The isolation and cloning of an Arabidopsis thaliana DNA fragment which hybridizes to the yeast cytochrome c gene

by Shelley Kay Watters

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Biological Sciences
Montana State University

Abstract:
Cytochrome c protein sequences have been studied extensively to determine evolutionary relationships among eukaryotic organisms. More recently, comparisons of cytochrome c gene sequences from animal and yeast sources have been made, but these analyses have not included plants. This study was conducted because the isolation of a cytochrome c gene from a plant has not been reported. Since Arabidopsis thaliana, a small Crucifer, has the smallest genome known among flowering plants, we chose to attempt to isolate a cytochrome c gene from this plant.

We have isolated a DNA fragment, approximately 900 basepairs in length, containing a putative cytochrome c gene from the plant, A. thaliana. It was isolated from an A. thaliana genomic library using the protein-coding region from the yeast iso-1-cytochrome c gene as a specific hybridization probe. The DNA sequence of 485 basepairs of the fragment has been determined, but the rest of the DNA fragment (approximately 400 basepairs) presented sequencing difficulties and therefore remains to be sequenced. Comparisons of the known sequence with cytochrome c gene sequences of other eukaryotic organisms indicate there are no regions of similarity.

THE ISOLATION AND CLONING OF AN ARABIDOPSIS THALIANA DNA FRAGMENT WHICH HYBRIDIZES TO THE YEAST CYTOCHROME C GENE

by

Shelley Kay Watters

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

of

Master of Science

in

Biological Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

December 1987

APPROVAL

of this thesis submitted by

Shelley Kay Watters

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.

Dec 31/87
Date

E R Nye
Chairperson, Graduate Committee

Approved for the Major Department

12/31/87
Date

Peter F. Brussard
Head, Major Department

Approved for the College of Graduate Studies

1-6-88
Date

W B Malone
Graduate Dean

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. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made.

Permission for extensive quotation from or reproduction of this thesis may be granted by my major professor, or in his absence, by the Director of Libraries when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for finanacial gain shall not be allowed without my written permission.

Signature Shelley Matt

Date 1/5/88

ACKNOWLEDGEMENTS

I'd like to thank Ernie Vyse for his friendship, guidance, and the many hours of effort he put into this project; Tom Blake for donations of supplies and ideas; Cliff Bond for his suggestions and encouragement; Ken Tindall for help with DNA sequencing; the members of my committee: Dave Cameron, Sam Rogers and Dave Sands for their useful comments; and Steve Malmberg for his patience and understanding.

TABLE OF CONTENTS

TABLE OF CONTENTS--Continued

LIST OF FIGURES

LIST OF FIGURES--Continued

ABSTRACT

Cytochrome c protein sequences have been studied extensively to determine evolutionary relationships among eukaryotic organisms. More recently, comparisons of cytochrome c gene sequences from animal and yeast sources have been made, but these analyses have not included plants. This study was conducted because the isolation of a cytochrome c gene from a plant has not been reported. Since Arabidopsis thaliana, a small Crucifer, has the smallest genome known among flowering plants, we chose to attempt to isolate a cytochrome c gene from this plant.

We have isolated a DNA fragment, approximately 900 basepairs in length, containing a putative cytochrome c gene from the plant, A. thaliana. It was isolated from an A. thaliana genomic library using the protein-coding region from the yeast iso-1-cytochrome c gene as a specific hybridization probe. The DNA sequence of 485 basepairs of the fragment has been determined, but the rest of the DNA fragment (approximately 400 basepairs) presented sequencing difficulties and therefore remains to be sequenced. Comparisons of the known sequence with cytochrome c gene sequences of other eukaryotic organisms indicate there are no regions of similarity.

INTRODUCTION

Cytochrome c is a small protein, encoded in the nucleus, which functions as an electron carrier in the respiratory chain of all eukaryotic organisms. Many amino acid residues have been conserved in the cytochrome c proteins of diverse species (Margoliash and Schejter, 1966), probably because the protein acts in such a specific capacity. By exploiting the existence of these conserved sequences, Scarpulla et al. (1981) and Limbach and Wu (1983, 1985a, b) have demonstrated that the deoxyribonucleic acid (DNA) coding for cytochrome c from one organism can be used as a specific hybridization probe for the cytochrome c gene of a different organism. We therefore assumed that the iso-1-cytochrome c gene from the yeast, Saccharomyces cerevisiae, could be used as a hybridization probe to isolate the cytochrome c gene from the plant, Arabidopsis thaliana.

Cytochrome c amino acid sequences have been compared extensively to determine evolutionary relationships among organisms (Margoliash and Schejter, 1966). A comparison of the DNA sequences of cytochrome c genes may further aid in the determination of these relationships and may also provide information about the evolution of codon usage, regulatory sequences and intron and exon arrangements.

