MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 00788 3972
LIBRARY
Michigan State
University

28568529

This is to certify that the

dissertation entitled

CYTOCHROME B_5 GENE IN CHICKEN

presented by

Hong Zhang

has been accepted towards fulfillment
of the requirements for

Ph. D degree in Genetics

Major professor

Date Aug. 2, 1989

MSU is an Affirmative Action/Equal Opportunity Institution

O-12771

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE

MSU Is An Affirmative Action/Equal Opportunity Institution

c:\circ\datedue.pm3-p.1

CYTOCHROME B_5 GENE IN CHICKEN

By

Hong Zhang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
For the degree of

DOCTOR OF PHILOSOPHY

Genetics Program

1989

ABSTRACT

CYTOCHROME B_5 GENE IN CHICKEN

By

Hong Zhang

Cytochrome b_5 functions as an electron transport carrier in fatty acid desaturation in animal liver, in methemoglobin reduction in erythrocytes, and in cytochrome P-450 reduction. It exists in two forms: an amphipathic form, and a cytosolic form. The amphipathic form consists of a N-terminal hydrophilic domain which contains a functional heme as a catalytic site, and a C-terminal hydrophobic domain which anchors the protein in the microsomal membrane. The cytosolic form is equivalent to the hydrophilic domain of the amphipathic form. The possibility that the cytosolic form is derived from the amphipathic form by proteolytical processing was proposed in the literature. This dissertation describes the isolation and characterization of the chicken cytochrome b_5 gene, and indirect support for the above hypothesis.

Cytochrome b_5 cDNA clones were isolated from chicken liver and

erythrocytes by probing cDNA libraries with synthetic oligonucleotides designed from the chicken cytochrome b_5 protein. The cDNA clones from liver and erythrocytes showed 100% homology and encoded a protein with an amphipathic form. The Northern analysis indicated only one kind of message present in liver total RNA, and this message is about the same size as the cDNA from erythrocytes. Furthermore, a lambda genomic clone was shown to contain a cytochrome b_5 gene with an amphipathic form, and this clone produced the same hybridization pattern as the chicken genomic DNA did in Southern analysis. All the data suggested that there is only one copy of the cytochrome b_5 gene in chicken.

The presence of one gene excluded the possibility that different genes are reponsible for the two forms of cytochrome b_5 protein in chicken. The complete identity of cDNA clones from liver and erythrocytes excluded the possibility that differential RNA splicing is the reason for two kinds of cytochrome b_5 from a single gene. Posttranslational modifications appear to be the mechanism for synthesis of the cytosolic cytochrome b_5. There may be one or more erythroid proteases which are responsible for the solubilization of amphipathic cytochrome b_5 in erythrocytes to give a cytosolic protein. The data presented in this dissertation supports the existence of such proteases in erythrocytes.

Copyright by

HONG ZHANG

1989

To China

Who Is Fighting For

Democracy

Acknoledgements

I would like to thank all members of my guidance committee, Drs Jerry Dodgson, Tom Friedman, Chris Somerville, and Edward E. Tolbert, for their continuously generous support and invaluble suggestions during my stay as a graduate student.

I would like to thank Dr. Barry Chelm for his contribution in serving as a member of my guidance committee before he left us two years ago.

I thank all my friends and colleagues in PRL and Genetics Program; their support made my life so much easier. There is no need to mention a particular person in C. Somerville's lab for me to say "thank you"; you all made this lab a wonderful place to work and to learn.

But, I would like to thank again to my major professor, Chris, for his patience, advices, and everything he provided in guiding me toward my Ph.D.

Finally, I would like to thank my grandmother and my parents, their love, support and understanding have always been with me.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

LITERATURE REVIEW ON CYTOCHROME B_5

1. Introduction

Cytochrome b_5 is a protohemoprotein which is present in high amounts in the microsomes of animal liver cells. In the reduced state, it shows an asymmetrical α-absorption band with a peak at 556 nm and a shoulder around 560 nm. See figure 1 below.

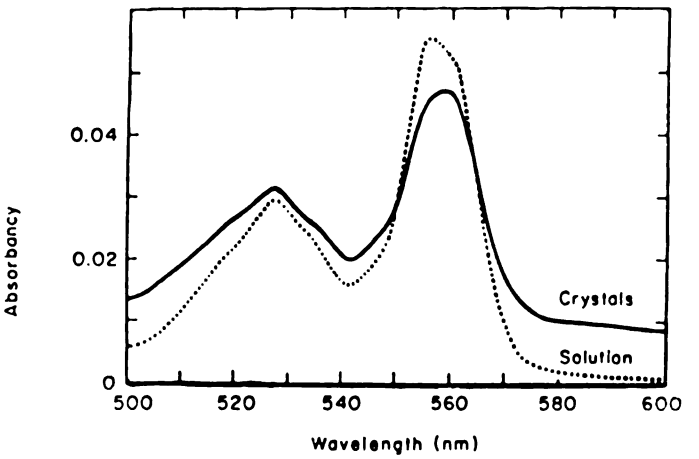

Figure 1. The absorption spectra in α and β region of the reduced cytochrome b_5 in solution and crystals (3).

This cytochrome is bound to the membrane and reduced with NADH by a flavoprotein (cytochrome b_5 reductase) which is also bound to the membrane. The primary structures and ternary structure of cytochrome b_5 from several animal sources have been determined (1,2). Hagihara et al. (3) suggested that use of the term cytochrome b_5 in referring to hemoproteins of biological materials other than liver microsomes, based simply on similarity in the wavelength of the α peak, may not be desirable unless: a, there is similarity of the low temperature spectrum in the reduced state ($77^{o}K$ or lower); b, there is similarity in amino acid sequence or immunochemical similarity; c, a similar reactivity to the microsomal cytochrome b_5 reductase is shown.

Besides the microsomes of animal liver, cytochrome b_5 is contained in the outer membranes of mitochondria of the same tissue (4,5). A similar cytochrome contained in erythrocytes has been purified and sequenced (1, 6-8). Spectrally similar pigments are also found in yeast (9) and plants (10, 49, 50).

2. Functions of cytochrome b_5

1). An electron carrier in fatty acid desaturation

Studies of an in vitro system employing microsomal membranes from animal liver capable of desaturating fatty acids showed that the overall reaction had a requirement for oxygen and NADH (11). The first studies examined the requirements for the introduction of the $\Delta 9$ double bond by the microsomal fraction of rat liver, using stearic and palmitic acids as substrates. The same system was shown to be

able to introduce a $\Delta 6$ double bond into oleic acid to form $18:2^{\Delta 6,9}$ (12). Oshino et al. suggested that the desaturation was associated with the microsomal electron transport chain and possibly involved cytochrome b_5, not cytochrome P-450, because cyanide inhibited the desaturation, whereas carbon monoxide did not (13). The first definitive report implicating the NADH-specific microsomal electron transport chain showed that the rate of reoxidation of reduced cytochrome b_5 was increased by stearoyl-CoA (51). The absolute requirement for cytochrome b_5 was shown by removing endogenous cytochrome b_5 from detergent-solubilized microsomes and observing the restoration of desaturase activity upon addition of the purified cytochrome b_5 (14-16). The successful in vitro reconstitution of lipid desaturation was done by adding the purified cytochrome b_5 reductase, cytochrome b_5, and stearoyl-CoA desaturase together plus substrates in an artificial membrane (17, 18). The involvement of these microsomal electron transport components in other desaturase reactions has been demonstrated by the inhibition of the particular desaturase reaction by antibodies raised to the purified cytochrome b_5. These results have led to the scheme in figure 2 below.

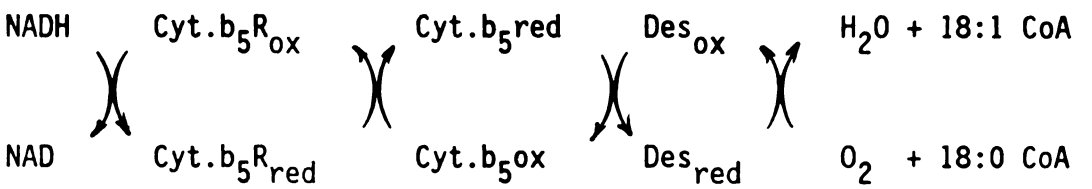

Figure 2. The role of cytochrome b_5 in fatty acid desaturation. Cyt.b_5R=Cytochrome b_5 reductase. Cyt.b_5=Cytochrome b_5. Des=Desaturase.

2). Reduction of methemoglobin in erythrocytes

The methemoglobin reduction system of red blood cells catalyzes the reduction of the four ferric ions of methemoglobin to ferrous ions (21). This reduction proceeds at a rate that is sufficient to maintain approximately 99% of the hemoglobin in its ferrous state, despite the continuous conversion of hemoglobin to methemoglobin by various oxidants of the cells. Under normal conditions, most of the methemoglobin reduction can be attributed to catalysis by an NADH-utilizing system, an NADH-dependent reductase. Since the rate of methemoglobin reduction catalyzed by purified reductase was slow relative to the rate observed in intact cells, the existence of a second component of the system was suggested. Also, there is no correlation between the rate of methemoglobin reduction in intact cells and the amount of NADH-specific reductase that can be detected in these cells (20). A soluble cytochrome b_5 in erythrocytes markedly stimulates the catalysis of methemoglobin reduction by the reductase (21). At concentrations present in erythrocytes, cytochrome b_5 serves as an effective substrate for erythrocyte NADH-reductase, and the resulting ferrocytochrome b_5 then transfers an electron to methemoglobin as follows:

$$1/2\ NADH + Cyt.b_5\ (Fe^{3+}) \xrightarrow{\text{reductase}} 1/2\ NAD^+ + Cyt.b_5\ (Fe^{2+})$$

$$4\ Cyt.b_5\ (Fe^{2+}) + Hemoglobin\ (Fe^{3+}) \longrightarrow 4\ Cyt.b_5\ (Fe^{3+}) + Hemoglobin\ (Fe^{2+})$$

The name erythrocyte cytochrome b_5 arises from the spectral and structural similarity of the protein to microsomal cytochrome b_5 (22, 23). The reductase has been termed erythrocyte cytochrome b_5 reductase because it acts upon erythrocyte cytochrome b_5 and because it is enzymically similar to microsomal cytochrome b_5 reductase (24).

3). Reduction of cytochrome P-450

The scheme in figure 3 has been proposed for the mechanism of cytochrome P-450 in hydroxylation reactions (25).

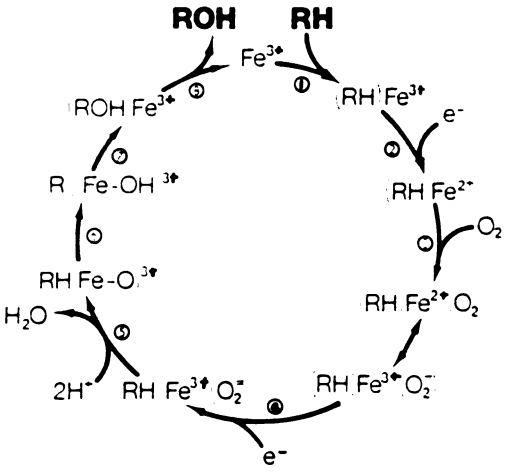

Figure 3. The pathway of oxygenation by cytochrome P-450. Cytochrome b_5 was proposed to provide the second electron for the P-450 catalyzed hydroxylation reactions (25).

As indicated in the scheme, substrate binding to native, ferric P-450 is followed by reduction to the ferrous state, thereby allowing oxygen binding. A second reduction results in spliting of the oxygen-oxygen bond, one atom being lost as water. The other oxygen atom, presumably now an "activated oxygen," is inserted into a carbon-hydrogen bond of the substrate to produce the corresponding alcohol, which is then released with regeneration of the ferric form of the enzyme and completion of the catalytic cycle. Hildebrandt and Estabrook (26) suggested that cytochrome b_5 may supply the second electron. Miki et al. purified a form of cytochrome P-450 with a high affinity for cytochrome b_5 and showed that reduction of the P-450 by NADH required NADH-cytochrome b_5 reductase, cytochrome b_5, and suitable concentrations of detergents (27). More recently, Pompon and Coon proposed a new model for the involvement of cytochrome b_5 in the P-450 related reactions (28) shown in figure 4.

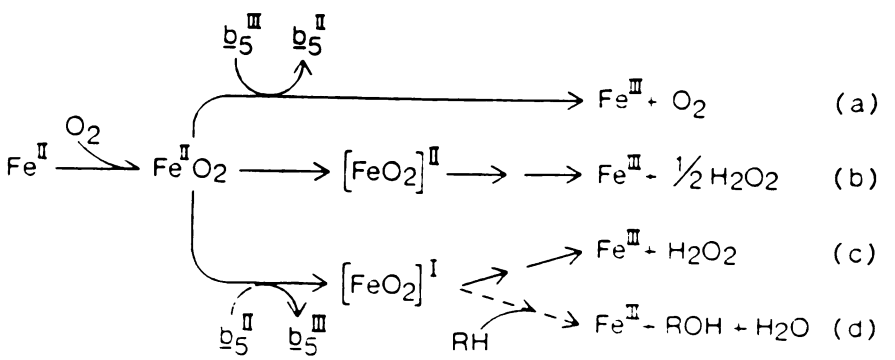

Figure 4. Model for cytochrome b_5 in cytochrome P-450 related reactions (28).

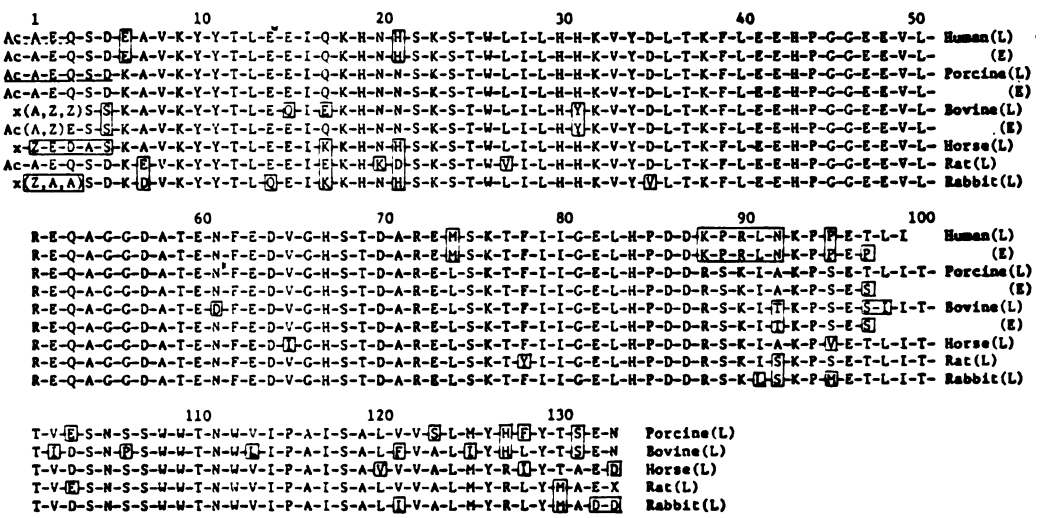

Figure 5. The primary structures of cytochrome b_5 proteins from different animal sources (1). [L] designates a protein from liver and [E] indicates an erythrocyte protein.

On the basis of this model, competition between spontaneous decomposition of the ferrous dioxygen intermediate and its reduction by cytochrome b_5 is believed to contribute to the partition between abortive hydrogen peroxide production and substrate hydroxylation in this enzyme system.

