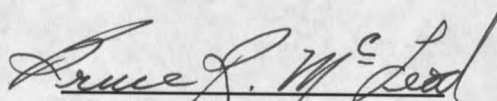
Dr. Norm F. Weeden

  
(Signature)

6/30/00  
Date

Approved for the College of Graduate Studies

Dr. Bruce M. McLeod

  
(Signature)

7-7-00  
Date

## STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the library shall make it available to borrowers under rules of the library.

If I have indicated my intention to copyright this thesis by including a copyright notice, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Signature Russell A. Love

Date July 6, 2000



## TABLE OF CONTENTS

|                                                                                                   | Page |
|---------------------------------------------------------------------------------------------------|------|
| IDENTIFICATION AND PHYLOGENETIC ANALYSIS<br>OF PLANT TELOMERASE GENES.....                        | 1    |
| Introduction.....                                                                                 | 1    |
| Materials and Methods.....                                                                        | 10   |
| Database Identification of Putative <i>Arabidopsis thaliana</i> TERT.....                         | 10   |
| Collection of <i>Arabidopsis thaliana</i> DNA.....                                                | 10   |
| Identification of <i>Arabidopsis thaliana</i> TERT ( <i>AtTERT</i> ).....                         | 10   |
| Sequencing of Clones.....                                                                         | 13   |
| Database Identification of Putative <i>Oryza sativa</i> (Rice) TERT.....                          | 13   |
| Collection of <i>Oryza sativa</i> (Rice) DNA.....                                                 | 14   |
| Confirmation of <i>Oryza sativa</i> (Rice) TERT Fragment Sequence.....                            | 15   |
| Intron Prediction of the Rice BAC sequence.....                                                   | 16   |
| cDNA Library Screening for <i>Oryza sativa</i> (Rice) TERT.....                                   | 16   |
| Phylogenetic Analysis and Sequence Alignments<br>of Known TERTs.....                              | 17   |
| Phylogenetic Analysis of the identified TERTs including<br><i>Oryza sativa</i> putative TERT..... | 18   |
| Results and Discussion.....                                                                       | 19   |
| Identification of <i>Arabidopsis thaliana</i> TERT Gene.....                                      | 19   |
| Structural Characterization of <i>AtTERT</i> .....                                                | 23   |
| Identification of <i>Oryza sativa</i> (Rice) Putative Gene.....                                   | 31   |
| Structural Characterization of the Putative <i>Oryza sativa</i><br>TERT Gene.....                 | 33   |
| IDENTIFICATION OF RIBULOSEBISPHOSPHATE CARBOXYLASE/<br>OXYGENASE IN WHEAT.....                    | 37   |
| Introduction.....                                                                                 | 37   |
| Photosynthesis.....                                                                               | 37   |
| Structure of Rubisco.....                                                                         | 40   |
| Rubisco and the Calvin Cycle.....                                                                 | 41   |

TABLE OF CONTENTS - Continued

|                                                                                                     | Page |
|-----------------------------------------------------------------------------------------------------|------|
| Rubisco and Photorespiration.....                                                                   | 41   |
| Plant Strategies for Reducing Photorespiration.....                                                 | 43   |
| Regulation of Rubisco.....                                                                          | 46   |
| The Role of Rubisco Activase.....                                                                   | 47   |
| Structural Features of Rubisco Activase.....                                                        | 49   |
| Materials and Methods.....                                                                          | 53   |
| Cloning of Three Partial Wheat Rubisco Activase "B" Genes<br>from Wheat Genomic DNA.....            | 53   |
| Sequencing of Clones.....                                                                           | 55   |
| Cloning of the Complete Rubisco Activase "B" Gene<br>(WRcaB) from a Wheat cDNA Library.....         | 55   |
| <i>In vivo</i> Excision of cDNA from Phage and Sequencing.....                                      | 58   |
| Detection of the Complete Wheat Rubisco Activase "A" Gene<br>(WRcaA) from a Wheat cDNA Library..... | 59   |
| Multiple Sequence Alignment of Three Partial WRcaB Genes.....                                       | 60   |
| Sequence Analysis of WRcaA and WRcaB.....                                                           | 60   |
| Results and Discussion.....                                                                         | 62   |
| Identification of Three Partial Rubisco Activase "B"<br>Genes from Wheat.....                       | 62   |
| Identification of the WRcaB Gene in Wheat.....                                                      | 62   |
| Identification of the WRcaA Gene in Wheat.....                                                      | 67   |
| Multiple Sequence Alignment and Sequence Analysis<br>of Rubisco Activase Genes.....                 | 69   |
| REFERENCES CITED.....                                                                               | 73   |
| APPENDIX A.....                                                                                     | 79   |

## LIST OF TABLES

|                                                                                                | Page |
|------------------------------------------------------------------------------------------------|------|
| Table                                                                                          |      |
| 1-1. Amino Acid Sequence Homology for 466 Positions<br>among the Nine Known TERT Proteins..... | 29   |

## LIST OF FIGURES

| Figure                                                                                                                          | Page |
|---------------------------------------------------------------------------------------------------------------------------------|------|
| 1-1. End Replication Problem.....                                                                                               | 5    |
| 1-2. Telomerase Elongation Model.....                                                                                           | 7    |
| 1-3. PCR Amplification of <i>Arabidopsis thaliana</i> Genomic DNA<br>using Exact and Degenerate Primers.....                    | 20   |
| 1-4. Genomic <i>Arabidopsis thaliana</i> TERT Fragment.....                                                                     | 22   |
| 1-5. Multiple Sequence Alignment of Motifs of Nine<br>Full-length Telomerase RTs.....                                           | 24   |
| 1-6. 15 Motifs Identified in TERT Genes.....                                                                                    | 26   |
| 1-7. Phylogenetic Tree of all Known TERTs.....                                                                                  | 30   |
| 1-8. Genomic <i>Oryza sativa</i> (Rice) TERT Fragment.....                                                                      | 32   |
| 1-9. Multiple Sequence Alignment of Motifs of Ten TERT Fragments<br>(270 nucleotides) including <i>Oryza sativa</i> (Rice)..... | 33   |
| 1-10. Phylogenetic Tree of Known TERT Fragments and<br><i>Oryza sativa</i> Fragment.....                                        | 35   |

LIST OF FIGURES - Continued

| Figure                                                                                 | Page |
|----------------------------------------------------------------------------------------|------|
| 2-1. Photosynthetic (Calvin Cycle) and Photorespiration Cycles.....                    | 39   |
| 2-2. C <sub>3</sub> and C <sub>4</sub> Leaf Anatomy.....                               | 44   |
| 2-3. C <sub>4</sub> Metabolic Pathway for Concentrating CO <sub>2</sub> .....          | 45   |
| 2-4. Three Different WRcaB Gene Fragments.....                                         | 63   |
| 2-5. Amino Acid and Nucleotide Sequence of Rubisco Activase<br>from Wheat (WRcaB)..... | 65   |
| 2-6. Amino Acid and Nucleotide Sequence of Rubisco Activase<br>from Wheat (WRcaA)..... | 69   |
| 2-7. Multiple Sequence Alignment of Known Rca Genes.....                               | 71   |

## ABSTRACT

Telomerase is a ribonucleoprotein complex that adds telomeres to the ends of eukaryotic chromosomes. Two subunits are associated with telomerase, the catalytic subunit, commonly referred to as TERT, and the RNA subunit. Telomerase has been extensively studied in mammalian and yeast systems, but little is known about the roles of telomeres and telomerase in plants or their effect on plant growth and development. This thesis reports the identification of two plant TERT genes, *Arabidopsis thaliana* and *Oryza sativa* (rice). From the multiple sequence alignments it was concluded that, as with previously identified TERTs, plant TERTs contain the conserved reverse transcriptase motifs 1, 2 and A-E as well as the TERT specific motif T. In addition, the alignment of all known TERTs showed seven additional conserved motifs, five upstream of motif T (TEL1-5) and two downstream of the RT motif E (F and G). Phylogenetic analysis revealed that the TERT catalytic subunit in plants is conserved throughout evolution, and, that plant TERT proteins resemble those of higher eukaryotes (human, mouse and hamster) more closely than to those of lower eukaryotes (yeast or ciliates).

Ribulose bis-phosphate carboxylase activase (Rubisco activase) is an enzyme that catalyzes the activation of Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). Rubisco activase has been identified in a number of plant species, including barley, rice, maize, *Arabidopsis*, tobacco, tomato and spinach, but only within the monocot barley have two distinct Rubisco activase genes been identified. All other plants have one gene associated with Rubisco activase. In this thesis, we set out to identify the Rubisco activase gene(s) in an additional monocot, wheat. Screening of a cDNA library resulted in identification of two wheat Rubisco activase genes, WRcaA and WRcaB. The 432 amino acid WRcaB polypeptide was found to be 96% identical to the barley BRcaB polypeptide and the 363 amino acid WRcaA polypeptide was found to be 95% identical to the barley BRcaA polypeptide. Multiple sequence alignment also showed WRcaA and WRcaB contain the same structural features found through studies on other plant species.

## IDENTIFICATION AND PHYLOGENETIC ANALYSIS OF PLANT TELOMERASE GENES

### Introduction

Telomeres, or the specialized structures found at the ends of eukaryotic chromosomes, were first characterized by Muller in 1927 (Blackburn, 1995). In *Drosophila*, Muller produced genetic mutations via X-ray irradiation and closely studied the breakage and rejoining of chromosomes. Analysis of the mutations led to the finding that breakage and rejoining (chromosomal rearrangements) never occurred at the terminal end of chromosomes, thus leading to the concept of a "cap" at the end of the chromosome. These discoveries were supported by maize chromosome breakage studies done by Barbara McClintock in 1941 (Blackburn, 1995). McClintock used a breakage-fusion-bridge cycle to study newly formed chromosomes ends. The breakage-fusion-bridge cycle consisted of the movement of two centromeres of a dicentric chromosome to opposite poles of the mitotic spindle, thus forming a bridge between poles. The breaking of the bridge, during anaphase or telophase, produced two monocentric chromosomes with newly formed ends. Throughout these studies, McClintock was able to draw several conclusions concerning broken chromosomes, including the idea that newly formed chromosome ends were "sticky" and would fuse with other newly formed chromosome ends (normal, unbroken, chromosomes were stable and would not fuse to other ends) (Blackburn, 1995).

These initial studies, on chromosome rearrangement, were followed by cytological

studies of chromosome ends, showing many cases in which the ends of chromosomes behaved in special ways, and molecular studies of telomeres (Blackburn, 1995). The molecular analysis of telomeres began with the discovery that the most distal portion of the *Tetrahymena thermophila* chromosomes consisted of tandem arrays of repeated sequence (Blackburn, 1995). Since then, progress has been made in the characterization of telomere structure and function. It was soon discovered, by looking at various eukaryotic organisms, that telomeres, regardless of species, were structurally and functionally similar. All telomeres consist of repeated sequences with a G-rich strand oriented in the 5' to 3' direction towards the chromosome terminus (Blackburn, 1995). Each terminus contains a 3' overhang of two repeat units (Regad, 1994; Richards, 1991; Zentgraf, 1995) which enables telomerase, an RNA-dependent DNA polymerase, to synthesize additional G-rich repeats (Blackburn, 1995). The telomeric DNA usually consists of tandem repeats of 5-8 bp sequence elements. Some common sequence motifs include: TTAGGG (vertebrates); TTTAGGG (angiosperms); TTAGGG (fungi); and  $T_{2-4}GGGG$  (ciliates) (Blackburn, 1995). Other organisms have telomeric repeats that are much longer and more highly variable. *Saccharomyces cerevisiae*, for example, has an imperfect telomeric repeat,  $T(G)_{2-3}(TG)_{1-6}$  (Zakian, 1996), and *Candida albicans* has a unique 23 bp repeat (Blackburn, 1995). The length of the telomeric tandem arrays for the above species has been shown to vary, ranging from 5-20kb in humans, 100-150kb mice, 200-250bp in *Arabidopsis*, to very short stretches of 18-20bp in *Oxytricha* and *Euplotes* (Zentgraf, 1994 and Richards, 1988).

An additional interesting feature of telomere structure is the presence of bound proteins associated with them. Proteins associated with telomeres have been isolated and



characterized for different systems and are associated with different parts of the telomeric tract, such as the single-stranded 3' overhang or the double-stranded telomeric repeat tracts (Zentgraf, 1995, Shippen, 1998 and Zentgraf, 2000). The first group of telomere-associated proteins interact with the telomeric single-stranded 3' overhang in ciliates. They are assumed to function as molecular chaperones, capping the G-quartets formed by the single-stranded 3' overhang (the G-rich overhanging strands of telomeres that dimerize to form stable complexes in solution via the formation of G-quartet, or cyclic, tetramers). The second group of telomere-binding proteins, double-stranded telomere-binding proteins, has been characterized in human and yeast and are classified as a family of Myb-related (protooncogene) telomeric DNA-binding proteins. Rap1p, one example of a double-stranded telomere-binding protein, is involved in telomeric position effects as well as in telomere length regulation. The telomeric position effect refers to the ability of telomeres to repress the transcription of genes in their vicinity (Stavenhagen, 1998). Plant telomere-binding proteins have also recently been discovered (Zentgraf, 1995). In contrast to human/yeast, they appear to form complexes with both double-stranded and single-stranded telomeric DNA effectively. Thus, the plant telomere-associated proteins appear to fit simultaneously into both categories described above, suggesting that they may have multiple functions.

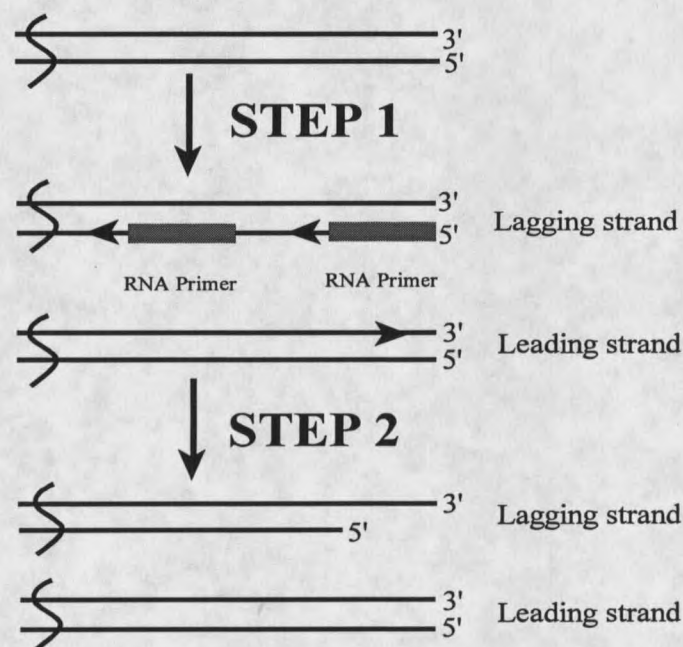
The functional roles of telomeres have been well established. It is known that telomeres perform at least three basic functions. First, telomeres may maintain genome integrity by forming protective caps at the ends of chromosomes that prevent chromosome end to end fusion, recombination, or exonucleolytic degradation (McEachern, 1996 and Gall, 1995). Second, they may serve a function in nuclear organization, as they seem to be

associated with the nuclear matrix (Zentgraf, 1995). Third, telomeres are needed for the complete replication of chromosomal termini as a solution to the "end replication problem" (discussed below). Because RNA primers are required to initiate DNA strand replication, conventional DNA polymerases cannot completely replicate eukaryotic chromosomes containing 3' terminal extensions (Lingner, 1997). Thus, the enzyme telomerase, a ribonucleoprotein, is needed for replication of chromosomal termini.

Telomerase, first purified from *Tetrahymena* (Lingner, 1997), is an RNA-dependent DNA polymerase containing an RNA and a protein subunit. The RNA subunit provides the template for addition of short sequence repeats to the 3' end of the telomere and was first identified in *Tetrahymena* by Greider and Blackburn (Greider, 1989). To date, a variety of telomerase RNA subunits have been identified from different organisms, including ciliates, yeasts and mammals (reviewed in Nugent, 1998). The telomerase catalytic protein component, also first identified and sequenced in *Tetrahymena*, has now been identified in ciliated protozoans, mammals, yeasts and the plant-*Arabidopsis thaliana*. The protein component serves as the catalyst for telomere repeat synthesis and contains sequence similarity to reverse transcriptase, including the presence of seven reverse transcriptase (RT) motifs: 1,2, and A-E (Bryan, 1999). In addition to containing the RT sequence motifs, a telomerase-specific motif, T, is also present in all telomerases identified to date, but its function has not yet been elucidated. Identification of the eight (T, 1, 2, and A-E) telomerase reverse transcriptase motifs led to the naming of the catalytic subunit as "TERT" (Telomerase Reverse Transcriptase) by Nakamura et al., 1997.

Genes in the RT family encode a wide variety of elements including telomerases,

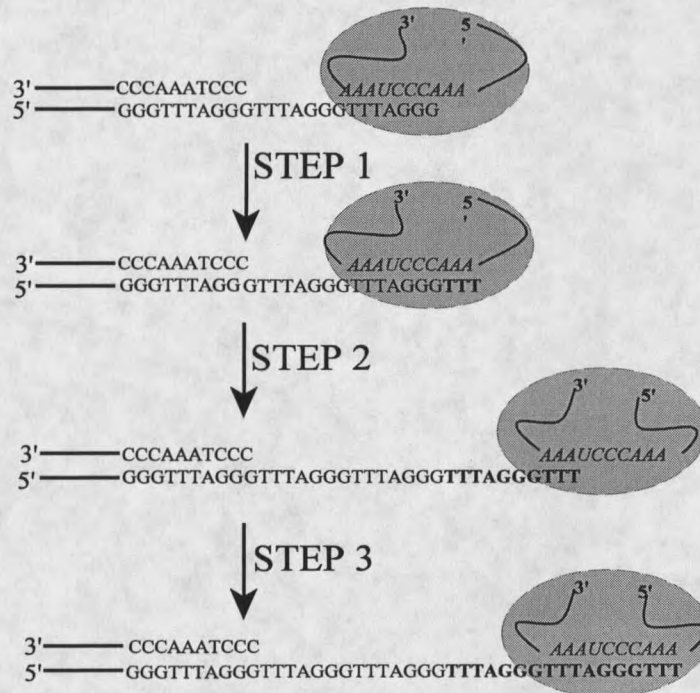
retroviruses, long terminal repeat (LTR) retrotransposons and group II introns to name a few (Nakamura, 1998). And although they are a diverse group, they all contain RT motifs. The seven canonical RT sequence elements have been found in the catalytic subunit of telomerase, which is not surprising, as telomerase polymerizes DNA using an RNA template, itself the definition of an RT (Nakamura, 1998). Telomerase has been found in evolutionarily diverse organisms, and in all cases contain an RNA template and a conserved TERT subunit, suggesting that telomere maintenance by telomerase is an ancient mechanism.



**Figure 1-1. End Replication Problem.** STEP 1: Denaturation of the template occurs first. The leading strand synthesis results in complete replication of the 5' template sequence. Both lagging and leading strand synthesis are initiated by RNA primers. STEP 2: RNA primers are removed from the lagging strand, resulting in an incompletely replicated end. A 3' overhang at one end of each "new" DNA strand (chromosome) is thus created. Only one end of the chromosome is illustrated. The left-hand curved line indicates the rest of the chromosome.

Chromosome replication poses a problem for eukaryotic cells. Replication by DNA polymerase does not work for replication of chromosome ends, as the 3'-end of chromosomes are not completely copied via this mechanism. Thus, each successive round of conventional DNA replication results in a loss of telomeric sequence as DNA polymerase fails to complete synthesis of the lagging strand (Figure 1-1). To prevent continuous loss of sequence, telomerase has evolved to add sequence onto 3' chromosomal termini. The model of telomere repeat addition by telomerase (Figure 1-2) was proposed by Greider and Blackburn (Greider, 1989) and has since been supported by additional studies. Telomerase binds to the 3' overhang of the chromosome end, with the RNA template base pairing with the most terminal telomeric repeat. The telomere is then elongated by the addition of nucleotides to the 3' terminus. Once the end of the templating domain is reached, a translocation step is necessary to reposition the DNA back to the beginning of the RNA template for another round of nucleotide synthesis (Blackburn, 1995 and Shippen, 1998).

Thus, telomere maintenance is dependent on telomerase and therefore telomerase often determines the future of the cell. In healthy cells, telomerase activity appears to be tightly regulated and maintains proper telomere length. However, in cancer cells or aging cells, telomeres are abnormally long or short. In human somatic cells, telomeres are found to shorten with each successive cell division. Furthermore, in both mammalian and fungal systems (Nakamura, 1997 and Ligner, 1997), severe telomere degradation is implicated in cellular senescence and aging. On the other hand, human immortalized cell lines and cancer cell lines are often found to have higher levels of telomerase, suggesting that unchecked



**Figure 1-2. Telomerase Elongation Model.** STEP 1: The TTTAGGG repeat of the telomere is base paired with the RNA template sequence of telomerase. Elongation occurs as the RNA is copied to the end of the template region. STEP 2: Translocation repositions the telomere sequence and exposes additional template sequence. STEP 3: Another round of template copying produces additional TTTAGGG repeats. New bases are seen in bold and letters in italics indicate RNA template.

addition of telomeres by telomerase may result in immortality of cell lines (Zakian, 1997).

Despite increasing interest in telomeres and telomerase over the past decade, studies of plant telomeres and telomerase have been limited. This is somewhat surprising as the telomere concept was developed in maize (Blackburn, 1995) and the first higher eukaryotic telomere was cloned from *Arabidopsis thaliana* (Richards, 1988).

Current plant telomere and telomerase research includes focus on the developmental

control of telomere length and of telomerase expression in plants. Plant telomeres, as stated above, were first identified in the plant *Arabidopsis thaliana* by Richards and Ausubel (Richards, 1988). The plant telomere sequence appears to be TTTAGGG for many identified plant species, including *Arabidopsis*, barley, tomato, maize, wheat, and *Melandrium album* (white campion) (Shippen, 1998, Riha, 1998). Interestingly, plants of the onion family apparently lack the TTTAGGG telomeric repeats, as neither *in situ* hybridization nor Southern blotting have been successful in detection of this telomeric sequence (Shippen, 1998). Although the telomere repeat itself is conserved within most plants, telomere length varies to a large degree between species as well as between different tissues of the same species. For example, *Arabidopsis* telomeres average 2-4 kb (Richards, 1988) whereas tobacco telomeres extend up to 130 kb (Shippen, 1998). It has also been determined that telomere length varies between tissues of the same plant. For example, plant telomeres are shortened in embryos and inflorescences during differentiation in barley. Moreover, the telomere shortening is believed to be due to an absence of telomerase, as seen in the similar process in humans (Kilian, 1995). The telomeres in barley are especially long during undifferentiated growth in suspension cultures, likely the result of higher levels of telomerase activity (Kilian, 1995).

Telomerase expression is also believed to be developmentally regulated. Telomerase expression was first reported in tobacco using a biochemical telomerase (oligo extension) assay method of detection (Fajkus, 1996). More recently, the PCR-based Telomere Repeat Amplification Protocol (TRAP) assay has been successfully used to detect telomerase activity in a variety of plants including: barley, *Arabidopsis thaliana*, cauliflower, tobacco, soybean

and *M. album* (McKnight, 1997; Fajkus, 1998; Shippen, 1998; Riha, 1998). Utilizing such methods, expression of plant telomerase activity in various tissues has been explored (Killian, 1998, McKnight, 1998, Riha, 1998, and Heller, 1996). Telomerase activity was easily detected in organs containing rapidly proliferating meristems, such as germinating seedlings, root tips, embryos, anthers, carpels, flowers and the floral buds. Telomerase activity, however, was either low or absent in such tissues as the shoot apex, leaf, or stem. The above observations suggest that telomere length in plants, as in human and yeasts, is maintained by a telomerase-mediated mechanism.

Despite recent progress in investigating plant telomeres and telomerase, much work must still be done before a solid understanding of the role of telomerase and its controls are achieved. Identification and study of both the TERT and RNA components of plant telomerase will be an essential step towards this goal. In this thesis, with the goal of identifying plant TERT genes, we utilized the higher plant preliminary sequence data made available by the Stanford Sequencing Center (Stanford, California) to identify putative plant *TERT* gene fragments. Using sequence information from the putative *Oryza sativa* (rice) and *Arabidopsis thaliana* *TERT* gene fragments, primers were made with the goal of identifying full length *TERT* genes by either PCR or by screening a cDNA library. The deduced polypeptide sequence from resulting genes were subject to phylogenetic analysis, against known TERT sequences, in an attempt to reinforce the findings that the TERT catalytic subunit is a conserved component of telomerase. Surprisingly, it was found that *Arabidopsis* TERT and *O. sativa* TERT more closely resemble the TERT proteins of higher eukaryotes than TERTs from lower eukaryotes.