The structure of cytochrome c is important to consider when discussing its evolutionary significance. It is a small peripheral protein found on the inside of the inner mitochondrial membrane. It

contains from 104 residues, in mammals, to 112 residues in some plants (Lemberg and Barrett, 1973). Of the eukaryotic cytochrome c protein sequences known, residues in positions 14, 17, 18 and 80 are bound to a non-protein heme group (Lemberg and Barret, 1973). It is this heme group which makes it possible for cytochrome c to act as an electron carrier between cytochrome c_1 and cytochrome oxidase in the respiratory chain. The amino acids which are associated with the heme group and those which are responsible for the proper tertiary stucture of the protein appear to remain highly conserved. Any residues which are responsible for correct positioning of the protein in the mitochondria may be conserved as well. There are 35 amino acid residues which have been highly conserved in the known sequences of eukaryotic cytochrome c proteins (Smith and Boyer, 1970). These are shown in Figure 1. Because there is conservation of amino acid residues, many of the nucleotides coding for these amino acids will also be conserved.

We cannot expect complete identity of nucleotides between two cytochrome c genes, even within regions which code for identical amino acids, because of the degeneracy of the genetic code. However, if two cytochrome c proteins exibit a number of regions with identical amino acids, there may be some areas of their corresponding genes which have complete nucleotide identity.

In the hybridization analysis done by Limbach and Wu (1983, 1985a, b), the DNAs used for the hybridization probes coded for sequences displaying more than 50 percent (%) amino acid similarity with the cytochrome c proteins from the studied organisms. When

probing a chicken genomic library for the cytochrome c gene, they used the protein-coding region of the yeast cytochrome c gene (Limbach and Wu, 1983). The cytochrome c proteins from these two organisms have 62 identical amino acids in homologous sites. The two proteins also share five regions containing five or more identical, sequential amino acid residues.

Isolation of the mouse cytochrome c gene was done by hybridizing the DNA from a mouse genomic library with the cytochrome c gene from rat. Rat cytochrome c DNA codes for an amino acid sequence identical to that of mouse (Limbach and Wu, 1985a).

The Drosophila melanogaster gene was identified by using cytochrome c DNA from mouse as a heterologous hybridization probe (Limbach and Wu, 1985b). When the cytochrome c amino acid residues from these two organisms are compared, 86 identical pairs of amino acids are found.

The cytochrome c genes from chicken, mouse and Drosophila melanogaster have been sequenced by Limbach and Wu (1983, 1985a, b). The cytochrome c genes from the following organisms have also been sequenced: Saccharomyces cerevisiae, CYC1 (Smith et al., 1975), CYC7 (Montgomery et al., 1980); Schizosaccharomyces pombe, PoCYC (Russell and Hall, 1982); and rat, RC4 (Scarpulla et al., 1981). An evolutionary comparison has been made of the above cytochrome c genes as well as some pseudogenes from rodents and humans by Wu et al. (1986).

As of yet, the isolation of a cytochrome c gene from a plant has not been reported. If a plant cytochrome c gene were isolated it could