3. Properties of cytochrome b_5

1). Primary structure of the protein from vertebrates

The microsomal cytochrome b_5 in its native state is an amphipathic protein with an Mr of 16,000. It contains two domains: a N-hydrophilic catalytic segment consisting of about 80 amino acid residues, and a C-hydrophobic segment that is required for binding the cytochrome b_5 to the microsomal membrane (29, 30). Controlled proteolytic digestion of the native protein yields a water soluble cytochrome b_5 with a Mr of 11,000. The sequence of this soluble cytochrome b_5 has been determined (31) and is very similar to that of cytochrome b_5 found in the supernatant fraction of erythrocytes. The complete microsomal cytochrome b_5 sequences have been determined from 6 different animal species (1). Whereas the microsomal polypeptide is 133 amino acids long, the erythrocyte b_5 is 97 residues long (1). Figure 5 shows a comparison of some known cytochrome b_5 sequences (see the opposing page).

The sequence homology between the various forms of cytochrome b_5 is very striking. The sequence of residues 104-126 contains only hydrophobic or uncharged hydrophilic side chains. The basic and

acidic amino acid residues occur at the COOH-terminus and the peptide linking the membranous segment to the globular heme-carrying segment. Such residues may be expected to be present outside the hydrophobic milieu of the lipid membrane, thereby suggesting that the hydrophobic segment either penetrates the membrane or folds back on itself so that the COOH-terminus is near the cytoplasmic surface of the membrane. Although the soluble cytochrome b_5 accepts electrons from cytochrome b_5 reductase (33), the complete cytochrome b_5, including the membranous segment, is necessary for functional reconstruction of the stearyl-CoA desaturase system (17, 18). The spectral properties of complete cytochrome b_5 are essentially the same as those of the soluble form (heme peptide segment) (34).

2) Plant cytochrome b_5 structure

Microsomal cytochrome b_5 has also been discovered and characterized in plants (10, 49, 50). Bonnerot et al. first purified cytochrome b_5 from potato tubers by 350 fold and this protein is very similar to animal cytochrome b_5 in terms of its Mr (16700) and its absorption spectrum (49). Later, Madyastha et al. reported the purification of a very similar cytochrome protein to 30% homogeneity from Catharanthus roseus, and this protein has a Mr of 16500 (50). Jollie et al. (10) purified the microsomal cytochrome b_5 from Pisum sativum, and sequenced the N-terminal part of this cytochrome b_5 (figure 6). There is no similarity between this sequence and any animal cytochrome b_5 protein. However, they presented results which indicate that the antibody raised against rat cytochrome b_5

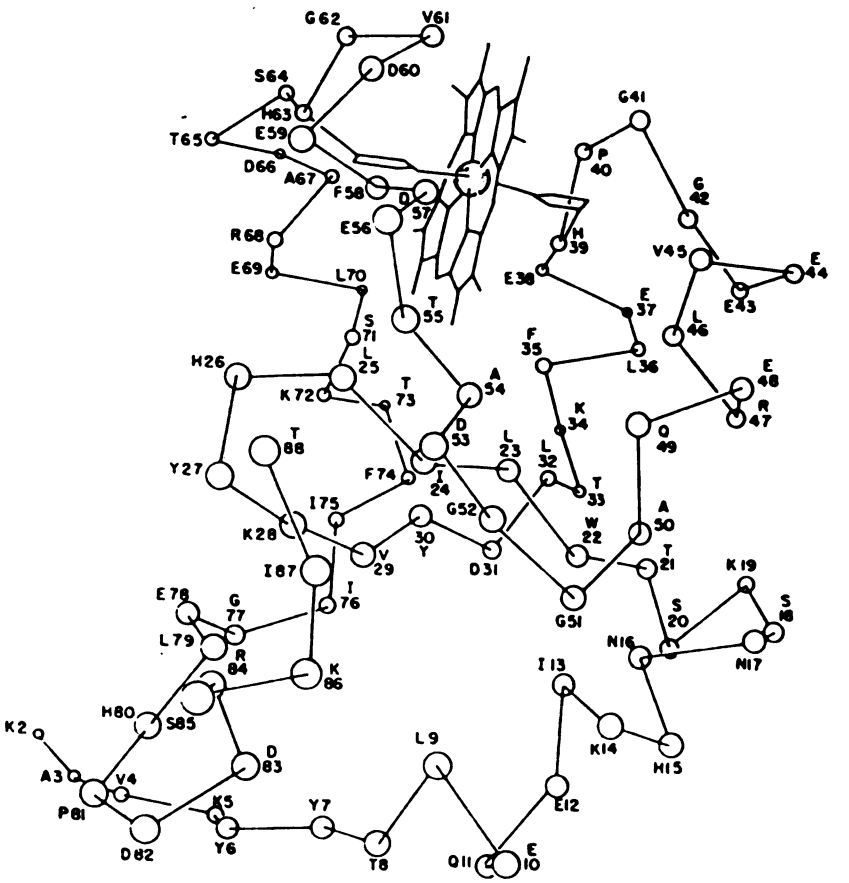

Figure 7. Schematic diagram of the backbone chain of cytochrome b_5 which was solubilized from liver microsome with panccreatic lipase (36).

recognized the pea cytochrome b_5 protein on western blots. This suggests that the conservation of this protein extends into the plant kingdom, at least at the epitope level.

NH_2-Ala-Leu-Leu-Gln-Glu-Asp-Glu-Ala-Ile-Asp-Asp
-Phe-Asp-Phe-Asp-Asp-Gly-Ala-Lys-Asp-Asp-Asp-Gly

Figure 6. NH_2-terminal amino acid sequence of pea cytochrome b_5 (10).

Even though the protein was purified and partially sequenced, its function in plants is still not known. Possibly it is involved in reactions like those in animals, since the microsomal electron transport systems of higher plants are involved in a variety of central metabolic transformations including fatty acid desaturation and the mixed function oxidase activities of the cytochrome P-450 dependent monooxygenases. Microsomal cytochrome b_5 is certainly a good candidate for a member of the electron transport system.

3) Ternary structure

The X-ray crystallographic studies of calf liver cytochrome b_5 solubilized by pancreatic lipase (with 93 amino acid residues) at 2.8 Å and 2.0 Å were carried out by Mathews et al (2, 35, 36). The ternary structure based on their analysis is presented in figure 7 on the opposing page.

Like many other proteins, the interior of the molecule is distinctly nonpolar. The heme is buried in a hydrophobic crevice

with two vinyl groups lying deep in the interior of the molecule. One of the propionic acid groups lies on the surface of the molecule and is hydrogen-bonded to the γ-oxygen and the peptide nitrogen of Ser-64, and the other projects outward into solution. The walls of the heme crevice are formed by two pairs of roughly antiparallel helices and the floor by the pleated sheet structure. The iron atom is coordinated by His-39 and His-63 which extend from the wall of the crevice. The nitrogens of the two histidines are hydrogen-bonded to the main chain carbonyl oxygens of Gly-41 and Phe-58, respectively. Furthermore, His-39 is in van der Waals contact with Leu-46 and His-63 lies close and paralled to Phe-58, suggesting a $\pi - \pi$ interaction between the latter pair of residues. Thus, the histidine residues are held firmly in place by the rigidity of the backbone structure and by a variety of interactions with the main and side chains. The core part (residues 3 to 86) contains the heme group at the top, lying in the hydrophobic crevice, and also has a narrow hydrophobic group open to the aqueous environment. The residues principally involved in this latter group are Phe-35, Leu-70, and Phe-74. The site of the action of the NADH-cytochrome b_5 reductase may be this group, and the two phenylalanines in the vicinity of the group may provide a path for an electron to the heme.

von Bodman et al (44) chemically synthesized a gene coding for rat liver cytochrome b_5 and expressed it in Escherichia coli. Transformants containing the soluble core of cytochrome b_5 produced holoprotein containing the protoporphyrin IX prosthetic group in amounts up to 8% of the total cellular protein. The complete cytochrome b_5 gene including the membrane anchor domain was also

efficiently expressed in E. coli with incorporation of the holoprotein into the membrane fraction of the cell. The successful expression of cytochrome b_5 in E. coli means that it is possible to construct mutant cytochrome b_5 forms with alterations at specifically selected amino acids within the product protein. von Bodman et al. (44) replaced histidine-63 with alanine by cassette mutagenesis. The resulting protein failed to incorporate heme during fermentative growth and was not reconstituted with exogenous heme after purification of the apoprotein. Mutant cytochrome b_5 protein with a methionine substituted at position 63 resulted in the production of the apo form of the cytochrome in high yield. Purification of this apoprotein following the identical procedure for the wild-type holoprotein allowed reconstitution with heme to form the intact mutant protein. This methionine-63 cytochrome b_5 displayed an axial high spin ESR signal (g=6) and optical spectra in the ferric form, which was interpreted as evidence that the methionine sulfur was not bonded to the heme iron. Consistent with this interpretation, the reduced protein was found to readily bind carbon monoxide with a 420 nm Soret maxima similar to that observed for myoglobin and hemoglobin. It appears that the methionine-63 protein is a state five-coordinate heme protein. It will be very informative to see the results after the surface charge distributions, the composition of the hydrophobic membrane anchor domain, and the residues that potentially control the redox potential and electron transfer rate have been altered.

4. Cytochrome b_5-lipid interaction

There are two distinct mechanisms for the integration of de novo-synthesized polypeptides into cell membranes. One is specified by an "insertion" sequence and proceeds unassisted into any exposed cell membrane, resulting in the anchorage of a hairpin-loop domain of the polypeptide chain into the lipid bilayer; such a hairpin loop could easily extend into the hydrophilic milieu on the other side of the membrane. The other one is mediated by a "signal" sequence and is dependent on a signal-specific receptor that effects the translocation of a domain of the polypeptide from the biosynthetic compartment to the other side of a specific cell membrane. The membrane bound cytochrome b_5 is first synthesized on free ribosomes (45), then bound to microsomal membranes without being recognized by any receptors (37).

Using the binding of the cytochrome b_5 to artificial phospholipid vesicles as a model system, Enoch et al (38) found two types of protein binding: one was capable of intermembrane transfer, the other was not. Based on these properties, they proposed a model for two different orientations of the protein in the membrane (figure 8).

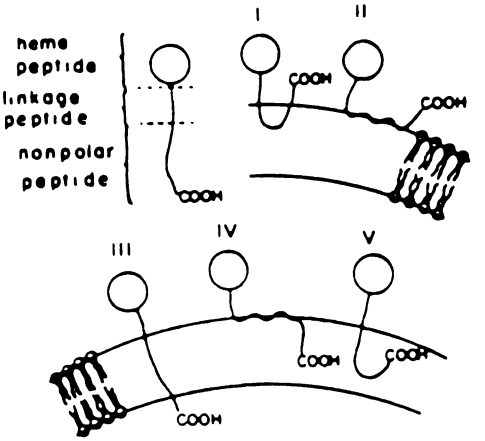

Figure 8. Models for insertion of cytochrome b_5 in artificial membranes (38).

The ability of cytochrome b_5 to transfer from artificial membranes to other membranes, but not from biological membranes, may reflect a difference in the nature of the protein binding to the membrane. A nontransferable form of cytochrome b_5, which may represent the microsomal type of binding, was obtained when cytochrome b_5 was bound during the formation of phosphatidylcholine vesicles. A soluble, heme peptide fragment of cytochrome b_5 was released when vesicles containing cytochrome b_5 in the transferable form were incubated with carboxypeptidase Y. In contrast, the nontransferable form of cytochrome b_5 in microsomes and artificial vesicles was not released by carboxypeptidase Y treatment (39, 40). When cytochrome b_5 binds to pure, unperturbed bilayers, the loose binding form is predominately obtained. However, if the bilayer is in a perturbed state due to the presence of deoxycholate or another integral membrane protein (i.e., desaturase) in the bilayer, then

cytochrome b_5 is inserted in the tight binding form. On the basis of their studies of cytochrome b_5, Enoch et al. concluded that integral membrane proteins in general do not readily undergo intermembrane transfer between biological membranes (38).

As to the mechanism and topology of the interaction between the hydrophobic domain of cytochrome b_5 and the membrane, we know very little. The results of a predictive analysis for conformational features, according to the rules of Chou and Fasman (41,42), are similar for the amino acid sequence of the membranous segment from equine and bovine proteins. The sequence from 103 to 112, which contains a cluster of three tryptophanyl residues, seems to consist of 3-4 overlapping β-turns. However, Jagow and Sebald stressed that the prediction must be viewed with caution, because the conformational parameters employed were derived from studies on globular hydrophilic proteins and may not be necessarily extendable to membranous peptides (43).

5. Erythrocyte cytochrome b_5

On the basis of studies done with bovine erythrocyte cytochrome b_5, Hultquist et al. (32) proposed that the soluble erythrocyte cytochrome b_5 is derived during erythropoiesis by proteolytic cleavage of the membrane-bound cytochrome b_5 present in the endoplasmic reticulum of the proerythroblasts. Bovine erythrocyte cytochrome b_5 is indistinguishable from protease-solubilized liver microsomal cytochrome b_5 on the basis of spectral properties (22) and ability to react with other redox proteins (24). Cytosolic

cytochrome b_5 is not present in an immature erythroid cell, but instead, a membranous form of the cytochrome b_5 is present (45). An electron microscope study has shown that the endoplasmic reticulum disappears during erythroid maturation (32). There are reports of the existence of stromal proteases which are activated after hemolysis (46, 47). It is possible that a particular class of proteases convert membrane-bound cytochrome b_5 into cytosolic cytochrome b_5 during erythroid maturation. Since bovine liver lysosomal proteases can digest microsomal cytochrome b_5 to produce hydrophilic segments which correspond to erythrocyte cytochrome b_5 in vitro, these proteases can serve as a good model for the putative erythroid proteases which solubilize microsomal cytochrome b_5 during erythroid maturation.

Comparison of the cytochrome b_5 sequences of both erythrocyte and liver forms in species such as bovine support the hypothesis of Hultquist and coworkers (32), because the sequence of erythrocyte cytochrome b_5 is identical to liver cytochrome b_5 from residue 1 to 97. The problem is that residue 97 is proline for human erythrocyte cytochrome b_5 and serine for the porcine protein, while residues 97 for human and porcine liver cytochrome b_5 are threonine. Three possibilities exist to explain the above problem. 1. There are two or more cytochrome b_5 genes in those species, and the cytosolic cytochrome b_5 and microsomal cytochrome b_5 are encoded by two different, but closely related genes; 2. There is only one cytochrome b_5 gene which gives rise to more than one form of cytochrome b_5 protein by an alternative RNA splicing mechanism. It has been shown in mouse that four forms of myelin basic protein are

encoded by a single gene, and all four mRNAs are produced through an alternative splicing mechanism (48). 3. There is only one cytochrome b_5 gene, but different cytochrome b_5 proteins are due to posttranslational modifications. Proteolytic cleavage of the membrane-bound cytochrome b_5 to produce the cytosolic cytochrome b_5 in erythrocyte cells can explain the bovine case. In human and porcine, there may be one more modification after the proteolytic cleavage, the addition of one amino acid to the C-terminal of the proteolytically processed protein. I am not aware of any precedents for the third possibility. The investigation of this problem represents one of the main goals of the work described in this study.

6. Acknowledgments

I thank the authors and publishers for letting me use their figures in this chapter (figures 3 to 8 from 25, 28, 1, 10, 36, and 38).

7. References

1). K. Abe, S. Kimura, R. Kizawa, F. K. Anan, Y. Sugita (1985). J. Biochem. (Tokyo) 97:1659-1668

2). F. S. Mathews, P. Argos, and M. Levine (1972). Cold Spring Harbor Symp. Quant. Biol. 36:387-393

3). B. Hagihara, N. Sato, and T. Yamanaka (1975). The Enzymes. 11:549-593

4). I. Raw, N. Petragnani, and O. C. Nogueira (1960). J. Biol. Chem. 235:1517-1520

5). D. F. Parsons, G. R. Willians, W. Thompson, D. F. Wilson, and B. Chance (1967). Mitochondrial Structure and Compartmentation (E. Quagliariello et al., eds.) p.5. Adriatice Editrice, Bari.