## Materials and Methods

### Database Identification of Putative *Arabidopsis thaliana* TERT

A putative *Arabidopsis thaliana* TERT gene was found in the GenBank *Arabidopsis* (incomplete genome) database via the “TBLASTN” algorithm using a 76 amino acid fragment of the known human TERT (hTERT) amino acid sequence (motifs B, C and D) as query. Parameters were set at default and the hTERT query pulled up multiple hits including a 625 bp *Arabidopsis thaliana* genomic DNA fragment, accession number B27802. The deduced polypeptide sequence of the unidentified *Arabidopsis* gene was 41% identical to hTERT with 65% similarity. This protein had regions which showed substantial similarity to reverse transcriptase RT motifs C, D and E of hTERT, and was therefore considered a putative TERT.

### Collection of *Arabidopsis thaliana* DNA

RNAse treated *Arabidopsis thaliana* DNA (Colombia ecotype) was obtained from B. Sharrock, Montana State University. The DNA was concentrated by ethanol precipitation and then resuspended in TE buffer (10mM Tris pH8, 1mM EDTA). The concentration was quantitated via the Beckman Du-50 spectrophotometer at  $A_{260}$ , and DNA was inspected by running a 1% agarose gel to ensure it was not degraded.

### Identification of *Arabidopsis thaliana* TERT (*At*TERT)

Exact and degenerate primers were constructed using sequence information from the putative *Arabidopsis* TERT fragment found in GenBank and from *Saccharomyces cerevisiae*,



*Schizosaccharomyces pombe*, *Tetrahymena thermophila*, and *Oxytricha trifallax* TERT sequences. The primers used for identification of TERT from genomic *Arabidopsis* DNA were: Arab C1 (rev)-exact, 5'-AGA CAC AAA ATG TAG TCA TC-3'; Arab C2 (rev)-exact, 5'-GTC ATCA ATA AAT CTC AGT AA-3'; Arab C1 (rev)-deg., 5'-GAY GAY TAY ATH TTY GTN TCN-3'; Arab C2 (rev)-deg., 5'-TYG NYG NSA NTT MTT VAT SAY-3'; B (for)-yeast combo, 5'-MRA RAA GWT GGT MTY YYT CAR GG-3'; B (rev)-yeast combo, 5'-NSW NCC YTG NGG DAT NCC-3'; and T (for)-yeast combo, 5'-RWC STT TYT TYT AYD KCA CBG A-3'. IUB group codes, seen above, to identify degenerate sequences are as follows: R= A+G, Y= C+T, M= A+C, K= G+T, S= G+C, W= A+T, H= A+T+C, B= G+T+C, D= G+A+T, V= G+A+C, and N= A+G+C+T.

The first round of PCR was performed in a Perkin Elmer DNA thermal cycler 480 using 50µl reactions containing the following: 2µg *Arabidopsis* genomic DNA, 1 unit of *tag* polymerase (Promega), 0.1 mM dNTPs, 250 µM of each primer and buffer C (provided by the manufacturer) with [Mg<sup>++</sup>] at 1.5mM. A 5 minute denaturation step at 94° was followed by 30 cycles of 94°C for 1 minute, 50°C for 2 minutes, and 72°C for 1 1/2 minutes. The appropriate number of cycles was determined experimentally by quantitating the amount of PCR product in a cycle-dependent manner, comparing product yield after 21, 23, 25, 27, 29, 31, 33, and 35 cycles. Primer sets used, in separate attempts to identify the TERT gene, for the first round of PCR were: Arab C1 (rev)-exact and T (for)-yeast combo, Arab C1 (rev)-exact alone, T (for)-yeast combo alone, Arab C1 (rev)-deg. and T (for)-yeast combo, and Arab C1 (rev)-deg. The approximately 1600 bp band, produced by Arab C1 (rev)-exact primer alone, was gel purified. The band was cut from the gel, purified by the pulp-spin

method (spin DNA through eppendorf tube containing paper pulp in TE buffer for 10 minutes at 14 K and collect supernatant), ethanol precipitated and resuspended in 20  $\mu$ l of water. The gel purified DNA (1  $\mu$ l) was electrophoresed on a 1% agarose gel for quantitation and to determine purity, and was then used as template in the second round of PCR.

The second round of PCR, nested PCR, used 0.5ng of the ~1600 bp Arab C1 (rev) alone fragment as template. All other PCR conditions were kept the same as first round PCR. Nested primers sets used in the nested PCR reactions were as follows: Arab C1 (rev)-exact and T (for)-yeast combo, Arab C1 (rev)-exact alone, T (for)-yeast combo alone, Arab C2 (rev)-exact and T (for)-yeast combo, Arab C2 (rev)-exact alone, Arab C2 (rev)-deg. and T (for)-yeast combo, Arab C2 (rev)-deg. alone, Arab C2 (rev)-exact and B (for)- yeast combo, B (for)-yeast combo alone, B (rev)-yeast combo and T (for)-yeast combo, and B (rev)-yeast combo alone. From these primer sets, various PCR product sizes were separated on a 1.5% agarose gel stained with 0.5  $\mu$ g/ml ethidium bromide and visualized under UV<sub>254</sub>. Resulting bands of the correct size, estimated from the size of other telomerases, were seen for Arab C2 (rev)-exact and T (for)-yeast combo (1002 bp), Arab C2 (rev)-exact and B (for)-yeast combo (153 bp), and B (rev)-yeast combo and T (for)-yeast combo (849 bp) (Figure 1-3). The bands were cloned into the pGem-T cloning vector using the pGem-T Vector System II (Promega), as per the manufacturers directions, and sequenced. Each PCR product (2  $\mu$ l) was ligated into the pGem-T vector and transformed into *E. coli* JM109 competent cells (Promega). The clones were screened for plasmids with inserts of the desired size and the plasmids containing correct sized inserts were used directly in sequencing reactions.

### Sequencing of Clones

Sequencing reactions were set up for the ABI PRISM 377 as described by the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer). For each 20 $\mu$ l reaction, 0.2 $\mu$ g of DNA, 3.2pmole primer, and 4 $\mu$ l of BIGDYE mix (A-Dye Terminator, C-Dye Terminator, G-Dye Terminator, T-Dye Terminator, dITP, dATP, dCTP, dTTP, Tris-HCl (pH 9.0), MgCl<sub>2</sub> thermal stable pyrophosphatase, and AmpliTaq DNA Polymerase) were mixed and placed in the GeneAmp PCR system 9600. Thermal cycling was as follows: 96°C for 15 seconds, 50°C for 10 seconds, and 60°C for 4 minutes. This cycle was repeated 25 times and PCR products were precipitated with 75% isopropanol and pellets were washed with 70% ethanol. The resulting pellet was dried and sequenced by the Montana State University center for sequencing.

The clones were sequenced using universal primers SP6 (5'-ATT TAG GTG ACA CTA TAG-3') and T7 (5'-TAA TAC GAC TCA CTA TAG GG-3'). Internal primers were constructed to verify the sequence in both directions. The primers were exact sequences taken from SP6 and T7 sequencing results and included: ArabNB1 (rev) 5'-TAT CCA AAC CAT TAG ATC-3', ArabNC1 (rev) 5'-CAA AGT CCT CTC GAG ATG-3', ArabNT1 (for) 5'-TGG GAA AGA TTA ATA AGC-3', and ArabNT2 (for) 5'-GAA CCA GAT GTT CTT GG-3'.

### Database Identification of Putative *Oryza sativa* (Rice) TERT

A segment of DNA containing a potential *Oryza sativa* TERT was identified by searching the *Arabidopsis thaliana* database (Stanford University Home Page) via the

“TBLASTN” algorithm. The search utilized a 28 amino acid segment of the *Arabidopsis thaliana* TERT protein sequence in the region identified as the “C motif” as query against the higher plant sequence database with the expect parameter set at 100. The second match, with a match score of 74, was accession number AQ510589 from the *Oryza sativa* sequencing project at Clemson University. AQ510589 was a 531 base pair genomic fragment. Sequence alignment of this deduced polypeptide sequence to the TERT protein sequence of *Arabidopsis thaliana* identified multiple regions of sequence similarity, including the canonical reverse transcriptase motifs C, D, and E.

#### Collection of *Oryza sativa* (Rice) DNA

Clemson University provided the *Oryza sativa* Rice BAC clone, in the pBeloBACII vector, (catalog # nbxb0095N19f) as a stab culture in *E. Coli* strain DH10B. One liter of culture was grown in LB (Lennox L Broth Base) at 37°C with overnight shaking. Cells were harvested and DNA extracted via the Wizard Plus Maxipreps DNA purification system (Promega) using the vacuum filtration protocol, according to the manufacturers directions. Recovered DNA was resuspended in 0.75ml TE buffer and the final yield was 20µg of DNA. DNA yield was determined by running a 1% agarose gel with 0.5µg/ml ethidium bromide and visualizing DNA concentration by ultraviolet light. The fragment was sequenced using universal SP6 and T7 primers to ensure the correct sequence. Cultures were prepared for long term storage by: 1) directly storing recovered DNA from the above purification at -70°C and 2) preparing a glycerol stab containing culture in LB and 8% glycerol and storing at -70°C.

*Oryza sativa* (California ecotype) genomic DNA,  $1\mu\text{g}/\mu\text{l}$ , was obtained from M. Giroux, Montana State University. Concentration and quality of the DNA was checked on a 1% agarose gel.

#### Confirmation of *Oryza sativa* (Rice) TERT Fragment Sequence

The BAC genomic fragment, nbxb0095N19f (GenBank accession AQ510589), containing the Rice TERT was amplified by PCR and sequenced to verify the sequence obtained from the database. The purified DNA, from Wizard Plus Maxipreps DNA purification system (described above), was used as template in a PCR reaction utilizing two end primers, made from the database sequence. One internal primer, to amplify just the 270 bp TERT fragment, was also made. Primers used for PCR were: Rice EP-2 (for) 5'-CCT GAA TAT TTG TTA ATG AGG-3, Rice EP-3 (rev) 5'-GGA AAC ACA TCG TCC AAG TCA AT-3', and Rice ER-1 (rev) 5'-GTC ATA CCT CGT ATA ATC AGC-3'. Internal primer Rice ER-1 (rev) and primer EP-2 (for) amplified the 270 bp TERT fragment whereas the other 261 bp of the 531 bp fragment did not align with known TERT sequences. The PCR reaction, to obtain the fragment of interest from the BAC DNA, was done using  $2\mu\text{g}$  BAC DNA, 1 unit of *Tfl* polymerase, 0.1mM dNTPs,  $250\mu\text{M}$  primers, and a 1.5mM concentration of  $\text{MgSO}_4$  in *Tfl* buffer. A 5 minute denaturation step at  $94^\circ\text{C}$  was followed by 40 cycles of  $94^\circ\text{C}$  for 1 minute,  $52^\circ\text{C}$  for 1 minute, and  $68^\circ\text{C}$  for 2 minutes. The resulting fragments were purified using the QIAquick PCR product purification protocol (Qiagen) and the products were used as template in the sequencing reactions. The sequencing was done, as described above, using Rice EP-2 (for), Rice EP-3 (rev) and Rice ER-1 (rev) as primers.

### Intron Prediction of the Rice BAC Sequence

The program NetGene2 at the Center for Biological Sequence Analysis (Hebsgaard, 1996) was used to predict possible introns in the 531 bp rice fragment. *Arabidopsis thaliana* was chosen as organism of interest and submission of the rice nucleotide sequence was via FASTA format.

### cDNA Library Screening for *Oryza sativa* (Rice) TERT

The Rice 5'-STRETCH cDNA Library was purchased (Clontech) and contained  $\lambda$ gt11 as the cloning vector and *E. coli* Y1090- as the host strain. The library was screened using two distinct methods, PCR and nucleotide hybridization.

Host cells, for PCR screening of the library, were prepared by adding 0.1g maltose and 1M  $\text{MgSO}_4$  to 50 ml of LB. The culture was shaken overnight at 37° or until a turbid suspension was achieved. After preparation of the host cells, by centrifuging down cells and resuspending them in 25ml 10mM  $\text{MgSO}_4$ , the library was titered to determine the phage particle concentration. A series of dilutions were set up in SM buffer (5.8g NaCl, 2.0g  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 1M Tris pH 7.5, 2% gelatin, per liter of water) to make 10-fold serial dilutions ranging from 10 to  $10^6$  fold diluted. Each dilution (10 $\mu$ l) was added to 200 $\mu$ l host cells and incubated at 37°C for 15 minutes. After incubation, 3ml of top agarose (0.75% agarose in NZCYM media (Life Technologies)) was added to each sample and poured onto NZCYM plates. The plates were incubated overnight at 37°C and the resulting plaques were counted to obtain titer. The titer was determined using the following equation: pfu (plaque forming units)/ml = #plaques/ $\mu$ l used  $\cdot$  dilution factor  $\cdot$  1000 $\mu$ l/ml.

Screening of the library by PCR consisted of multiple rounds of plate dilutions and PCR to screen for the correct insert size. The two primers used in the screening were: Rice ER-1 (rev) and Rice EP-2 (for) (see above for sequences). To begin,  $10^3$  pfu/ $\mu$ l ( $10^3$  pfu/100mm) plates were made as described above in the titering protocol. After plaque formation, 2 ml SM buffer was added to each plate and placed at 4°C overnight. The SM buffer was removed, and a drop of chloroform was added to inhibit bacterial growth. For each PCR reaction, 2  $\mu$ l of template (phage suspended in SM buffer) was added (after heating at 100°C to denature phage protein coat), 250  $\mu$ M primers, 0.1mM dNTPs, 1 unit of *Tfl* polymerase, and 1.5mM concentration of MgSO<sub>4</sub> and *Tfl* buffer. PCR conditions included a 2 minute denaturation step at 94°C followed by 40 cycles of 94° for 1 minute, 55°C for 1 minute, and 68°C for 2 minutes. The sample showing the correct sized band, as determined by the previously sequenced BAC clone, was used as template to make the second dilution series on NZCYM plates.

#### Phylogenetic Analyses and Sequence Alignments of Known TERTs

A multiple sequence alignment was prepared using the deduced polypeptide sequence of *Arabidopsis* and known TERT polypeptides. The multiple sequence alignment of all TERT polypeptides was constructed using ClustalW (v. 1.74) (Higgins, 1991) with gap openings and extension penalties set at 30 and all other parameters set at default. The Blossum series of similarity matrices was chosen. The resulting alignment was obtained and minor adjustments were made by hand. GenBank accession numbers for sequences used in the alignment were: human, AF015950; mouse, AF051911; *O. trifallax*, AF060230; *E.*

*aediculatus*, U95964; *T. thermophila*, AF062652; *S. pombe*, AF015783; *S. cerevisiae*, U20618; *C. albicans*, AF216871; hamster, AF149012; and *A. thaliana*, AF172097.

The ClustalW TERT alignment was then analyzed using Dayhoff PAM distance matrix and Neighbor Joining in the PHYLIP (v. 3.5) package (Felsenstein, 1989). 1000 Bootstrap value replicates were performed using the SeqBoot program and the resulting data was fed into TreeView (v. 1.5) for the phylogenetic tree output (Page, 1996).

#### Phylogenetic Analysis of the Identified TERTs including *Oryza sativa* putative TERT

As a full length *OsTERT* gene was not obtained, only the 270 bp segment pulled from the higher plant database was used in comparison against the corresponding fragments of ten existing TERT genes. A multiple alignment and phylogenetic analysis of the 270 bp TERT fragments (including rice) were performed as described above for the *Arabidopsis* analysis.



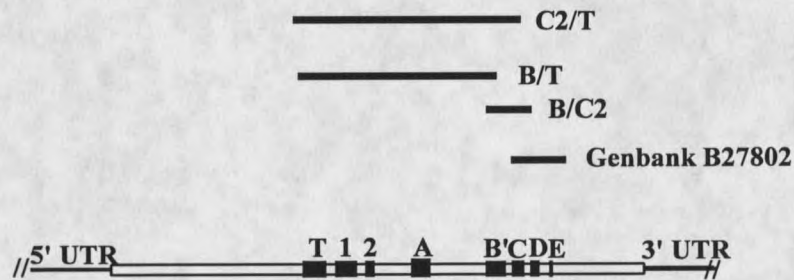
## Results and Discussion

### Identification of *Arabidopsis thaliana* TERT Gene

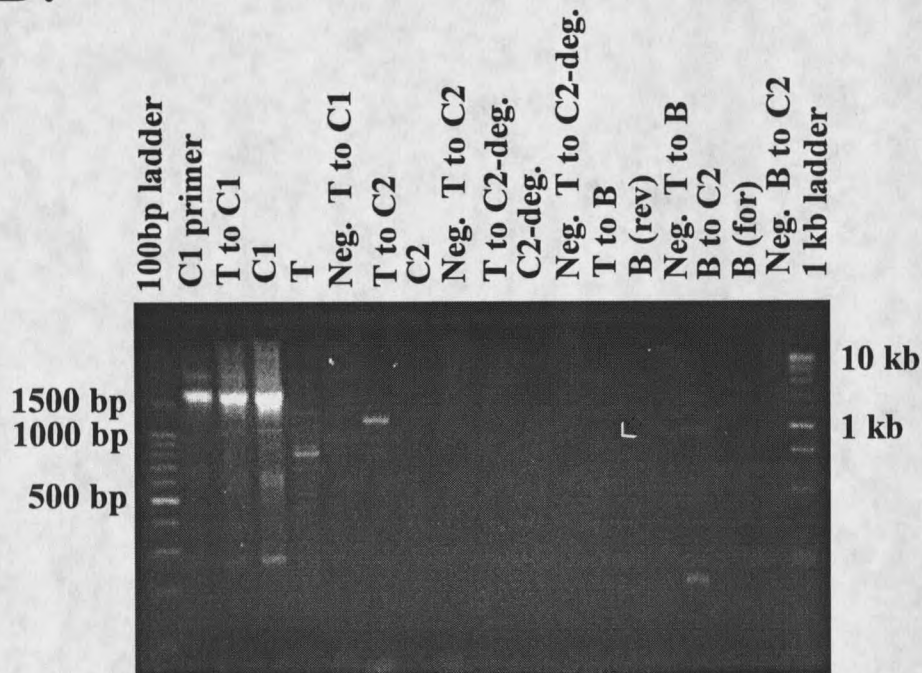
The primary goal of this study was to identify plant *TERT* genes. The GenBank *Arabidopsis thaliana* database was searched (via the BLAST algorithm) using a 76 amino acid fragment of the known human TERT protein as the query sequence. From this BLAST output, a putative *A. thaliana* *TERT* coding sequence was identified (GenBank # B27802). The 625 bp genomic sequence encoded an amino acid sequence with 57% similarity to the region of hTERT containing motifs C, D and E.

Exact primers were made using sequence information from the 625 bp putative *TERT* gene fragment, and degenerate primers made using sequence information from previously cloned TERT genes (*S. cerevisiae*, *S. pombe*, *T. thermophila* and *O. trifallax*). Various combinations of the exact and degenerate primers were used in a PCR reaction utilizing *A. thaliana* genomic DNA as the template. These PCR products, as well as control reactions (containing either water as template or one primer alone), were electrophoresed on an agarose gel. No bands of the expected size, as estimated by known TERT sequences, were visible by ethidium bromide staining. Interestingly, one control sample, amplified with the Arab C1 (rev)-exact primer alone (see materials and methods), resulted in strong amplification of an approximately 1600 bp DNA fragment (Figure 1-3B). We assumed that the primer bound a cryptic C-like binding site upstream of the known motif C, and this sample was subsequently used as template in a set of nested PCR reactions. These PCR reactions, using the Arab C1

A.



B.



**Figure 1-3. PCR Amplification of *Arabidopsis thaliana* Genomic DNA using Exact and Degenerate Primers.** A. Schematic of telomere gene with PCR products shown as black bars. GenBank B27802 indicates fragment retrieved from GenBank database. B. Agarose gel showing PCR products produced by each primer set (as discussed in Materials and Methods). *Arabidopsis* TERT products of the expected size are seen in lanes T to C2, T to B (indicated by arrow) and B to C2.

(rev)-exact 1600 bp fragment as template, were set up using different primer sets and similar controls as seen in the first round of PCR. PCR products were resolved on an agarose gel. The gel (Figure 1-3B) shows bands of the correct size (as estimated by known TERT sequences) for three reactions: T (for)-yeast combo to Arab C2 (rev)-exact, T (for)-yeast combo to B (rev)-yeast combo, and B (for)-yeast combo to Arab C2 (rev)-exact.

Sequencing of the three DNA fragments gave us additional segments of the *Arabidopsis* gene, spanning the gene from motif T to motif C (Figure 1-3A). Since the telomerase-specific motif T was now identified, we were confident that the gene fragment we had obtained was indeed telomerase, rather than another member of the RT gene family.

Sequencing both strands from the DNA fragments obtained by PCR gave us an 876 bp fragment of the *A. thaliana* TERT gene (Figure 1-4). The fragment contained the previously described RT motifs 1, 2, and A-C as well as the telomerase-specific motif T. Soon after these data were obtained, the entire *Arabidopsis thaliana* TERT (designated *AtTERT*) gene was published (Oguchi, 1999). The sequence of the full *AtTERT* gene, corresponding to the TERT fragment we identified, was identical with no variation at either the amino acid or nucleotide level. The *Arabidopsis* TERT gene was found to be 3372 bp, with eleven introns, encoding an 1124 amino acid polypeptide. The calculated molecular mass was determined to be 131 kDa, similar to previously described TERTs which range from 103 to 133 kDa, and the calculated isoelectric point was pH 9.9. Previously described TERT isoelectric points ranged from 9.5 to 11.3 (Oguchi, 1999).

It is noteworthy, that a fragment was detected when using the T (for)-yeast combo primer alone (Figure 1-3B). We chose not to follow up on this fragment as even if it had been

**MOTIF T**

R L N I Y Y Y R K R S W E R L I S K  
GACGGCTAAATATTTATTATTACCGGAAAAGGAGCTGGGAAAGATTAATAAGCAAAG

E I S K A L D G Y V L V D D A E A E S  
AAATTAGCAAAGCCCTTGATGGATATGTCCTAGTAGACGATGCTGAAGCTGAAAGTA

**MOTIF 1**

S R K K L S K F R F L P K A N G V R M  
GCAGGAAGAAGCTATCAAAGTTTAGATTTTTTACCAAAGGCCAATGGTGTGAGGATGG

**MOTIF 2**

V L D F S S S S R S Q S L R D T H A V L  
TGTTAGACTTTAGTTCTTCGTCAAGGTCGCAATCTCTTCGTGATACACATGCTGTTTT

K D I Q L K E P D V L G S S V F D H D  
GAAGGACATCCAGCTCAAAGAACCAGATGTTCTTGGGTCTTCTGTCTTTGACCATGAT

D F Y R N L C P Y L I H L R S Q S G E  
GATTTCTACAGAAACCTATGCCCATATCTGATCCATTTAAGAAGTCAATCTGGAGAAC

**MOTIF A**

L P P L Y F V V A D V F K A F D S V D Q  
TTCCTCCTTTGTACTTTGTGGTTGCGGATGTATTCAAAGCATTTGATTCAGTCGACCA

G K L L H V I Q S F L K D E Y I L N R C  
GGGTAAGCTGCTTCATGTCAATCAAAGTTTTCTGAAAGATGAATACATCTTAAACAGA

R L V C C G K R S N W V N K I L V S S  
TGTAGGCTGGTCTGCTGTGGGAAGAGATCCAATTGGGTAAACAAATACTAGTCTCGA

D K N S N F S R F T S T V P Y N A L Q  
GTGACAAAATTCTAACTTTTCAAGATTCACATCAACTGTTCCATATAATGCACTGCA

S I V V D K G E N H R V R K K D L M V  
AAGTATCGTGGTTGATAAGGGAGAAAACCATCGAGTGAGGAAAAAGGATCTAATGG

W I G N M L K N N M L Q L D K S F Y V  
TTTGGATAGGAAATATGCTAAAGAACACATGCTGCAGTTGGATAAAAGCTTCTATG

**MOTIF B**

Q I A G I P Q G H R L S S L L C C F Y  
TACAAATAGCTGGAATACCTCAGGGACACAGATTATCATCCTTGTTATGTTGTTTTTA

Y G H L E R T L I Y P F L E E A S K D  
CTACGGGCATCTCGAGAGGACTTTGATCTACCCATTCTCGAAGAAGCCTCTAAAGA

V S S K E C S R E E E L I I P T S Y K  
TGTGTCTTCTAAAGAATGCAGTAGAGAGGAAGAGCTTATAATTCCCACGAGCTATAA

**MOTIF C**

L L R F I  
GTTACTGAGATTTATT

**Figure 1-4. Genomic *Arabidopsis thaliana* TERT Fragment.** 876 bp fragment resulting from PCR amplified products. Motifs are underlined and labeled (T, 1, 2 and A-C).

part of the TERT fragment it would have given us sequence information downstream of motif T, sequence information we had obtained already.