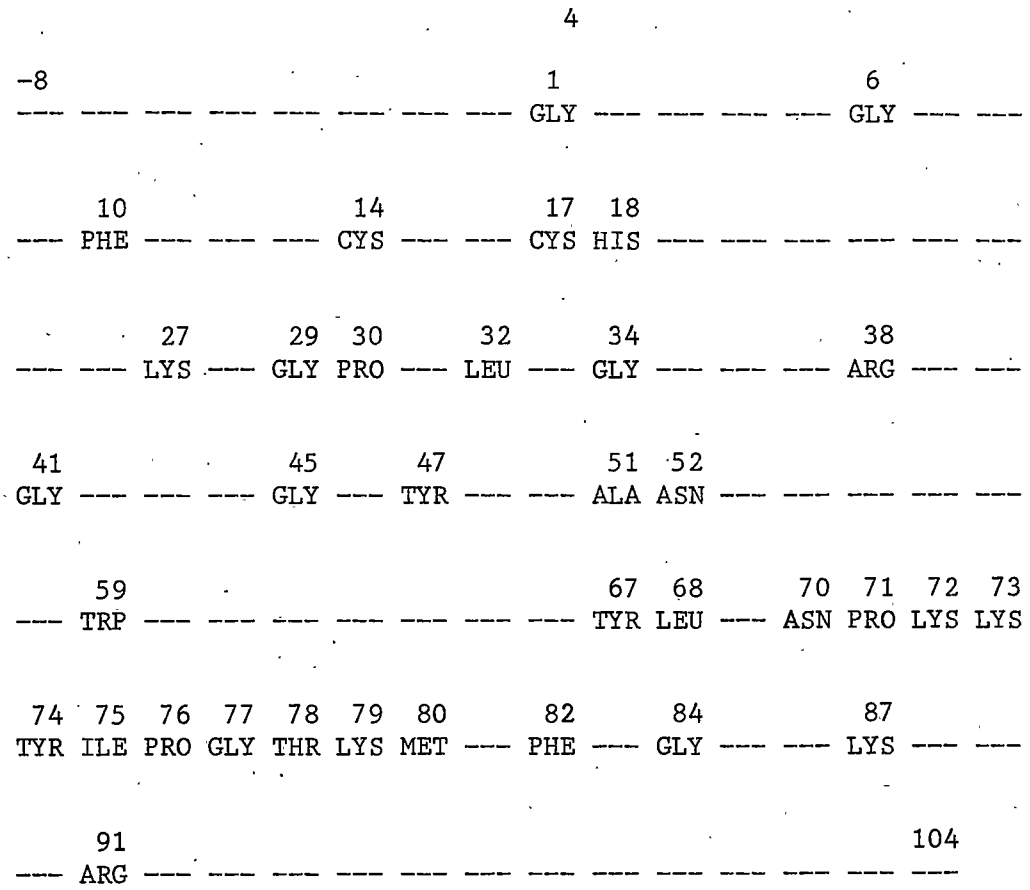

Figure 1. Conserved amino acid residues in the sequences of known eukaryotic cytochrome c proteins.

be used as a specific hybridization probe for other plant cytochrome c genes. Comparisons of a plant cytochrome c gene with the same genes from other plants and unrelated organisms should provide interesting and informative data about evolutionary relationships. Also, valuable information on the control, regulation and structure of plant genes could be obtained from sequencing data of cytochrome c genes.

One of the easiest flowering plants with which to do genetic studies is Arabidopsis thaliana, a member of the mustard family. This plant is rapidly being adopted as the organism of choice for molecular genetics of plant growth and development (Meyerowitz and Pruitt, 1985;

Pang and Meyerowitz, 1987). It contains a haploid chromosome complement of five and has the smallest genome size of any known flowering plant with only 7×10^7 basepairs (Leutwiler et al., 1984). A. thaliana, then, is an advantagous plant from which to isolate specific genes because the construction and screening of a genomic library is easier when the genome size is small.

The amino acid sequence for the cytochrome c protein of A. thaliana is not known but the sequences are known for the following five plants: wheat, Triticum aestivum (Stevens et al., 1967; Boulter et al., 1970); bean, Phaseolus aureus L. (Thompson et al., 1970; Boulter et al., 1970); sunflower, Helianthus annuis L. (Ramshaw et al., 1970; Boulter et al., 1970); sesame, Sesamum indicum L. and castor, Ricinus communis L. (Boulter et al., 1970). When comparing the cytochrome c sequences of all five plants and the protein sequence of yeast, 57 identical amino acids are found at homologous sites (underlined by solid lines in Figure 2). There are eight more amino acids which are identical to yeast cytochrome c amino acid residues, at homologous sites, from one or more plant cytochrome c proteins (underlined by dotted lines in Figure 2).

Plants and yeast probably diverged between 1.0 and 1.1 billion years ago (Hightower and Meagher, 1986) but the cytochrome c sequences have retained a considerable amount of amino acid similarity. By making use of this homology we will isolate a fragment of DNA from an A. thaliana library which hybridizes to the protein-coding region of yeast cytochrome c DNA. Through DNA sequence analysis we may be able

to predict the sequence of the A. thaliana cytochrome c protein and describe the first general structure of a plant cytochrome c gene.

```
                                                    1
Plant     ALA SER PHE   X  GLU ALA PRO  X  GLY  X   X   X   X  GLY
Yeast              THR GLU PHE LYS ALA GLY SER ALA LYS LYS GLY

                        10                                      20
Plant      X  LYS ILE PHE LYS THR LYS CYS ALA GLN CYS HIS THR VAL
Yeast     GLU THR LEU PHE LYS THR ARG CYS GLU LEU CYS HIS THR VAL

                                                30
Plant      X   X  GLY ALA GLY HIS LYS GLU GLY PRO ASN LEU  X  GLY
Yeast     GLU LYS GLY GLY PRO HIS LYS VAL GLY PRO ASN LEU HIS GLY

                            40
Plant     LEU PHE GLY ARG GLN SER GLY THR THR  X  GLY THY SER TYR
Yeast     ILE PHE GLY ARG HIS SER GLY GLN ALA GLU GLY TYR SER TYR

              50                                        60
Plant     SER  X  ALA ASN SYS ASN  X  ALA VAL  X  TRY  X GLU  X
Yeast     THR ASP ALA ASN ILE LYS LYS ASN VAL LEU TRP ASP GLU ASN

                                    70
Plant     THR LEU TYR ASP TYR LEU LEU ASN PRO LYS LYS TYR ILE PRO
Yeast     ASN MET SER GLU TYR LEU THR ASN PRO LYS LYS TRY ILE PRO

                        80                                      90
Plant     GLY THR LYS MET VAL PHE PRO GLY LEU LYS LYS PRO GLN  X
Yeast     GLY THR LYS MET ALA PHE GLY GLY LEU LYS LYS GLU LYS ASP

                                                100
Plant     ARG ALA ASP LEU ILE ALA TYR LEU LYS  X   X  THR  X  SER
Yeast     ARG ASN ASP LEU ILE THR TYR LEU LYS LYS ALA CYS GLU
```