6). D. E. Hultquist, D. W. Reed, P. G. Passon and W. E. Andrews (1971). Biochem. Biophys. Acta 229: 33-41

7). D. E. Hultquist, R. T. Dean, and R. H. Douglas (1974). Biochem. Biophys. Res. Commun. 60:28-34

8). D. A. Schafer, and D. E. Hultquist (1983). Biochem. Biophys. Res. Commun. 115:807-813

9). Y. Yoshida, Y. Tamura-Higashimaki, and R. Sato (1983). Arch. Biochem. Biophys. 220:467-476

10). D. R. Jollie, S. G. Sligar, and M. Schuler (1987). Plant Physiol. 8:457-462

11). J. B. Marsh, and A. T. James (1962). Biochem. Biophys. Acta 60:320-328

12). P. W. Hollway, R. Peluffo, and S. J. Wakil (1963). Biochem. Biophys. Res. Commun. 12:300-304

13). N. Oshino, Y. Imai, and R. Sato (1966). Biochem. Biophys. Acta 128:13-28

14). P. W. Hollow (1971). Biochemistry. 10:1556-1560

15). P. W. Hollow, and J. T. Wakil (1972). Biochemistry 11:3689-3696

16). T. Shimakata, K. Mihara, and R. Sato (1972). J. Biochem. (Tokyo) 72:1163-1174

17). P. Strittmatter, L. Spatz, D. Rogers, M. J. Setlow, and R. Redline (1974). Proc. Natl. Acad. Sci. USA 71:4565-4569

18). M. R. Prasad, and V. C. Joshi (1979). J. Biol. Chem. 254:6362-6369

19). D. Hultquist (1979). Methods Enzymol. 52:463-473

20). Y. Kanazawa, M. Hattori, K. Kosaka, and K. Nakao (1968). Clin. Chim. Acta 19:524

21). D. Hultquist, and P. G. Passon (1971). Nature New Biol. 229:252-254

22). P. G. Passon, D. W. Reed, D. W. Reed, and D. E. Hultquist (1972). Biochem. Biophys. Acta 275:51-61

23). D. E. Hultquist, R. H. Doulas, and R. T. Dean (1975). In "Erythrocyte Structure and Function" (G. Brewer, ed.), p.297. Alan R. Liss, Inc., New York.

24). P. G. Passon, and D. E. Hultquist (1972). Biochem. Biophys. Acta 275:62-73

25). R. E. White, and M. J. Coon (1980). Ann. Rev. Biochem. 49:315-356

26). A. Hildebrandt, R. W. Estabrook (1971). Arch. Biochem. Biophys. 143:66-79

27). N. Miki, T. Sugiyama, and T. Yamano (1980). In "Microsomes, Drug Oxidations, and Chemical Carcinogenesis", ed. M. J. Coon, A. H. Conney, R. W. Estabrook, H. Gelboin, J. R. Gillette, and P. J. O'Brien, pp. 469-478. New York: Academic.

28). D. Pompon, and M. J. Coon (1984). J. Biol. Chem. 259:15377-15385

29). P. Strittmatter, M. J. Rogers, and L. Spatz (1972). J. Biol. Chem. 247:7188-7194

30). L. Spatz, and P. Strittmatter (1971). Proc. Natl. Acad. Sci.

68:1042-1046

31). F.G. Nobrega, and J. Ozols (1971). J. Biol. Chem. 246:1706-1717

32). S. R. Slaughter, C. H. Williams, and D. E. Hultquist (1982). Biochem. Biophys. Acta 705:228-237

33). Y. Imai, and R. Sato (1977). Biochem. Biophys. Res. Commun. 75:420-426

34). P. Strittmater, P. Fleming, M. Connors, and D. Corcoran (1978). Methods Enzymol. 52:97-101

35). F. S. Mathews, M. Levine, and P. Argos (1971). Nature (London), New Biol. 233:15-16

36). F. S. Mathews, M. Levine, and P. Argos (1972). J. Mol. Biol. 64:449-464

37). D. J. Anderson, K. E. Mostov, and G. Blobel (1983). Proc. Natl. Acad. Sci. 80:7249-7253

38). H. G. Enoch, P. J. Fleming, and P. Strittmater (1979). J. Biol. Chem. 254:6483-6488

39). S. Tajima, and R. Sato (1980). J. Biochem (Tokyo) 87:123-134

40). K. Christiansen, and J. Carlsen (1985). Biochem. Biophys. Acta 815:215-222

41). P. Y. Chou, and G. D. Fasman (1977). J. Mol. Biol. 115:135-175

42). P. Y. Chou, and G. D. Fasman (1977). Trends Biochem. Sci. 2:128-131

43). G. von Jagow, and W. Sebald (1980). Ann Rev. Biochem. 49:281-314

44). S. B. von Bodman, M. A. Schuler, D. R. Jollie, and S. G. Sligar (1986). Proc. Natl. Acad. Sci. 83:9443-9447

45). S.R. Slaughter, and D.E. Hultquist (1979). J. Cell Biol.

83:231-239

46). S.L. Morrison, and H. Neurath (1953). J. Biol. Chem. 200:39-51

47). K.I. Altman (1959). Am. J. Med. 27:936

48). F. Ferra, H. Engh, L. Hudson, J. Kamholz, C. Puckett, S. Molineaux, and R.A. Lazzarini (1985). Cell 43:721-727

49). C. Bonnerot, A. Galle, A. Jolliot, and J.C. Kader (1985). Biochem. J. 226:331-334

50). K. Madyastha, and N. Krishnamachary (1986). Biochem. Biophys. Res. Commun. 136:570-576

51). N. Oshino, Y. Imai, and R. Sato (1971). J. Biochem. (Tokyo) 69:155-167

CHAPTER II

THE PRIMARY STRUCTURE OF CHICKEN LIVER CYTOCHROME B_5 DEDUCED FROM THE DNA SEQUENCE OF A cDNA CLONE

------Reprint from ARCHIVES OF BIOCHEMISTRY
AND BIOPHYSICS, Vol. 264, No. 1, July,
pp343-347, 1988

ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS
Vol. 264, No. 1, July, pp. 343-347, 1988

COMMUNICATION

The Primary Structure of Chicken Liver Cytochrome b_5 Deduced from the DNA Sequence of a cDNA Clone

HONG ZHANG AND CHRIS SOMERVILLE[1]

Genetics Program and DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824

Received January 25, 1988, and in revised form April 11, 1988

A cDNA clone encoding the chicken liver cytochrome b_5 was isolated by probing a library with synthetic oligonucleotides based on a partial amino acid sequence of the protein. Determination of the DNA sequence indicated a 414-nucleotide open reading frame which encodes a 138-amino acid residue polypeptide. The open reading frame contains 6 amino acids at the amino terminus which were not present on any of the cytochrome b_5 polypeptides for which the amino acid sequence has been determined directly, suggesting that the protein is proteolytically processed to the mature form. The results of genomic Southern analysis were consistent with the presence of two structurally different genes in the chicken genome, raising the possibility that the soluble and membrane-bound forms of the protein are the products of separate genes.

© 1988 Academic Press, Inc.

Liver microsomal cytochrome b_5 is an amphipathic membrane protein consisting of an N-terminal hydrophilic domain which contains a functional heme as a catalytic site and a C-terminal hydrophobic domain which anchors the protein in the microsomal membrane (1, 2). It functions as a component of the microsomal stearyl-CoA desaturase (3, 4), and is also involved in liver cytochrome *P*-450 reduction (5, 6). Determination of part or all of the amino acid sequences of liver cytochrome b_5 from six different vertebrate species (7-14) has revealed that the primary structures are highly conserved (15). The proteins characterized to date have a molecular mass of about 16 kDa and contain 133 amino acid residues. The protein has been extensively studied as a model for protein-protein interaction, protein-membrane interaction, and the dynamics of heme protein folding (16-19). However, many questions remain concerning the mechanism and topology of the interaction of the cytochrome with membranes, the structure and regulation of the genes which code for cytochrome b_5, and the structure of the protein in nonvertebrates.

[1] To whom correspondence should be addressed.

Although no genes encoding cytochrome b_5 have been cloned previously, a gene encoding rat liver cytochrome b_5 has been synthesized and expressed in *Escherichia coli* (20). Since there are several questions which cannot be addressed with a synthetic gene we have undertaken the cloning and sequencing of a chicken liver cDNA which encodes the membrane-associated cytochrome b_5.

EXPERIMENTAL PROCEDURES

Materials. A chicken liver λ-gt11 cDNA library, constructed by blunt-end ligating *Eco*RI linkers to cDNA, was kindly provided by J. Dodgson (Michigan State University). Oligonucleotides were synthesized by the phosphoramidite method on an Applied Biosystems 380A instrument. The plasmid pBluescript (KS$^+$) was purchased from Stratagene (San Diego, CA).

Plaque screening. The cDNA library was plated on *E. coli* Y1090 and nitrocellulose plaque lifts were screened with the oligonucleotide mixtures b_5-1 and b_5-4 (Fig. 1) which were end-labeled to an average specific activity of 10^9 dpm μg^{-1} with [γ-^{32}P]ATP (3000 Ci mmol^{-1}) and T4 polynucleotide kinase (21). Filters were prehybridized 3 to 5 h at 42°C in 6× SSC

0003-9861/88 $3.00
Copyright © 1988 by Academic Press, Inc.
All rights of reproduction in any form reserved.

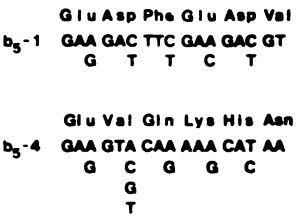

FIG. 1. The oligonucleotide mixtures used as probes for the cytochrome b_5 gene.

(1X SSC is 150 mM NaCl, 15 mM sodium citrate adjusted to pH 7.0), 50 mM $NaPO_4$ (pH 6.8), 5× Denhardt's solution (0.1% (w/v) Ficoll, 0.1% (w/v) polyvinylpyrrolidone, and 0.1% (w/v) bovine serum albumin), 100 $\mu g\ ml^{-1}$ of sonicated herring DNA (Sigma). The hybridizations were carried out at 37°C for 24 to 30 h in the same solutions with the addition of 10% (w/v) dextran sulfate and 1-2 pmol ml^{-1} of labeled oligonucleotide. The temperature for final washing was based on an empirical formula (22) by assuming that all ambiguous positions contained A or T bases. The filters were washed four times for 5 min in 6X SSC at 22°C, twice for 30 min at 37°C, and once for 20 min at 42°C (for b_5-1) or 40°C (for b_5-4).

DNA sequence analysis. The 0.8-kb insert in a recombinant phage was excised as a single fragment by cleavage with *Eco*RI and subcloned in both orientations into the *Eco*RI site of pBluescript to produce the recombinant plasmids pHZ5-1 and pHZ5-2. A series of overlapping unidirectional deletions were made in plasmids pHZ5-1 and pHZ5-2 with exonuclease III and mung bean nuclease essentially as described (23). The deleted inserts from both orientations were self-ligated to produce a series of overlapping plasmids which were then transformed into *E. coli* DH5α. The plasmids were sequenced on both strands as double-stranded DNA by the chain termination method as described (24).

Genomic Southern analysis. The chicken DNA was purchased from Clontech Laboratories (Palo Alto, CA). DNA (5 μg per lane) was digested to completion with restriction endonucleases, resolved by electrophoresis in 0.8% agarose gels containing 89 mM Tris-borate (pH 8.2), 2 mM EDTA, and transferred to nitrocellulose filters as described (21). The filters were prehybridized for 4 to 6 h at 42°C in 50% (v/v) formamide, 5X SSC, 50 mM $NaPO_4$ (pH 6.8), 5X Denhardt's solution, 250 $\mu g\ ml^{-1}$ of sonicated herring DNA. The hybridization was carried out at 42°C for about 20 h in the same solutions with the addition of 0.8 μg of probe DNA which was nick translated with [α-^{32}P]dCTP to a specific activity of about 10^8 dpm μg^{-1} (21). After hybridization, the filters were washed three times for 20 min at 42°C in 2X SSC, 0.1% SDS, then twice for 30 min at 65°C in 0.1X SSC, 0.1% SDS.

RESULTS AND DISCUSSION

From a partial amino acid sequence of the chicken liver cytochrome b_5 (13) we designed two nonoverlapping oligonucleotide mixtures of 17-mers (Fig. 1) which had the lowest possible degree of ambiguity. It was subsequently learned that the region of amino acid sequence used to design oligonucleotide mixture b_5-1 erroneously contained an Asp residue instead of an Asn residue (14). However, this resulted in only one incorrect nucleotide in the oligonucleotide mixture and did not prevent the effective use of the mixture as a hybridization probe. These oligonucleotides were used to screen a λgt11 cDNA library constructed from chicken liver poly(A)$^+$ RNA. Among the 200,000 plaques screened, 41 hybridized to oligonucleotide mixture b_5-1. Only 13 out of the 41 clones were also recognized by oligonucleotide mixture b_5-4 which was derived from a sequence near the amino terminus of chicken cytochrome b_5. The size of the inserts in these 13 phage were determined by restriction analysis, the largest insert, a 0.8-kb *Eco*RI fragment, from one of the 13 phage was subcloned in both orientations into the *Eco*RI site of pBluescript, and the DNA sequence was determined.

The DNA sequence of the cDNA clone encoding chicken liver cytochrome b_5 and the deduced amino acid sequence of the open reading frame is shown in Fig. 2. The clone lacked a 3′-poly(A) sequence, suggesting incomplete methylation during library construction. Because the clone had only 20 nucleotides upstream of the first ATG codon, it appears likely that some of the mRNA leader sequence is also missing. Thus, it is not possible to exclude the possibility that translation begins at a codon further upstream. However, the deduced amino acid sequence is in agreement at each residue with at least one of the two independently obtained partial amino acid sequences of the chicken protein which were previously obtained from residue 8 to 91 (13, 14). The open reading frame of 414 nucleotides encodes a polypeptide of 15,544 Da containing 138 amino acid residues. All of the previously determined liver cytochrome b_5 sequences contain 133 amino acid residues and, where it has been unambiguously determined (15), begin with an N-acetylated alanine (designated Ala 1 in Fig. 2). Although the amino terminus for the chicken protein was not previously determined, the apparently ubiquitous presence of an N-terminal alanine on the other vertebrate proteins raises the possibility that the chicken and other forms of the protein are proteolytically processed from a larger precursor.