#### Structural Characterization of *At*TERT

A multiple alignment of the amino acid sequences of conserved motifs was done for all known TERTs, including *At*TERT. In previously published papers (reviewed in Bryan, 1999), multiple sequence alignments of TERTs included only eight conserved motifs: T, 1, 2, and A-E, representing only 226 amino acids of the full gene which ranges from 867 (*C.albicans*) to 1126 (human) amino acids. In our multiple sequence alignment, however, we aligned the entire TERT polypeptides, resulting in 15 highly conserved regions (representing 466 amino acids), TEL1-5, T, 1, 2 and A-G. The multiple sequence alignment of all 15 conserved regions is seen in Figure 1-5. Here, only the motifs are shown (denoted by black underbars), rather than the entire sequence (Appendix A), including the eight previously



|      |   | I                                                               | II        |
|------|---|-----------------------------------------------------------------|-----------|
| ham. | 1 | RNVLTGFGFALLN-SVTESLRVSGAWMLLNRRVGGDLLVYLL-LQPSLTGARRLVEILFLG   |           |
| m.   | 1 | RNVLAFCFELLN-TVETILRVSGAWMLLSRVGGDLLVYLL-LQPNLTGARRLVEIIFLG     |           |
| h.   | 1 | KNVLAFCFALLD-TVTDALRGSGAWGLLRRVGGDVLVHLL-LRPSLTGARRLVETIFLG     |           |
| Ea.  | 1 | KLQCFGFQLKGN----KQYFFQDEWNQVRAMIGNELFRHLY-NKNNISAMDRAQT-IFTN    |           |
| Ot.  | 1 | KLQTFGYKLVNN-NLNKQCMITKNWEKVYFYLGDHLMHIY-NSKTLSSQQQIRDN-IYKE    |           |
| Tt.  | 1 | KQVLDSDSIIILG-NLSRQTLYSSNERNLLNIEGEEDEKYEEL-IPDPHPQTIIKKT-ILKE  |           |
| ar.  | 1 | REIVERIIEEMO-SPILELITSSWEFLLRVGHVDMVYLL-LRPNCSDSKLLMNHIFGE      |           |
| sp.  | 1 | RNLMLKGFSMNH-NYLISILESKNQILLLEITIGSDAMHYLL-SICDRNTVHMWLQWIFPR   |           |
| Sc.  | 1 | NNVLTFG-YKLA-NNVNLTLLKGAAMKMFHSIVGTYA FVDLL-ILPSSSSIKKLTLD-LREA |           |
| Ca.  | 1 | NNVLAAG-YKTS-NFILTHITSHNWSTIFSLIGPKKFLELL-IRPSIQITILQDIL-SFSG   |           |
|      |   | TEL5                                                            | TEL4 TEL3 |

|      |    | II                                                           | Y    |
|------|----|--------------------------------------------------------------|------|
| ham. | 59 | MRP-KRYWQMRPLFQQLLVNHARCPYVRLLRSHCRF-LRLHSSPWQVYGFQACVGLV    |      |
| m.   | 59 | SRP-ERYWQMRPLFQQLLVNHAECQYVRLLRSHCRF-LRLHSSPWQVYGFRLACLCKVV  |      |
| h.   | 59 | SRP-QRYWQMRPLFLELLGNHAQCPYGVLLKTHCPL-LLRQHSSPWQVYGFVRACLRRLV |      |
| Ea.  | 55 | IFR-KLKDKVIEKIAYMLEKVKDFNENYILTSCPL-LFSYTTDNKCVTQFINEFFYNIL  |      |
| Ot.  | 58 | VFG-ELKGKVMSEIEQVIVNQKKFDFKYYSKNCPL-LFEYQQDQROI SNFLTEFVANVF |      |
| Tt.  | 58 | YQS-EERDLFLEFTEKIVQNFHNINFNLLKKFCKL-LYDTEISYKQITNFLRQIIQNCV  |      |
| ar.  | 59 | VNV-CLYHSLKSLKNNLIGKTKSSHLKMLLDKHCPL-ERKLYCTNDQVVSFIWAICRYIV |      |
| sp.  | 59 | QFG-KRLKVPYPLIEQTAKRLHRISLSKVYNHYCPY-ILSYSLKPNQVFAFLRSILVRVF |      |
| Sc.  | 57 | IFP-RLKVRINLTQKLLKRHKRLNYSILNSICPP-HLSRQSPKERVCLKFTIIVILQKLL |      |
| Ca.  | 57 | LNP-KRYRGFKSLLSRIIAND-----KKCRY-VVSNKTEISQVIOQFVLLVLGKLL     |      |
|      |    | TEL2                                                         | TEL1 |

|      |     | III                                                            |
|------|-----|----------------------------------------------------------------|
| ham. | 117 | PPGLWGSRRHNQRRFEKKNVRFISLGKY-DKLSLQELTWKMKVQDCRWL-WLMDAYVVELL  |
| m.   | 117 | SASLWGTRHNERRFEKKNLKKFISLGKY-GKLSLQELMWKMKVEDCHWL-WLMDTYVQOLL  |
| h.   | 117 | PPGLWGSRRHNERRFEKKNLKKFISLGKH-AKLSLQELTWKMSVRDCAWL-WLMSVYVVELL |
| Ea.  | 113 | PKDFLTGR-NRKNFQKKVKKYVELNKH-ELIHKNNLLEKINTREISWM-WIFEDLVVSLI   |
| Ot.  | 116 | PKNFLEGK-NKKIFNKKMLQFVKFNRF-ESFTKISLKNFRVNEVSWL-WVFEDIAITLM    |
| Tt.  | 116 | PNQLLGKK-NFKVLEKLYEFVQMKRF-ENQKVLDYICFMDVDFVEWF-FIINKIVIPVL    |
| ar.  | 117 | PESILGITHQMRVLRKNIAWFSRRRN-EKCTVNQFLHKVKPSDFPFF-WLFEITVKKLV    |
| sp.  | 117 | PKLIWGNQRIFEIILKDLTFLKLSRY-ESFSIHYLMSNIKISEIEWL-WLYNSFIIPIL    |
| Sc.  | 115 | PQEMFGSKKNKGKIIKNLNLISLPLN-GYLPFDSLKKLRKDFRWL-WLFRQLIPKII      |
| Ca.  | 105 | PLDAWGGVSNKKI IKDRVDFLLGAN-EKIHMDLFRGIRLKDFKWL-WLFEHLKNIL      |
|      |     | TEL1 Motif T                                                   |

|      |     | IV                                                            |
|------|-----|---------------------------------------------------------------|
| ham. | 175 | RSFFYVTETTFQKNRIFFYRKSMWRRLQSIGVRHHLE-EVRQRQEAWPAMPICRLRFIPK  |
| m.   | 175 | RSFFYITESTTFQKNRIFFYRKSVWSKLQSIGVRQHLE-EVRHHQDTWLAMPICRLRFIPK |
| h.   | 175 | RSFFYVTETTFQKNRIFFYRKSVWSKLQSIGIRQHLK-EVRQHREARPALLTSRLRFIPK  |
| Ea.  | 170 | RCFFYVTEQOKSYSKTYYYRKNIWDVIMKMSIADLKK-EVEEWKKS LGFAPG-KLRLIPK |
| Ot.  | 173 | RCYFYSTEKAKKEYQRIFFYRKNIWNMIMRLSIDDLLK-EMRIFCESQNFAPG-KLRLIPK |
| Tt.  | 173 | RYNFYITEKHKEGSQIFYYRKPIWKLVSKITIVKLEE-LIPE--DSFQKYPQGKLRIIPK  |
| ar.  | 175 | HFNFYATESQGGRLNIYYRKRSWERLISKEISKALD-ES----SR--KK-LSKFRFLPK   |
| sp.  | 175 | QSFFYITESSDLNRNTVYFRKDIWKLRCRPFITSMKM-NVRMD-TQKTTLP PAVIRLIPK |
| Sc.  | 173 | QTFYFCTEISSTVT-IVYFRHDTWNKLITPFIVEYFK-RNHNS-YTLSNFNHSKMRIIPK  |
| Ca.  | 163 | RSFYITETSSIVSSINYYFPQYLWKELYESWVSKYAK-QRE-----QLPCGKIKLIPK    |
|      |     | Motif T Motif 1                                               |

|      |     |                |                         |                                |
|------|-----|----------------|-------------------------|--------------------------------|
| ham. | 234 | PSG-LRPIVNM    | SY-MGTRA-FDKGKQAQHF     | TQCL-PRMYFVKADVTGAYDAIPQDKLVEV |
| m.   | 234 | PNG-LRPIVNM    | SYSGTRA-LGRRKQAQHF      | TQRL-PRMYFVKADVTGAYDAIPQDKLVEV |
| h.   | 234 | PDG-LRPIVNM    | DYVVGART-FRREKRAERLTSRV | -PELYFVKVDVTGAYDTIPQDKLVEV     |
| Ea.  | 228 | KTI-FRPIMTFNRK | IVNSD-RKTTKLTNTNKL      | LN-PKLEFATMDIEKCYDSVNREKLSTF   |
| Ot.  | 231 | GDT-FRPIMTFNRK | IPNOVGKFQSRMTTNNKL      | QT-PKLYFVAMDEKCYDNVDCERVVNF    |
| Tt.  | 230 | KGS-FRPIMTFNRK | QKQKNI-KLNLNQILMDSQ     | L-POLYVVTLDIKKCYDSIDQMKLINF    |
| ar.  | 227 | ANG-VR--MV     | DFSSSSRS-----SQ----     | SI-PELYFVVADVFKAFDSVDQGKLLHV   |
| sp.  | 233 | KNT-FRLITNIRK  | RFLIKM-GSNKKMLVSTNQ     | TL-RKKYFVRIDIKSCYDRIKQDLMFRI   |
| Sc.  | 230 | KSNNFRIIAIPCR  | GADEEE-FTIYKENHNKAIQ    | P-PELYFMKFQVKSICYDSIPRMECMRI   |
| Ca.  | 215 | RSS-FRVICVPIKE | KEKRE-FERYKEVLSVPVGQ    | I-PKLEFILKFQDMKECYDRISQPVLMKK  |

Motif 2

Motif A

|      |     |                  |            |                     |              |              |
|------|-----|------------------|------------|---------------------|--------------|--------------|
| ham. | 290 | IANMIR-ROYVQCQGI | POGSSLS    | STIL-CSLCFGDMEN---- | KLFAEV--LLLR | FVDDFLI      |
| m.   | 291 | VANMIR-ROYVQCQGI | POGSSLS    | STIL-CSLCFGDMEN---- | KLFAEV--LLLR | FVDDFLI      |
| h.   | 291 | IASILK-KSYVQCQGI | POGSIL     | STIL-CSLCYGMEN----  | KLFAET--LLLR | LVDDFLI      |
| Ea.  | 285 | LKTTKL-KFYKOTKGI | POGLCVSS   | IL-SSFYATLEE----    | SSLGFL--LLMR | LTDYLL       |
| Ot.  | 289 | LQKSDL-KQYKQMKGI | POGLCVSY   | IL-SSFYANLEE----    | NALQFL--LLMR | LTDYLL       |
| Tt.  | 287 | FNQSDL-VKTRQKRG  | IPQGLNISGV | L-CSFYEGKLEE----    | EYTQFL--LLMR | LTDYLF       |
| ar.  | 272 | LQSFLK-SFYVQIAG  | IPQGHRLSS  | IL-CCFYGHLE         | RTL          | IYPFLEE      |
| sp.  | 290 | VKKKLLK-SQYLQKVG | IPQGSILSS  | FL-CHFYMED          | LID----      | EYLSFT--VLLR |
| Sc.  | 288 | LKDALK-KCYREDGL  | FOGSSLSA   | PI-VDLVYDD          | LLE----      | FYSEFK--LILK |
| Ca.  | 272 | LEELFE-ALYKRKR   | GWFQGFSLSS | IF-CDILYSAM         | VH----       | DCFQFL--LFLR |

Motif B

Motif C

|      |     |                  |           |         |                |                 |                 |           |
|------|-----|------------------|-----------|---------|----------------|-----------------|-----------------|-----------|
| ham. | 342 | VTPHLVQA-EAFIRAL | VRGIP     | EYGC    | MINIQKT        | VVNF-HCLFPWCGL  | LID-IKASL       | TFQRT     |
| m.   | 343 | VTPHLDQA-KTFLSTL | VHGVPEYGC | MINIQKT | VVNF-HCLFPWCGL | LID-IKTSL       | TFQSV           |           |
| h.   | 343 | VTPHLTHA-KTFLRTL | VRGVP     | EYGC    | VVNL           | RKT             | VVNF-HGLFPWCGL  | LID-IRASL |
| Ea.  | 337 | ITQENNA-VLFEK    | LNVSR     | ENGFKFN | MKKL           | QTSF-QDYCDWIG   | ISID-LCTL       | NLNMQT    |
| Ot.  | 341 | MTTEKNA-MLFEK    | LYQLSLGN  | FFKFH   | MKKL           | KTNE-DDLFWHIG   | ISID-LCTL       | NLNMQT    |
| Tt.  | 339 | ISDSQONA-LNLIV   | QONCANN   | NGFM    | ENDQ           | KITTNE-QNECQWIG | KSID-NQTIN      | VAISI     |
| ar.  | 328 | VSTSRDOA-SSFYHRL | KHGF      | DYCNF   | NETK           | FCINE-VPFVR     | WTGLLIN-ISST    | FSAVQ     |
| sp.  | 342 | ITVNKKDA-KKFNLS  | LRGFEK    | HNFTS   | LEKTV          | INE-KKRM        | PFTGLLIN-FNST   | SVELT     |
| Sc.  | 340 | ISTDQQQV-INIKKL  | AMGCFQ    | KYN     | AKANRDK        | ILAVS-DTVIQ     | ECAMHIF-IR----- |           |
| Ca.  | 324 | VTPDSDQV-HNI--   | LSGKILES  | YGA     | VN             | KDKTVMVN-KTSIDE | VGLEVN-TN-----  |           |

Motif D

Motif E

Motif F

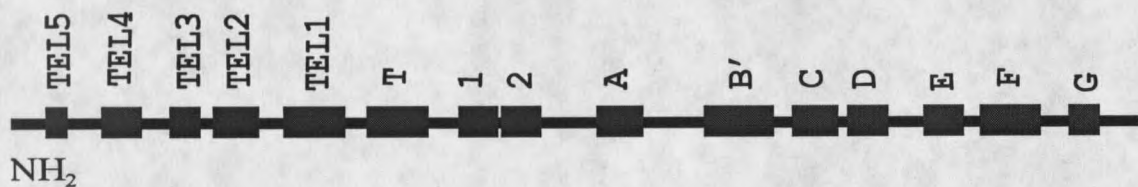
|      |     |              |           |        |            |              |               |
|------|-----|--------------|-----------|--------|------------|--------------|---------------|
| ham. | 399 | FKAGNMROKLLA | VRLKCHSL  | FIDLQ  | MNSLQ-TVCT | INVY--KIFILQ | AYRFHACALQ    |
| m.   | 400 | FKAGTMRNKLLS | VRLKCHGL  | FIDLQ  | VNSLQ-TVCT | INIY--KIFILQ | AYRFHACVQL    |
| h.   | 400 | FKAGNMRRKLE  | GVRLKCHSL | FIDLQ  | VNSLQ-TVCT | NIY--KILLLQ  | AYRFHACVQL    |
| Ea.  | 394 | KKASMWLKKKL  | KSEFLMNN  | THYER  | KT-ITTED   | -FANK        | TIN--KLFI     |
| Ot.  | 398 | NESILWLKKKL  | KSEFLMNN  | SFYEK  | ST-INTK    | -FANIT       | LS--KLYIAA    |
| Tt.  | 396 | KNLKSQLNKLR  | SFFLNQ    | LIDYEN | PN-INSFE   | -GLCRO       | LYHHSKATVM    |
| ar.  | 385 | NKPVNLRQKLC  | YFLPKCH   | PLEDSN | INSGE-IVR  | INIY--QIFLLA | AMKFFHCYVYEV  |
| sp.  | 398 | KHMGKSFFYK   | ILRSSLAS  | FAQVF  | IDITHNS    | KF-NSCC      | NIY-----RLGYS |
| Sc.  | 389 | SKSSKGLFRSL  | IAFNTR    | SYKT   | IDTNIN     | STN-TVLM     | QID--HVVKN    |
| Ca.  | 371 | FRTFTLVRYL   | KTFYQLN   | EGFLD  | CSFGVLE    | -NVLEN       | MG--SLRL      |

Motif F

|      |     |                  |
|------|-----|------------------|
| ham. | 455 | -YQAFLLKLAGHSVTY |
| m.   | 456 | -YQAFLLKLAASVIY  |
| h.   | 456 | -HQAFLLKLTRHRVTY |
| Ea.  | 449 | -LKHFTEIFSTKYYIF |
| Ot.  | 453 | -LKFFITRFKQKNEE  |
| Tt.  | 454 | -LNKLLOKTRKSRFFY |
| ar.  | 441 | -LDAYIQVLKKNSRY  |
| sp.  | 448 | -LKPSF-KYHPCFEQL |
| Sc.  | 445 | -LNGFLESLSNTSKE  |
| Ca.  | 427 | -LNCIKQITVKEFSSE |
|      |     | <b>Motif G</b>   |

**Figure 1-5. Multiple Sequence Alignment of Motifs of Nine Full-length Telomerase RTs.** Motifs are denoted by black underbars. Black boxes represent an amino acid conserved in at least five or more proteins and the grey boxes represent similar amino acids in at least five proteins. Species abbreviations are as follows: ham., hamster; m., mouse; h., human; Ea., *Euplotes aediculatus*; Ot., *Oxytricha trifallax*; Tt., *Tetrahymena thermophila*; At., *Arabidopsis thaliana*; Sp., *Schizosaccharomyces pombe*; Sc., *Saccharomyces cerevisiae*; and Ca., *Candida albicans*. Regions showing hypomutability (Friedman, 1999) are indicated by: I(---), II(.....), III(— —), and IV(——). The non-essential region,  $\gamma$ , is indicated by .

described motifs T, 1, 2, and A-E and seven of our newly described motifs, TEL1-5, F and G. Black boxes represent amino acids conserved in at least five or more proteins whereas the grey boxes represent similar amino acids in at least five proteins. Of the newly described motifs, five were found to be upstream of the telomerase-specific motif T, TEL1-TEL5, and



**Figure 1-6. 15 Sequence Motifs Identified in Known TERT Proteins.** Reverse transcriptase (1,2 and A-E) and telomerase-specific (T) motifs identified in all TERT deduced polypeptides with the addition of seven new motifs described here (TEL1-5, F and G).



two were found downstream of the motif E, motifs F and G (Figure 1-6). No conserved motifs were located at the far N-terminus (Appendix A).

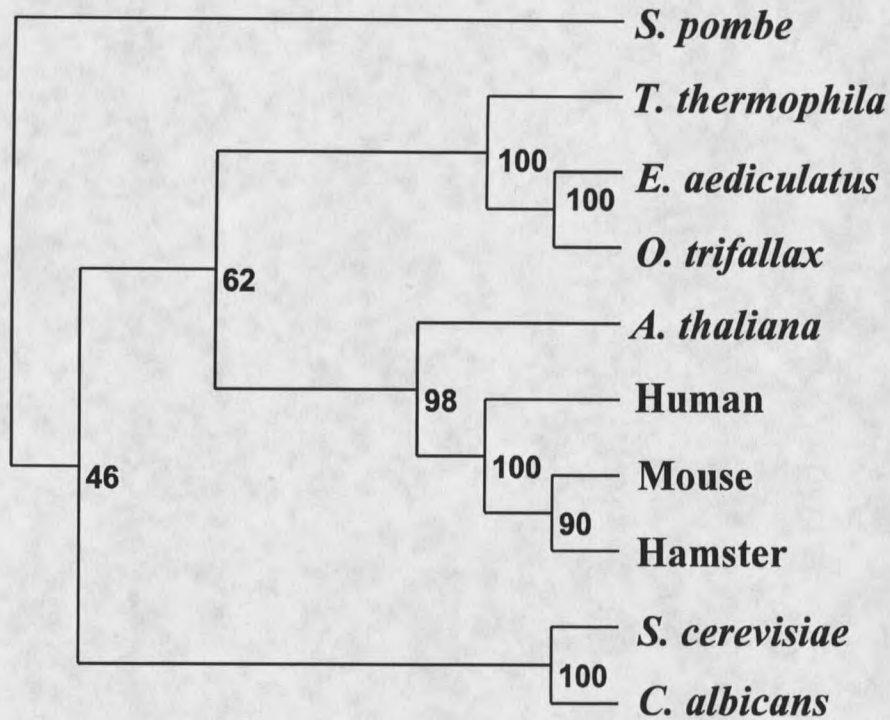
Since multiple sequence alignments can be somewhat subjective, we compared our motif alignment with results from mutation studies done on the N-terminal domain of Yeast TERT (Friedman, 1999). Unigenic mutation studies work under the assumption that analyzing distribution of mutations within complementing alleles of a gene could distinguish between essential and non-essential regions of the encoded protein. Non-essential protein regions are assumed to tolerate mutations well, while essential regions are predicted to contain few or no amino acid substitutions (Friedman, 1999). A total of 33 mutations were made between the N-terminal and the conserved telomerase specific motif T (Friedman, 1999). Friedman's results showed that there were four regions of hypomutability, meaning essential regions for Sc\_Est2p function (*S. cerevisiae* TERT), and three non-essential regions. Motifs TEL5 and TEL4 were included in region I of the hypomutable regions, motifs TEL3 and TEL2 were included in region II, a fragment of TEL1 and a fragment of the conserved motif T were included in region III, and the rest of motif T was included in hypomutable region IV (Figure 1-5). We did, however, include one non-essential region in our alignment, a twelve amino acid block at the beginning of the TEL1 motif, corresponding to the  $\gamma$  non-essential region. Of the 12 amino acids in this region, however, mutations were only made on five of the 12 residues (Friedman, 1999), none of which were the highly conserved amino acids (H, L, S, P, R or V) as seen in our alignment. Overall, the point mutation study identified four essential regions, all of which corresponded to the five highly conserved N-terminal regions identified by our alignment, which showed amino acid sequence conservation

in all four of the hypomutable regions. Thus, this work supports our hypothesis that the newly identified and highly conserved motifs TEL1-TEL5 most likely have been evolutionary conserved, and thus have functional importance.

The sequence alignments from all 15 highly conserved domains were subsequently used to determine amino acid sequence similarity/identity between TERTs and to construct a possible phylogenetic tree (Table 1-1 and Figure 1-7). As *AtTERT* was the first plant telomerase to be identified, it was important to determine its relationship to other known TERTs. The motifs of *A. thaliana* show greater sequence identity and similarity to those of human, mouse and hamster than they do to those of TERTs from yeast or ciliates (Table 1-1). Similarly, *A. thaliana* clusters closely with human, mouse and hamster in the phylogenetic tree (Figure 1-7). The tree contains three main clusters of TERT proteins, corresponding to those of yeast (*S. cerevisiae*, *C. albicans* and *S. pombe*), ciliates (*T. thermophila*, *O. trifallax* and *E. aediculatus*) and higher eukaryotes (human, mouse, hamster and *A. thaliana*). The bootstrap values, indicated at each node, indicate the percentage of 1000 bootstrap replicates that would place that branch in that particular position on the tree. As each value is close to 100 (between 90-100) within clusters, we have high confidence that this is a valid phylogenetic tree. Overall, the identification of *AtTERT* reinforces the conclusion that the TERT catalytic subunit is a phylogenetically conserved component of telomerase throughout evolution. The high level of conservation, between TERTs from distinct organisms, suggests strong conservation over time. Telomerase, without doubt, has been maintained over time and is of evolutionary significance.

**Table 1-1. Amino Acid Sequence Homology for 466 Positions Among the Nine Known TERT Proteins**

| Identity (Similarity) | <i>A. thaliana</i> | <i>O. trifall.</i> | <i>T. thermo.</i> | <i>E. aedicul.</i> | mouse  | human  | hamster | <i>S. pombe</i> | <i>S.cerevisiae</i> |
|-----------------------|--------------------|--------------------|-------------------|--------------------|--------|--------|---------|-----------------|---------------------|
| <i>C. albicans</i>    | 19(45)             | 18(45)             | 19(49)            | 21(46)             | 24(53) | 24(49) | 24(54)  | 23(48)          | 29(57)              |
| <i>S. cerevisiae</i>  | 21(49)             | 21(53)             | 19(48)            | 21(50)             | 25(55) | 25(53) | 25(55)  | 24(51)          |                     |
| <i>S. pombe</i>       | 24(48)             | 24(51)             | 23(49)            | 23(49)             | 29(54) | 27(52) | 28(55)  |                 |                     |
| hamster               | 32(58)             | 22(53)             | 24(53)            | 22(49)             | 85(94) | 76(91) |         |                 |                     |
| human                 | 32(60)             | 23(51)             | 23(52)            | 23(48)             | 72(87) |        |         |                 |                     |
| mouse                 | 34(60)             | 22(53)             | 25(54)            | 22(49)             |        |        |         |                 |                     |
| <i>E. aediculatus</i> | 23(46)             | 54(75)             | 32(62)            |                    |        |        |         |                 |                     |
| <i>T. thermophila</i> | 25(52)             | 34(63)             |                   |                    |        |        |         |                 |                     |
| <i>O. trifallax</i>   | 23(49)             |                    |                   |                    |        |        |         |                 |                     |



**Figure 1-7. Phylogenetic Tree of All Known TERTs.** TERTs were aligned at 466 positions, including previously described conserved motifs T, 1, 2 and A-E (Nakamura, 1997) and motifs TEL1-5, F and G (as described in this thesis). The number at each node indicates the percentage of 1000 bootstrap replicates for statistical support.