Figure 2. Comparison of plant and yeast cytochrome c amino acid sequences. The above plant sequence is a consensus amino acid sequence for the following plants: wheat, bean, sunflower, sesame and castor. "X" indicates an amino acid which varies between the five plants. The yeast sequence is that of Saccharomyces cerevisiae. Amino acids which are underlined by solid lines indicate amino acids which are the same in yeast and all five plant cytochrome c proteins. Dotted lines indicate amino acids which are the same in yeast and one or more plant cytochrome c proteins.

MATERIALS AND METHODS

Outline of Procedures

To obtain DNA from Arabidopsis thaliana which may contain a cytochrome c gene, total genomic DNA or a genomic library (bacteriophage lambda containing random fragments of the A. thaliana genome) must be screened. We chose to screen a genomic library by using the protein-coding region of the yeast iso-1-cytochrome c gene as a specific hybridization probe. Outlined below is the sequence of procedures which were performed in this study.

1) Isolation of yeast cytochrome c DNA to be used as a hybridization probe.
2) Hybridization of the yeast cytochrome c probe with an A. thaliana library.
3) Isolation of lambda-A. thaliana plaques which hybridize to the yeast cytochrome c probe. (Bacteriophage lambda forms individual plaques when grown on plates with an appropriate host.)
4) Isolation of DNA from plaques which hybridize to the yeast cytochrome c probe.
5) Digestion of DNA from hybridizing plaques with restriction enzymes and rehybridization with the yeast cytochrome c probe.

6) Isolation of a fragment of A. thaliana DNA which hybridizes to the yeast cytochrome c probe.

7) Subcloning of a hybridizing DNA fragment into the pBR322 plasmid in order to localize a smaller hybridizing fragment of A. thaliana DNA. (This step is necessary so that a reasonably sized fragment may be obtained for DNA sequencing.)

8) Subcloning of a smaller fragment into M13 viral DNA. (M13 is used as a vector for DNA sequencing).

9) Sequencing of cloned A. thaliana DNA fragment.

(Descriptions of the bacterial strains, viruses, plasmids, media, buffers and solutions referred to in this section can be found in the Appendix.)

Source of PAY9 Plasmid and Arabidopsis thaliana Library

PAY9 Plasmid. The plasmid, pAY9, was a gift provided by the laboratory of John Stiles (University of Hawaii). It consists of a SalI-HindIII fragment from the yeast iso-1-cytochrome c gene (CYC1) inserted into the SalI-HindIII site of the plasmid, pUC9 (personal communication).

A. thaliana Library. An amplified A. thaliana library was received from the laboratory of Elliot Meyerowitz (California Institute of Technology). It was prepared from whole plant DNA of the A. thaliana strain Landsberg erecta by partially digesting with the restriction endonuclease MboI and ligating into the BamHI arms of the bacteriophage lambda vector EMBL4.

DNA Isolations and Purification

Plasmid and M13 DNA Isolation. Large scale isolation of DNA was done to obtain 100-500 micrograms (µg) of DNA. Large scale isolations of plasmid pAY9, plasmid pBR322 and bacteriophage M13 DNAs were performed using the alkaline lysis method (Birnboim and Doly, 1979). DNA isolated in this manner was purified by ultra-centifugation in cesium chloride-ethidium bromide gradients. Cesium cloride was added to the samples at a final concentration of 1.55 grams/milliliter (g/ml) and ethidium bromide at 600 micrograms/milliliter (µg/ml). Centrifugation took place at 250,000 x gravity (g) for 18-24 hours at 20 degrees celsius (C).

Small scale isolation of recombinant DNA (from pBR322 and M13 recombinants) was done by the method described by Rodriguez and Tait (1983). This method requires less time than the large scale isolations and is adequate when smaller quantities (1-2 µg DNA) are required.

DNA was quantified using the spectrophotometric method (Maniatis et al., 1982) or by estimation on an agarose gel using known amounts of DNA in the same size range.