Alignment of all available microsomal cytochrome b_5 sequences indicates striking similarity between the chicken and other sources of cytochrome b_5 (Fig. 3). As with the cytochrome b_5 from other vertebrates, most of the sequence heterogeneity is located at the N- and C-terminal ends. The sequence from residues 42 to 72, which forms the heme-binding site (25), is

```
          10         20         30         40         50         60
GGCTGTCTGTCAAGCGAGGAT ATG GTG GGC TCC AGT GAA GCC GGC GGT GAG GCG TGG CGG GGC
                      Met Val Gly Ser Ser Glu Ala Gly Gly Glu Ala Trp Arg Gly
                         -5                   1                   5

       70         80         90         100        110        120
CGC TAC TAT CGG CTG GAG GAG GTG CAG AAG CAT AAC AAC AGC CAG AGC ACC TGG ATC ATC
Arg Tyr Tyr Arg Leu Glu Glu Val Gln Lys His Asn Asn Ser Gln Ser Thr Trp Ile Ile
    10                 15                 20                 25

       130        140        150        160        170        180
GTG CAC CAC CGT ATC TAC GAC ATC ACC AAG TTC CTG GAT GAG CAC CCT GGT GGA GAA GAA
Val His His Arg Ile Tyr Asp Ile Thr Lys Phe Leu Asp Glu His Pro Gly Gly Glu Glu
    30                 35                 40                 45

       190        200        210        220        230        240
GTC CTT AGG GAG CAA GCT GGG GGA GAT GCT ACT GAG AAC TTT GAA GAT GTT GGC CAC TCT
Val Leu Arg Glu Gln Ala Gly Gly Asp Ala Thr Glu Asn Phe Glu Asp Val Gly His Ser
    50                 55                 60                 65

       250        260        270        280        290        300
ACA GAT GCA AGG GCG CTG TCG GAA ACA TTT ATT ATT GGG GAG CTT CAC CCG GAT GAT AGA
Thr Asp Ala Arg Ala Leu Ser Glu Thr Phe Ile Ile Gly Glu Leu His Pro Asp Asp Arg
    70                 75                 80                 85

       310        320        330        340        350        360
CCG AAG CTT CAG AAA CCA GCA GAA ACT CTT ATT ACC ACT GTG CAG TCT AAT TCC AGT TCA
Pro Lys Leu Gln Lys Pro Ala Glu Thr Leu Ile Thr Thr Val Gln Ser Asn Ser Ser Ser
    90                 95                 100                105

       370        380        390        400        410        420
TGG TCC AAC TGG GTG ATC CCG GCA ATA GCA GCA ATT ATT GTG GCC CTG ATG TAT CGT TCC
Trp Ser Asn Trp Val Ile Pro Ala Ile Ala Ala Ile Ile Val Ala Leu Met Tyr Arg Ser
    110                115                120                125

       430        440       450       460       470       480       490
TAC ATG TCA GAG TGA GCACCTTACTGAGAACTAATGCAAGAAGAGACTGATCTGGGAGAGAATAGAAGCAATCC
Tyr Met Ser Glu ---
    130

  500       510       520       530       540       550       560       570
TAACCCAATATATTTCCTGACAAAAGCCTGATGTCTGAAGATAAATTCAACTTTTTCAGAAAACTGAACAATTCTTTTC

   580       590       600       610       620       630       640       650
TGCTGTGCACTTTTCTTGATGTTGCCTTCTTATTTGCTGCACTGAAGTAATAAAAAGGCAGCATTTCTTTTCGTATAAC

    660       670       680       690       700       710       720       730
AATATATTCTCTAATGAATGATTTGATAACTGTATTAGTTGCTGTATTAAAATAGTTTTGTAAGTAGCATTCTGATTCT

     740       750       760       770
GGTTATATCTTTTTAATCTGTAATGGAGTCTGTCTTGCA
```

FIG. 2. Composite nucleotide sequence of the cDNA for chicken cytochrome b_{5} and the deduced amino acid sequence. The amino acid sequence is numbered from the alanine residue most commonly found at the amino terminus of the vertebrate protein. The regions of homology to the oligonucleotide probes extend from nucleotides 82 to 98 and 217 to 233 for b_{5}-4 and b_{5}-1, respectively.

completely conserved. The chicken polypeptide lacks one amino acid at the C-terminus which is present on all other known sequences. The overall amino acid sequence homologies of chicken liver cytochrome b_{5} with the available sequences for this protein from other species are human 76.8%, porcine 77.4%, bovine 71.8%, rat 78.2%, and rabbit 79%.

In vertebrate erythrocytes, a cytochrome b_{5} is present in the soluble fraction where it is involved in the reduction of methemoglobin (26). The sequence of bovine erythrocyte cytochrome b_{5} was reported to be identical to the liver microsomal protein from residues 1 to 97, suggesting that the erythrocyte protein was derived from the same gene product as the microsomal protein by proteolytic processing during erythroid maturation (27). However, the presence of an amino acid difference at the C-terminal residue of the erythrocyte cytochrome b_{5} from human, porcine (15), and rabbit (7), suggests that mammalian erythrocyte cytochrome b_{5} is encoded by a different mRNA. Since only one amino acid difference was observed between the two forms, the two kinds of mRNA could arise from a single gene by differential mRNA splicing, or could be the products of highly conserved separate genes. In order to examine these possibilities we probed filters containing restriction digests of total chicken DNA with the complete cDNA and with a 3′ *Hin*dIII fragment of the cDNA (nucleotides 307 to 774 in Fig. 2) which encodes only the hydrophobic domain (residues 91 to 132). When

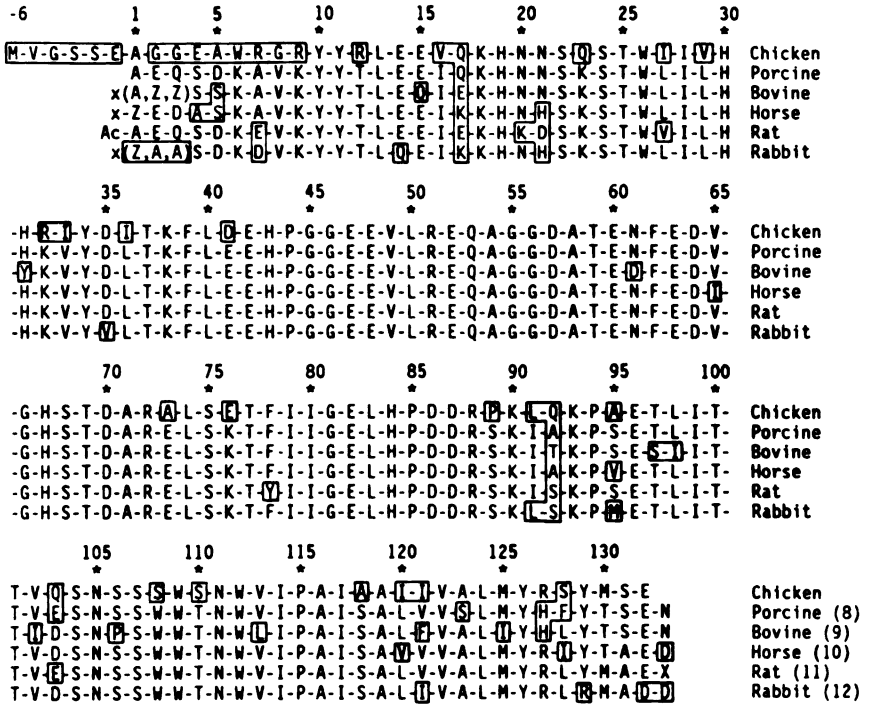

FIG. 3. A comparison of the deduced amino acid sequence of chicken cytochrome b_5 with the sequences obtained by direct amino acid sequencing of the microsomal cytochrome b_5 from other vertebrates. The numbers in parentheses give the references for the sequences. The residues which differ from the most consensus sequence are enclosed in boxes.

the entire cDNA was used as a probe we observed two *Eco*RI bands of 18 and 8.7 kb, three *Hin*dIII fragments of 3.5, 2.3, and 1.4 kb, and two *Bgl*II fragments of 16.5 and 2.3 kb (Fig. 4A). On the basis of other experiments (results not presented), we consider it likely that the slightly reduced intensity of the 18-kb *Eco*RI band was due to incomplete fragmentation by the acid treatment which resulted in incomplete transfer to the nitrocellulose filter. By contrast, when we probed the filters with the region of cDNA encoding the hydrophobic domain, we observed homology only to one *Eco*RI fragment of 8.7, one *Hin*dIII fragment of 3.5 kg, and one *Bgl*II fragment of 16.5 kb (Fig. 4B). There are no internal *Eco*RI or *Bgl*II sites and only one *Hin*dIII site in the cDNA clone. Thus, these results could be explained by the presence of two genes, one of which lacks homology to the region of the cDNA which encodes the hydrophobic domain. These results are also consistent with the presence of one gene containing intron sequences of less than about 5.3 kb total length with one site each for *Eco*RI, *Bgl*II, and *Hin*dIII. An unequivocal resolution of this problem will require the cloning and characterization of a cDNA for the soluble cytochrome b_5 from erythroid cells. In this respect, the

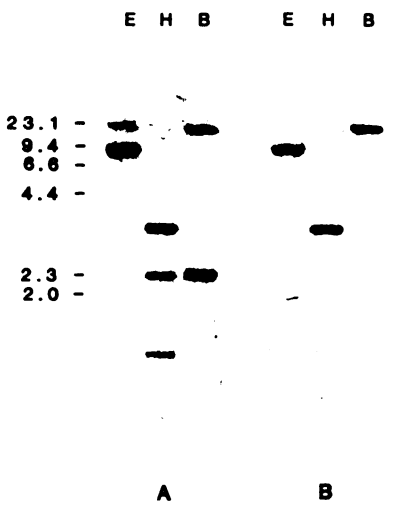

FIG. 4. Hybridization of cytochrome b_5 cDNA probes to chicken genomic DNA. The genomic DNA was digested with *Eco*RI (E), *Hin*dIII (H), and *Bgl*II (B). [A] The entire cDNA was used as a probe. [B] The *Hin*dIII fragment encoding the hydrophobic domain was used as a probe.

chicken is a favorable experimental organism because it has nucleated erythroid cells. The availability of the cDNA clone described here should facilitate a resolution of this and several other problems concerning the structure and function of cytochrome b_5.

ACKNOWLEDGMENTS

We are grateful to J. Dodgson for the gift of the cDNA library. This work was supported in part by grants from the U.S. Department of Agriculture (86-CRCR-1-2046) and the U.S. Department of Energy (AC02-76ER01338).

REFERENCES

1. SPATZ, L., AND STRITTMATTER, P. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 1042-1046.
2. SPATZ, L., AND STRITTMATTER, P. (1973) *J. Biol. Chem.* **248**, 793-799.
3. OSHINO, N., IMAI, Y., AND SATO, J. (1971) *J. Biochem. (Tokyo)* **69**, 155-167.
4. STRITTMATTER, P., SPATZ, L., CORCORAN, D., ROGERS, M. J., SETLOW, B., AND REDLINE, R. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4565-4569.
5. WHITE, R. E., AND COON, M. J. (1980) *Annu. Rev. Biochem.* **49**, 315-356.
6. HILDEBRANT, A., AND ESTABROOK, R. W. (1971) *Arch. Biochem. Biophys.* **143**, 66-79.
7. KIMURA, S., ABE, K., AND SUGITA, Y. (1984) *FEBS Lett.* **169**, 143-146.
8. OZOLS, L., AND GERARD, C. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3725-3729.
9. FLEMING, P. J., DAILEY, H. A., CORCORAN, D., AND STRITTMATTER, P. (1978) *J. Biol. Chem.* **253**, 5369-5372.
10. OZOLS, J., AND GERARD, C. (1977) *J. Biol. Chem.* **252**, 8549-8553.
11. OZOLS, J., AND HEINEMANN, F. S. (1982) *Biochim. Biophys. Acta* **704**, 163-173.
12. KONDO, K., TAJIMA, S., SATO, R., AND NARITA, K. (1979) *J. Biochem. (Tokyo)* **86**, 1119-1128.
13. TSUGITA, A., KOBAYASHI, M., TANI, S., KYO, S., RASHID, M. A., YOSHIDA, Y., KAJIHARA, T., AND HAGIHARA, B. (1970) *Proc. Natl. Acad. Sci. USA* **67**, 442-447.
14. NOBREGA, F. G., AND OZOLS, J. (1971) *J. Biol. Chem.* **246**, 1706-1717.
15. ABE, K., KIMURA, S., KIZAWA, R., ANAN, F. K., AND SUGITA, Y. (1985) *J. Biochem. (Tokyo)* **97**, 1659-1668.
16. DAILY, H. A., AND STRITTMATTER, P. (1979) *J. Biol. Chem.* **254**, 5388-5396.
17. ENOCH, H. G., FLEMING, P. J., AND STRITTMATTER, P. (1979) *J. Biol. Chem.* **254**, 6438-6488.
18. INOKO, Y. (1980) *Biochim. Biophys. Acta* **599**, 359-369.
19. BENDZKO, P., AND PFEIL, W. (1983) *Biochim. Biophys. Acta* **742**, 669-676.
20. BECK VON BODMAN, S., SCHULER, M. A., JOLLIE, D. R., AND SLIGAR, S. G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9443-9447.
21. MANIATIS, T., FRITSCH, E. F., AND SAMBROOK, J. (1982) Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
22. SMITH, M. (1983) *in* Methods of DNA and RNA Sequencing (Weissman, S. M., Ed.), Praeger, New York.
23. HENIKOFF, S. (1984) *Gene* **28**, 351-359.
24. ZHANG, H., SCHOLL, R., BROWSE, J., AND SOMERVILLE, C. (1988) *Nucleic Acids Res.* **16**, 1220.
25. MATHEWS, F. S., ARGOS, P., AND LEVINE, M. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 387-393.
26. HULTQUIST, D. E., AND PASSON, P. G. (1971) *Nature New Biol.* **229**, 252-254.
27. SLAUGHTER, S. R., WILLIAMS, C H., AND HULTQUIST, D. E. (1982) *Biochim. Biophys. Acta* **705**, 228-237.

CHAPTER III

CYTOCHROME B_5 GENE IN CHICKEN

Abstract

A cDNA clone coding for chicken cytochrome b_5 has been isolated from an erythrocyte cDNA library using synthetic oligonucleotides based on a partial amino acid sequence of the protein and the DNA sequence of the previously described chicken liver cytochrome b_5 cDNA clone. The complete homology between the erythrocyte cDNA and the liver cDNA suggests that they are transcribed from the same gene. Both genomic blotting data and the mapping of cytochrome b_5 genomic clones support the notion that there is only one cytochrome b_5 gene in chicken. This gene appears to be responsible for the two forms of cytochrome b_5 protein discovered in different organisms. This suggests that, at least in chicken, the formation of soluble erythrocyte cytochrome b_5 occurs by proteolytic processing of membrane-bound cytochrome b_5 found in liver.

Introduction

Cytochrome b_5 is a heme protein (1,2) which is involved in the fatty acid desaturation in animal liver (3), methemoglobin reduction in erythrocytes (4,5) and cytochrome P-450 reduction (6). It exists in two forms: an amphipathic form in the microsomal membrane of animal liver, and a cytosolic form in erythrocytes. The amphipathic form, which is 133 amino acid residues long, consists of an N-terminal hydrophilic domain which contains a functional heme as a catalytic site and a C-terminal hydrophobic domain which anchors the protein in the microsomal membrane. The cytosolic form is equivalent to the hydrophilic domain of the amphipathic form. Both forms have been purified and sequenced from several species (1, 7). The amino acid sequences of the two forms in a given species are either the same or differ by only one amino acid residue at the C-terminus of the cytosolic form. For example, the primary structure of bovine erythrocyte cytochrome b_5 is identical to its liver form from residues 1 to 97. However, residue 97 is proline for human erythrocyte and serine for porcine erythrocyte forms, whereas residue 97 of both human and porcine liver forms is threonine. This raises possibility that two forms of cytochrome b_5 come from two different mRNAs. But the question as to the whether those mRNAs are transcribed from a single gene or two different genes cannot be answered with available data.