The finding that plant telomerases cluster with those of mammals is interesting for two reasons. First of all, previous phylogenetic analysis studies, on the divergence times of biological organisms such as plants, fungi, animals, protists and bacteria, suggest contrasting information to ours. To date, the most widely accepted phylogenetic analyses (comparing protein sequences) indicates that animals and fungi group together, excluding plants to a different branch. This suggests that animals and fungi are sister groups while plants constitute an independent evolutionary lineage (Baldauf, 1993 and Doolittle, 1996). Our data, however,

fail to show these same results. Rather, phylogenetic analysis of TERT proteins group animals and plants together, with fungi more distantly related. It is possible that previous sequence analysis studies have falsely grouped fungi and animals together, excluding plants, or that plant and animal *TERT* genes have for some reason evolved in parallel.

Secondly, our results indicate that plants may be a useful model system in which to study the role of telomerase and to address fundamental questions concerning telomere function in higher eukaryotes. The high similarity between the plant and mammalian TERT proteins, and the advantages of using plant system over a mammalian system, make this a strong possibility. Plants have many advantages over mammalian systems. For example, plants have significantly shorter telomeres than mice, making changes in telomere maintenance easily detectable as large changes in a short sequence (Fitzgerald, 1999). Also, certain plants, such as *Arabidopsis*, have benefits such as a short generation time (6 weeks), the ability to be grown in a test tube, and a small genome that has almost been completely sequenced.

#### Identification of *Oryza sativa* (Rice) Putative TERT Gene

After identification of *AtTERT*, we set out to identify additional plant *TERT* genes. Using a 28 amino acid segment of the *Arabidopsis thaliana* TERT protein sequence as query, we searched the "higher plant" database (Stanford University, *Arabidopsis thaliana* sequencing project, Home Page) via the "TBLASTN" algorithm. The search detected various potential TERT sequences, including a 531 bp fragment (GenBank #AQ510589), 270 bp of which were determined by re-sequencing to be a putative rice TERT gene (Figure 1-8) with the canonical reverse transcriptase motifs C, D and E. The 270 bp rice fragment was

MOTIF C

L M R F I D D F I F I S F S L E H A Q  
TTAATGAGGTTTCATTGATGATTTTCATATTTATCTCTTTCTCACTGGAGCATGCTCAA

MOTIF D

K F L N R M R R G F V F Y N C Y M N D  
AAATTCCTCAATAGGATGAGAAGAGGTTTTGTGTTCTACAATTGCTACATGAACGACA

S K Y G F N F C A G N S E P S S N R L Y  
GCAAATATGGCTTTAATTTCTGTGCTGGAAATAGTGAGCCTTCCTCTAATAGACTCTA

MOTIF E

R G D D G V S F M P W S G L L I N C E  
CAGGGGTGATGATGGAGTCTCATTCATGCCATGGAGTGGTTTGCTAATAAATTGTGAA

T L E I Q A D Y T R Y D C  
ACTTTGGAAATTCAAGCTGATTATACGAGGTATGACTGTT

**Figure 1-8. Genomic *Oryza sativa* (Rice) TERT Fragment.** 270 bp fragment retrieved from the *Arabidopsis thaliana* “higher plant” database. Motifs are underlined and labeled (C, D and E).

aligned with the corresponding region of *AtTERT* and determined to be 53% identical.

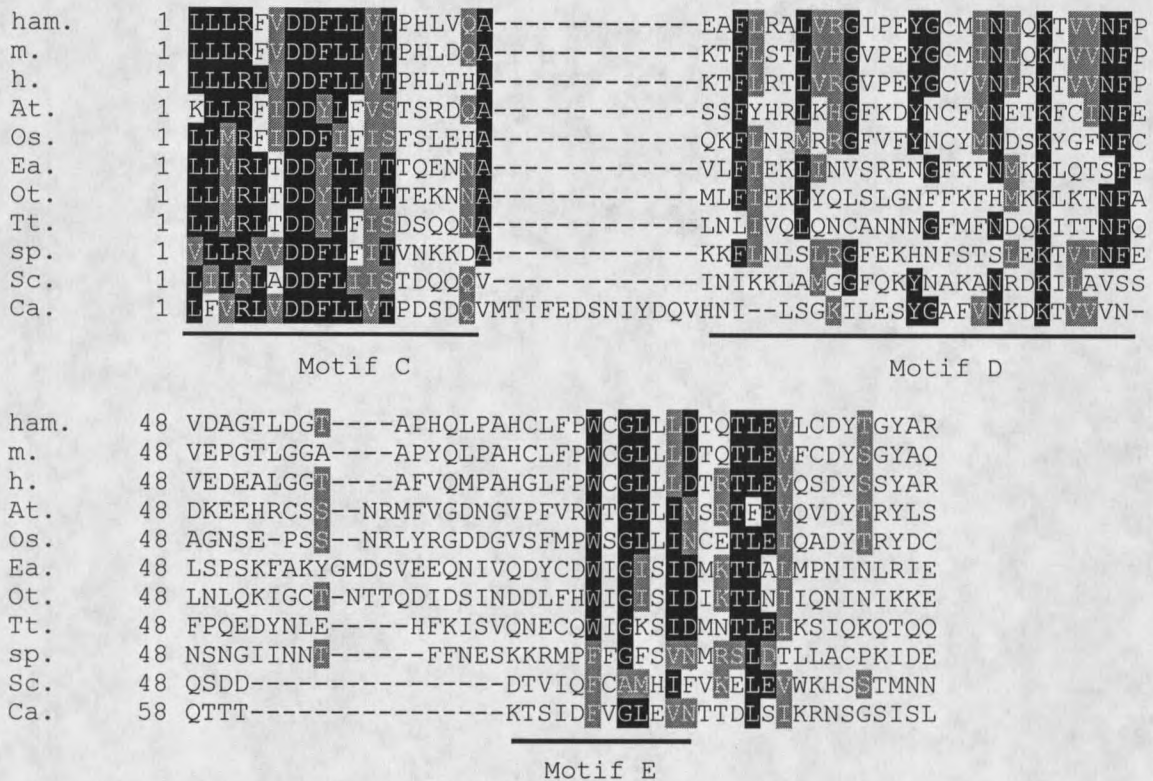
To identify the entire rice *TERT* gene, the 270 bp fragment was used as a probe to screen a rice cDNA library. The library was screened, both via hybridization and via PCR, but all attempts to obtain a TERT clone were unsuccessful. We hypothesized that screening the library may have been unsuccessful as a result of either using a probe that was too short to detect the telomerase gene under the hybridization conditions used or, alternatively, the telomerase RNA was present in low amounts when the library was constructed and thus only a very small percentage of the cDNA are telomerase. The next step, identifying the entire TERT gene, may include making additional attempts at screening the cDNA library via



different hybridization conditions, screening the cDNA library using antibodies made to a rice TERT peptide, or testing different primer sets that may be successful in rice TERT detection.

### Structural Characterization of the Putative *Oryza sativa* TERT Gene

Without the full length gene, we were unable to include the rice TERT protein in the multiple sequence alignment as seen for *Arabidopsis* (Figure 1-5). However, we did attempt



**Figure 1-9. Multiple Sequence Alignment of Motifs of Ten TERT Fragments (270 Amino Acid Residues) Including *Oryza sativa* (Rice).** Motifs are denoted by black underbars. Black boxes represent an amino acid conserved in at least five of more proteins and grey boxes represent similar amino acids in at least five proteins. Species abbreviations are as follow: ham., hamster; m., mouse; h., human; At., *Arabidopsis thaliana*; Os., *Oryza sativa*; Ea., *Euplotes aediculatus*; Ot., *Oxytricha trifallax*; Tt., *Tetrahymena thermophila*; Sp., *Saccharomyces pombe*; Sc., *Schizosaccharomyces cerevisiae*; and Ca., *Candida albicans*.

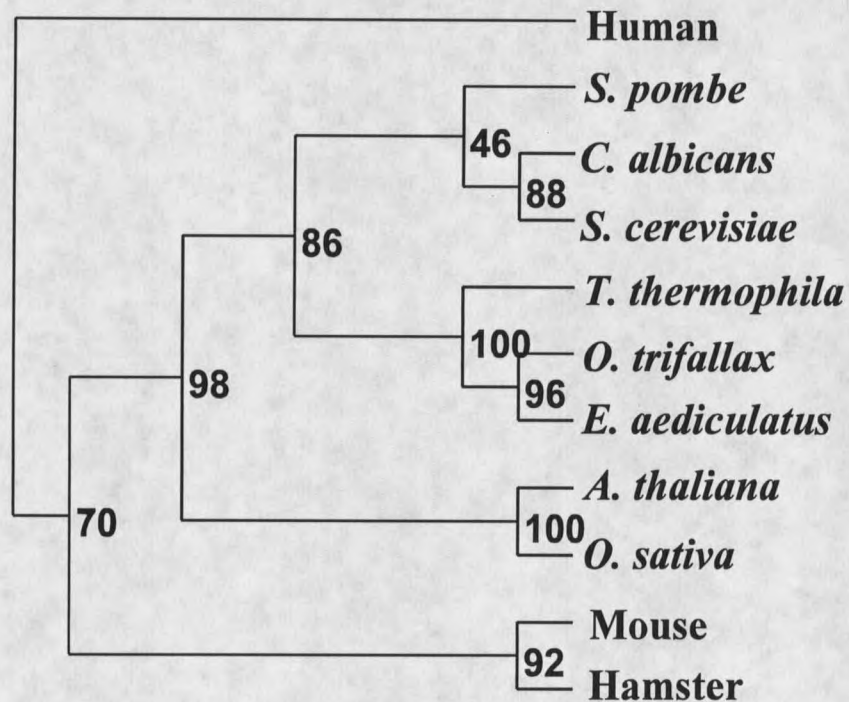
a modified alignment, aligning 270 amino acids of all known TERTs to the 270 amino acid deduced rice polypeptide sequence. This information, we felt, would support previous data clustering *Arabidopsis thaliana* TERT with mammalian TERTs rather than to yeast or ciliate TERT proteins.

The modified, 270 amino acid multiple sequence alignment (Figure 1-9) includes all eleven known TERT proteins. As with the *Arabidopsis* motif alignment, the three previously defined motifs (C, D and E) are denoted by black underbars and, similarly, black boxes signify a conserved amino acid in five or more proteins whereas the grey boxes represent similar amino acids in at least five proteins.

The multiple sequence alignment was then subject to phylogenetic analysis (Figure 1-10). The resulting tree, unlike the one obtained for *Arabidopsis* (Figure 1-7), consisted of four clusterings: yeast (*S. pombe*, *C. albicans* and *S. cerevisiae*), ciliates (*T. thermophila*, *O. trifallax* and *E. aediculatus*), plants (*A. thaliana* and *O. sativa*) and mammals (mouse, hamster and human), with the plant cluster most closely related to the mammalian cluster (as seen previously). The reason the higher eukaryotes separate into two groups after the addition of rice TERT (*OsTERT*) is that the addition of each new sequence increases the similarity among one group, and simultaneously makes divergence more apparent against other groups. Bootstrap values range between 70 and 100 (with the exception of the *S. pombe* branch) thus lending statistical support to the alignment and phylogenetic tree. Recent findings show that under a wide range of conditions, groups supported by bootstrap values above 70% have a >95% probability of representing true clades (Baldauf, 1993). The accumulation of additional plant sequences (including the full length *OsTERT* protein) will



no doubt strengthen support of the plant cluster. More important than looking within a cluster, is comparison between clusters. This new data supports the finding, previously discussed, that the motifs of plant TERTs show a greater sequence similarity to human, mouse and hamster than they do to TERTs from yeast or ciliates. This data also suggests important evolutionary relations among species. As the plant motifs are most similar to mammalian motifs, it is a logical step to focus on plants as a model system both to determine specific roles of telomerase in growth and development and to answer fundamental questions concerning telomere and telomerase function in higher eukaryotes.



**Figure 1-10. Phylogenetic Tree of Known TERT Fragments and *Oryza sativa* Fragment.** TERTs were aligned at 270 positions, including previously described motifs C, D and E (Nakamura, 1997). The number at each node indicates the percentage of 1000 bootstrap replicates for statistical support.

In addition to studying evolutionary relations among species, the discovery of plant *TERT* genes also makes possible avenues of research aimed at understanding the effects of structure and function of TERT on plant life cycles. Potential future studies may include, for example, examining implications for plant cell proliferative capacity by either up-regulating or down-regulating telomerase expression. Down-regulating telomerase expression may be a successful preventative measure against the growth of weeds (inhibiting root or flower growth) or up-regulating telomerase expression may result in increased productivity (an increased telomerase activity leading to a larger endosperm and thus improved grain yield over time). The importance of plants as staple foods to world economics and potential of telomerase research make investigating plant telomerases a promising field of research.

## **IDENTIFICATION OF RIBULOSEBISPHOSPHATE CARBOXYLASE/OXYGENASE ACTIVASE IN WHEAT**

### Introduction

Plants and cyanobacteria have evolved a mechanism of utilizing light energy to reduce  $\text{CO}_2$  and  $\text{H}_2\text{O}$  to produce carbohydrates and  $\text{O}_2$ . This process, known as photosynthesis, produces the carbohydrates that serve as the energy source for not only the plant itself, but also for organisms that directly or indirectly consume photosynthetic life.  $\text{CO}_2$  fixation, an essential step in photosynthesis, involves a conversion of the enzyme Ribulose-bisphosphate carboxylase/oxygenase (Rubisco) to an activated state for catalytic competency. The activation process has been shown to involve  $\text{CO}_2$ ,  $\text{Mg}^{2+}$  and Rubisco activase. For general interest, the role of Rubisco in photosynthesis and photorespiration is discussed in greater detail in the sections that follow. However, background information most relevant to this thesis is discussed in the later sections on regulation of Rubisco by Rubisco activase.

### Photosynthesis

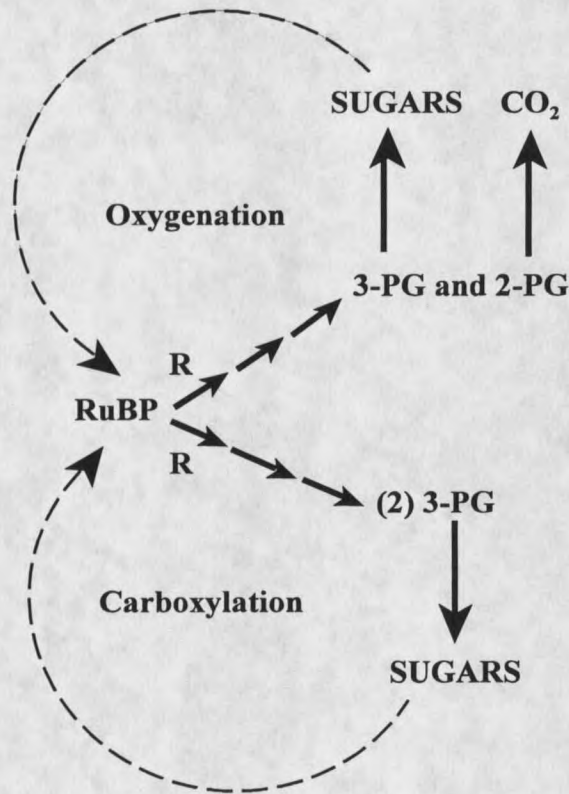
Photosynthesis, in plants, takes place in the chloroplast. There are approximately 1 to 1000 chloroplasts per cell (Voet, 1995) and about half a million chloroplasts per square millimeter of leaf surface (Campbell, 1993). Chloroplasts typically are ellipsoid in shape,  $\sim 5\mu\text{m}$  in length and are mainly found in cells of the mesophyll. They are known to have a highly permeable outer membrane and an impermeable inner membrane, known to enclose the stroma and various solutions (of enzymes, DNA, RNA and ribosomes), all of which are

involved in the synthesis of chloroplast proteins. A third compartment, the thylakoid, is surrounded by the stroma and consists of stacks of disklike sacs called grana. The grana, in turn, are interconnected by unstacked stroma lamellae. The thylakoid membrane, arising from invaginations in the inner membrane of developing chloroplasts, contains distinct lipids to give the thylakoid membrane fluidity (Voet, 1995 and Campbell, 1993).

Photosynthesis occurs in two phases in plants. The first phase of photosynthesis, or the light reaction, occurs in the thylakoids and uses light energy to generate chemical energy (NADPH and ATP). Specifically, the light reaction utilizes  $2\text{H}_2\text{O}$  and light energy to produce  $2\text{O}$  and  $4\text{H}$ . The resulting  $4\text{H}$  is then used in the dark reaction. The dark reaction, occurring in the stroma, uses the NADPH and ATP ( $4\text{H}$ ) from the light reaction to drive the synthesis of carbohydrates. The dark reaction includes:  $4\text{H}$  and  $\text{CO}_2$  reducing to  $\text{CH}_2\text{O}$  and  $\text{H}_2\text{O}$  (Voet, 1995 and Campbell, 1993).

The principal photoreceptor in the light reaction of photosynthesis is chlorophyll. Chlorophyll molecules act as antennas to collect light and then pass the energy (via electron transfer) to a photosynthetic reaction center. The photosynthetic reaction centers, or light-harvesting complexes, consist of arrays of membrane-bound proteins, each containing several pigment molecules, including chlorophyll and accessory pigments (Voet, 1995). In plants, the light reaction occurs in two reaction centers, Photosystem I (PSI) and Photosystem II (PSII). PSI and PSII are connected in a series, thus enabling the system to generate enough force to form NADPH by oxidizing  $\text{H}_2\text{O}$  in a noncyclic pathway. PSII contains an MN (Manganese-containing oxygen-evolving) complex that oxidizes two  $\text{H}_2\text{O}$ s to four  $\text{H}^+$  and  $\text{O}_2$ . The electrons are passed individually through a carrier to P680, the reaction center's photon-

absorbing species. Once ejected from P680, the electron passes to a pool of plastoquinone molecules, and then enters the cytochrome  $b_6f$  complex, which transports the electrons directly to the photon-absorbing pigment, P700, of PSI. Once released from P700, the electron migrates through a chain of chlorophyll and ferredoxin molecules and then is either returned cyclically to the plastoquinone pool (to translocate protons across the thylakoid membrane) or acts to reduce  $\text{NADP}^+$  in a noncyclic process by ferredoxin- $\text{NADP}^+$  reductase (Voet, 1995).



**Figure 2-1. Photosynthetic (Calvin Cycle) and Photorespiration Cycles.** Oxygenation step (photorespiration): Under low  $\text{CO}_2$  conditions,  $\text{O}_2$  reacts with RuBP to form 3-PG and 2-PG. The 2-PG is hydrolyzed to glycolate and is partially hydrolyzed to form  $\text{CO}_2$ . Carboxylation step (photosynthesis):  $\text{CO}_2$  interacts with RuBP to form two molecules of 3-PG. 3-PG, in both cycles, generates sugars. Arrows indicate multiple steps, the first of which is catalyzed by Rubisco.

The dark reaction, or the reaction utilizing products from the light reaction to synthesize carbohydrates, consists of a metabolic cycle that can take one of two pathways. The substrate, in this case Ribulose-1,5-bisphosphate (RuBP), can be used in either the photosynthetic (carboxylation) pathway or photorespiration (oxygenation) pathway (Figure 2-1). Rubisco is the enzyme that can take the substrate (RuBP) down either the carboxylation or oxygenation pathway.  $\text{CO}_2$  and  $\text{O}_2$  levels determine the pathway of RuBP, high  $\text{CO}_2$  leading to the carboxylation (Calvin cycle) or low  $\text{CO}_2$  (high  $\text{O}_2$ ) resulting in activation of the oxygenation pathway (photorespiration).

### Structure of Rubisco

Rubisco is found at very high levels in leaf tissues (comprising up to 50% of leaf proteins) and is considered the most abundant protein in the biosphere (Voet, 1995; Buchanan, 1973; and Horecker, 1973). It was the extremely high concentration of Rubisco that allowed Wildman and Bonner (Horecker, 1973) to identify Rubisco seven years before the discovery of the Rubisco carboxylation reaction. It has been hypothesized (Horecker, 1973) that the reason for the presence of such a large amount of the enzyme is compensation for low catalytic activity.

Rubisco from higher plants is a large enzyme (consisting of eight large subunits, encoded by chloroplast DNA, and eight small subunits, encoded by a nuclear gene) with a molecular mass of 560,000 daltons (Buchanan, 1973). The 16 subunit enzyme has the symmetry of a square prism with the catalytic site situated in the large subunit. The chloroplast *rbcL* gene encodes the 55,000-dalton large subunit, which contain two domains.

The first domain consists of a  $\beta$  sheet and the second domain is a folded  $\alpha\beta$  barrel enzyme, containing the enzyme's active site for  $\text{CO}_2$  fixation (Voet, 1995 and Buchanan, 1973). The 15,000-dalton small subunit is encoded by the *rbcS* gene and the function of this subunit has yet to be determined (Buchanan, 1973).

### Rubisco and the Calvin Cycle

The photosynthetic  $\text{CO}_2$  fixation pathway, or the Calvin cycle (Figure 2-1), is used by plants to incorporate  $\text{CO}_2$  into carbohydrates. The primary step of the Calvin cycle takes three molecules of Ribulose-5-phosphate (Ru5P) and produces three molecules of Ribulose-1,5-bisphosphate (RuBP), with the help of ATP and phosphoribulokinase (phosphorylation event). The RuBP, then, is carboxylated (using three molecules of  $\text{CO}_2$  and Rubisco) to yield two molecules of 3-phosphoglycerate (3-PG). The 3-PG molecules continue on in the Calvin cycle to produce several photosynthetic intermediates. Sugar intermediates produced by this pathway include: Erythrose-4-phosphate, Sedoheptulose-1-7-bisphosphate, Sedoheptulose-7-phosphatate, Fructose-1,6-bisphosphate, Fructose-6-phosphate (exported from the cell as sucrose), Xylulose-5-phosphate, and Ribose-5-phosphate. The final step of this pathway consists of either an isomerase step converting Ribose-5-phosphate or an epimerase step converting Xylulose-5-phosphate back to the precursor molecule of the pathway, Ru5P. The Ru5P, in turn, is converted to RuBP and the cycle begins again (Voet, 1995).

### Rubisco and Photorespiration

At low  $\text{CO}_2$  and high  $\text{O}_2$ , however, photorespiration occurs, not photosynthesis (Figure 2-1). In this cycle,  $\text{O}_2$  competes with  $\text{CO}_2$  as a substrate for Rubisco. The oxygenase

activity of photorespiration, similar to the Calvin cycle, begins with Ru5P as the initial substrate. Also, similar to the Calvin cycle, Ru5P undergoes phosphorylation to produce RuBP. It is at this step that entry into the photorespiratory cycle may occur. Rather than a carboxylation event, RuBP is subject to an oxygenation reaction, catalyzed by Rubisco. In the oxygenation reaction,  $O_2$  reacts with RuBP to form one molecule of 3-PG and two molecules of 2-phosphoglycolate (2-PG). The 3-PG (rather than two 3-PG molecules produced by the Calvin cycle) is the continuing substrate in sugar production whereas the 2-PG is concurrently hydrolyzed to form one molecule of 3-PG and  $CO_2$ , both of which are used immediately for the regeneration of RuBP via the Calvin cycle (Douce, 1999 and Voet, 1995).

Photorespiration is often referred to as a “wasteful, nonessential pathway” as it requires great energy and overall leads to a loss of  $CO_2$  that could potentially be fixed by photosynthesis (Spreitzer, 1993; and Tolbert, 1973). In a healthy plant, photosynthetic  $CO_2$  fixation usually dominates over photorespiratory  $CO_2$  release. However, it has been determined that when temperatures increase, photorespiration may be especially high, matching or surpassing the photosynthetic rate, as the oxygenase activity of Rubisco increases more rapidly with increased temperatures than does the carboxylase activity (Tolbert, 1973 and Voet, 1995). Loss of  $CO_2$  is one limiting factor photorespiration poses on the growth of many plants. An additional problem with photorespiration is energy used. The recycling of 2-PG into 3-PG is a costly reaction, requiring a large machinery consisting of 16 enzymes and six translocators (Douce, 1999) as well as a great deal of energy. RuBP oxygenation does not lead to significantly more energy consumption than does RuBP carboxylation, but

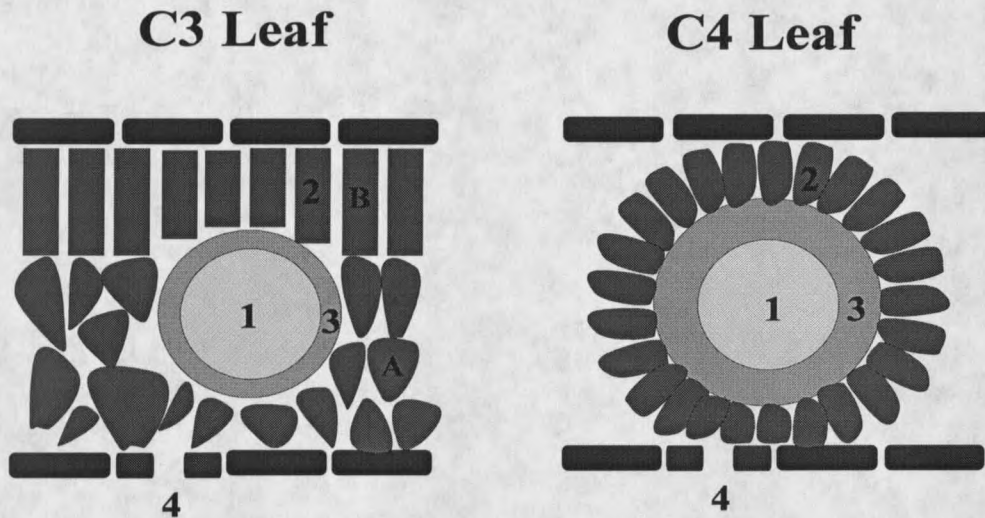


rather considerably less energy is conserved (Ogren, 1984). Questions arising from these studies include: what is the function of photorespiration (is its purpose to alleviate the damage that oxygen radicals can cause in green leaves) and how can photorespiration be lessened or stopped to increase plant productivity?