Bacteriophage Lambda DNA Isolation. Bacteriophage lambda DNAs containing Arabidopsis thaliana inserts were isolated by an adaption of the plate lysate method described by Maniatis et al. (1982). The bacteriophage were grown in the Escherichia coli strain, Su6, on NZYCM agarose plates as described above. Five milliliters (mls) of SM solution was added to each plate and the plates were shaken at room temperature from 1-2 hours. The SM solution was transferred to

centrifuge tubes and centrifuged at 8,000 x g for 10 minutes. After recovering the supernatant ribonuclease A (RNase) and deoxyribonuclease I (DNase) were each added to achieve a final concentration of 1 µg/ml. The viral particles were then precipitated by adding an equal volume of a solution containing 20% polyethelene glycol and 2.0 molar (M) sodium chloride (NaCl) in SM. The samples were allowed to incubate 1 hour at 0 C and centrifuged at 10,000 x g for 20 minutes. The bacteriophage particles were resuspended in 0.5 ml SM. Twenty-five microliters (µl) 10% sodium dodecyl sulfate (SDS), 5 µl 0.5 M disodium ethylene diamine tetra-acetate (EDTA) and 5.0 mls of 10 milligrams/milliliter (mg/ml) protienase K were added and samples were incubated at 68 C for 1 hour. Samples were then incubated at 0 C for 30 minutes after the addition of 0.2 ml of 5.0 M potassium acetate (KoAc). The debris was pelleted at 10,000 x g for 20 minutes and the DNA in the supernatant was precipitated using 2 volumes of 95% ethanol (EtOH). This procedure was used for both large and small scale isolation of lambda DNA.

Purification of DNA Fragments by Electroelution. DNA fragments separated on agarose gels were isolated by an adaption of the electroelution method (McDonnel et al., 1977). Because of the damaging effect of ultraviolet (UV) light on DNA, the DNA fragments of choice were cut from agarose gels with little or no exposure to UV light. In this method of isolation ISCO cups, pictured in Figure 3, were employed (Allington et al., 1978). The piece of agarose containing the DNA was placed inside chamber B on top of a fine mesh screen. One percent agarose was poured over the piece(s) to keep it in

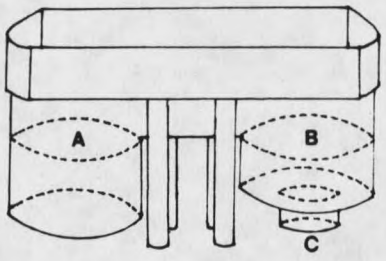

Figure 3. Drawing depicting an ISCO cup. Letters designate the chambers within the apparatus.

place. A separate apparatus was set up on which dialysis tubing was secured tightly to the bottoms of chambers B and C. Chambers B and C were then filled with 1X tris-borate (TBE) buffer. The agarose "plug" from the first apparatus was then gently slipped into the B chamber of the second apparatus, displacing all of the buffer except that contained in chamber C. The apparatus was placed in a divided, horizontal electrophoresis tank so that chambers B and C were on the positive side and chamber A on the negative side. Chamber A was then filled with enough 1X TBE to go over the top of the divider and on to the agarose in chamber B. Current was then applied at 50 volts (V), 25 milliamps (mAmps) for 3-4 hours. After this period the current was reversed for 1 minute. The buffer was removed from the top of the

agarose plug, the plug was carefully taken from chamber B and the DNA containing buffer from chamber C was transferred to an eppendorf tube. The DNA solution was then extracted using phenol, chloroform/phenol, chloroform/isoamyl alcohol, ether, and precipitated with ethanol as described (Maniatis et al., 1982).

Construction of PBR322 and M13 Recombinants

PBR322-A. thaliana Recombinant. A 1.5 kilobase (kb) fragment from a bacteriophage lambda-A. thaliana isolate (designated lambda-3A) was previously shown to hybridize with the yeast cytochrome c (CYC1) probe. Lambda-3A DNA and pBR322 DNA were then cut with the enzyme BamHI. All cleavages using restriction endonucleases were performed as described by Maniatis et al. (1982). In order to ensure that all of the DNA had been cleaved it was separated electophoretically on 0.7% agarose gels, stained with ethidium bromide and visualized using short wave-length ultraviolet light. The BamHI cut DNAs were then phenol extracted and ethanol precipitated. The two DNAs were ligated using a two-fold molar excess of lambda-3A DNA. The ligations were done from 12-16 hours at 4 C using T4 DNA ligase using the method described by Maniatis et al. (1982).

M13-A. thaliana Recombinant. After further localizing the hybridizing fragment in a pBR322-A. thaliana recombinant (322-3A), a smaller fragment was subcloned into M13mp18 and M13mp19. DNA from 322-3A was cleaved first with BalI, then with HindIII. M13mp18 and M13mp19 DNA were cleaved with SmaI and HindIII. The DNAs were again visualized by gel electrophoresis to ensure that cleavage was

complete. They were then phenol extracted and ethanol precipitated. Total DNA from the 322-3A cleavage was added at a two-fold molar excess to M13mp18 DNA and in the same manner to M13mp19 DNA. Both samples were ligated using T4 DNA ligase at 4 C for 12-16 hours.