Isolation of a cytochrome b_5 cDNA clone from chicken liver has provided new information about cytochrome b_5 primary structure (8). The genomic Southern analysis using this liver cDNA clone as a probe

```
      Glu Asp Phe Glu Asp Val
b5-1  GAA GAC TTC GAA GAC GT
        G   T   T   C   T

      Glu Val Gln Lys His Asn
b5-4  GAA GTA CAA AAA CAT AA
        G   C   G   G   C
            G
            T

         Gly Arg Tyr Tyr Arg Leu Glu
b5-7 (3')CCG GCG ATG ATA GCC GAC CTC(5')

         Val Ile Pro Ala Ile Ala Ala
b5-8 (3')CAC TAG GGC CGT TAT CGT CGT(5')
```

Figure 1. Oligonucleotides designed from chicken cytochrome b_5 protein (b_5-1, and b_5-4), and from chicken liver cytochrome b_5 clone (b_5-7, and b_5-8).

indicated that either one gene or at most two genes encode cytochrome b_5 in chicken. In this chapter, we describe the isolation and characterization of cDNA clones from erythrocyte cells and genomic clones of cytochrome b_5. The results presented here suggest that there is only one gene in chicken which is responsible for all forms of cytochrome b_5. This suggests that posttranslational modification is the mechanism responsible for the synthesis of the two forms of cytochrome b_5 in chicken.

Materials and methods

Materials:

A chicken erythrocyte lambda gt11 cDNA library was constructed by blunt-end ligating EcoRI linkers to cDNA (26). A chicken Charon 4A genomic library was constructed by collecting 7-23 kb fragments from a partial digestion of genomic DNA with enzymes AluI and HaeIII, then ligating them to EcoRI linkers (25). Both libraries were kind gifts from Dr. J. Dodgson (Department Of Microbilogy and Public Health, Michigan State University). The total RNA used for Northern blots and the beta-globin gene used as an internal standard were from D. Browne and J. Dodgson. The plasmid pBluescript (KS^+) was purchased from Stratagene (San Diego, CA). Oligonucleotides (figure 1) were synthesized by the phosphoramidite method on an Applied Biosystems 380A instrument.

Plaque screening:

The cDNA library was plated on E. coli Y1090 and nitrocellulose plaque lifts were screened with the oligonucleotide mixtures b_5-1, b_5-4, b_5-7, and b_5-8 (fig. 1) which were end-labeled to an average specific activity of 10^9 dpm ug^{-1} with (γ-^{32}P)ATP (3000 Ci $mmol^{-1}$) and T4 polynucleotide kinase (9). Filters were prehybridized 3 to 5 h at 42°C in 6 X SSC (1 X SSC is 150 mM NaCl, 15 mM sodium citrate adjusted to pH 7.0), 50 mM $NaPO_4$ (pH 6.8), 5 X Denhardt'solution (0.1% (w/v) Ficoll, 0.1% (w/v) polyvinylpyrrolidone, and 0.1% (w/v) bovine serum albumin) and 100 ug ml^{-1} of sonicated herring DNA (Sigma). The hybridizations were carried out at 37°C for 24 to 36 h in the same solution with the addition of 1-2 pmol ml^{-1} of labeled oligonucleotide. The temperature for final washing was based on an empirical formula (10) by assuming that all ambiguous positions contained A or T bases. The filters were washed four times for 5 min in 6 X SSC at 37°C, twice for 30 min at 42°C (for b_5-1), 40°C (for b_5-4), and 60°C (for b_5-7 and b_5-8).

The charon 4A library was plated on E. coli 803 supF and nitrocellulose plaque lifts were screened with the chicken liver cDNA clone which was labelled by random-priming (24) to at least 10^9 dpm ug^{-1} with (α-^{32}P)dCTP (3000 Ci $mmol^{-1}$). Filters were prehybridized 2 to 6 h at 42°C in 6 X SSC, 30% formamide (v/v), 50 mM $NaPO_4$ (pH 6.8), 5 X Denhardt's solution, 250 ug ml^{-1} sonicated hering DNA. The hybridizations were carried at 42°C for 24-46 h in the same solution with the addition of 0.2 ug of labeled probe. The washing conditions were as follows: once in 6 X SSC plus 30% formamide at 42°C for 10

min, twice in 2 X SSC plus 0.5% SDS at 55°C for 20 min each time, twice in 0.1 X SSC plus 0.1% SDS at 55°C for 20 min each time.

DNA sequence analysis:

The 1.5 kb insert in a recombinant phage was excised as a single fragmentby cleavage with EcoRI and subcloned in both orientations into the EcoRI site of pBluscript to produce the recombinant plasmids pHZ-3 and pHZ-4. A series of overlapping unidirectional deletions were made in plasmids pHZ-3 and pHZ-4 with exonuclease III and mung bean nuclease essentially as described (11). The deleted inserts from both orientaions were self-ligated to produce a series of overlapping plasmids which were sequenced on both strands as double-stranded DNA by the chain termination method as described (12).

Genomic Southern analysis:

The chicken DNA was purchased from Clontech Laboratories (Palo Alto, CA). DNA (5 ug per lane) was digested to completion with restriction endonucleases, resolved by electrophoresis in 0.8% agarose gels containing 89 mM Tris-borate (pH 8.2), 2 mM EDTA, and transferred to nitrocellulose filters as described (9). The filter was prehybridized for 4 to 6 h at 42°C in 50% (v/v) formamide, 5 X SSC, 50 mM $NaPO_4$ (pH 6.8), 5 X Denhardt's solution, 250 ug ml^{-1} sonicated herring DNA. The hybridization was carried out at 42°C for about 20 h in the same solutions with the addition of 0.8 ug of probe

DNA which was nick translated with (α-^{32}P)dCTP to a specific activity of about 10^8 dpm ug^{-1} (9). After hybridization, the filter was washed three times for 20 min at 42^{o}C in 2 X SSC, 0.1% SDS, then twice for 30 min at 65^{o}C in 0.1 X SSC, 0.1% SDS.

Northern analysis:

Total RNA (10 ug per lane) was electrophoresed in a 0.8% agarose gel containing formaldehyde (13), then blotted onto a nitrocellulose filter. The filter was prehybridized and hybridized under exactly the same conditions as in the genomic Southern analysis. The filter was first probed with the cytochrome b_5 cDNA clone of erythrocytes, exposed to films, then rehybridized to a chicken β-globin gene (25).

Results

1. Cloning and sequencing of a cytochrome b_5 gene from erythrocyte cells

Oligonucleotides b_5-1 and b_5-4 were previously used to obtain a cDNA clone for cytochrome b_5 from chicken liver (8). These oligonucleotides were also used to screen a chicken erythrocyte cDNA library. Out of 200,000 plaques screened, 19 hybridized to b_5-1, and two of these also hybridized to b_5-4. These two clones, designated lambda HZ-3 and lambda HZ-5, were also recognized by oligonucleotides b_5-7 and b_5-8 which were based on regions of sequence from the hydrophilic domain and hydrophobic domain of

```
   1  GTG TGG TGA GTC GCG GCG GCG TTG GGC TGT CTG TCA AGC GAG GAT ATG     48
                                                                Met      1

                                                          B5-7
  49  GTG GGC TCC AGT GAA GCC GGC GGT GAG GCG TGG CGG GGC CGC TAC TAT     96
   2  Val Gly Ser Ser Glu Ala Gly Gly Glu Ala Trp Arg Gly Arg Tyr Tyr     17

                    B5-4
  97  CGG CTG GAG GAG GTG CAG AAG CAT AAC AAC AGC CAG AGC ACC TGG ATC    144
  18  Arg Leu Glu Glu Val Gln Lys His Asn Asn Ser Gln Ser Thr Trp Ile     33

 145  ATC GTG CAC CAC CGT ATC TAC GAC ATC ACC AAG TTC CTG GAT GAG CAC    192
  34  Ile Val His His Arg Ile Tyr Asp Ile Thr Lys Phe Leu Asp Glu His     49

 193  CCT GGT GGA GAA GAA GTC CTT AGG GAG CAA GCT GGG GGA GAT GCT ACT    240
  50  Pro Gly Gly Glu Glu Val Leu Arg Glu Gln Ala Gly Gly Asp Ala Thr     65

          B5-1
 241  GAG AAC TTT GAA GAT GTT GGC CAC TCT ACA GAT GCA AGG GCG CTG TCG    288
  66  Glu Asn Phe Glu Asp Val Gly His Ser Thr Asp Ala Arg Ala Leu Ser     81

 289  GAA ACA TTT ATT ATT GGG GAG CTT CAC CCG GAT GAT AGA CCG AAG CTT    336
  82  Glu Thr Phe Ile Ile Gly Glu Leu His Pro Asp Asp Arg Pro Lys Leu     97

 337  CAG AAA CCA GCA GAA ACT CTT ATT ACC ACT GTG CAG TCT AAT TCC AGT    384
  98  Gln Lys Pro Ala Glu Thr Leu Ile Thr Thr Val Gln Ser Asn Ser Ser    113

                          B5-8
 385  TCA TGG TCC AAC TGG GTG ATC CCG GCA ATA GCA GCA ATT ATT GTG GCC    432
 114  Ser Trp Ser Asn Trp Val Ile Pro Ala Ile Ala Ala Ile Ile Val Ala    129

 433  CTG ATG TAT CGT TCC TAC ATG TCA GAG TGA GCA CCT TAC TGA GAA CTA    480
 130  Leu Met Tyr Arg Ser Tyr Met Ser Glu ***

 481  ATG CAA GAA GAG ACT GAT CTG GGA GAG AAT AGA AGC AAT CCT AAC CCA    528
 529  ATA TAT TTC CTG ACA AAA GCC TGA TGT CTG AAG ATA AAT TCA ACT TTT    576
 577  TCA GAA AAC TGA ACA ATT CTT TTC TGC TGT GCA CTT TTC TTG ATG TTG    624
 625  CCT TCT TAT TTG CTG CAC TGA AGT AAT AAA AAG GCA GCA TTT CTT TTC    672
 673  GTA TAA CAA TAT ATT CTC TAA TGA ATG ATT TGA TAA CTG TAT TAG TTG    720
 721  CTG TAT TAA AAT AGT TTT GTA AGT AGC ATT CTG ATT CTG GTT ATA TCT    768
 769  TTT TAA TCT GTA ATG GAG TCT GTC TTG CAT ATG AAT TTT ATA GCT TTA    816
 817  AAT TAC TAC CAA AAC TTT GTA CAT GTA TTT GTC CAT GTA CAC AAC CTA    864
 865  ACT TAA AAA TCA TGT TGT CGT CTT AAA TCT AGA ATG TTT GAG TAA GAG    912
 913  GCT AAT TAA AAT AAA CAT AAT GGA AGA AGC TGA GTA TAG TAA TGA GTA    960
 961  CAG GTG CCT GTA AAT GGT TGG GTC CTG CCA GTC AGG CTA TAA GAA GAT   1008
1009  AAC TTT CCT TCC CTC CTG CCA TGT GGT CTT AGA GTT GTT ACA GGT ACT   1056
1057  CCT GCT GGC AAG CTG TTG TTT GAC TGC CAT GGG AAA ATT AAA GTA AAA   1104
1105  TAT GAA ATC CAC TGG CCC AGT TAT GTC CAT CTC CGT TTT GTG AAC TGT   1152
1153  TGA ACT GTT CTG CAA AAA AGG CAG AAA GTG CTG TGT AAA TTC CAC TAC   1200
1201  AGG TAA TAT AAC TGC TAC TAA TAC TGT TCT TGC CAA GCA CTC AGG TGA   1248
1249  CTC TGA AAC TTG TTC TGG AAC TTC TAG ACT TGT ATA CAA TCT TCA ACT   1296
1297  TTA TCA TGG TAT GTC CTG ATG GGG TGG AAA AAG TGA TGC TTC TGA CTG   1344
1345  TTC TGT TAT GTG CTC CTT GGT GCT TTA CTA TGG AGA GAT GAC CAT TTT   1392
1393  CTG TGC TAA ATA CAG GAC AAC TGA AAG TCT GCA TTT TGT GGT GAA TTT   1440
1441  TTT TTT TAT TTT TAT TTT TTA GTC ATG CAT AAA TGA TCA TGA ATA AAA   1488
1489  GTT TAA TTG CTT ACT CTT T
```

Figure 2. Sequence of a cDNA clone and the deduced amino acid sequence for cytochrome b_5 from erythrocytes. Regions where oligonucleotides recognize are marked.

cytochrome b_5 respectively. The fact that the hydrophobic domain probe (b_5-8) hybridized to the two clones suggested that cytochrome b_5 mRNA in erythrocytes encodes a hydrophobic domain comparable to that in cytochrome b_5 mRNA from liver.

The inserts in lambda HZ-3 lambda HZ-5 were subcloned into the EcoRI site of pBluescript to produce plasmids designated pHZ-3 and pHZ-5. The larger insert (1.5 kb) was in pHZ-3. The complete sequence of the insert in pHZ-3 is shown in figure 2. This cDNA clone was approximately twice as large as the cDNA clone from chicken liver (8). The region of the cDNA from nucleotides 27 to 799 was 100% homologous to the liver cDNA clone. The open reading frame extends from nucleotide 48 to 462, so most of the extra sequence in pHZ-3 is at the 3' untranslated region.

2. Expression of the cytochrome b_5 gene in liver and erythrocyte cells

The complete sequence identity between the cDNAs from erythrocyte and liver cells suggested that there is only one kind of cytochrome b_5 message in both liver and erythrocyte cells. In order to examine this, a Northern blot of total RNA from liver and erythrocyte cells was probed with the cDNA clone from liver (figure 3).

The cytochrome b_5 probe hybridized to a 1.6 kb mRNA from liver which is approximately the same size as the cDNA clone from erythrocytes. The intensity of signal was relatively high as indicated by the fact that panel A in Figure 3 is an 24 h exposure. By contrast, there was no apparant hybridization of the probe to the

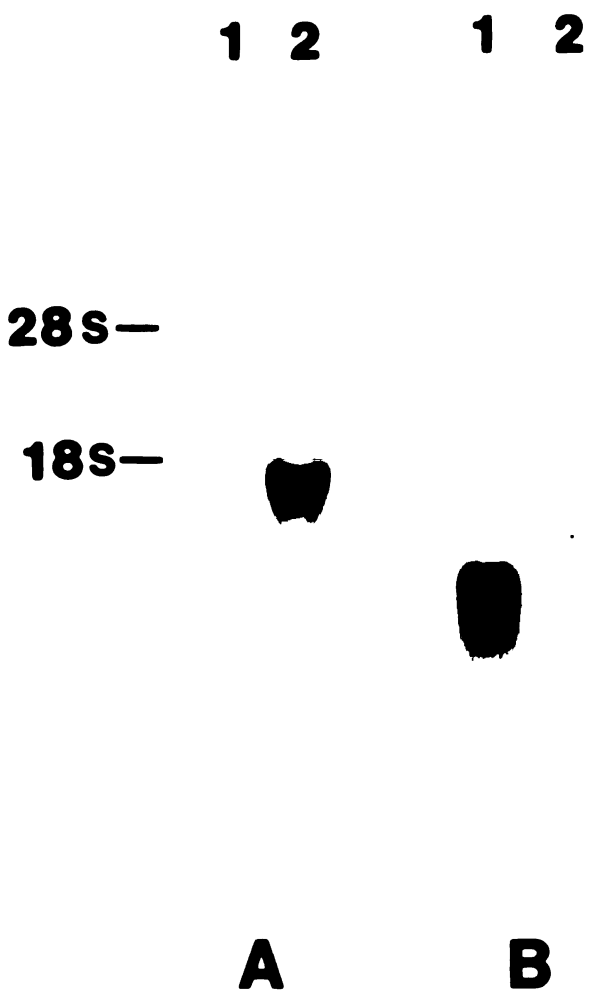

Figure 3. The Northern Blotting of liver total RNA (lane 1) and erythrocyte total RNA (lane 2). Filter was first hybridized to cytochrome b_5 gene (A), then the filter was washed before it was rehybridized to bete-globin gene (B).