### Plant Strategies for Reducing Photorespiration

Three distinct modes of carbon fixation have evolved. Plants utilizing C<sub>3</sub>, C<sub>4</sub> and CAM photosynthetic pathways all possess differences in adaptation and productivity (Brown, 1993). The C<sub>3</sub> pathway, the most primitive of the groups, is utilized by most plants, including grains and most flowering plants (Campbell, 1993). The C<sub>3</sub> plants use the method of CO<sub>2</sub> fixation described above, using the Calvin cycle to incorporate CO<sub>2</sub> into organic material, with high CO<sub>2</sub> (within the air space of the leaf) concentrations leading to photosynthesis and low CO<sub>2</sub> concentrations leading to photorespiration.

To avoid the photorepiration problems of C<sub>3</sub> plants, C<sub>4</sub> plants have evolved a unique solution. C<sub>4</sub> plants are derived from the more primitive C<sub>3</sub> photosynthetic pathway and contain complex modifications that confer advantages in certain environments (Brown, 1993). C<sub>4</sub> plants are typically found in tropical regions and include such plants as sugar cane, corn, and weeds (Campbell, 1993 and Voet, 1995). C<sub>4</sub> plants differ from C<sub>3</sub> plants in two distinct ways: 1) anatomy and 2) photosynthetic metabolic pathway. The leaves of C<sub>4</sub> cycle plants (Figure 2-2) consist of mesophyll cells arranged in a concentric fashion around the vascular bundle sheath cells. The bundle sheath cells, in turn, envelop the fine veins of the vascular

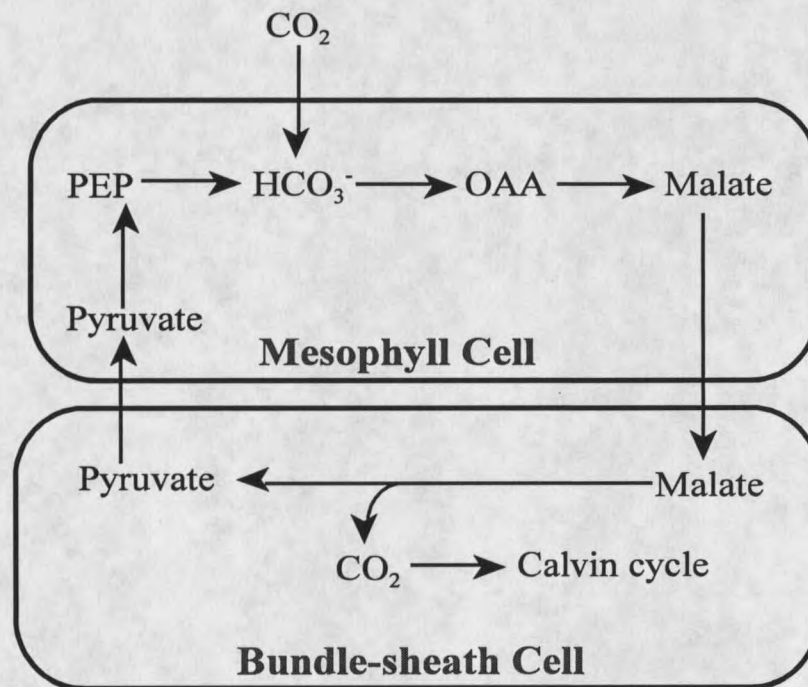


**Figure 2-2. C3 and C4 Leaf Anatomy.** 1. Vein (consists of vascular bundles with the xylem and phloem); 2. Mesophyll Cells (in C3 plants, mesophyll includes spongy mesophyll cells (A) and palisade mesophyll cells (B)); 3. Bundle-sheath Cells; and 4. Stoma (interchange of gases for respiration and photosynthesis occur here).

tissue (Brown, 1993 and Voet, 1995). Thus, we see cooperation of  $\text{CO}_2$  assimilation between two distinct cell types, mesophyll cells and bundle sheath cells.

The second difference between C3 and the more recently evolved C4 plants is the use of a distinct metabolic pathway in C4 plants (Figure 2-3). The C4 metabolic cycle concentrates  $\text{CO}_2$  from the atmosphere in specialized photosynthetic cells (mesophyll and bundle sheath cells), thus preventing photorespiration (Voet, 1995; Brown, 1993; and Ogren, 1984).  $\text{CO}_2$  from the atmosphere enters mesophyll cells and is fixed by a reaction with phosphoenolpyruvate (PEP) to yield oxaloacetate (OAA). The OAA, in turn, is reduced to malate which is then exported to the bundle-sheath cells. Once in the bundle-sheath cells, the malate is de-carboxylated to form  $\text{CO}_2$ , pyruvate, and NADPH. The concentrated  $\text{CO}_2$ , at

this point, proceeds to the Calvin cycle (in the bundle-sheath cells), where photosynthesis resumes and  $\text{CO}_2$  is fixed by Rubisco. As the rate of PEP carboxylation and malate decarboxylation exceeds the rate of RuBP carboxylation, the  $\text{CO}_2$  concentration is maintained and RuBP oxygenation is decreased, minimizing or altogether eliminating photorespiration. Thus, C4 plants have evolved a more efficient mechanism of photosynthesis, although at the expense of consuming five ATPs per  $\text{CO}_2$  fixed versus three ATPs required by the Calvin cycle. In addition to decreasing photorespiration, the C4 cycle also results in higher



**Figure 2-3. C4 Metabolic Pathway for Concentrating  $\text{CO}_2$ .** Mesophyll Cell:  $\text{CO}_2$  enters the mesophyll cells and is converted to oxaloacetate (OAA). The OAA is converted to malate and exported to the bundle-sheath cells (via plasmodesmata). Bundle-sheath Cell: Malate is decarboxylated by an NADPH-specific malate dehydrogenase to form  $\text{CO}_2$  and pyruvate. The  $\text{CO}_2$  then enters the Calvin cycle where 3-PG is produced.

photosynthetic rates, more efficient use of water (resulting from higher photosynthetic rates and stomatal conductance), and more efficient use of nutrients (due to increased photosynthetic rates and a lower investment of synthesizing nitrogen in Rubisco) (Brown, 1993).

CAM (Crassulacean acid metabolism) plants have evolved a different version of the C4 pathway. CAM plants are often desert-dwelling succulent, or water storing, plants (Voet, 1995). Rather than separating CO<sub>2</sub> acquisition and the Calvin cycle by tissues (as in the C4 pathway), CAM plants separate CO<sub>2</sub> acquisition and the Calvin cycle by time (Voet, 1995). In C3 and C4 plants, the stomata open during the day to acquire CO<sub>2</sub>. Desert-dwelling plants, however, avoid this mechanism as the opening of stomata during high temperatures would result in excess water loss. Rather, CAM plants absorb CO<sub>2</sub> at night and store it in the Crassulacean acid metabolism pathway in the form of malate, similar to the C4 pathway (Figure 2-3). The stored malate is broken down to CO<sub>2</sub>, during the day, and is then free to enter the Calvin cycle.

### Regulation of Rubisco

Rubisco must first be activated by a number of factors before it can participate in either photorespiration or photosynthesis. The Rubisco holoenzyme is assembled in a catalytically inactive form and is activated by the binding of an activator (CO<sub>2</sub>) followed by addition of Mg<sup>2+</sup> on the Rubisco large subunit, near the active site (Lorimer, 1980). Also necessary for Rubisco activation is catalysis of the above process by Ribulose-bisphosphate carboxylase activase, or Rubisco activase, in an ATP-dependent reaction (Streusand, 1987).

Kinetic and physical studies have established that the activation of Rubisco involves the addition of  $\text{CO}_2$  and  $\text{Mg}^{2+}$  (Lorimer, 1980). The activator  $\text{CO}_2$  was found to bind, in the form of a carbamate, to the  $\epsilon$ -amino group of the lysine 201 residue of the large subunit of Rubisco (Lorimer, 1980). The carbamylation reaction is promoted by a slightly alkaline pH and completes the formation of a binding site for the  $\text{Mg}^{2+}$ , also necessary for activation (Garrett, 1999). Carbamylation is spontaneous, but cannot occur if the active site of Rubisco is occupied by a sugar phosphate (such as Ru5P). An increase of Ru5P concentration (in the cell) is actually known to inhibit carboxylation, as Ru5P binds tightly to decarbamylated and carbamylated sites. Thus, after  $\text{Mg}^{2+}$  binding, the release of Ru5P from the active site of Rubisco is necessary. This action is mediated by Rubisco activase, a regulatory protein that binds in an ATP-dependent reaction. Rubisco activase couples the energy of ATP hydrolysis to the release of inhibitory sugar phosphates (Ru5P) bound to carbamylated or decarbamylated Rubisco active sites.

#### The Role of Rubisco Activase

Rubisco activase was first isolated from spinach extracts (Werneke, 1988) and is known to increase the proportion of active Rubisco by promoting dissociation of tightly bound sugar-phosphates (including Ru5P) from the active site of Rubisco (Salvucci, 1993). Without Rubisco activase, the sugar-phosphates would not be removed, resulting in either blocked carbamylation or inhibition of catalysis by binding to the active site of carbamylated Rubisco (Salvucci, 1993). Hydrolysis of ATP by Rubisco activase is required for activating Rubisco, presumably to supply energy for inducing conformational changes that alter the

binding affinity of Rubisco for sugar-phosphates (Streusand, 1987).

Rubisco activase has been studied in a variety of higher plants, including *Arabidopsis thaliana*, barley, spinach, cucumber, and tobacco (Werneke, 1989; Rundle, 1991; Werneke, 1988; Preisig-Muller, 1992; and Qian, 1993), which has provided important structural information. Rubisco activase has been shown, in every plant except barley and maize, to be comprised of two polypeptides of 46 and 42 kDA (Werneke, 1989). Further evidence showed that the two polypeptides arise from alternative splicing of Rubisco activase transcripts, and thus are the result of a single gene (Werneke, 1989). The resulting gene was designated Rca. The two Rca polypeptides were found to differ only in their carboxy-terminal amino acid regions and are both independently capable of catalyzing Rubisco activation *in vitro* (Zielinski, 1989). In maize leaves, however, only the smaller polypeptide is present (Salvucci, 1987) and in barley, Rubisco activase has been found to be encoded by two closely linked genes, RcaA and RcaB (Rundle, 1991).

Additionally, concerning barley rubisco activase, the RcaA gene is equivalent to the Rca gene other plant Rubisco activase genes examined and likewise displays alternate splicing seen within the Rubisco activase genes of previously identified species (Rundle, 1991). In previously identified dicots (*Arabidopsis* and spinach), splicing occurs in the last intron of the Rca gene, thus producing two distinct gene products. Barley, however, is unique in that it contains three gene products. One gene product from the RcaB gene (BRcaB) and two gene products from the RcaA gene (BRcaA). Since the same splicing pattern is found in the monocot (barley) studied, as in the Rca genes of the dicots examined, it is possible that, in regard to Rubisco activase, some component of pre-mRNA splicing has been conserved

among both monocots and dicots and thus pressure to preserve the structure of the polypeptides derived from the RcaA gene is strong (Rundle, 1991).

### Structural Features of Rubisco Activase

Comparison of known Rubisco activase sequences, along with structural studies, reveal conserved structural features of higher plant Rubisco activase. Conserved structural features include: 1) nucleotide binding regions, 2) DNA sequence elements similar to light regulated elements seen in other genes, 3) ATP binding domains (ATP  $\gamma$ -phosphate and adenine binding domain for ATP), 4) role of the N-terminal (specifically amino acid Trp16), and 5) role of Asp amino acid at positions 231 and 174.

As Rubisco activase requires ATP for activity, it is not surprising that multiple regions of nucleotide binding (P-loops) have been found. Two regions of nucleotide binding have been found and are suggested to represent nucleotide-binding sites in ATP-utilizing enzymes (Werneke, 1988). Within the first region, the lysine residue 171 of spinach Rubisco activase was found to interact with phosphate groups of bound nucleotides, and has been found to be essential for both Rubisco activase and ATPase activities. Point mutation studies, replacing Lys171 with Arginine, Isoleucine, or Threonine, resulted in nonfunctional Rubisco activase and ATPase activity (Shen, 1991). This conserved lysine residue was also found in additional diverse ATP utilizing enzymes including adenylate kinase, bacterial ATPase, and various oncogenes (Shen, 1991).

Zielinski et al., 1989, showed that the accumulation of Rca mRNA in barley was stimulated by white light. DNA sequences resembling light-regulatory elements in the 5'

flanking regions of RcaB and RcaA of barley have been identified (Rundle, 1991). RcaA contains putative BoxII/BoxIII-like sequences at positions -87, -110, and -185, previously identified as sites of interaction of DNA binding proteins in the promoter of the RbcS-3A (Rubisco small subunit) gene of pea (Green, 1988) which have been shown to convey light responsiveness to chimeric genes in transgenic plants (Rundle, 1991). Another light responsive element, at positions -36, -106, -167, and -177 (Mitra, 1989), has been found that is similar to the element that is responsible for the light regulated expression of an *Arabidopsis Cab* gene. The *Cab* gene, or chlorophyll a/b-binding protein, is a major structural component of the light-harvesting complex. The gene is primarily induced to express in the photosynthetic cells by light. Thus making it a potentially significant element in Rubisco activase genes that may lead to a better understanding the photosynthetic reaction. A third possible light regulation element, found at position -132, is a G-box sequence motif thought to be important in light regulated expression of RbcS (Rubisco small subunit) in a number of organisms (Giuliano, 1988). Finally, the RcaA gene also has been found to contain an AT-1-like sequence at positions -580, to -620. This sequence has previously been identified to facilitate interaction of light regulated *Cab* and *RbcS* gene promoters with transcription factors in a phosphorylation-dependent manner (Rundle, 1991). Similar to the RcaA gene, the RcaB gene also contains putative Box II/III sequences and two possible G-box sequence motifs (Rundle, 1991).

Three ATP binding domains have also been found in tobacco (Salvucci, 1993 and Salvucci, 1994). From these studies, it was determined that the N-terminus (55 amino acids) was not required for ATP binding and that two ATP $\gamma$ BP modified peptides (indicative of



ATP binding domains) were present, one adjacent to the P-loop (residues 155-164) and the second from a distinct region of the protein (residues 210-220) (Salvucci, 1993). Two additional ATP binding sites were subsequently found in tobacco (Salvucci, 1994). One of the sites was the adenine binding domain for ATP (residues 68-74), and the second site found was not a nucleotide binding domain, but a domain with the ability to bind with high affinity to adenine nucleotides containing an azido substitution on the base (residues 10-14).

Site directed mutagenesis of the N-terminus led to the discovery that: 1) deletion of the first 50 amino acids from the N-terminus led to the inability of Rubisco activase to activate Rubisco, and 2) modification of Trp16 (previously identified by Salvucci, 1994 to be conserved among all Rubisco activase species) with Ala or Cys also led to an inability of Rubisco activase to activate Rubisco (van de Loo, 1996). It has thus been hypothesized that Trp16 may either mediate a conformational change within the activase protein that is indirectly required for interaction with Rubisco or may directly mediate interaction between the two proteins (van de Loo, 1996).

Finally, the roles of Asp231 and Asp174 in binding of ATP/Mg<sup>2+</sup> have been ascribed (van de Loo, 1998) via point mutation studies in tobacco. From these studies, it was determined that mutations of Asp231 and Asp174 did not prevent ATP binding, but did prevent aggregation into higher-ordered oligomers and consequently prevented the catalysis of ATP hydrolysis, regardless of binding to ATP. Thus, point mutations at Asp231 and Asp174 prevent the enzyme from undergoing the conformational changes that commit the enzyme to aggregation, followed by catalysis (van de Loo, 1998).

We have set out to identify and characterize Rubisco activase in an additional

monocot, wheat. In this report, we present evidence showing that wheat Rubisco activase is encoded by two genes, WRcaB and WRcaA. These genes are closely related to each other (with the addition of a C-terminal tail in WRcaA) and to the other Rubisco activase species previously characterized. The additional monocot sequences, provided by this study, support previous evidence that specific structural features of Rubisco activase are conserved among monocots and dicots and that these may be linked to the function of the protein.

### Materials and Methods

#### Detection of Three Partial Wheat Rubisco Activase "B" Genes from Wheat Genomic DNA

A 500ng/ $\mu$ l sample of Wheat genomic DNA (Chinese Spring ecotype) was obtained from L. Talbert, Montana State University. The DNA was RNase treated by adding 1 $\mu$ l of boiled RNase to the sample and incubating at 37°C for one hour. The DNA was electrophoresed on a 1% agarose gel with 0.5 $\mu$ g/ml ethidium bromide and was visualized by UV<sub>254</sub> light to check quality.

The resulting RNase-treated DNA was used as template in PCR reactions designed to pull out a fragment of the Rubisco activase (Rca) gene in wheat. The following primers were made (exact and degenerate) from the known nucleotide sequences of Rubisco activase genes (BRcaB and BRcaA) in barley: Bacact-1 (for)-exact, 5'-GAC GAC CAG CAG GAC ATC AC-3'; Bacact-2 (for)-exact, 5'-TAC GAG TAC ATC AGC CAG GG-3'; Bacact-2B-29 (rev)-exact, 5'-CTT GCG CAC CTC GTC GTC GTA-3'; Cact-1 (for)-deg., 5'-GAY GAY CAR CAR GAY ATH AC-3'; Cact-2 (rev)-deg., 5'-RAA RAA RTC DAT NSW YTG NCC; and Cact-2B (rev)-deg., 5'-TTN CKN ACY TCR TCR TCR TA-3'. IUB group codes, to signify degenerate sequences, are as follows: R= A+G, Y= C+T, K= G+T, S= G+C, W= A+T, H= A+T+C, D= G+A+T, and N= A+T+G+C.

First round PCR was carried out using the following primer sets: Bacact-1 (for)-exact and Bacact-2B-29 (rev)-exact, Bacact-2 (for)-exact and Bacact-2B-29 (rev)-exact, Cact-1 (for)-deg. and Cact-2 (rev)-deg., Cact-1 (for)-deg. and Cact-2B (rev)-deg., Bacact-2 (for)-

exact and Cact-2 (rev)-deg., and Bacact-2 (for)-exact and Cact-2B (rev)-deg. A 50 $\mu$ l PCR reaction containing 0.5 $\mu$ g wheat DNA, 250 $\mu$ M of each primer, 1 unit of elongase enzyme (Life Technologies), 0.1mM dNTPs, 1.5mM Mg<sup>++</sup>, in buffer B (provided by manufacturer) was made for each of the above primer sets. PCR was carried out on a Perkin Elmer DNA thermal cycler 480, with a 94°C denaturation step for 2 minutes followed by 30 cycles of 94°C for 1 minute, 48°C for 1 minute, and 68°C for 2 minutes. PCR products from Bacact-1 (for)-exact/Bacact-2B-29 (rev)-exact and Cact-1 (for)-deg./Cact-2B (rev)-deg. were used as template in round 2, nested PCR.

Round two, nested PCR, was set up similar to round one PCR with the exception of different primer sets and a change in initial template concentration. Precisely 5 $\mu$ l of PCR product from primer sets 1) Bacact-1 (for)-exact/ Bacact-2B-29 (rev)-exact and 2) Cact-1 (for)-deg./Cact-2B (rev)-deg were used as template. The primer sets used to nest, into both of the above PCR products, were: Bacact-2 (for)-exact and Bacact-2B-29 (rev)-exact, Bacact-2 (for)-exact and Cact-2B (rev)-deg., Bacact-2 (for)-exact and Cact-2 (for)-deg., and Cact-1 (for)-deg. and Cact-2 (rev)-deg. The primer set Bacact-2 (for)-exact and Cact-2 (rev)-deg. produced a band that closely matched the expected size for Rubisco activase, as estimated from the known barley sequence. The band was gel purified, cloned, and sequenced. The entire PCR product was electrophoresed on a 0.8% agarose gel and the desired band was removed and washed via the pulp-spin method. After ethanol precipitation, the pellet was dried and resuspended in 10 $\mu$ l of water. 1 $\mu$ l of the resulting DNA was re-checked for purity on a 0.8% agarose gel. A 3.5 $\mu$ l aliquot of gel-purified DNA was added in the ligation reaction with the pGem-T vector (according to the manufacturers instructions)

and transformed into JM109 competent cells (Promega). The clones were screened for inserts of the desired size and used directly in sequencing reactions.

### Sequencing of Clones

Sequencing reactions were set up for the ABI PRISM 377 as described by the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer). For each 20  $\mu$ l reaction, 0.2  $\mu$ g of DNA, 3.2 pmole primer, and 4  $\mu$ l BIGDYE mix (A-Dye Terminator, C-Dye Terminator, G-Dye Terminator, T-Dye Terminator, dTTP, dATP, dCTP, dTTP, Tris-HCl (pH 9.0), MgCl<sub>2</sub>, thermal stable pyrophosphatase, and AmpliTaq DNA Polymerase) were mixed and placed in the GeneAmp PCR system 9600. Thermal cycling was as follows: 96°C for 15 seconds, 50°C for 10 seconds, and 60°C for 4 minutes. This cycle was repeated 25 times and extension products were then purified. PCR products were precipitated with 75% isopropanol and washed with 70% ethanol. The resulting pellet was sequenced by the Montana State University sequencing.

The clones were sequenced using universal primers SP6 (5'-ATT TAG GTG ACA CTA TAG-3') and T7 (5'-TAA TAC GAC TCA CTA TAG GG-3'). One additional primer, an internal primer, was needed to sequence the entire 651 bp fragment, Wheat RA-1 (for), 5'-KCA TCA ACC CCA TCA TKA TKA KCK-3' (K=G + T).

### Cloning of the Complete Rubisco Activase "B" Gene (WRcaB) from a Wheat cDNA Library

A lambda ZAPII wheat cDNA library was generously donated by Dr. Karen Browning, University of Texas. The library contained pBluescript KSII as the cloning vector and an *E. coli* XL1-Blue host cell line. The host cells were prepared by placing a colony into a flask

containing 50ml LB, 0.1g maltose, and 0.5mL 1M  $\text{MgSO}_4$  with shaking overnight at 37°C. The cells were centrifuged and resuspended in 25ml cold 10mM  $\text{MgSO}_4$ .

The library was first titrated to determine phage particle concentration. A series of dilutions were set up in SM buffer (5.8g NaCl, 2.0g  $\text{MgSO}_4/7\text{H}_2\text{O}$ , 50ml 1M Tris pH 7.5, 5mL 2% gelatin, and 1L of water) to make 10-fold serial dilutions ranging from 10 to  $10^6$  fold diluted. The above dilutions ( $10\mu\text{l}$ ) were then mixed with  $200\mu\text{l}$  XL1-Blue host cells and incubated at 37°C for 15 minutes. They were then added to 3mL top agarose (0.75% agarose in NZCYM media) and poured onto NZCYM plates. The plates were incubated overnight at 37°C and resulting plaques were counted and used in the following equation to determine titer:  $\text{pfu/ml} = \# \text{ plaques}/\mu\text{l used} \cdot \text{dilution factor} \cdot 1000\mu\text{l/ml}$ .

The wheat cDNA library was screened by PCR. The primers used for PCR screening were Bacact-2 (for)-exact and Bacact-2B-29 (rev)-exact, previously shown to be successful in detecting the presence of the WRcaB gene. To begin, six  $10^6/\mu\text{l}$  ( $10^6$  pfu/100mM) plates were prepared as described above in the titering protocol. After plaque formation, 2ml SM buffer was added to each plate and placed at 4°C overnight. The SM buffer was recovered and a drop of chloroform was added to each microfuge tube to inhibit bacterial growth. The resulting plaque/SM buffer mix was used as template for PCR.  $2\mu\text{l}$  plaque/SM buffer mix aliquot was first heated at 100°C for 5 minutes to denature the phage protein coat. The first round PCR reaction was comprised of  $2\mu\text{l}$  denatured plaque/SM mix as template,  $250\mu\text{M}$  primers, 0.1mM dNTPs, 1 unit of *Tfl* polymerase, and 1.5mM concentration of  $\text{MgSO}_4$  in *Tfl* buffer. PCR conditions included a 2 minute denaturation step carried out at 94°C followed by 40 cycles of 94°C for 1 minute, 55°C for 1 minute, and 68°C for 2 minutes. The sample

showing the band of the correct size was then used as template to make the NZCYM plates in the second dilution series.