Transformations and Infections

Transformation with Plasmid DNA. Transformation of the above ligation mixture (BamHI pBR322 and BamHI lambda-3A DNAs) into the bacterial strain DH5 was done according to the method described by Himeno et al. (1984). The transformants were spread on LB media containing 50 µg/ml ampicillin. This ensured that only transformants would grow since pBR322 carries a gene coding for ampicillin resistance. A recombinant from this transformation which contained the 1.5 kb fragment from the bacteriophage lambda-A. thaliana DNA and hybridized to the yeast CYC1 probe was designated 322-3A.

Transformation with M13 DNA. Recombinant M13 DNA was used to transform the JM107 bacterial strain using an adaption of the procedure described by Messing (1979). JM107 was grown and added to the DNA as described by Hackett et al. (1984). Tubes containing 3.0 mls of YT soft agar were melted and maintained at 45 C. Fifty microliters 2% 5-bromo-4-chloro-3-indolyl-B-D-galactoside (X-gal) and 10 µl 100 millimolar (mM) isopropyl-B-D-thiogalactoside (IPTG) were added to the agar. After incubating the DNA-JM107 mix for 40 minutes, it was added to the agarose and heat-shocked (in the agarose) at 42 C for 2 minutes. The mixture was immediately poured onto warmed YT

plates. After overnight incubation at 37 C, plaques which appeared clear were chosen for further analysis.

Infections Using the Bacteriophage Lambda-A. thaliana Library. The A. thaliana library was grown on the bacterial strain, Su6. Plaques were obtained by the plating method described (Maniatis et al., 1982). The medium employed was either LB, when the bacteriophage were to be grown for hybridization analysis, or NZYCM medium when they were grown for DNA isolation.

Restriction Endonuclease Analysis

PAY9 Analysis. PAY9 was cleaved with the enzymes TaqI and EcoRI. This cleavage generated a 343 basepair (bp) fragment containing the protein-coding region of yeast CYC1 gene and an additional 20 basepairs of 3' non-coding DNA. The digestions were run on a 1.2% agarose gel in order to adequately separate fragments in the 6.0-0.3 kb range. Samples were run from 3-4 hours at 50 volts, 20 milliamps. The gel was stained with ethidium bromide and visualized on a short wave-length ultraviolet lamp. A band which migrated a distance which corresponded to a size of approximately 340 bp was cut from the gel and purified by electroelution. The nucleic acid sequence of the fragment, the yeast CYC1 probe is shown in Figure 4.

Analysis of DNA Isolated from Lambda-A. thaliana Bacteriophage. DNA was isolated from lambda-A. thaliana plaques which appeared to hybridize to the yeast CYC1 probe. DNA from each isolate was digested with BamH1. Samples from each digestion were run on 1.0% agarose gels for 2-3 hours at 50 volts, 25 milliamps. The gels were stained with ethidium bromide and visualized on an UV lamp.

```
        EcoR1    HaeIII
ATG ACT GAA TTC AAG GCC GGT TCT GCT AAG AAA GGT GCT ACA CTT TTC
Met Thr Glu Phe Lys Ala Gly Ser Ala Lys Lys Gly Ala Thr Leu Phe
                    30

                 AccI                                      AsuI
AAG ACT AGA TGT CTA CAA TGC CAC ACC GTG GAA AAG GGT GGC CCA CAT
Lys Thr Arg Cys Leu Gln Cys His Thr Val Glu Lys Gly Gly Pro His

AAG GIT GGT CCA AAC TTG CAT GGT ATC TTT GGC AGA CAC TCT GGT CAA
Lys Val Gly Pro Asn Leu His Gly Ile Phe Gly Arg His Ser Gly Gln

AluI
GCT GAA GGG TAT TCG TAC ACA GAT GCC AAT ATC AAG AAA AAC ATG TTG
Ala Glu Gly Tyr Ser Tyr Thr Asp Ala Asn Ile Lys Lys Asn Val Leu

TGG GAC GAA AAT AAC ATG TCA GAG TAC TTG ACT AAC CCA AAG AAA TAT
Trp Asp Glu Asn Asn Met Ser Gly Tyr Leu Thr Asn Pro Lys Lys Tyr

         KpnI        HaeIII                MboII
ATT CCT GGT ACC AAG ATG GCC TTT GGT GGG TTG AAG AAG GAA AAA GAC
Ile Pro Gly Thr Lys Met Ala Phe Gly Gly Leu Lys Lys Gly Lys Asp

                                                        AsuI
AGA AAC GAC TTA ATT ACC TAC TTG AAA AAA GCC TGT GAG TAAACAGGCCC
Arg Asn Asp Leu Ile Thr Tyr Leu Lys Lys Ala Cys Glu       HaeIII

           TaqI
CTTTTCCTTTGTCGA
```

Figure 4. Nucleic acid sequence and restriction sites of the yeast CYC1 probe.

Analysis of 322-3A. This analysis was done to locate a smaller fragment of DNA containing the hybridizing region of DNA. By hybridizing the 322-3A DNA cleaved with various enzymes to the yeast CYC1 probe, a more exact region of hybridization can be found.