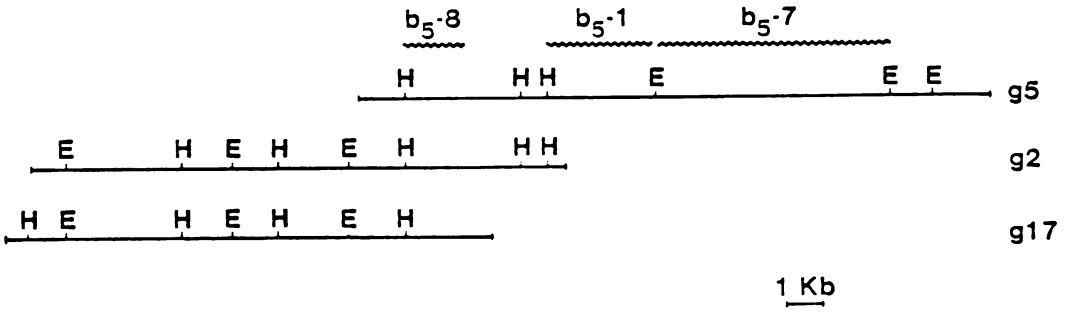

Figure 4. Restriction maps of three genomic clones of cytochrome b_5. Wavy lines designate the fragments which hybridized to the indicated oligonucleotides.

erythrocyte RNA lane, even after 3 days exposure. Presumably the abundance of cytochrome b_5 mRNA is relatively low in this cell type. This is consistant with the fact that only two clones out of 200,000 plaques screened were recovered from the erythrocyte cDNA library. In order to ensure that the mRNA was not degraded, the filter was rehybridized to the beta-globin gene. In this case, the probe hybridized strongly to an mRNA of the correct size (0.6 kb). This internal standard indicates that the RNA from erythrocytes was intact and was accurately quantitated. These results are consistent with the hypothesis that the membrane-bound form (amphipathic form) and the cytosolic form (hydrophilic form) of cytochrome b_5 are the products of posttranslational modification of a polypeptide produced from a common mRNA.

3. Cloning and mapping the genomic sequences

The chicken liver cDNA clone was used as a probe to screen a chicken Charon 4A genomic library. Twelve clones were isolated from 600,000 plaques. Analysis of the restriction pattern of these phage indicated that only three were independent. A partial restriction map for each of these 3 clones, designated g2, g5, and g17, is presented in figure 4. The devised restriction maps overlapped, indicating that they all contain a common region of the chicken genome.

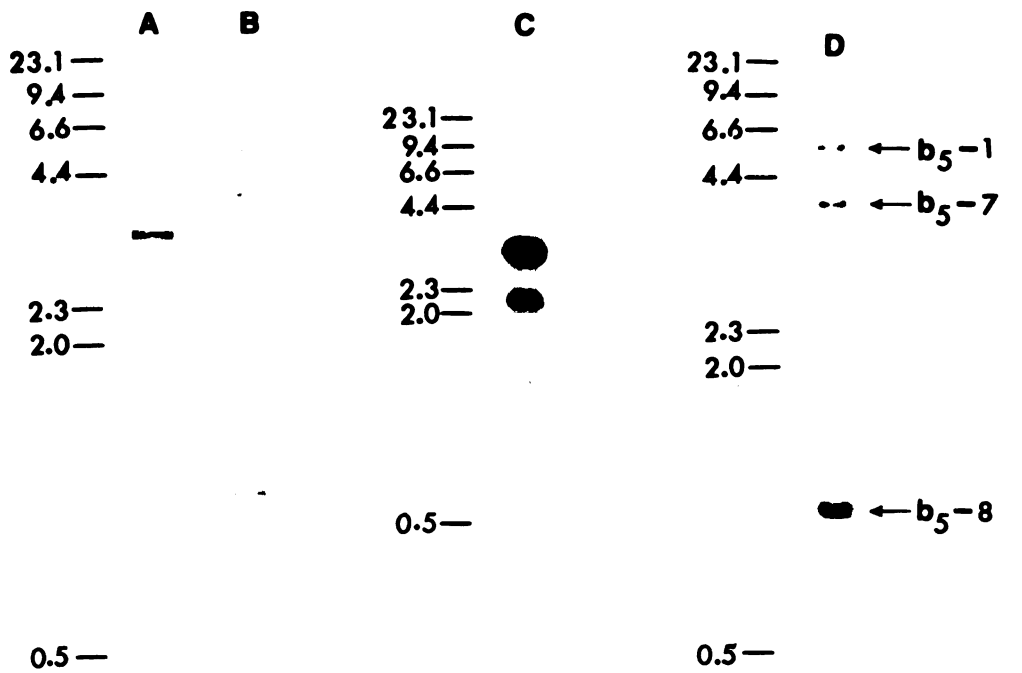

Figure 5. Restriction digests of chicken genomic DNA and the g5 clone probed with liver cytochrome b_5 cDNA or oligonucleotides. Lanes A and B are genomic digests, C and D are g5 clone digests. A and C were digested with Hind III, B and D were with Xba I. A, B, and C were probed with cytochrome b_5 gene, and D was probed with oligonucleotides b_5-1, b_5-7, and b_5-8 sequentially. The bands recognized by specific oligonucleotides in lane D are marked.

4. Genomic southern analysis

In a previous experiment (see Chapter II), chicken genomic DNA was digested with 3 restriction enzymes and probed with the chicken liver cytochrome b_5 cDNA clone. The results were consistent with 2 possible explanations. 1. There are two genes in the chicken genome which give rise to two kinds of cytochrome b_5; 2. There is only one gene in the chicken genome, but within the gene, there is approximately 5.3 kb of intron DNA which harbors one site each for EcoRI, BglII, and HindIII. When clone g5 was digested with HindIII and XbaI and probed with the chicken liver cDNA clone, the hybridization pattern produced was exactly like that of genomic digests (figure 5).

The results of this experiment indicate that there is only one cytochrome b_5 gene in the chicken genome which is entirely contained in clone g5, because the 18 kb fragment of lambda clone g5 generated tha same hybridization pattern as the total chicken genomic DNA did. Clone g5 encodes one copy of microsomal cytochrome b_5 gene. From analysing the size of the fragments which hybridize to the oligonucleotides, it is apparent that there are introns within the gene, and the introns are more than 5.3 kb long, but less than 9.3 kb.

Discussion

It has been suggested that the bovine erythrocyte cytochrome b_5 protein is derived from the same gene product as the microsomal

protein by proteolytic processing during erythroid maturation (7). This is based on the observation that the cytosolic cytochrome b_5 is not present in an immature erythroid cell. Instead, a membrane-bound form of cytochrome b_5 is present (23). Electron microscopy also showed that the endoplasmic reticulum disappears during erythroid maturation. Microsomal cytochrome b_5 from liver cells can, when treated with liver lysosomal proteases, produces two hydrophilic segments one of which was identical to the form II of bovine erythrocyte cytochrome b_5 (7). Erythrocyte cytochrome b_5 I and II are equivalent to residues 1-97 and 1-95, respectively, of microsomal cytochrome b_5 in bovine. The existence of lysosomal proteases capable of converting microsomal cytochrome b_5 to the cytosolic protein nurtured the idea that the putative erythroid proteases are responsible for the solubilization of microsomal cytochrome b_5 in erythrocyte cells. Several lines of evidence presented here support the above hypothesis. First, a cDNA clone from erythrocytes was about 1.5 kb long. This is comparable in size to the homologous RNA from liver. Second, the cDNA clones from erythrocyte and liver cells are 100% homologous. Both of the erythrocyte cDNA clones contained the hydrophobic membrane-binding domain, indicating that they both encode a protein which is exactly the same as the microsomal cytochrome b_5 found in liver. Third, all the fragments of chicken genomic DNA with homology to the cDNA clone are carried on a single clone (g5). Furthermore, analysis of the structure of this clone by probing with oligonucleotides from various regions of the coding sequence are consistent with the existence of only one cytochrome b_5

gene in chicken. Therefore, this gene must be responsible for synthesis of two forms of cytochrome b_5.

Unfortunately, there has been no direct characterization of the cytochrome b_5 protein in the chicken erythrocyte. Thus, it is not possible to directly compare the sequence of the erythrocyte protein to the deduced sequence of the cDNAs. However, if, as seems certain, a soluble form exists in chicken, it must be derived by posttranslational modification. As to the analogous situations in human and porcine, we cannot rule out the possibilities that there are two different genes, or one gene with alternative RNA splicing products, or posttranslational modification. The level of gene expression is regulated in an organ-specific manner, presumably reflecting the degree of necessity. Liver is the place where fatty acids, cholesterol, and other products are being actively synthesized (3, 14-17). As a component of those biosynthetic processes, cytochrome b_5 plays an obligatory role. In erythrocytes, however, only a relatively small quantity of cytochrome b_5 may be required for methemoglobin reduction, and thus the level of its mRNA is much lower in red cells as compared to liver. No other functions have been documented for cytochrome b_5 in erythrocytes. Of course, it is also possible that the rate of turnover of cytochrome b_5 may be much lower in erythrocyte cells than that in liver, so that the actual concentration of cytochrome b_5 molecules could be comparable.

There are several recent reports concerning the cytochrome b_5 cDNA structure in rabbit and bovine (18, 19). The rabbit and chicken liver cytochrome b_5 mRNAs are 63% homologous. No studies of the regulation of the cytochrome b_5 gene are available. Since terminal

desaturase activity was shown to be regulated by hormones and diets (20-22), it would be interesting to test whether or not cytochrome b_5 is subject to any kind of regulation besides organ-specific regulation. It is unlikely that the cytochrome b_5 gene is subject to a simple "On or Off" form of regulation, since it is involved in so many vitally important biological processes.

Acknowledgments

We thank Dr. J. Dodgson for his generous gifts of plasmid, libraries, andespecially the advice.

References

1). K. Abe, S. Kimura, R. Kizawa, F. K. Anan, Y. Sugita (1985). J. Biochem. (Tokyo) 97:1659-1668

2). F. S. Mathews, P. Argos, and M. Levine (1972). Cold Spring Harbor Symp. Quant. Biol. 36:387-393

3). N. Oshino, Y. Imai, and R. Sato (1966). Biochem. Biophys. Acta. 128:13-28

4). D. Hultquist, and P. G. Passon (1971). Nature (London). 229:252-254

5). D. Hultquist (1979). Methods Enzymol. 52:463-473

6). R. E. White, and M. J. Coon (1980). Ann. Rev. Biochem. 49:315-356

7). S. R. Slaughter, C. H. Williams, and D. E. Hultquist (1982). Biochem. Biophys. Acta 705:228-237

8). H. Zhang, and C. Somerville (1988). Arch. Biochem. Biophys. 264:343-347

9). T. Maniatis, E. F. Fritsch, J. Sambrook (1982). Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

10). M. Smith (1983). In Methods of DNA and RNA Sequencing (Weissman, S. M., Ed.), Praeger, New York.

11). S. Henikoff (1984). Gene 28:351-359

12). H. Zhang, R. Scholl, J. Browse, and C. Somerville (1988). Nucleic Acid Res. 16:1220

13). F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, K. Struhl (1988). Current Protocols In Molecular Biology. 4.9.1-4.9.4

14). S. R. Keyes, J. A. Alfano, I. Jansson, and D. L. Cinti (1979). J. Biol. Chem. 254:7778-7784

15). V. R. Reddy, D. Kupfer, E. Caspi (1977). J. Biol. Chem. 252:2797-2801

16). F. Paultauf, R.A. Porugh, B.S.S. Masters, and J.M. Johnson (1974). J. Biol. Chem. 249:2661-2662

17). F.F. Kadlubar, and D.M. Ziegler (1974). Arch. Biochem. Biophys. 162:83-92

18). N. Dariush, C.W. Fisher, and A.W. Steggles (1988). Prot. Seq. Data Anal. 1:351-353

19). R.J. Cristiano, and A.W. Steggles (1989). Nucleic Acids Res. 17:799

20). P.M. Lippiello, C.T. Holloway, S.A. Garfield, and P.W. Holloway (1979). J. Biol. Chem. 254:2004-2009

21). V.C. Joshi, and L.P. Aranda (1979). J. Biol. Chem. 254:11779-11782

22). N. Oshino, and R. Sato (1972). Arch. Biochem. Biophys. 149:369-377

23). S.R. Slaughter, and D.E. Hultquist (1978). J. Cell Biol. 83:231-239

24). A.P. Feinberg, and B. Vogelstein (1983). Anal. Biochem. 132:6-13

25). J.B. Dodgson, J. Strommer, and J.D. Engel (1979). Cell 17:879-887

26). N.S. Yew, H.-R. Choi, J.L. Gallarda, and J.D. Engel (1987). Proc. Natl. Acad. Sci. 84:1035-1039

CHAPTER IV

SEARCHING FOR THE PLANT CYTOCHROME B_5

1. Introduction

The presence of cytochrome b_5 in microsomal membranes of higher plants has been documented in the literature, and several groups have reported the purification of cytochrome b_5 proteins (1-3). Jollie et al. determined a short amino acid sequence from the N-terminus of pea cytochrome b_5 which is completely unrelated to any known sequences of cytochrome b_5 from animals.

Since the sequence of cytochrome b_5 is so highly conserved among animals (4), we considered it possible that the plant cytochrome b_5 genes may be isolated by exploiting sequence homology instead of purifing the protein first. Toward this end, two approaches were tried. Oligonucleotides corresponding to the most conserved regions of the cytochrome b_5 protein were used as probes to screen plant genomic or cDNA libraries, and the chicken cytochrome b_5 gene was used as a heterologous probe.

	Trp	Trp	Thr	Asn	Trp	Val	
b_5-2	TGG	TGG	ACN	AAC	TGG	GT	
				T			
	His	His	Lys	Val	Tyr	Asp	
b_5-3	CAT	CAT	AAA	GTN	TAT	GA	
	C	C	G		C		
	Glu	Glu	Ile	Lys	Lys	His	Asn
b_5-5	GAA	GAA	ATT	CAA	AAA	CAT	AA
	G	G	C	G	G	C	
			A				

Figure 1. Oligonucleotides designed from animal cytochrome b_5 protein (see figure 2). N represents T, A, C, and G.

2. Materials and methods

The Arabidopsis lambda gt10 cDNA library, constructed from leaf mRNA, was from N. Crawford (Stanford, CA). The EMBL4 genomic library was from E. Meyerowitz (Caltech, CA).

The conditions for oligonucleotide end-labelling, hybridization, and screening were the same as in chapter II. The heterologous probing was carried out as follows: prehybridization was done in 30% formamide, 5 X SSPE, 5 X Denhardt's solution, 50 mM $NaPO_4$ (pH 6.5), 0.2 mg/ml sonicated herring DNA for 12 hrs at 42°C. Hybridization was done in the same solution except using E. coli DNA as carrier DNA and adding dextran sulfate to 5% for 48 hrs at 42°C. The filters were washed twice in 5 X SSPE plus 0.5% SDS for 20 min each time at 37°C, then twice in 2 X SSPE plus 0.1% SDS for 20 min each time at 37°C and finally once in 0.2 X SSPE, 0.1% SDS at 42°C for 30 min (1 X SSPE is 150 mM NaCl, 10 mM NaH_2PO_4, and 1.3 mM EDTA adjusted to pH 7.4).

The construction of overlapping deletions and DNA sequencing were done as described in (6, 7). All the other DNA manipulations were performed following the standard methods (5).

The oligonucleotide mixtures in figure 1 were deduced from animal protein sequences, as indicated in figure 2, and they were used as probes to screen Arabidopsis cDNA libraries.