The second dilution series consisted of adding 5 $\mu$ l phage stock (which produced a positive signal in round one PCR) to make three 10<sup>4</sup> pfu/plate and three 10<sup>3</sup> pfu/plate dilutions. As before, 1 $\mu$ l of each dilution was added to host cell and incubated, poured into top agarose and incubated overnight on NZCYM plates. From the second dilution plates, plaques in groups of 100 were isolated using the small end of a disposable pipette and placed in 500 $\mu$ l SM buffer overnight. A drop of chloroform was also added to inhibit bacterial growth. The six samples were then used as template in a PCR reaction. Each 50 $\mu$ l PCR reaction contained the same components as the first round screening, with the exception of template. PCR conditions were also the same, with the exception of 20 cycles rather the 40 cycles as before. The phage stock that produced a positive PCR result was used as template in the third dilution series.

The plates for the third dilution series were made the same as for the second dilution series, three 10<sup>4</sup>/plate and three at 10<sup>3</sup>/plate samples. After incubating overnight, plaques were isolated in groups of 10 and placed into 100 $\mu$ l of SM buffer overnight at 4°C. PCR was identical to the second dilution series PCR. The phage stock producing a positive result was then used in the final dilution series, the isolation of phage from individual plaques.

Plates were made from the desired phage stock from the third dilutions series as stated above. From the plates, individual plaques were isolated and placed in 100 $\mu$ l SM buffer. The plaques were PCR screened (as above) and the individual plaque containing wheat Rca was identified.

### *In vivo* Excision of cDNA from Phage and Sequencing

The *in vivo* excision of the individual phage lysate was achieved by adding 25 $\mu$ l individual phage lysate in SM buffer to 200 $\mu$ l XL1-Blue host cells and 1 $\mu$ l R408 helper phage (Stratagene). This mixture was incubated 15 minutes at 37°C and added to 2.5mL LB broth. After an additional 3 hour incubation at 37°C, with shaking, the sample was heated at 70°C for 20 minutes to kill the host cells, centrifuged at 5K RPM for 5 minutes (to remove cell debris), and the resulting Phagemid stock (supernatant) was recovered. After diluting the phagemid stock (1:100), 5 $\mu$ l was added to 200 $\mu$ l host cells and incubated at 37°C for 15 minutes. A total of 25 $\mu$ l was plated on an LB-amp plate and incubated overnight. Plasmids were prepared from the resulting colonies via the Wizard Plus SV Minipreps DNA Purification System Protocol (Promega) as per the manufacturers directions. Purified plasmids were electrophoresed on a 1% agarose gel, to check for inserts, and plasmids with inserts were sequenced.

For initial sequencing, the universal primers M13 forward (5'-CCC AGT CAC GAC GTT GTA AAA CG-3') and M13 reverse (5'-AGC GGA TAA CAA TTT CAC ACA GG-3') were used to confirm sequence. To sequence the entire 2,044 bp gene in both directions, additional internal primers were used. Internal primers used to sequence the entire clone were: WRA1 (for), 5'-GAC CAC CTT CCT CGG GAA GAA-3'; WRA2 (for), 5'-GCA TGG AGA AGT TCT ACT GG-3'; WRA5 (rev), 5'-TTC ACC GTG TAC TGC GTC GTC C-3'; WRA6 (rev), 5'-AAG GTG TCC ACC AGC CTC A-3'; WRA7 (rev), 5'-CTC GCT GAG GTA CTT GTC GG-3'; WRA8 (rev), 5'-GAT GTC GTA AGC GAG ACC-3'; and WRA9 (for), 5'-TCC AGG AGC AGG AGA ACG-3'.



The resulting sequences were aligned by overlapping using the program Sequencher (v. 3.1.1) and the entire WRcaB gene was submitted to GenBank, accession number AF251264.

#### Cloning of the Complete Wheat Rubisco Activase "A" Gene (WRcaA) from a cDNA Library

To screen the cDNA library for a potential WRcaA gene by PCR, a primer was made from the 3' terminal barley sequence, specific to BRcaA (5'-CGT CGC TCC TTG CCG TTG G-3'). This primer was used in conjunction with other primers (that anneal either to WRcaB or WRcaA) and resulting PCR analysis showed WRA2 (for) and Barley RA-2 (rev) was the only set to produce a band estimated from the BRcaA gene data to be the correct size. PCR conditions, to test primer sets, were as follows: 3  $\mu$ l of the wheat cDNA library was added to 7  $\mu$ l SM buffer and heated at 100°C for 5 minutes prior to use. After heating, the sample was spun down and 1  $\mu$ l was used as template in each reaction. Along with template, 1 unit of *Tfl* polymerase, 0.1mM dNTPs, 250  $\mu$ M primers, and 1.5mM MgSO<sub>4</sub> and *Tfl* buffer were added. PCR conditions included a 2 minute denaturation step at 94°C followed by 30 cycles of 94°C for 1 minute, 48°C for 1 minute, and 68°C for 2 minutes. The different primer sets tested were: WRA1 (for) and Barley RA-2 (rev), WRA2 (for) and Barley RA-2 (rev), WRA3 (for) and Barley RA-2 (rev), Bacact-2 (for)-exact and Barley RA-2 (rev), and Bacact-1 (for)-exact and Barley RA-2 (rev). As the primer set WRA2 (for) and Barley RA-2 (rev) was the only set to produce a band of the expected size, they were chosen for use in PCR screening of the cDNA library.

The DNA fragment, produced by WRA2 (for) and Barley RA-2 (rev), was sequenced

to confirm it encoded Rubisco activase. The 507 bp band of interest was gel purified (via pulp-spin method), cloned and sequenced. A 3.5  $\mu$ l aliquot of the gel purified product was used in the ligation reaction during cloning with the pGem-T vector system (Promega), along with 5  $\mu$ l 2X buffer, 0.5  $\mu$ l pGem vector and 1  $\mu$ l of T4 DNA ligase. The ligation reaction was then used in the transformation process into JM109 competent cells (Promega). The clones were screened for plasmids containing the 507 bp inserts and the purified plasmids were used directly in sequencing reactions (as described above) using primers T7 and M13 (reverse).

The PCR screening of the wheat cDNA library for the entire WRcaA gene was similar to the screening protocol seen above for detection of the WRcaB gene. The primers used for PCR screening the cDNA library were WRA2 (for) and Barley RA-2 (rev). Plates were made as described for the screening of the WRcaB gene and similarly three rounds of screening were performed, resulting in the identification of an individual phage containing the insert of interest. The *in vivo* excision of the individual phage lysate is described above (*In vivo* Excision of cDNA from Phage and Sequencing).

#### Multiple Sequence Alignment of Three Partial WRcaB Genes

The three partial WRcaB genes, identified by PCR as seen above, were aligned using the program Clustal W (v.1.74) (Higgins, 1991) with gap openings and extension penalties set at 30 and all other parameters set at default. The Blossom series of similarity matrices was chosen and minor adjustments were made by hand.

#### Sequence Analysis of WRcaA and WRcaB

As with the partial WRcaB genes, WRcaA and WRcaB genes were aligned using

Clustal W (v. 1.74) (Higgins, 1991): All parameters were set as above, and GenBank accession numbers for sequences used in the alignment were as follows: BRcaA1, M55449; BrcaB, M55449; BRcaA2, M55447; WRcaA, (unpublished fragment); WRcaB, AF251264.

## Results and Discussion

### Identification of Three Partial Rubisco Activase "B" Genes from Wheat

To begin the identification and characterization of Rubisco activase in wheat, we screened wheat genomic DNA, by PCR, using exact and degenerate primers derived from the barley BRcaB gene. Of the six different primer sets used, two separate samples contained PCR products estimated to be of the correct size. The two samples were then used in nested PCR to further deduce which sample contained the Rubisco activase gene. Nested PCR reaction were performed on both samples above using two different primer sets. One primer set (Bacact-2 (for)-exact and Cact-2 (rev)-deg.) successfully produced a product estimated to be the correct size in both PCR samples. Sequencing both PCR products resulted in two distinct 208 amino acid (626 nucleotide) fragments of the WRcaB gene. The two fragments, WRcaB1 and WRcaB2, differed at 17 nucleotide positions (these differences did not result in any amino acid substitutions). To further check these results, and rule out sequencing error, additional PCR reactions were performed and products were cloned and sequenced. Sequence analysis of eight total clones revealed three distinct fragments, WRcaB1, WRcaB2 and WRcaB3 (Figure 2-4). This finding, however, is not surprising as the wheat genome is hexaploid (AABBDD) and gene variations from the three genomes have previously been noted (Sallares, 1995).

### Identification of the WRcaB Gene in Wheat

After identification of the partial WRcaB fragments, we set out to clone the full-length

|     |                                                                     |     |       |     |       |        |
|-----|---------------------------------------------------------------------|-----|-------|-----|-------|--------|
| 1   | ATGCCCTTGCACACACCCCGATGCGGTCCTCCGGCGGGTGGCGGCCCGCAGTAAGAACTTCTCCATG | 1   | ..... | 1   | ..... | WRCAB2 |
| 1   | .....                                                               | 1   | ..... | 1   | ..... | WRCAB3 |
| 1   | .....                                                               | 1   | ..... | 1   | ..... | WRCAB1 |
| 61  | CGGCCGTCGGCGGATGAGGGCGCGGTACAGCGGTCCGAGAAAGTCCGTTGCCGGTGACGATGATG   | 61  | ..... | 61  | ..... | WRCAB2 |
| 61  | .....                                                               | 61  | ..... | 61  | ..... | WRCAB3 |
| 61  | .....                                                               | 61  | ..... | 61  | ..... | WRCAB1 |
| 121 | GGCAGCGGTGGGTTCTCTCTCTTGTGTTGTACATCCCGGGGAGGCTGCACGTTGGTGGCGCG      | 121 | ..... | 121 | ..... | WRCAB2 |
| 121 | .....                                                               | 121 | ..... | 121 | ..... | WRCAB3 |
| 121 | .....                                                               | 121 | ..... | 121 | ..... | WRCAB1 |
| 181 | TCCGCGATGTTTCATCAAGGTGGCGTTCCACCATCTGTTGTTCTTCCACCGGTGTACTGCGTCTC   | 181 | ..... | 181 | ..... | WRCAB2 |
| 181 | .....                                                               | 181 | ..... | 181 | ..... | WRCAB3 |
| 181 | .....                                                               | 181 | ..... | 181 | ..... | WRCAB1 |
| 241 | CCGCCCATCCGCCCGCGCGCGCGCGGTCCAGGTCGTTGATGAAAGAGGCACACATCTTGCCC      | 241 | ..... | 241 | ..... | WRCAB2 |
| 241 | .....                                                               | 241 | ..... | 241 | ..... | WRCAB3 |
| 241 | .....                                                               | 241 | ..... | 241 | ..... | WRCAB1 |
| 301 | TTCTTGATTAATGTCGGCGCGCGCTCCGGTACCTCTCGCGGATCAGCTTGGCGCGCTCGCCG      | 301 | ..... | 301 | ..... | WRCAB2 |
| 301 | .....                                                               | 301 | ..... | 301 | ..... | WRCAB3 |
| 301 | .....                                                               | 301 | ..... | 301 | ..... | WRCAB1 |
| 361 | GCCTTGGCCGCTCTCCAGCTCTCCGGCGCTCATCATGATGGGTTGATGCCCATCTTGCGC        | 361 | ..... | 361 | ..... | WRCAB2 |
| 361 | .....                                                               | 361 | ..... | 361 | ..... | WRCAB3 |
| 361 | .....                                                               | 361 | ..... | 361 | ..... | WRCAB1 |

|        |     |                                                              |
|--------|-----|--------------------------------------------------------------|
| WRcaB2 | 421 | AACACCAGCTCGCACTGGAACGACTTGCCCTGTCCCTTGCCTCCCCAGATGCCCAGGATG |
| WRcaB3 | 421 | .....                                                        |
| WRcaB1 | 421 | .....                                                        |
| WRcaB2 | 481 | AGAGGGACCTTGATGTTGGGGAGCGTCATGAAGTTCTTGGCGAGGTGGACGATGAGCTTG |
| WRcaB3 | 481 | ..... <b>T</b> .....                                         |
| WRcaB1 | 481 | ..... <b>A</b> ..... <b>T</b> .....                          |
| WRcaB2 | 541 | TCCATGAATGCCGGGGCGATGTAGAGCCCGTCCATGGTGTGTCGAAGTCGTACTTCCGC  |
| WRcaB3 | 541 | .....                                                        |
| WRcaB1 | 541 | ..... <b>G</b> ..... <b>T</b> .....                          |
| WRcaB2 | 601 | AGGCCCTGGCTGATGTACTCGTAAAT                                   |
| WRcaB3 | 601 | .....                                                        |
| WRcaB1 | 601 | .....                                                        |

**Figure 2-4. Three Different WRcaB Gene Fragments.** Dots indicate amino acids identical to those seen in the complete WRcaB2 sequence (above). Numbers at the left indicate the nucleotide number of each gene.

M A S A F S S T V G A P  
 GAGCGGGGAGAGAACAATCGACACGATGGCTTCTGCTTTCTCGTCCACCGTTGGAGCTCCGG  


---

 A S T P T T F L G K K V K K Q A G A L N Y  
 CGTCGACCCCGACCACCTTCCTCGGGAAGAAGGTGAAGAAGCAGGCCGGTTCGTTGAACTAC  
  
 Y H G G N K I N N R V V R A M A A K K E L  
 TACCATGGTGGCAACAAGATCAACAATAGGGTGGTCAGGGCCATGGCGGCCAAAAAGGAACT  
  
 D E G K Q T D A D R W K G L A Y D I S D D  
 TGACGAGGGCAAGCAGACCGATGCCGATCGGTGGAAGGGTCTCGCTTACGACATCTCCGATG  
  
 Q Q D I T R G K G I V D S L F Q A P M G D  
 ACCAGCAGGACATCACGAGGGGGAAAGGCATCGTGGACTCCCTGTTCCAGGCCCCCATGGGC  
  
 G T H E A I L S S Y E Y I S Q G L R K Y  
 GACGGCACCCACGAGGCCATCCTGAGCTCCTACGAGTACATCAGCCAGGGCCTGCGCAAGTA  
  
 D F D N T M D G L Y I A P A F M D K L I V  
 CGACTTCGACAACACCATGGACGGGCTGTACATCGCCCCGGCGTTCATGGACAAGCTCATCG  
  
 H L A K N F M T L P N I K V P L I L G I W  
 TCCACCTCGCCAAGAACTTCATGACACTCCCCAACATCAAGGTCCCTCTCATCCTGGGTATC  
  
 G G K G Q G K S F Q C E L V F A K M G I  
 TGGGGAGGGCAAGGGACAGGGCAAGTCGTTCCAGTGCGAGCTGGTGTTCGCCAAGATGGGCAT  
  
 N P I M M S A G E L E S G N A G E P A K L  
 CAACCCCATCATGATGAGCGCCGGAGAGCTGGAGAGCGGCAACGCCGGCGAGCCGGCCAAGC  
  
 I R Q R Y R E A A D I I K K G K M C C L F  
 TGATCCGGCAGAGGTACCGCGAGGCTGCCGACATTATCAAGAAGGGCAAGATGTGCTGCCTC  
  
 I N D L D A G A G R M G G T T Q Y T V N  
 TTCATCAACGACCTGGACGCCGGCGCGGGGCGGATGGGCGGGACGACGACGAGTACACGGTGAA  
  
 N Q M V N A T L M N I A D A P T N V Q F P  
 CAACCAGATGGTGAACGCCACCCTGATGAACATCGCGGACGCGCCACCAACGTGCAGTTCC  
  
 G M Y N K E E N P R V P I I V T G N D F S  
 CGGGGATGTACAACAAGGAGGAGAACCCACGCGTGCCCATCATCGTACCGGCAACGACTTC  
  
 T L Y A P L I R D G R M E K F Y W A P T  
 TCGACGCTGTACGCGCCCCTCATCCGGGACGGCCGCATGGAGAAGTTCTACTGGGCGCCCAC  
  
 R E D R I G V C K G I F R T D N V P D E A  
 CCGGGAGGACCGCATCGGCGTGTGCAAGGGCATCTTCCGCACCGACAACGTCCCCGACGAGG  
  
 V V R L V D T F P G Q S I D F F G A L R  
 CCGTGGTGAGGCTGGTGGACACCTTCCGGGGCAGTCCATCGACTTCTTCGGCGCGCTGCGG



```

A R V Y D D E V R K W V G E I G V E N I S
GCGCGGGTGTACGACGACGAGGTGCGCAAGTGGGTGCGGAGATCGGCGTCGAGAACATCTC

K R L V N S R E G P P T F D Q P K M T I E
CAAGCGGCTCGTCAACTCCAGGGAGGGGCCGCGACGTTGACCAGCCCAAGATGACCATCG

K L M E Y G H M L V Q E Q E N V K R V Q
AGAAGCTCATGGAGTACGGCCACATGCTGGTCCAGGAGCAGGAGAACGTGAAGCGCGTGCAG

L A D K Y L S E A A L G Q A N D D A M A T
CTCGCCGACAAGTACCTCAGCGAGGCGGCGCTCGGCCAAGCCAACGACGACGCCATGGCGAC

G A F Y G K @
CGGCGCCTTCTACGGCAAGTAGAAAGTCCTATACTTAAGATGCATGCGTGCATGCATGCACT

```

---

```

ATATATATGCTGGAATATTTTGGACTCGAATCCACTCAAACTTTTTGATTCATCGCCGCACA

```

---

```

GGTAGCCCAATTGTAAGTCTCTTCTTGAAAGAGAATCCAATTGTATATGTTTTCGTATTGCC

```

---

```

TATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

```

**Figure 2-5. Amino Acid and Nucleotide Sequence of Rubisco Activase from Wheat (WRcaB).** The full-length WRcaB gene is 432 amino acids (1296 nucleotides) in length. The 5' and 3'-UTR's are underlined in black.

WRcaB gene. Using primers made from sequence information from the 626 nucleotide WRcaB fragment, we screened a wheat cDNA library by PCR. The screening resulted in the identification of one positive clone. The clone was sequenced on both strands and determined to be the full-length WRcaB gene (Figure 2-5). The 3'-UTR was determined to be 168 nucleotides and the 5'-UTR was found to be 24 nucleotides. Primer extension, however, was necessary to determine full 5'-UTR.

The WRcaB mRNA clone encoded a polypeptide of 432 amino acids and was found to be 96% identical to the BRcaB gene. As we had only screened a cDNA library, there was no information regarding WRcaB introns. Further analysis will include screening a genomic



library to obtain the entire gene. This will allow us characterize the Rca genes and polypeptide sequences. This will include making a physical map of the wheat Rca locus, looking at introns/exon arrangements (BrcaB is known to contain the longest intron, 12.2 kb, yet identified in plants) and to look at possible alternate splicing (Rundle, 1991).

#### Identification of the WRcaA Gene in Wheat

Since we were unsure if a WRcaA gene existed in wheat (barley is the only plant determined to date to have both an RcaA and RcaB gene), we first conducted a PCR experiment on the cDNA library to see if we could detect a WRcaA gene. The PCR was done using a BRcaA specific primer (as BRcaA has a 36 amino acid C-terminal extension not found in the BRcaB gene) and an N-terminal primer non-specific to RcaA or RcaB. The PCR results indicated a product estimated to be the correct size, and sequencing confirmed the fragment was indeed WRcaA gene. Using the primer set that had successfully pulled out the WRcaA fragment, we screened the cDNA library by PCR and isolated the WRcaA gene. The clone isolated from the cDNA library, unfortunately, was a partial clone, lacking approximately 99 amino acids from the N-terminus.

The WRcaA gene fragment (Figure 2-6) was sequenced on both strands and found to be 1089 nucleotides, encoding a polypeptide of 363 amino acids. As stated above, approximately 99 amino acids from the N-terminus were missing, but the C-terminus (determined in barley Rubisco activase to be the A gene specific region) was present, clearly indicating we had found a second gene encoding Rubisco activase in wheat (WRcaA). Comparison of the two wheat genes showed that they were 82% identical. The difference

S S Y E Y V S Q G L R K Y D F D N T M G G  
GCTCCTACGAGTACGTACGCCAGGGCCTGCGCAAGTACGACTTCGACAACACCATGGGCGGG

L Y I A P A F M D K L S V H L A K N F M T  
CTGTACATCGCCCCGGCGTTCATGGACAAGCTCAGCGTCCACCTCGCCAAGAACTTCATGAC

I P N I K I P L I L G I W G G K G Q G K  
AATCCCCAACATCAAGATCCATCTCATCTTGGGTATCTGGGGAGGCAAGGGTCAAGGAAAAT

S F Q C E L V F A K M G I N P I M M S A G  
CCTTCCAGTGTGAGCTTGTCTTCGCCAAGATGGGCATCAACCCAATCATGATGAGTGCCGGA

E L E S G N A G E P A K L I R Q R Y R E A  
GAGCTGGAGAGTGGCAACGCCGAGAGCCAGCCAAGCTGATCAGGCAGCGGTACCGTGAGGC

A D M I K K G K M C C L F I N D L D A G  
TGCCGACATGATCAAGAAGGGTAAGATGTGCTGCCTCTTCATCAACGATCTTGACGCCGGTG

A G R M G G T T Q Y T V N N Q M V N A T L  
CGGGTCGGATGGGCGGGACCACACAGTACACCGTCAACAACCAGATGGTGAACGCCACCCTA

M N I A D A P T H V Q L P G M Y N K R E N  
ATGAACATCGCCGATGCCCCACCACGTGCAGCTTCCAGGCCATGTACAACAAGCGGGAGAA

P R V P I I V T G N D F S T L Y A P L I  
CCCACGCGTGCCCATCATCGTCACCGGCAACGACTTCTCGACGCTGTACGCGCCCCTCATCC

R D G R M E K F Y W A P T R E D R I G V C  
GGGACGGCCGCATGGAGAAGTTCTACTGGGCGCCACCCGGGAGGACCGCATCGGCGTGTGC

K G I F Q T D N V S D E S V V K I V D T F  
AAGGGTATCTTCCAGACCGACAATGTGAGCGACGAGTCCGTCGTCAAGATCGTCGACACCTT

P G Q S I D F F G A L R A R V Y D D E V  
CCCAGGACAATCCATCGACTTTTTTCGGTGCTCTGCGTGCTCGGGTGTACGACGACGAGGTGC

R K W V T S T G I E N I G K K L V N S R D  
GCAAGTGGGTGACCTCTACCGGTATCGAGAACATTGGCAAGAAGCTTGTGAACCTCGCGGGAC

G P V T F E Q P K M T V E K L L E Y G H M  
GGACCAGTGACGTTTGAGCAGCCAAAGATGACAGTCGAGAAGCTGCTAGAGTACGGGCACAT

L V Q E Q D N V K R V Q L A D T Y M S Q  
GCTCGTCCAGGAGCAGGACAATGTCAAGCGTGTGCAGCTTGCTGACACCTACATGAGCCAGG

A A L G D A N Q D A M K T G S F Y G K G A  
CAGCTCTGGGTGATGCTAACCAGGATGCGATGAAGACTGGTTCCTTCTACGGCTAAGGGGCA

Q Q G T L P V P A G C T D Q T A K N F A  
 CAGCAAGGTACTTTGCCTGTGCCGGCAGGATGCACCGACCAGACTGCCAAGTCAGCTCCCC

P T A R S D S C L Y T F  
 AACGGGAGGAGAGTGATCGACGTTTCCATCCATTT

**Figure 2-6. Amino Acid and Nucleotide Sequence of Rubisco Activase from Wheat (WRcaA).** Partial WRcaA gene is 363 amino acids (1089 nucleotides) in length.

between WrcaB and WrcaA, for the most part, was due to the 37 amino acid C-terminal tail, found only in the WRcaA gene. The genes were otherwise highly similar. A comparison of the WRcaA gene with BRcaA showed them to be 95% identical, similar to the WRcaB and BRcaB genes which were determined to be 96% identical.

#### Multiple Sequence Alignment and Sequence Analysis of Rubisco Activase Genes

After identification of both the WRcaA and WRcaB genes, we conducted a multiple sequence alignment to existing Rca genes of barley and wheat, including WRcaA (363 amino acid fragment), WRcaB, BRcaA1, BRcaA2 and BRcaB (Figure 2-7). The multiple sequence alignment showed a high degree of conservation of sequence identities of the encoded polypeptides between the two wheat Rca genes and when comparing the wheat Rca genes to Rca genes isolated from barley. Interestingly, the identity was greater between WRcaA/BRcaA (95%) and WRcaB/BRcaB (96%) than between the WRcaA and WRca B genes (82%). A similar comparison between the WRcaA and previously identified dicots, *Arabidopsis* (ARca) and Spinach (SRca) (Rundle, 1991), found that WRcaA was 80-82% identical to ARca and SRca. As WRcaA shows 80+% amino acid sequence identity with dicot Rca genes, but higher identity with WRcaB and barley Rca genes (an additional monocot).