Plasmid DNA designated 322-3A (PBR322 containing the hybridizing fragment) was digested using the following enzymes: BamHI, HindIII, EcoRI, XbaI, PstI, SalI, KpnI, SmaI, BglII, ClaI, PvuI, BalI, AvaI, HincII, AvaI, Sau96I and DdeI. Most were done as single digestions, but some double digestions were performed using BamHI as a second enzyme in order to localize restriction sites more precisely. All samples were run on 0.7%-1.0% agarose gels from 2-3 hours at 50 volts, 25 milliamps. The gels were stained with ethidium bromide and visualized on an UV lamp.

Southern Hybridization Analysis

Plaque Hybridizations. The mean size of the A. thaliana inserts in the lambda library were 15-20 kb. With a genome size of 7×10^7 bp, it was estimated that approximately 20,000 plaques were necessary in order to adequately screen the entire genome (Clarke and Carbon, 1976). However, approximately 50% of the library contained bacteriophage lambda which lacked A. thaliana inserts so more than 40,000 plaques were actually screened. The A. thaliana library was grown on the bacterial strain, Su6, and transfered to nitrocellulose filters (Southern, 1975). The filters were baked and prewashed as described by Maniatis et al. (1982). Prehybridizations were done at 23 C from 4-24 hours in a solution containing 50% formamide, 5X Denhardt's, 0.1% SDS and 100 µg/ml denatured salmon sperm DNA. Hybridizations were allowed to proceed at 23 C from 36-48 hours in freshly prepared prehybridization solution containing 32P-labeled yeast CYC1 probe at 10^5 counts per minute (cpm) per filter. The probe

was labeled by the random priming technique first described by Taylor et al. (1976) and modified by Fienberg and Vogelstein (1983). The filters were washed at 23 C under the following conditions: 5X SSC, 0.2% SDS rinse; 5X SSC, 0.2% SDS for 30 minutes (2 times); 2X SSC, 0.2%SDS for 15 minutes (2 times); 1X SSC, 0.2% SDS for 15 minutes (1 time).

DNA Hybridizations. DNAs from the plaques which appeared to hybridize to the yeast CYC1 gene were isolated, run on 0.7% agarose gels and transfered to nylon transfer membrane (Zeta-Probe by Bio-Rad). The transfer was done by the alkaline blotting method as described (Reed and Mann, 1985). Filters were prehybridized in a solution containing 1.5X SSPE, 1% SDS, and 0.5% Blotto at 53 C for 4 hours, then hybridized in the same solution (freshly prepared) containing 10^6–10^7 cpm yeast CYC1 probe at 53 C for 24 hours. Washes were done for 15 minutes at room temperature with each of the following solutions: 2X SSC, 0.1% SDS; 0.5X SSC, 0.1% SDS; 0.1% SSC, 0.1% SDS. The final wash was done at 40 C for 15 minutes in 0.1X SSC, 1.0% SDS.

Dot-Blot Hybridizations. After the hybridizing fragment had been cloned into pBR322 and localized in a smaller fragment, it was cloned into M13mp18 and M13mp19 in preparation for sequencing. In order to avoid isolating the fragment, whole plasmid DNA was cut and ligated to the DNA of the M13 derivatives. The M13 clones were screened using a variation of the alkaline blotting method. Nylon membrane was rinsed in 0.4 M sodium hydroxide (NaOH) and placed between the two plates of a dot-blot apparatus. Then 50 µl of M13 recombinant virus in YT

growth medium and 5 ul of 2% SDS were added to the wells of the apparatus. The samples were drawn into the membrane using a vacuum pump. The wells were washed with 0.4 M NaOH which was drawn through in the same manner. The membranes were then prehybridized and hybridized as described above.

DNA Sequencing

Preparation of Template. M13mp18 and M13mp19 single-stranded recombinant DNA was isolated by precipitating the phage particles in the presence of polyethylene glycol and NaCl and extracting with phenol according to the method described by Yamamoto et al. (1970).

Sequencing Reactions. Sequencing was carried out using the Sanger chain termination method (Sanger et al., 1977).

The reactions were performed using a sequencing kit from P-L Biochemicals (catalog # 27-1510). The manufacturers directions were followed with one exception: A different polymerase reaction buffer (5X) was used which contains 70 mM Tris-HCl pH 7.5, 70 mM magnesium chloride (MgCl), 500 mM NaCl.

Electrophoresis and Autoradiography. Thin, 5.0%-8.0% polyacrylamide gels containing 8 M urea and Tris-borate buffer were used for electrophoretic analysis (Sanger and Coulson, 1978). Shark's tooth combs were used instead of conventional combs. Samples were loaded using a Pipetman and plastic pipet tip and run from 2-8 hours at approximately 1000 V, 25 mAmps. Autoradiography took place overnight at -40 C.