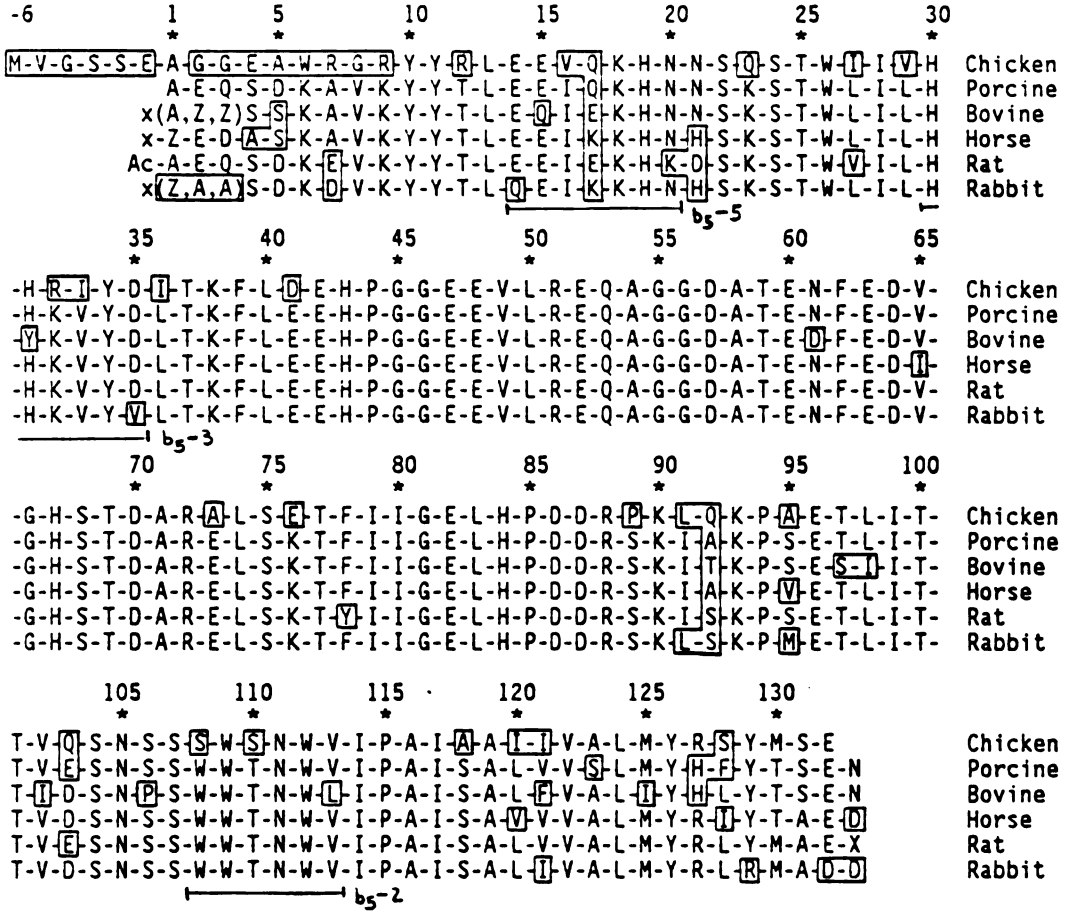

Figure 2. A comparison of the deduced amino acid sequence of chicken cytochrome b_5 with the sequences obtained by direct amino acid sequencing of the microsomal cytochrome b_5 from other vertebrates. Regions where the oligonucleotides were designed are marked.

3. Results:

1). Chicken liver cytochrome b_5 clone as a probe

When the chicken liver b_5 cDNA was used as a probe to hybridize to an Arabidopsis genomic EcoRI digest, several fragments hybridized to the probe (figure 3).

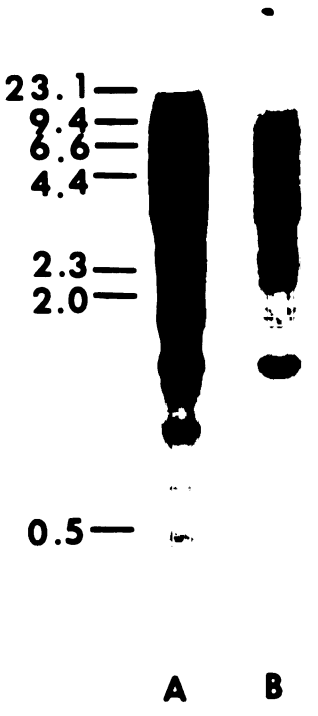

Figure 3. The Eco RI digest of Arabidopsis genomic DNA was probed with chicken liver cytochrome b_5 cDNA clone. Lane A was washed at 37°C for the final wash, and Lane B was washed once more at 46°C.

After a 46°C wash, the 1.7 kb and 4.0 kb bands still gave strong signals, suggesting significant homology to the probe. Therefore, the cDNA clone was used as a probe to screen both genomic and cDNA libraries. Three clones out of 200,000 plaques from the genomic library, and four out of 150,000 plaques from the cDNA library were

isolated which hybridized to the probe. The DNAs were made from all these clones, and digested with EcoRI. It turned out that all three genomic clones covered a common fragment (1.7 kb) which hybridized to the probe. This fragment (AB5-10-0) together with other two larger cDNA clones (AB5-11 and AB5-13-0) were subcloned into pBluescript plasmids, and AB5-10-0 and AB5-13-0 were sequenced in one orientation as double stranded DNA (6, 7). Figure 4 shows the sequencing strategies these two clones.

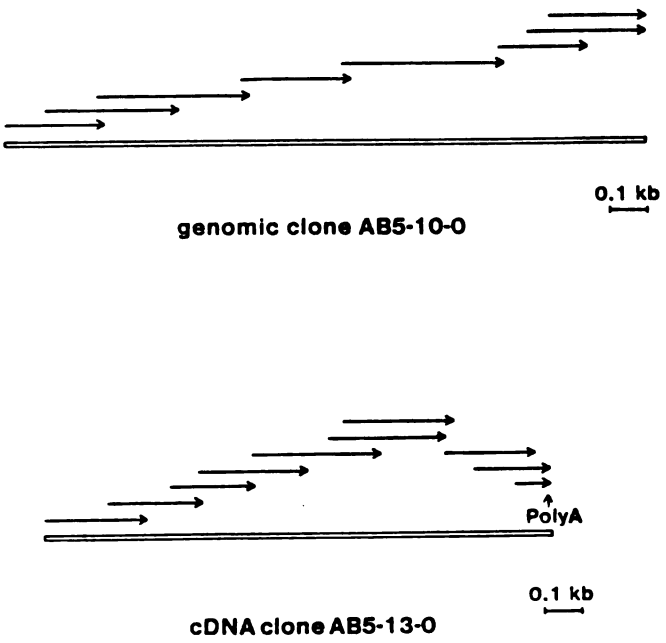

Figure 4. Sequencing strategies for clones AB5-10-0 and AB5-13-0.

There was no particular reason to choose the cDNA clone AB5-13-0 to sequence first, since the sequencing of cDNA clone AB5-11 was also planned (later the plan was droped). Both cDNA clones hybridized to the chicken probe equally well under low stringency condition.

```
         10         20         30         40         50         60         70         80
5' AAACGCTCAG GACAGGGACG GGTGGACTCC ACTGCACGTT GCAGTACAGG CAAGAAGAAT CGGCGTGGCG GGGCTAAGAG

         90        100        110        120        130        140        150        160
   ATTAGATGAA TCCGAGATTG AGCCCGAGAA CCTCGTGCCA GAGGAATGGA GGGATATCAG GCGGAGGTGA ATCTGACGAA

        170        180        190        200        210        220        230        240
   GAAGGACAAG CGCAAAATAG CGCAGGAGAT GGAGTTCGGG GTTCGGGTGG AGAAGAAGAG GCAAGGGCTA ATTCCGCTGA

        250        260        270        280        290        300        310        320
   GGAAAGTTGA CTTGAATGAC TTTCTCACGT ACAAGGAAGC CAAGTTGGCT CAATTGAGGC CTGTCATTCT CGATAAACCG

        330        340        350        360        370        380        390        400
   GGAAATTTCT CCGACGACAG TGGAGCGTCA AGATTGGAGA GACCGCTGTA TCATCTCCCA GCGAGCGAGT GGCTCCTAAG

        410        420        430        440        450        460        470        480
   AACCCTAGAT GGGCAGTTTA CGGAAAGGGA TTCGACCACG TTGCCAAGTT CTTCAATAGC GACAAGTACG ATCCCAGCGA

        490        500        510        520        530        540        550        560
   CAAGAAATCC GACGGCCCTC GAAAGCTGCT TTCAAAAGAA GAGAAGTTTA TGCTCAATAG CCGGAATCCT GACCTAGCCG

        570        580        590        600        610        620        630        640
   TTGCCACATC AAAAAAATGG CTTCCTCTTC ACACACTGGC AGCATGTGGA GAGTTTTATC TGGTTGATTC CTTGCTAAAG

        650        660        670        680        690        700        710        720
   CACAATCTTG ATATCAATGC AACCGATGTG GGCGGCTTGA CAGTACTTCA CCGAGCAATC ATTGGTAAGA AGCAGGCTAT

        730        740        750        760        770        780        790        800
   TACTAACTAC CTGCTGAGGG AATCGGCAAA TCCATTTGTT CTTGATGACG AAGGTGCGAC CTTGACGCAC TATGCTGTGA

        810        820        830        840        850        860        870        880
   AACACATCAG CTCCCACAAT AAAACTTCTC CTACTGTATA ACGCTGATAT AAACGCTCAG GACAGGGACG GGTGGACTCC

        890        900        910        920        930        940        950        960
   ACTGCACGTT GCAGTACAGG CCAGAAGAAG CGACATTGTA AAGCTTCTTT TGATAAAAGG GGCGGACATA GAAGTGAAGA

        970        980        990       1000       1010       1020       1030       1040
   ACAAGGATGG GTTAACTCCG CTTGGGCTTT GCTCTACCTT GGAAGAGAGA TAAGGACGTA TGAGGTGATG AAGCTGTTGA

       1050       1060       1070       1080       1090       1100       1110       1120
   AAGAGTTTCC ACTTAGCAGA CACAAGAAGA GATTGGTAAC AACAGATGAA GATATTGAAT AGTCCTTCAA TTTCAGCTTG

       1130       1140       1150       1160       1170       1180       1190       1200
   AAGTACACTC ACTTATGAGA ACCTGAGAAA AGGAGATGGA GGTAAAGGTG ATGATTAGGG CATTGGAACC TCGGAGTCGG

       1210       1220       1230       1240       1250       1260       1270       1280
   AGTGGGTCCA CTGTCTCACT TCCTTAAATT TGGTTTGCTG TTAGTCTTAT CCATCGATTT TGGATATTTA TCACAACTTG

       1290       1300       1310       1320
   ATCCATTCTT AAAGAAAATA TCTGAAAATA AATAAAAAGT AATACATAT 3'
```

Figure 5. DNA Sequence of the clone AB5-13-0.

Figure 5 shows the complete sequence of AB5-13-0. There might be some mistakes in this sequence, since it was only sequenced in one orientation. The DNA data bank (Gene Bank R58.0, December, 1988 and EMBL Bank R17.0, November 1988) was searched with HIBIO DNASIS^{TM} program. No sequences were identified which had similarity more than 50% to AB5-13-0. According to Doolittle (12), the range of chance similarities between two unrelated sequences can exceed 50% if we allow gaps in the sequences. Similarities below 50% were not considered informative. All three reading frames have been used to screen the protein data bank (Protein Identification Resource R18.0 September 1988) to look for homologous proteins with HIBIO PROSIS^{TM} program. No cytochrome b_5 protein from any sources was found to be homologous. However, there is one protein from Drosophila called Notch in which part of its sequence is similar to reading frames 2 and 3 at about 25% identity (figures 6 and 7). Possibly there are some mistakes in the sequence of AB5-13-0 which shifted the reading frame and split the similarities into two reading frames. However either match extends over 100 residues. It is very likely they are related. The significance of this observation will be discussed later.

The genomic clone AB5-10-0 was also sequenced in one orientation (figure 4), and the sequence was analyzed the same way as the clone AB5-13-0 except all 6 reading frames were analyzed (3 for each direction). Because it was a genomic clone, we don't know which was the sense strand. No information can be deduced concerning this clone after searching the DNA bank and the protein bank. The chance of finding a similar DNA sequence in the DNA bank is very low,

```
                                PROSIS HOMOLOGY SEARCH
                            24.6% identity in 130 aa overlap

                      10        20        30        40        50        60
AB513RF2.AMI  NAQDRDGWTPLHVAVQARRIGVAGLRDMNPRLSPRTSCQRNGGISGGGESDEEGQAQNSA

A24768        PKRQRSDPVSGVGLGNNGGYASDHTMVSEYEEADQRVWSQAHLDVVDVRAIMTPPAHQDG
                    1840      1850      1860      1870      1880      1890

                      70        80        90       100       110       120
AB513RF2.AMI  GDGVRGSGGEEEARANSAEESLELSHVQGSQVGSIEACHSRTGKFLRRQWSVKIGETAVS

A24768        GKHDVDARGPCGLTPLMIAAVRGGGLDTGEDIENNEDSTAQVISDLLAQGAELNATMQKT
                    1900      1910      1920      1930      1940      1950

                     130       140       150       160       170       180
AB513RF2.AMI  SPSERVAPKNPRWAVYGKGFDHVAKFFNSDKYDPSDKKSDGPRKLLSKEEKFMLNSRNPD
                                                                ...:..:...
A24768        GETSLHLAARFARADAAKRLFHAGADANCQDNTGRTPLHAAVAADAMGVFQILLRNRATN
                    1960      1970      1980      1990      2000      2010

                     190       200       210       220       230       240
AB513RF2.AMI  LAVATSKKWLPLHTLAACGEFYLVDSLLKHNLDINATDVGGLTVLHRAIIGKKQAITNYL
              :..   .   ::   : .   .:..:.. . ::::.: .: :.::.:   .. . .: :
A24768        LNARMHDGTTPLILAARLAIEGMVEDLITADADINAADNSGKTALHWAAAVNNTEAVNIL
                    2020      2030      2040      2050      2060      2070

                     250       260       270       280       290
AB513RF2.AMI  LRESANPFVLDDEGATLTHYAVKHISSHNKTSPTVRY-KRSGQGRVDSTARCSTGQKKRH
              :.. ::. . ::...:    :... : .. ..   .. .:.  ...:. .:  .... .:
A24768        LMHHANRDAQDDKDETPLFLAAREGSYEACKALLDNFANREITDHMDRLPRDVASERLHH
                    2080      2090      2100      2110      2120      2130

           300       310       320       330       340       350
AB513RF2.AMI  CKASFDKRGGHRSEEQGWVNSAWALLYLGREIRTYEVMKLLKEFPLSRHKKRLVTTDEDI

A24768        DIVRLLDEHVPRSPQMLSMTPQAMIGSPPPGQQQPQLITQPTVISAGNGGNNGNGNASGK
                    2140      2150      2160      2170      2180      2190

           360       370       380       390       400       410
AB513RF2.AMI  ESFNFSLKYTHLEPEKRRWRRLGHWNLGVGVGPLSHFLKFGLLLVLSIDFGYLSQLDPFL

A24768        QSNQTAKQKAAKKAKLIEGSPDNGLDATGSLRRKASSKKTSAASKKAANLNGLNPGQLTG
                    2200      2210      2220      2230      2240      2250

           420       430
AB513RF2.AMI  KKISENKKVIH

A24768        GVSGVPGVPPT
                    2260
```