This data supports Rundles (Rundle, 1991) contention that RcaA and RcaB may have been derived from a common ancestor that post-dated the monocot-dicot divergence (Rundle, 1991).

Sequence similarity between BRca genes and WRca genes is great and, not unexpectedly, structural motifs previously found in barley and tobacco (Werneke, 1988, Salvucci, 1993, and Rundle, 1991) also appear to be present in the wheat genes. For example, two potential ATP binding domains are found conserved among barley and wheat Rca genes and are represented by dotted lines in Figure 2-7 at residues 155-164 and 210-220 in Figure 2-7. The ATP binding domains have thus been found in all deduced Rubisco activase polypeptide sequences, consistent with the observation that activation of Rubisco activase *in vitro* requires ATP (Rundle, 1991).

In this initial study, we have determined that two genes encode Rubisco activase in wheat and that both genes have a high degree of similarity to previously identified Rubisco activase genes. Further studies may look at alternative splice patterns in wheat Rca genes, with the intent of determining physiological significance for two forms of Rubisco activase. As Rubisco activase activity has important implications in photosynthesis, the alternative splicing may impart an advantage or be essential to photosynthesis in higher plants. Further analysis of structure and function of Rubisco activase may one day enable us to modulate the activation level of Rubisco leading to a far improved ratio of the Calvin cycle to photorespiration. As photorespiration drains away as much as 50% of the carbon fixed by the Calvin cycle (Cambell, 1995), reducing photorespiration by modulating Rubisco activase and (thus Rubisco) activity may lead to increased crop yields and food supplies.

|        |     |                                                               |
|--------|-----|---------------------------------------------------------------|
| BRcaA1 | 1   | MAAFSSTVGAPASTPTNFLGKKLKKQVTSAVNYHGKSSKANRFT-VMAA--ENIDEKR-   |
| BRcaA2 | 1   | .....-.....--.....-                                           |
| WRcaA  | 1   |                                                               |
| WRcaB  | 1   | ..S.....T.....V...AGALNY...GNKIN..VVRA...KK.LDEG.QT           |
| BRcaB  | 1   | ..S.....I.....V.-----NY...GNKMKS.VVR....KK.LDQG.QT            |
| BRcaA1 | 57  | NTDKWKGLAYDISDDQQDITRGKGIVDSLFFQAPTGHGTHEAVLSSYEYVSQGLRKYDFDN |
| BRcaA2 | 57  | .....T.....D.....                                             |
| WRcaA  | 1   | .....                                                         |
| WRcaB  | 61  | DA.R.....M.D....I.....I.....                                  |
| BRcaB  | 55  | DA.R.....M.D....I.....I.....                                  |
| BRcaA1 | 117 | TMGGFYIAPAFMDKLVVHLSKNFMTLPNIKIPLILGIWGGKGQGSFQCELVFAKMGINP   |
| BRcaA2 | 117 | .....                                                         |
| WRcaA  | 18  | ...L.....S...A....I.....                                      |
| WRcaB  | 121 | ..D.L.....I...A.....V.....                                    |
| BRcaB  | 115 | ..D.L.....I...A.....V.....                                    |
| BRcaA1 | 177 | IMMSAGELESGNAGEPAKLIRQRYREAADMIKKGKMCCLFINDLDAGAGRMGGTTQYTVN  |
| BRcaA2 | 177 | .....-----                                                    |
| WRcaA  | 78  | .....                                                         |
| WRcaB  | 181 | .....I.....                                                   |
| BRcaB  | 175 | .....G-.....I.N.....                                          |
| BRcaA1 | 237 | NQMVNATLMNIADAPTNVQLPGMYNKRENPRVPIVVTGNDFSTLYAPLIRDGRMEKFYWA  |
| BRcaA2 | 219 | -----                                                         |
| WRcaA  | 138 | .....H.....I.....                                             |
| WRcaB  | 241 | .....F.....E.....I.....                                       |
| BRcaB  | 234 | .....E.....I.....                                             |



|        |     |                                                              |
|--------|-----|--------------------------------------------------------------|
| BRcaA1 | 297 | PTRDDRIGVCKGIFQTDNVSDSVVKIVDTFPGQSIDFFGALRARVYDDEVKRWVGSTGI  |
| BRcaA2 | 239 | .....C.....                                                  |
| WRcaA  | 198 | ...E.....T...                                                |
| WRcaB  | 301 | ...E.....R...P..A..RL.....EI.V                               |
| BRcaB  | 294 | ...E.....R...P..A..RL.....EI.V                               |
|        |     |                                                              |
| BRcaA1 | 357 | ENIGKRLVNSRDGPVTFEQPKMTVEKLLEYGHMLVQEQDNVSRVQLADTYMSQAALGDAN |
| BRcaA2 | 299 | .....                                                        |
| WRcaA  | 258 | .....K.....                                                  |
| WRcaB  | 361 | ...S.....E..P..D.....I..M.....E.....K.L.E....Q..             |
| BRcaB  | 354 | ...S.....E..P..D.....I..M.....E.....K.L.E....Q..             |
|        |     |                                                              |
| BRcaA1 | 417 | QDAMKTGSFYGKGAQQGTLVPVEGCTDQNAKNYDPTARSDDGSCLYTF             |
| BRcaA2 | 359 | .....KGAQQGTLVPVEGCTDQNAKNYDPTARSDDGSCLYTF                   |
| WRcaA  | 318 | .....A.....T...FA.....SCLYTF                                 |
| WRcaB  | 421 | D...A..A....                                                 |
| BRcaB  | 414 | D.....A....                                                  |

**Figure 2-7. Multiple Sequence Alignment of Wheat and Barley Rca Genes.** Dots indicate amino acids identical to those seen in the complete BRcaA1 sequence (above). Dashes indicate an missing amino acid in that position in the alignment. Numbers at the left indicate the nucleotide number of each gene. The dotted line indicates ATP binding motifs.

## REFERENCES CITED

- Baldauf, S.L. and Palmer, J.D. 1993. Animals and fungi are each other's closest relatives: Congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci.* Vol. 90:11558-11562
- Blackburn, E.H., Greider, C.W., Henderson, E., Lee, M.S., Shampay, J., and Shippen-Lentz, D. 1989. Recognition and elongation of telomeres by telomerase. *Genome*. 31:553-560.
- Blackburn, E.H., Greider, C.W. 1995. *Telomeres*. Cold Spring Harbor Laboratory Press.
- Brown, H.R. and Bouton, J.H. 1993. Physiology and genetics of interspecific hybrids between photosynthetic types. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 44:435-56.
- Bryan, T.M., and Cech, T.R. 1999. Telomerase and the maintenance of chromosome ends. *Curr. Opin. Cell Biol.* 11:318-324.
- Buchanan, B.B. and Schurmann, P. 1973. Ribulose 1,5-Diphosphate carboxylase: A regulatory enzyme in the photosynthetic assimilation of carbon dioxide. *Current Topics in Cellular Regulation*. Vol.7, 1-20.
- Collins, K. and Gandhi, L. 1998. The reverse transcriptase component of the *Tetrahymena* telomerase ribonucleoprotein complex. *PNAS*. Vol.95, 8485-8490.
- Doolittle, R.F., Feng, D., Tsang, S., Cho, G., and Little, E. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science*. Vol. 271:470-477.
- Douce, R. and Neuburger, M. 1999. Biochemical dissection of photorespiration. *Curr. Opin. in Plt. Biol.* 2:214-222.
- Fajkus, J., Kovarik, A., and Kralovics, R. 1996. Telomerase activity in plant cells. *FEB*. 391: 307-309.
- Fajkus, J., Fulnecková, J., Hulanova, M., Berkova, K., Riha, K., and Matyasek, R. 1998. Plant cells express telomerase activity upon transfer to callus culture, without extensively changing telomere lengths. *Mol Gen Genet.* 260:470-474.

- Felsenstein, J. 1989. PHYLIP (Phylogeny Inference Package) version 3.2. *Cladistics* 5: 164-166.
- Fitzgerald, M.S., Riha, K., Gao, F., Ren, S., McKnight, T.D., and Shippen, D.E. 1999. Disruption of the telomerase catalytic subunit gene from *Arabidopsis* inactivates telomerase and leads to a slow loss of telomeric DNA. *PNAS*. Vol.96, No.26: 14813-14818.
- Friedman, K.L., and Cech, T.R. 1999. Essential functions of amino-terminal yeast telomerase catalytic subunit revealed for viable mutants. *Genes Dev.* 13(21):2863-74.
- Gall, J.G. 1995. In (eds.) C.W. Greider and E.H. Blackburn, *Telomeres*. Cold Spring Harbor Press, Cold Spring Harbor, NY.
- Garrett, R.H. and Grisham, C.M. 1999. *Biochemistry: Second ED*. Saunders College Publishing. 709-741.
- Giuliano, G., Pichersky, E., Malik, V.S., Timko, M.P., Scolnik, P.A., and Cashmore, A.R. 1988. An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. *Proc. Natl. Acad. Sci.* Vol. 85:7089-7093.
- Green, P.J., Yong, M., Cuozzo, M., Kano-Murakami, Y., Silverstein, P., and Chua, N. 1988. Binding site requirements for pea nuclear protein factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the *rbcS-3A* gene. *EMBO J.* Vol.7, No. 13:4035-4044.
- Greider, C.W., and Blackburn, E.H. 1989. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature*. 377:331-337.
- Hebsgaard, S.M., Korning, P.G., Tolstrup, N., Engelbrecht, J., Rouze, P., and Brunak, S. 1996. Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information. *Nucleic Acids Res.* Vol.24, No.17: 3439-3452
- Heller, K., Kilian, A., Piatyszek, M.A., and Kleinhofs, A. 1996. Telomerase activity in plant extracts. *Mol Gen Genet.* 252: 342-345.
- Higgins, D.G., Bleasby, A.J., and Fuchs, R. 1991. ClustalV: improved software for multiple sequence alignment. *CABIOS* 8: 189-191.
- Kilian, A., Stiff, C., and Kleinhofs, A. 1995. Barley telomeres shorten during differentiation but grow in callus culture. *PNAS*. Vol.92: 9555-9559.



- Killan, A., Heller, K., and Kleinhofs, A. 1998. Developmental patterns of telomerase activity in barley and maize. *Plant Molecular Biology*. 37:621-628.
- Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T.R. 1997. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science*. Vol.276: 561-567.
- Lorimer, G.H. and Mizziorko, H.M. 1980. Carbamate formation on the  $\epsilon$ -amino group of a lysyl residue as the basis for the activation of ribulosebiphosphate carboxylase by CO<sub>2</sub> and Mg<sup>2+</sup>. *Biochemistry*. 19:5321-5328.
- McEachern, M.J. and Blackburn, E.H. 1996. Cap-prevented recombination between terminal telomeric repeat arrays (telomere CPR) maintains telomeres in *Kluyveromyces lactis* lacking telomerase. *Genes and Development*. 10:1822-1834.
- McKnight, T.D., Fitzgerald, M.S., and Shippen, D.E. 1997. Plant telomeres and telomerases. A review. *Biochemistry*. Vol.62, No.11: 1224-1231.
- Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., and Cech, T.R. 1997. Telomerase catalytic subunit homologs from fission yeast and human. *Science*. Vol.277: 955-959.
- Nakamura, T.M. and Cech, T.R. 1998. Reversing time: origin of telomerase. *Cell*. Vol.92, 587-590.
- Nugent, C.I., and Lundblad, V. 1998. The telomerase reverse transcriptase: Components and regulation. *Genes Dev*. 12:1073-1085.
- Ogren, W.L. 1984. Photorespiration: Pathways, regulation, and modification. *Ann. Rev. Plant Physiol*. 35:415-42.
- Oguchi, K., Liu, H., Tamura, K., and Takahashi, H. 1999. Molecular cloning and characterization of AtTERT, a telomerase reverse transcriptase homolog in *Arabidopsis thaliana*. *FEBS Letters*. 457:465-469.
- Page, R.D.M.. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12: 357-358.
- Preisig-Muller, R. and Kindle, H. 1992. Sequence analysis of cucumber cotyledon ribulosebiphosphate carboxylase/oxygenase activase cDNA. *Biochimica et Biophysica Acta*. 1171:205-206.

- Qian, J., and Rodermel, S.R. 1993. Ribulose-1,5-bisphosphate carboxylase/oxygenase activase cDNAs from *Nicotiana tabacum*. *Plant Physiol.* 102:683-684.
- Regad, F., Matthieu, L., and Lescure, B. 1994. Interstitial telomeric repeats within the *Arabidopsis thaliana* genome. *J. Mol. Biol.* 239, 163-169.
- Richards, E.J., and Ausubel, F.M. 1988. Isolation of a higher Eukaryotic telomere from *Arabidopsis thaliana*. *Cell.* Vol.53, 127-136.
- Richards, E.J., Goodman, H.M., and Ausubel, F.M. 1991. The centromere region of *Arabidopsis thaliana* chromosome 1 contains telomere-similar sequence. *Nucleic Acids Research.* Vol.19, No.12, 3351-3357.
- Riha, K., Fajkus, J., Siroky, J., and Vyskot, B. 1998. Developmental control of telomere lengths and telomerase activity in plants. *The Plant Cell.* Vol.10: 1691-1698.
- Rundle, S.J. and Zielinski, R.E. 1991. Organization and expression of two tandemly oriented genes encoding ribulosebisphosphate carboxylase/oxygenase activase in barley. *The Journal of Biol. Chem.* Vol. 266, No. 8: 4677-4685.
- Sallares, R., Allaby, R.G., and Brown, T.A. 1995. PCR-based identification of wheat genomes. *Mol Ecol.* 4(4):509-14
- Salvucci, M.E., Werneke, J.M., Ogren, W.L., and Portis, A.R. 1987. Purification and species distribution of rubisco activase. *Plant Physiol.* 84:930-936.
- Salvucci, M.E., Rajagopalan, K., Sievert, G., Haley, B.E., and Watt, D.S. 1993. Photoaffinity labeling of ribulose-1,5-bisphosphate carboxylase/oxygenase activase with ATP  $\gamma$ -benzophenone. *Journal of Biol. Chem.* Vol. 268, No. 19:14239-14244.
- Salvucci, M.E., Chavan, A.J., Klein, R.R., Rajagopalan, K., and Haley, B.E. 1994. Photoaffinity labeling of the ATP binding domain of rubisco activase and a separate domain involved in the activation of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochemistry.* 33:14879-14886.
- Shen, J.B., Orozco, E.M., and Ogren, W.L. 1991. Expression of the two isoforms of spinach ribulose 1,5-bisphosphate carboxylase activase and essentiality of the conserved lysine in the consensus nucleotide-binding domain. *The Journal of Biol. Chem.* Vol. 266, No. 14:8963-8968.
- Shippen, D.E., and McKnight, T.D. 1998. Telomeres, telomerase and plant development. *Trends in Plant Science.* Vol.3, No.4, 126-129.

- Spreitzer, R.J. 1993. Genetic dissection of rubisco structure and function. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 44: 411-34.
- Stavenhagen, J.B. and Zakian, V.A. Internal tracts of telomeric DNA act as silencers in *S. cerevisiae*. *Genes Dev.* 8(12):1411-1422.
- Streusand, V.J., and Portis, A.R. 1987. Rubisco activase mediates ATP-dependent activation of ribulose biphosphate carboxylase. *Plant Physiol.* 85:152-154.
- Tolbert, N.E. 1973. Glycolate Biosynthesis. *Current Topics in Cellular Regulation.* Vol.7, 21-50.
- van de Loo, F.J., and Salvucci, M.E. 1996. Activation of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) involves rubisco activase Trp16. *Biochemistry.* 35:8143-8148.
- van de Loo, F.J. and Salvucci, M.E. 1998. Involvement of two aspartate residues of rubisco activase in coordination of the ATP  $\gamma$ -phosphate and subunit cooperativity. *Biochemistry.* 37:4621-4625.
- Voet, D. and Voet, J.G. 1995. *Biochemistry: Photosynthesis.* John Wiley & Sons, INC. New York, 626-661.
- Werneke, J.M., Chatfield, J.M., and Ogren, W.L. 1989. Alternative mRNA splicing generates the two ribulosebisphosphate carboxylase/oxygenase activase polypeptides in spinach and *Arabidopsis*. *Plant Cell.* Vol.1:815-825.
- Werneke, J.M., Chatfield, J.M., and Ogren, W.L. 1988. Catalysis of ribulosebisphosphate carboxylase/oxygenase activation by the product of a rubisco activase cDNA clone expressed in *Escherichia coli*. *Plant Physiol.* 87:917-920.
- Werneke, J.M., Zielinski, R.E., and Ogren, W.L. 1988. Structure and expression of spinach leaf cDNA encoding ribulosebisphosphate carboxylase/oxygenase activase. *Proc. Natl. Acad. Sci.* Vol.85:7870-791.
- Zakian, V.A. 1997. Structure, function, and replication of *Saccharomyces cerevisiae* telomeres. *Annu. Rev. Genet.* 30:141-172.
- Zakian, V.A. 1997. Life and cancer without telomerase. *Cell.* Vol.91:1-3.

Zentgraf, U. 1995. Telomere-binding proteins of *Arabidopsis thaliana*. Plant Molecular Biology. 27:467-475.

Zentgraf, U., Hinderhofer, K., and Kolb, D. 2000. Specific association of a small protein with the telomeric DNA-protein complex during the onset of leaf senescence in *Arabidopsis thaliana*. Plant Mol. Biol. 42:429-438.

## APPENDIX A

### Multiple Sequence Alignment of Ten Full-length TERT Genes.

Conserved motifs are denoted by black underbars. Black boxes represent an amino acid conserved within at least five or more proteins and the grey boxes represent similar amino acids. Species abbreviations are as follows: ham., hamster; m., mouse; h., human; Ea., *Euplotes aediculatus*; Ot., *Oxytricha trifallax*; Tt., *Tetrahymena thermophila*; At., *Arabidopsis thaliana*; Sp., *Schizosaccharomyces pombe*; Sc., *Saccharomyces cerevisiae*; and Ca., *Candida albicans*.

ham. 1 ---MPRAPRCRA-----VRALLRSQYQVWPLATFVRRIGPEGRLVQPGDPKVFRTL  
 m. 1 ---MTRAPRCPA-----VRSLLRSRYREVWPLATFVRRIGPEGRLVQPGDPKIYRTL  
 h. 1 ---MPRAPRCRA-----VRSLLRSYREVWPLATFVRRIGPEGRLVQPGDPAAFRAL  
 Ea. 1 ---MEVDVDNQADNHG-----IHSALKTCCEIKEAKTYSWIQKVIRCRNQSQS--  
 Ot. 1 MSAKKPVQSRLNIGNPTIPVTSNRSAPKPPGQPQFVNOEKKQSQNTTTGAFRSNQNN  
 Tt. 1 MQKINNINNNKQMLTR-----KEDLLTVLKQISALKYVSNIEFLLATEKIVQT-----  
 At. 1 ---MPRKPFRHVP-----EILWRLFGNRARNLNDALVDLIPN--RNIQPEQCRCRGQG  
 Sp. 1 ---MTEHHTPKSRILRFLENQY-V-----YLCTLNDYVQLVLRGSPASSY---SNICER  
 Sc. 1 MKILFEFIQDKLIDLQTN-----TYKEN-----LKCG-----  
 Ca. 1 -----MTVKVNEKATLLQYVLDN-----TSN-----DVPLL-

ham. 51 -----VARCIVCVPMDSQPPPADLSFHQVSSLKELVARV--VQRLCERG  
 m. 51 -----VAQCIVCMHWSQPPPADLSFHQVSSLKELVARV--VQRLCERN  
 h. 51 -----VAQCIVCVPMDSQPPPADLSFHQVSSLKELVARV--LQRLCERG  
 Ea. 47 -----HYKDLIEDIKIFAQTNIVATPRDYNEEDFKVIARKE--VFSTGLM  
 Ot. 61 ANSGGNGNFELDDLHLALKSCNEIGSAKTLFCFFMELQKINDKIPSKKNT---ELIKNDP  
 Tt. 50 -----SELDTQFQELTTTIIASEQNLVENYKQKYNQPN---FSQITI-  
 At. 49 -----CLGSSDKP-AFLLRSDDPHYRKLHRCFVVLHEQTPTPLDLS  
 Sp. 48 L-----RSDVQTS--SIFLHSTVVGFD--KPDEGVQFSSPKCS---QS  
 Sc. 30 -----HFNGIDEILTTFCALPNSRKIALPCL---PGDLSHKAVIDHCII  
 Ca. 27 -----PSLKEYMETVLVYKSIKRPLPAI--RPQESFDEFMKELVTRL

ham. 93 E-----RNVL  
 m. 93 E-----RNVL  
 h. 93 A-----KNVL  
 Ea. 89 IELIDKCLVELLSSSDVSDRQ-----KLQC  
 Ot. 118 QFYFCHNTILCTEQSYDEKTIEKLAKLEYKLEYKNSYIDMISKVIKELLIENKLNKLQT  
 Tt. 90 -----KQVI  
 At. 92 PTS-----WWSQ-----REIV  
 Sp. 86 ELIANVVKQMFDESFERR-----NNLL  
 Sc. 71 YLLTGELY-----NNVL  
 Ca. 67 VMEKS-----NNVI

TEL5

ham. 98 TEGFALLN-GAQQGPPMTFTTSVRSYLPN-----SVTESLRVSGAWMLLN  
 m. 98 AEGFELLN-EARGGPPMAFTSSVRSYLPN-----TVIETLRVSGAWMLLLS  
 h. 98 AEGFALLD-GARGGPPMAFTTSVRSYLPN-----TVTDALRGSGAWGLLLR  
 Ea. 114 FGFQLKGN-QLAKTHLTALSTQ-----KQYFFQDEWNQVRA  
 Ot. 178 FGYKLNN-EFGNQNHGMMQQNQDSSHNSNFMVKCDYI---NLNKQCMITKNWEKVYF  
 Tt. 94 DDSIILIG-NKQNYV-----QQIGTTTIGFYVEY---E-NINLSRQTIYSSNERNLLN  
 At. 103 ERIIEMMO-SGDCQNVICARYDKYDQS-----SPILELITSSWEFLK  
 Sp. 108 MKGFSNNH--EDFRAMHVNGVQNDLVSTFP-----NYLISILESKNWQLLLE  
 Sc. 83 TEG-YKTA-RNEDVN-----NSLFCHSA-----NNVNTLLKGAAWKMFHS  
 Ca. 76 AMG-YKTS-AMESRS-----IFTTFHSSG-----NFILTHITSHNWSTIFS

TEL5

TEL4

ham. 143 RVGDDLLVYLLARCA~~LYLL~~LVPPSCA-YQVCGSPLYQICATAETWPSVSRIYRPT-RPVGR  
 m. 143 RVGDDLLVYLLAHCA~~LYLL~~LVPPSCA-YQVCGSPLYQICATTDIWPSVSASYRPT-RPVGR  
 h. 143 RVGDDVLVHLLARCA~~LFV~~LVAPSCA-YQVCGPPLYQLGAATQARPPPHASG-PR-RRLG-  
 Ea. 150 MIGNELFRHLYKYLI~~FORT~~SEGLVQFCGNNVFDHL-----  
 Ot. 233 YLGDHLFMHLYKEYM~~IFL~~KTRDES~~LVQ~~ISGT-----NIFYCY  
 Tt. 142 IFGEEDFKYFLDFLVFTKVEQNGYLQVAG-----IVCLN  
 At. 147 RVGHDVMVYLLLLGKKH-QQVSGPPLCIKHKRTLSVHENKRKRDDNVQPPTKRQW-----  
 Sp. 153 IIGSDAMHYLLSKGST~~FEAL~~PNDNYLQISGIPLFKN-----N----  
 Sc. 121 LVGTYAFVDLLINYT~~VIQ~~FNGQ-----FFTQ  
 Ca. 115 LIGPKK~~TELL~~VNNKGFVSKV-----NGESV