RESULTS

Identification and Isolation of the Yeast CYC1 Probe

One microgram of the 343 basepair fragment containing the protein coding region of the yeast iso-1-cytochrome c gene was generated from approximately 25 µg of plasmid pAY9 DNA.

Following digestion of the plasmid with TaqI and subsequent separation on an agarose gel, twelve fragments were observed (Figure 5). The fragments were sized, using lambda DNA cut with HindIII as a reference, according to the method of Schaffer and Sederoff (1981). From the plasmid map, a total of nine fragments were expected of the indicated sizes (Figure 5). The undesignated fragments on the agarose gel pictured in Figure 5 were probably the result of partial digestion.

Cleavage of the TaqI digested plasmid with EcoRI would generate four new fragments of the following sizes: 15 bp, 845 bp, 253 bp and 343 bp. The 845 bp and 253 bp fragments would be indistinguishable on an agarose gel from other fragments which were produced from cleavage with TaqI. But the 343 bp fragment can be seen in Figure 5 in the lane containing EcoRI, TaqI digested pAY9. The 596 bp fragment, from which it was generated, is no longer evident. After cutting the piece of agarose containing the 343 bp fragment from the gel and electroeluting the DNA, a sample was run on a gel to determine the amount and purity. The fragment appeared to be free of contaminating DNA or RNA.

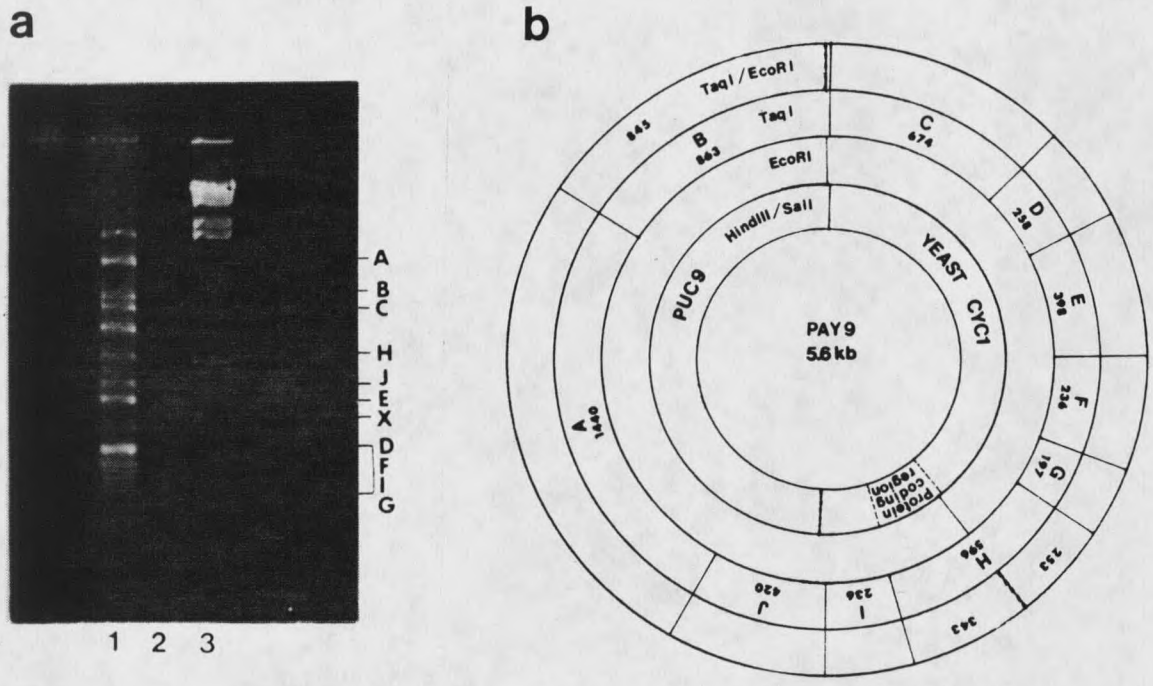

Figure 5. a) Photograph of an ethidium bromide-stained agarose gel containing restriction fragments of plasmid pAY9 DNA. Lane 1, pAY9, TaqI; Lane 2, pAY9, TaqI-EcoRI; Lane 3, lambda, HindIII. Letters assigned to individual fragments correspond to the letters in the map of pAY9. The letter "x" designates the 343 bp fragment containing the protein-coding region of the yeast iso-1-cytochrome c gene.
b) Restriction fragment map of the plasmid, pAY9.

Identification of A. thaliana DNAs which Hybridize to the Yeast CYC1 Probe

Plaques which appeared to hybridize to the yeast CYC1 probe were used to produce additional bacteriophage for the isolation of DNA. Using the plate lysate method, a plate containing about 10^4 plaques would yield approximately 2 µg of DNA. Five hundred nanograms of DNA