## TEL4

ham. 201 NFTHLGSTH~~VRN~~SSHQEA~~WKPP~~PLPSREAKRSL~~SIT~~NRSVPPSKKARCDLAPRLEKGPY  
 m. 201 NFTNLRFLQ~~QIK~~SSSRQEAPKPLALPSRGTKRHL~~SLT~~STSVPSAKKARCYPVPRVEEGPH  
 h. 199 -----CE~~RAW~~NHSVREAGVPLGLPAPGARRRGGASASRLPLPKRPRRGAAPERTPV  
 Ea. 186 KVNDKFD---KKQKG-----  
 Ot. 269 LNEKLGR~~LQAA~~FYEGPNKNAAN-----  
 Tt. 176 QYFS--VQV~~Q~~KKWYKN-NFNMNG-----  
 At. 201 LSSAVDDCP~~DD~~SATITPIVGEDVDQHREKKT~~TKRS~~RIYLKRRRKQ~~RKV~~NFKKVD~~CN~~APC  
 Sp. 190 -VFEETVS---K-KRK-----  
 Sc. 147 IVGNRCNEP~~H~~-----LPPKWV-----Q-----  
 Ca. 141 QIFG-DVNS~~HR~~-----

ham. 261 RQAVPTPSDK-TWVPNPAKSHAVPI~~SRTT~~KEDLSSGV--KAPGLS~~RS~~SG-SVCYKH~~K~~PSST  
 m. 261 RQVLPTPSGK-SWVPSPARSPEVP---TAEKDLSSKG--KV~~S~~DL~~SL~~SG-SVCKKH~~K~~PSST  
 h. 252 GQGSWAHPGR-TRGPSDRGFCV~~VSP~~ARPAEEATSLEG--ALSGTR~~SH~~SPVGRQH~~H~~AGPP  
 Ea. 198 -----GAADMNEP-----R-----CC~~ST~~CKYNV~~KNE~~KD~~H~~FLNNIN  
 Ot. 291 -----SAAQGSNPEANDLISAEQRKI--NTAIVMK~~TH~~KYNIKA~~ADE~~SYLT~~NQ~~E  
 Tt. 197 -----KATSNNNQNNANLSNEKKQENQYI~~Y~~PEIQ~~RS~~QIFYCNHMGREGPVFKSS  
 At. 261 ITPSTNGKVSTGNDEMNLHIG--INGSLTD---FVKQAKQV~~KRN~~K~~N~~FKGLSETYSTVIP  
 Sp. 201 -----RTIETSIT-----Q-----NKSARKEVSWNSISIS~~RS~~FSIF  
 Sc. 164 -----RSSSSSAT-----AAQIKQLTEPV~~TN~~KQFLHK  
 Ca. 151 -----KAVVVS~~KYIT~~KFNVL

ham. 317 SLQSPLCQNAFQLRP-YTET~~KR~~FLYSREGG--RERLNPSF~~LLNN~~LQPSLTGARRLVEI~~IF~~  
 m. 314 SLLSPPRQNAFQLRP-FIET~~RH~~FLYSRGDG--QERLNPSF~~LLSN~~LQPNLTGARRLVEI~~IF~~  
 h. 309 STSRPPRPWDTPCPPVYAET~~KH~~FLYSSGD--KEQLRPSF~~LLSS~~LRPSLTGARRLVEI~~IF~~  
 Ea. 228 VPNNWNMKS~~R~~-----T-RIFYCTHFNRNNQFFKKHEFVS--N~~KNN~~ISAM~~DRA~~QT-IF  
 Ot. 338 KGFWD~~DQ~~IKR-----N-RLFYCAHQNR--FFQKH--IL--NSK~~IL~~SQQQIRDN-IV  
 Tt. 246 FFNYSEIKKG-----FQFKVIQEKLQGR--QFINS~~D~~---K--TKPDHPQTIIKKT-IL  
 At. 315 PNHILKT-----LRPNC~~SDS~~LLLNHIF  
 Sp. 231 YRSSYKKFKQ-----DLYFNLH-----SICDRNTV~~HMW~~LQWIF  
 Sc. 191 LNINSSSFFP-----Y---S-K-----ILPSS~~SS~~IKKLT~~D~~-IR  
 Ca. 166 YNSYSRDFSR-----F---E-M-----IRPSIQ~~TI~~LQDIL-SF

## TEL3



ham. 374 LGMRPRTSGPLCGRRLS-----KRYWQMRPLFQQLLVNHARCFYVRLLRSHCR  
 m. 371 LGSRPRTSGPLCRTHRLS-----RRYWQMRPLFQQLLVNHAEQYVRLLRSHCR  
 h. 366 LGSRPWMPGTPRRLPLRP-----ORYWQMRPLFLELLGNHAQCPYGVLLKTHCP  
 Ea. 276 TNIFRFNRIRK-----ELKDKVIEKLAAYMLEKVDFNENYYLTKSCP  
 Ot. 381 KEVFGFNVRRA-----ELKGVKMSLIEQVIVNQKKFDEKYYSKNCP  
 Tt. 291 KEYQSKNFSCQ-----EERDLFLEFTEKIVQNFHNINENYLLKKFCK  
 At. 338 GEVNVWSTTPSHGKGNCPGSI-----CLVHSLKSLKNLLGKTKSSHLKMLLDKHCP  
 Sp. 264 PRQFGLINAFQVKQLHKVIPLVSQSTVVPKRLKQVYPLIEQTAKRLHRISLSKVYNHYCP  
 Sc. 219 EALFFPTNLVKIPQ-----RLKVRINLTQKLLKRHKRLNYSILNSICP  
 Ca. 194 SGLNPGRSS-----KRYRGFKSLLSRIIAND-----KKCR

TEL3

TEL2

ham. 423 FRTAAHQVAGALNT-----TSPQRLMNLRLHSSPWQVYGFILQ  
 m. 420 FRTANQQVTDALN-----TSPPHLMDLRLHSSPWQVYGFILR  
 h. 415 LRAAVTPAAGVCAREKPGGSVAAPEE-----EDTDPRLVQLLRQHSSPWQVYGFVR  
 Ea. 318 LPENWREKQKQIENLINKTREEK-----SKYEEELFSYTTDNKCVTQFIN  
 Ot. 423 LPENWKNLKSFLDAAVSGELR-----GOVFRQLFEYQQDQROIISNFLT  
 Tt. 333 LPENYQSLKSQVKQIVQSENKANQ---Q-----SCENLFNSLYDTEISYKQITNFLR  
 At. 391 VLLQEDALKSGTTSQSSRRQKADKLPHGSSSSQTGKPKCPSVEERKLYCINDQVVSFIW  
 Sp. 324 YIDTHDDEK-----TLSYSLKPNQVFAFLR  
 Sc. 263 PLEGTVDLS-----HLSRQSPKERVLFKFI  
 Ca. 224 YDILYAKFIGTSKCNFAN-----VVSNKTEISQVIOFVL

TEL1

ham. 461 ACVGLIVEPGLWGSRHNRFFKKNVRFISLGKYDKLSLQELTWKMKVQDCRWLRSSPGN  
 m. 457 ACLCKVVSASLWGTRHNERFFKKNLKKFISLGKYGKLSLQELMWKMKVEDCHWLRSSPGK  
 h. 467 ACLRRIVEPGLWGSRHNERRELNTKKFISLGKHAKLSLQELTWKMSVRDCAWLRSSPGV  
 Ea. 363 EFFYNILPKDFLTGR-NRKNFQKKVKKYVELNKHელიHKNLLLEKINTREISWQVETSA  
 Ot. 468 EFVANVEPKNFLEK-NKKIFNKKMLQFVKFNRFESFTKISLINKFRVNEVSWLSFKCKD  
 Tt. 382 QILQNCVPNQLLGKK-NFKVELEKLYEFVQMKREENQKVLDYICFMDVDFVDFVLDKNQ  
 At. 451 AICRYIVPESLLGTTHQMRVLRKNIAWFVSRRRNEKCTVNQFLHKVVKPSDFPFARKELC  
 Sp. 349 SILVRVFPKLIWGNQRIFEILKDLLETFLKLSRYESFSLHYLMSNIKISEIEWLVGKRS  
 Sc. 289 VILQKLLPQEMEGSKKNKGKIIKNLNLISLPLNGYLPFDSLKKLRLKDFRWLFISDIW  
 Ca. 258 LVLGKLLPLDAWGGVSNKKI IKDRVVDLILLGANEKIHMDLFRGIRLKDFFKWLGRAHQI

TEL1

ham. 521 NCVPAAEHRTR-----ERILAVFLFWLMDAYVVELLRSFFYVTETTFQKNRLLFFYRKSM  
 m. 517 DRVPAAEHRLR-----ERILATFLFWLMDTYVVQLLRSFFYITESTFQKNRLLFFYRKSV  
 h. 527 GCVPAAEHRLR-----EEILAKFLHWLMSVYVVELLRSFFYVTETTFQKNRLLFFYRKSV  
 Ea. 422 KHFFYYFHEN-----IYVLWKLRLWIFEDLVVSLIRCFYFVTEQQKSYSKTYYYRKNI  
 Ot. 527 ENKKFFMNEN-----EHVFFKVLKWVFEFLAITLMRCYFYSTEKAKKEYQRIFFYRKNI  
 Tt. 441 KFTQKRKYISDK-----RKILGDLIVEIINKIVIPVLRNFYITEKHKEGSQIFYYRKPI  
 At. 511 CMVNGHELQSESIRSTQOMLCTKWISWLFLEIVKKLVHFNFYATESOGGRLNIYYRKRS  
 Sp. 409 NAKMCLSDFEKR-----KQIFAEFIYWLNSFIPILOSFFYITESSDLNRNTVYFRKDI  
 Sc. 349 FTKHNFENL-----NQIAICFISWLFRLPKI IQIFFYCTEISSVT-IVYERHDT  
 Ca. 318 SSKQDFEMTIFTLR---TAFILKGYLWVLFELLLKNILRSFWYITETSSIVSSLNYFPQYL

Motif T



ham. 575 WRRIQSIGVRHHLERVRLQELSQE-----EVRQROEAWPAMPICRLRFIPKPSG--L  
 m. 571 WSKLQSIGVRQHLEVRRLRELSQE-----EVRHHQDTWLAAMPICRLRFIPKPNG--L  
 h. 581 WSKLQSIGIRQHLLKRVQLRELSQA-----EVRQHREARPAALLTSKRLRFIPKPDG--L  
 Ea. 475 WDVIMKMSIADLKKET-LAEVQEK-----EVEEWKKS LGFAPG-KLRLIPKKT--F  
 Ot. 580 WNMIMRLSIDDLLKQN-LKQVEKK-----EMRIFCESQNFAPG-KLRLIPKGD--F  
 Tt. 496 WKLVSKLTIVKLEEN-LEKVEEK-----LLEPE--DSFQKYPQGKLRIIPKKS--F  
 At. 571 WERLISKEISKALDGYVLVDAAEA-----ES----SR--KK-LSKRFILPKANG--V  
 Sp. 464 WKLLCRPFITSMKMEA-FEKLNNEN-----NVRMD-TQKTTLPFAVIRLLIPKNT--F  
 Sc. 400 WNKLTLPFIVEYFKTYLVENNVC-----RNENS-YILSNFNHSMRIIPKKSNEF  
 Ca. 375 WKELYESWWSKYAKMTIFMTLFNNLVKMPSKIDRE-----QLPCGKIKLIPKSS--F

Motif T

Motif 1

ham. 625 RPIVNMYSY-MGTRA-FDKGKQAQHFTQCLKTLFSVINYELTKHTNLLGASVLGLNDIYRT  
 m. 621 RPIVNMYSYMGTRA-LGRRKQAQHFTQRLKTLFSMLNYERTKHPHLMGSSVLGMNDIYRT  
 h. 631 RPIVNM DYVVGART-FRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLGLDDIHRA  
 Ea. 523 RPIMTFNKKTIVNSD-RKTTKLTNTKLLNSHLMKTLKLRMFKDPFGFAVFNY-DDVMKK  
 Ot. 628 RPIMTFNKKTINQVGKFQSEMTTNNKLQTAHMLKLNLSKMEKHSFGFAVFNY-DDIMKR  
 Tt. 543 RPIMTFLRKDKQKNI-KLNLNQILMDSQLVFRNLK-DMLGQK----IGYSVFDNKQISEK  
 At. 614 R--MVLDFSSSSRS-----SQ----SLRDTHAVLKDIQLKEPDVLGSSVFDHDDFYRN  
 Sp. 512 RLITNLRKRFILIM-GSNKKMLVSTNQTLRPVASILKHLINEESSGIP-FNL-EVY-MKL  
 Sc. 450 RIITAIPCRGADEEE-FTIYKENHKNAIQPTQKILEYLR--NKRPTSFTKIYSPTQIADRI  
 Ca. 426 RVICVPIKEKEKE-FERYKEVLSPVGQILRIKLSKLR-DTYESYRASVHSSSDVAEKIS

Motif 2

ham. 683 WRTF---VLRVRTLD-PAPRM YFVKADVTGAYDAIPQDKLVEVIANMIRHPDNS-YCIHQ  
 m. 680 WRAF---VLRVRALD-QTPRM YFVKADVTGAYDAIPQGKLVEVVANMIRHSESTYCIQY  
 h. 690 WRTF---VLRVRAQD-PPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTY-CVRRY  
 Ea. 581 YEEF---VCKWKQV--GQPKLEFATMDIEKCYDSVNREKLSTFLKTTKLLS--SDFWIMT  
 Ot. 687 YENF---VQWVKQI--NSPKLYFVAMDIEKCYDNVDCERVVNFLQKSDLMD--KEYFILN  
 Tt. 597 EAEF---IERWKNK--GRPQLYVTLDIKKCYDSIDQMKLLNFFNQSDLIQDTY-FINKY  
 At. 661 LCPY---LHLRSQSGELPPLYFVADVFAFDSVDQGLLHVIVQSFLK----DEYILNR  
 Sp. 568 LTFKKDLLKIRM----FGRKKYFVRIDIKSCYDRIKQDLMFRIVKKKIK----DPEFVIR  
 Sc. 507 KEFK--QRLLKKFNN-VLEPLYFMKFDVKSCYDSIPRMECMRIKDALKNE--NGEFVRS  
 Ca. 484 DYRDSLLTMTIFARFGEIPKLEILKFDKMECYDRISQPVLMKKLEELFENQDNK-TSYYV

Motif A

ham. 738 YAVVQR-----DRQ-----GQIHKSFR--QVSTLS--DLQPHMG-----  
 m. 736 AVVRR-----DSQ-----GQVHKSFR--QVTTLS--DLQPYMG-----  
 h. 745 AVVQK-----AAH-----GHVEKAFKS--HVSTLT--DLQPYMR-----  
 Ea. 634 AQILNRK-----NNI-----VIDSKNFRKK--EMKDYFRQKFQKIALEGGQY  
 Ot. 740 TFWLNRK-----NNI-----IVERSNFRKL--PIKQYFRYKFQKIGIDGSSY  
 Tt. 651 LLFQNKRP LLQIQQTNNLNSAMEIEEEKINKKPKMDNINFYYFNLKERQIAYS--LY  
 At. 714 CRIVCCG-----KRSNWN--KILVSS--DKNSNFS-----  
 Sp. 620 KYATIHA-----TS-----DRATKNEVS--EAFSYF--DMVPFEK-----  
 Sc. 562 QYFFNTN-----T--G--VLKLF--NVVNAS-----  
 Ca. 543 RYQAQLDA-----SHKLKKVKT--TIDTQYHNLNILSSS-----

ham. 769 -QFLKHLQD--S-DTSALRNSVVEIQSLSINEASSSLEDFDFF-----LRFVRNSVVKIGG-  
 m. 766 -QFLKHLQD--S-DASALRNSVVEIQSISMNESSSSLEDFDFF-----LHFLRHSVVKIGD-  
 h. 775 -QFVAHLQ-----ETSPLRDAVVEIQSSSINEASSGLEDFV-----LRFMCHHAVRIRG-  
 Ea. 674 PTLFSVLEN--EQNDLNAKKTLLIVEAKQRNYFKNDLLQPV-----INICQYNYINFNG-  
 Ot. 780 PTLFEILED--EFNDLNMKRTIIVEEQEQRKKFPKNDLLQPV-----LKICQNNYVTFNK-  
 Tt. 709 DDDDQILQKGFKEIQSDDRPFTVINODKPRCITKDIHNNH-----KHISQYNVISFNK-  
 At. 741 -RFTSTVP-----YNALQSTIVDKGENHRVRKDKLMVWI-----GNMKNKNMLQLDK-  
 Sp. 651 --VQQLS-----MKTSDTLFVDFVDYWTKSSSEIFKMI-----KEHLSGHIVKIGN-  
 Sc. 582 -----R-----VPKPYELYIDNVRTVHLSNODVINVV-----EMEIFKTALWVED-  
 Ca. 575 -RHLSNCKS-----LVDKTKTIALQKGNILEVCRSOLYDVVGSVKDAMTIFBMTIFCRGN

ham. 819 -RCYVQCQGIPOGSSSLSTLLCSLCFGDMEN---KLFAEVQOD-----  
 m. 816 -RCYVQCQGIPOGSSSLSTLLCSLCFGDMEN---KLFAEVQOD-----  
 h. 823 -KSYVQCQGIPOGSSSLSTLLCSLCFGDMEN---KLFAEIRRD-----  
 Ea. 726 -KIFYKOTKGIPQGLCVSSILSSFYATLEE---SSLGFLRDESMNPEN-----PNV  
 Ot. 832 -KOYKOMKGIPOGLCVSYILSSFYANLEE---NALQFLRKESSMDPEK-----PEI  
 Tt. 763 -VKTRQKRGIPOGLNLSGVLCSFYEGKLEE---EYTQFLKNAEQVN-----GSI  
 At. 787 -SFYVQIAGIPQGHRLSSILCCFYGHLERTLIYPFLEEASKDVSSKECSREEELIPTS  
 Sp. 696 -SOYLQKVGIPQGSILSSFLCHFYMEDLID---EYLSFTKKK-----G--  
 Sc. 622 -KCYHREDGLEFOGSSLSAPITVDLVYDDILE---FYSEFKASPSQD-----  
 Ca. 629 LALYKRKRGVFOGFSLSSTIFCDILYSAMVH---DCFQFLWKSQD-----

Motif B

ham. 857 GLLLRVDDFLLVTPHLVQA-----EAFRLALVRGIPPEYGCMINLQKTVVNF  
 m. 854 GLLLRVDDFLLVTPHLDOA-----KTFLSTLVHGVPEYGCMINLQKTVVNF  
 h. 861 GLLRLVDDFLLVTPHLTHA-----KTFLRTLVRGVPEYGCVMNLRKTVVNF  
 Ea. 773 NLLMRITDDYLLITTEQENNA-----VLFTIEKLINVSRENGFKFNMKKLQTSF  
 Ot. 879 NLLMRITDDYLLMTTEKNNA-----MLFTIEKLYQLSLGNFFKFHMKKLKTNF  
 Tt. 808 NLLMRITDDYLLFISDSQONA-----LNLIVQLQNCANNNGFMFNDQKITTNF  
 At. 846 YKLLRFIDDDYLFVSTSRDOA-----SSFYHRLKHGFKDYNCFMNETKFCINF  
 Sp. 735 SVLLRVDDFLEFIVNKKDA-----KKFLNLSLRGFEKHNFTSLEKTVINF  
 Sc. 663 TLTLRLADDFLEISTDQQQV-----INIKKLAMGGFQKYNANARDKILAVS  
 Ca. 671 ELEVRVDDFLLVTPDSQVMTIFEDSNIYDQVHNI--LSGKILESYGAFVKNKDKTVVNF

Motif C

Motif D

ham. 904 PVDAGTIDGT----APHQLPAHCLFPWCGLLLDTQTLEVLCDYTGARTSIKASLTFQRT  
 m. 901 PVEPGTIGGA----APYQLPAHCLFPWCGLLLDTQTLEVFCDYSGYAQTSIKTSLTFQSV  
 h. 908 PVEDEALGGT----AFVQMPAHGLFPWCGLLLDTQTLEVQSDYSSYARTSIRASLTFNRG  
 Ea. 820 PLSPSKFAKYGMDSVVEQNIVQDYCDWIGISIDMKTLALMPNINLRIEGILCTLNLMQT  
 Ot. 926 ALNLQKIGCT--NTTQDIDSINDDLFWHIGISIDIKTLNIIQNINIKKEGILCTLNVMQT  
 Tt. 855 QFPQEDYNLE-----HFKISVQNECQWIGKSIDMNTLEIKSIQKQTQQEINQINVAISI  
 At. 893 EDKEEHRCSS---NRMFVGDNGVPFVRWTGLLINSTFEVQVDYTRYLSGHISSTFSVAWQ  
 Sp. 782 ENSNGIINN-----FFNESKKRMPFTGLLINMRSLDTLLACPKIDEALFNST-SVELT  
 Sc. 710 SQSDD-----DTVIQFCAMHIFVKELEVWKHSSTMNNFHIR-----  
 Ca. 729 -QTTT-----KTSIDEVGLEVNTTDLSTIKRNSGSISLVTTN-----

Motif E

Motif F

ham. 960 FKAGRNMRQKLLAVLRRLKCHSLFVLDLQVNSLQTVCTNIVY---KIFILQAYRFHACALQLP  
 m. 957 FKAGKTMNRKLLSVLRRLKCHGLFVLDLQVNSLQTVCTNIIY---KIFILQAYRFHACVQLP  
 h. 964 FKAGRNMRRLFGVLRRLKCHSLFVLDLQVNSLQTVCTNIIY---KILLQAYRFHACVQLP  
 Ea. 880 KKASMWLKKKLKSLFMNNTIHYERKT-TTTFDFANKTIN---KLFTSGCYKYMCAKEYK  
 Ot. 985 NESILWLKKKLKSLFMNNTISFYFKST-TNKKFANITLS---KLYIAAAEKYVACCQEFK  
 Tt. 910 KNLSQLKKNKLRSLFLNQLIDYENPN-INSFEGLCROLYHHSKATVMKFPFMTKIFQID  
 At. 951 NKPVRNLROKLCYFLVPCHPLEDSNINSGETVRNIIY---QIFLLAAMKFFHCYVVEVS  
 Sp. 835 KHMGSFFYKILRSSLASFAQVFEDITHNSKFNSCCNIIY-----RLGYSMCMRAQA  
 Sc. 746 SKSSKGLFRSLIALFNTRISYKTIDTNLNSNTNTVLMQID---HVVKNISECYKSAFKDLS  
 Ca. 764 FRTFKTLVKYLYKTFYQLNLEGFLIDCSFGVLENVLENMG---SILRLVLRFEKTKFTSIV

---

Motif F

---

ham. 1017 FDQHVRRKNPAFFL-SILSNIASCOYSTILKVNAGMTLKAKG-ASGSFPPEAARWL-----  
 m. 1014 FDQVRKNLTFFL-GIISSQASCOYAILKVNPGMTLKA----SGSFPPEAAHWL-----  
 h. 1021 FHQQVWKNPTFFL-RVISTDASLCYSTILKAKNAGMSLGAKG-AAGPLPSEAVQWL-----  
 Ea. 936 DH--FKKNLAMSS-MIDLEVSKIIYSVTRAFFKYLVCNIK---DTIFGEEH-----  
 Ot. 1041 R---FHENTSLGGQNTDIKIIHIIYVVISFFKYLVCNVK---SPVFERDD-----  
 Tt. 969 LKKSQYSVQYKENTNENFLKDI LYTTVEDVCKILCYLQFEDEINSNIKEIFKNLYSWI  
 At. 1008 R---FWKLHPQTLFKFIIISVRYMERLINRRVRRINTGSSFRPVLLKLYKEEVIWIG----  
 Sp. 886 YLK-RMKDIFIPQRMFIIDILNVIGRKHWKKLAEIIGYTS---RRFLSSAEVKWL-----  
 Sc. 803 INV-TQNMQFHSFLQRIEMTV-----SGCPITK---CDPLIEYEVRF-----  
 Ca. 821 KYDTFHCY-----KFIKFLYDISNYTI-----VKYVETN---SDWEGAPEL-----

ham. 1070 -----C-----YQAFLLKLAGHSVTYKCLLGLPLRTAQKQLCRKLPR---ATM  
 m. 1064 -----C-----YQAFLLKLAHHSVIYKCLLGLPLRTAQKLLCRKLPE---ATM  
 h. 1074 -----C-----HQAFLLKLTIRHRVTYVPLLGLSLRTAQTQLSRKLPG---TTL  
 Ea. 981 -- YPDDFFLST-----LKHFIEIFSTTKYIENRVCMILK-----AKEAKLKSDQ---C  
 Ot. 1086 ---YQOFFFIYS-----LKFFITRFKQKQNEFAGVYKILK-----AKEKKLEVAK---I  
 Tt. 1029 MWDIIVSYLKKKKQFKGYLNKLLQKIRKSRRFFYLKEGCKSLQLILSQQKYQLNKKELEAI  
 At. 1061 -----LDAYIQVLKKKNSRYRMLLIYLSALSXHSLSQQ-----LSS  
 Sp. 937 -----FCLGMRDG---LKPSF-KYHPCFEQLIYQFQSLTDLIK-----PLRPVLRQVL  
 Sc. 843 -----I-----LNGFIESISSNTSKFKDNIILLRKEIQHLQAYIY-----  
 Ca. 859 -----LNCIKQIIVKEFSSFESEYSEIVEWVQTLNI-----

---

Motif G

---

ham. 1109 AILETAADPALSTDFQTILD-----  
 m. 1103 TILKAAAADPALSTDFQTILD-----  
 h. 1113 TALEAAAANPALPSDFKTILD-----  
 Ea. 1023 QSLIQYDA-----  
 Ot. 1128 EFQIQ-----  
 Tt. 1089 EFTDLNNLIQDIKTLIPKISAKSNQQNTN  
 At. 1098 -ELKYATDRSNSSSLWKLN-----  
 Sp. 981 -----FLHRRRIAD-----  
 Sc. 878 IYTHIVN-----  
 Ca. -----

MONTANA STATE UNIVERSITY - BOZEMAN



3 1762 10337675 0