Figure 6. Result of the protein data bank search with reading frame 2 of the clone AB5-13-0.

```
                    10        20        30        40        50        60
AB513RF3.AMI  TLRTGTGGLHCTLQYRQEESAWRGEIRIRDAREPRARGMEGYQAEVNLTKKDKRKIAQEM

A24768        AHGVTWFPEGFRAPAAVMSRRRRDPHGQEMRNLNKQVAMQSQGVGQPGAHWSDDESDMPL
                    1780      1790      1800      1810      1820      1830

                    70        80        90        100       110       120
AB513RF3.AMI  EFGVRVEKKRQGLIPLRKVDLNDFLTYKEAKLAQLRPVILDKPGNFSDDSGASRLERPLY

A24768        PKRQRSDPVSGVGLGNNGGYASDHTMVSEYEEADQRVWSQAHLDVVDVRAIMTPPAHQDG
                    1840      1850      1860      1870      1880      1890

                    130       140       150       160       170       180
AB513RF3.AMI  HLPASEWLLRTLDGQFTERDSTTLPSSSIATSTIPATRNPTALESCFQKKRSLCSIAGIL

A24768        GKHDVDARGPCGLTPLMIAAVRGGGLDTGEDIENNEDSTAQVISDLLAQGAELNATMQKT
                    1900      1910      1920      1930      1940      1950

                    190       200       210       220       230       240
AB513RF3.AMI  TPLPHQKNGFLFTHWQHVESFIWLIPCSTILISMQPMWAAQYFTEQSLVRSRLLLTTCGN

A24768        GETSLHLAARFARADAAKRLFHAGADANCQDNTGRTPLHAAVAADAMGVFQILLRNRATN
                    1960      1970      1980      1990      2000      2010

                    250       260       270       280       290       300
AB513RF3.AMI  RQIHLFLMTKVRPRTMLNTSAPTIKLLLLYNADINAQDRDGWTPLHVAVQARRSDIVKLL
                                        :.  .::::::.:..: :.:: :..... :..:
A24768        LNARMHDGTTPLILAARLAIEGMVEDLITADADINAADNSGKTALHWAAAVNNTEAVNIL
                    2020      2030      2040      2050      2060      2070

                    310       320       330       340       350       360
AB513RF3.AMI  LIKGADIEVKNKDGLTPLGLCSTLEERGRMRSCKSFHLADTRRDWQQMKILNSPSISAST
              :.. :. ....... ::: :    .  :.  .::.. .. ..:.  . .   .......
A24768        LMHHANRDAQDDKDETPLFL---AAREGSYEACKALLDNFANREITDHMDRLPRDVASER
                    2080      2090          2100      2110      2120

                    370       380       390       400       410       420
AB513RF3.AMI  LTYENLRKGDGGKGDDGIGTSESEWVHCLTSLNLVCCSYPSILDIYHNLIHSRKYLKINK
              : .. .:
A24768        LHHDIVRLLDEHVPRSPQMLSMTPQAMIGSPPPGQQQPQLITQPTVISAGNGGNNGNGNA
            2130      2140      2150      2160      2170      2180

AB513RF3.AMI  KYI

A24768         SGK
            2190
```

Figure 7. Result of the protein data bank search with reading frame 3 of the clone AB5-13-0.

```
         10         20         30         40         50         60         70         80
5' CATACATTAA CGTTAGCTTC GAAAGACTTT TCGGAATAAC TCCGCAAAAC CTGTTGGAAT TAACGTGAAA CAAGGCGACA

         90        100        110        120        130        140        150        160
   TCGGTCAATA AACCTAGCTC AGGCGGCAAG TATCCGGCTA TATCCGCATG GTTAAGGTCA ATCCCCGCCA CAACCAATAC

        170        180        190        200        210        220        230        240
   GCTTGGGTCG TCAAGTGCTG GTGCACAGAA AACTCCTTTG TAAGAACACA CGTCTGGACC AACCCAATTA GCCGCAGTGT

        250        260        270        280        290        300        310        320
   TGAACGGATC AGAGTAGAAT GCCTTTTTCC AAGCTTGAAG GGCAATGTAA GCCCTTTTAA GTCTGTTATT AGCAAATTTG

        330        340        350        360        370        380        390        400
   AGGTCTAAAT CGACTTCGTA CTCGATGTCG TCGGGAAGAT CTCCGTTTTC TGATAATGCC AAAAGTTGGC GTCTTGTAAG

        410        420        430        440        450        460        470        480
   AAAAGAAGCT TCCTCATCGG TGAGAGCCGC TGCGAATAGA AAAAAATGGA GATGGTGAAA CAAAAGAGGA GAAGGAAGCA

        490        500        510        520        530        540        550        560
   GCCAAAAGGC TTCTCCATTG TTCTTCGGGT CATGGTGATG AAAAAACTTG GAAATGGCTC TATGGCTTGC TTAATAGATG

        570        580        590        600        610        620        630        640
   AATGTTGTGG AGAGGAAGCC TATAATGAAA CGAAGGCAAA CAGAAACCAT TGGATTGTTT TGCTATTTTT GGTATAACAA

        650        660        670        680        690        700        710        720
   AATTTATATA GACTAACGTT TAAAGGGTAC GAAAATAGCA AACTATTGTA GCTAATTTAG AATTAGAGGT CGTTGACAAC

        730        740        750        760        770        780        790        800
   GTCTTTACAA GTTTAGTTTA GTTTTACCAT TGTTGAAATT TGGACGACCA ATATTTTTTA CAGTCTGGAA AAAAAAAGTA

        810        820        830        840        850        860        870        880
   AAAACTCTTT TGCTTAGTGT TGTGATTTGT GAATATGTCA AATTTTATTT GACGTTTGCG TTTTGAACTC TTTGGGAAAA

        890        900        910        920        930        940        950        960
   CAGTGACAAC AAATTAAAAC GTTTTCTTGT AAAATTGCTG ATACTGTAGG CTATATTTAT TGACACAACC AAAGACGTTT

        970        980        990       1000       1010       1020       1030       1040
   ACAAGTACGA AATTGTTGAA AATTTTAATT TTTGTTAAGA AAAAAGACTT TCTTAGTAGT ATAGTATTTG AAGAAAATTT

       1050       1060       1070       1080       1090       1100       1110       1120
   TCAAAGGATT AAATTAAGAA AATGATTCCA ACTAGTATAT CTGTTTTTAT AATCACTTTT AAGAATGATC AAATAATGTT

       1130       1140       1150       1160       1170       1180       1190       1200
   TTTGTGATAT TAACTGTTCA ATATGTCAAC ACCTCGTTAT TTAGTTTCAC TATATATTGT TTTGAACATT CCGTCATCAC

       1210       1220       1230       1240       1250       1260       1270       1280
   AAGCTCGCAA TAAAGTGAAA AGAAAAAATT CGGTTCGATC GAAAACAAAT AGTTACTTAA CCTCAATTCT TGTTATATTG

       1290       1300       1310       1320       1330       1340       1350       1360
   ACAGCCGATA CTACCGTCCT ATCTCCGCCA CGCCACAACC CGCAACATAA CCGGGCCCAT TTTGTGATGG TTCATGGAGC

       1370       1380       1390       1400       1410       1420       1430       1440
   TTGGTTTTTG GTATAAACTA TACCATTCTT GAAATCTCAA GGCCATAACG TAACTGCGGT CAACTTAGCG GCATCCAGAA

       1450       1460       1470       1480       1490       1500       1510       1520
   TCGATCAATA ACGGCCTGAG ACTCTACGGT TGGTTGCGGA GTCTATTGGG CCCTTGATGG GCTAATGGAG AGTCTAGGTG

       1530       1540       1550       1560       1570       1580       1590       1600
   AGGATGAGAA GGTGATTCTT CTGGCACATA GTTTGAGTGG GGCTTGCTAT TCCTAAGGCC ATGGAGTTGT TCTACAAAAA

       1610       1620       1630       1640       1650       1660       1670       1680
   GGTTCATATG GCTATTTTCG TTATTGGTGA AAGAGTTTCA GCACTGGATG ATCAAGAACA ACCCTCCTAA CCACGTGGAG


   CATATC 3'
```

Figure 8. DNA sequence of the clone AB5-10-0.

because on the average, 25% of the residues of any aligned sequences would be identical. Actually there would be a dispersion around that mean expectation, and a predictable fraction of random cases would be as much as 35% identical (12). The best match for protein similarity searches was 25% identity over 56 amino acid residues. The significance of this match is considered to be "improbable" according to Doolittle (12). The lack of similarity to any sequences in the protein bank could reflect the fact that no similar protein has been sequenced, or, perhaps much of the genomic fragment covers introns which substantially decreased the chances to find similar sequences in the protein bank with reasonable confidence. Figure 8 shows the sequence of the clone ab5-10-0. The reason why those two clones were isolated by using chicken cytochrome b_5 gene as a probe is not clear, possibly the very low stringency conditions for hybridization and washing allowed this happened. The sequence comparison showed 45.3% identity in 685 bp overlap between AB5-10-0 and chicken cytochrome b_5 gene, and 46% identity in 658 bp overlap between AB5-13-0 and chicken cytochrome b_5 gene. In both cases, there is a thirteen nucleotides identity continuously between the compared sequences.

2). Oligonucleotides as probes

Oligonucleotide mixtures b_5-2, b_5-3, b_5-5 were used to probe an Arabidopsis EcoRI digest of DNA (figure 9).

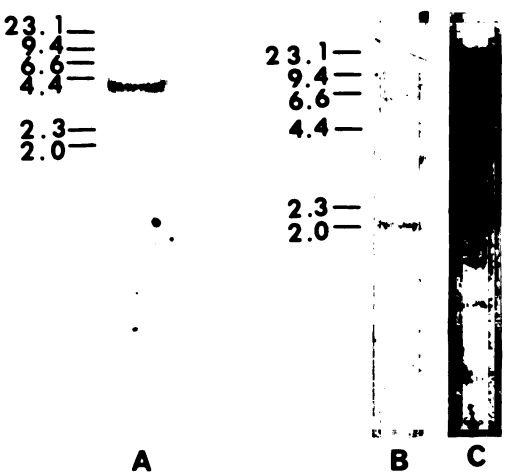

Figure 9. The Eco RI digests of Arabidopsis genomic DNA were probed with oligonucleotide mixture b_5-2 (lane A), b_5-3 (lane B), and b_5-5 (lane C).

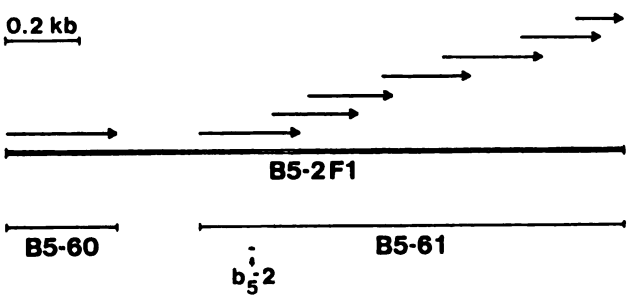

Figure 10. Sequencing strategy for clone B5-2F1.

As can be seen from lanes B and C, oligonucleotide mixtures b_5-3 and b_5-5 recognized a common fragment of approximately 2 kb. When both genomic and cDNA libraries were screened with the oligonucleotide mixtures as probes, no plaques hybridized with b5-3, but a few hybridized with b_5-5. Those clones isolated by b_5-5 alone were not sequenced, because the extremely strong hybridization to the 2.3 kb fragment would be expected to totally mask the signal from the 2.0 kb fragment during screening. The cDNA library was also screened with oligonucleotide b_5-2 and two clones were isolated out of 100,000 plaques. The inserts of these two cDNA clones were subcloned into pBluescript plasmids, and the smaller insert, designated b5-2F1, was deleted and sequenced as described (7, 8). Figure 10 shows the sequencing strategy of b5-2F1. Figure 11 shows the sequences of B5-60 and B5-61, partial sequences of B5-2F1. Even though this sequence is not complete, and sequenced in one orientation only, the message it carries is probably enough for us to do some computer analysis. Unfortunately, after searching both the DNA data bank and the protein data bank in the same way as that for clone AB5-13-0, no significantly similar sequences with known function can be found. The best match from the DNA data bank search is less than 50% identity, and the best match from the protein data bank search is 26.8% over 56 amino acid residues. Neither is informative. The sequence recognized by oligonucleotide mixture b_5-2 is marked in figure 10, and it is only 88% homologous to b_5-2.

B5-60

```
         10         20         30         40         50         60         70         80
5' CCATGGAATT AGATGATTCA CATGATGAAG AGGTATTGAA GTCTCTGGTT CCTGATCCCA TGAAACAAGA GCCTTTAGTA

         90        100        110        120        130        140        150        160
   ATTGAAAACA CTCCAGATCC TTTAGCAGGG GAACAGACAT GGCCANCAGA GGAAGAATGG CTGAGGCTGA CAAATCAAAG

        170        180        190        200        210        220        230        240
   CAGGGCAGGC TAAAGAAGAA GACTTGCGCG AGGACTCAGA TATAGCTGCT GATGTGATGA CAGTGAGAGA TCGTATGTAT

        250        260        270        280        290        300
   TCGAGTATGA TGATGTAGGA GTCACACAGA GTAGCAGATG AGTGACGATA CATGATCCGT AGTA 3'
```

B5-61

```
          10         20         30         40         50         60         70
5' CCTCGTGGGA TCCAAATGAG TCTTTACCTC AGGACTATGC TAGAATTTTT GCTTTTGATA ATGTCGCAAG
          80         90        100        110        120        130        140
   GACTCAAAAG CTTGTGCTCA AGCAAGCCCT GAAGATGGAA GAAGAAGACA GAGATGATTG TGTACCAATT
         150        160        170        180        190        200        210
   GGGTAATATG TACGGCTGCA TATTAAAGAA GTCCTCTTGG TGCTGCCTCT AACTATCTTC TCTTGTGAAC
         220        230        240        250        260        270        280
   ACACGAAACC AATTATAGGA TTTGGGCTTC TGCAGCATGA ATCCAAGATG TCGGTTCTGC ATTTTAGCGT
         290        300        310        320        330        340        350
   AAAGAAGTAT GATGGTTACG AAGCTCCTAT TAAAACAAAA GAAGAGCTAA TGTTCCATGT TGGTTTCCGT
         360        370        380        390        400        410        420
   CAATTCATTG CAAGGCCGGT ATTTGCGACC GACAATTTCA GCTCAGACAA GCACAAGATG GAGAGATTCC
         430        440        450        460        470        480        490
   TGCATCCAGG GTGTTGTCNN GTTGGCTTCA ATATACGGCC CCATATCTAT CCCACCCCTT CCTTTGGTAG
         500        510        520        530        540        550        560
   TTCTGAAGAT TTCTGAAGGG TCGGATCCTC CTGCCATTGC TGCCCTTGGT TCCTTGAAAA GCGTAGAATC
         570        580        590        600        610        620        630
   CAACAGAATA ATCTTAAAGA AGATAATATT AACTGGGTAC CCTCAGAGAG TATCGAAAAT GAAAGCTTCA
         640        650        660        670        680        690        700
   NTCAGATATA TGTTCCACAA CCCCGAAGAC GTGAAATGGT TCAAGCCTGT TGAAGTATGG TCAAAATGCG
         710        720        730        740        750        760        770
   GGCGTCGTGG TCGCGTGAAG GAACCAGTAG GCACACATGG GGCAATGAAA TGCATATTCA GTGGAGTGGT
         780        790        800        810        820        830        840
   TCAACAACAT GACGTAGTGT GCATGAACTT GTACAAGCGC GCTTATCCCA AGTNGCCTGA ACGTCTGTAC
         850        860        870        880        890        900        910
   CCATAGCTTC TCTGATTCGA ATTTTTTAAT GAAGATGATG ATGTTCAAGT TTCTCTGAAT TGACGACATT
         920        930        940        950        960        970        980
   CTTCTTTGGT AATTTATGGT GGTGGTTTAG TGGAAGTCCT CCAGTTTTGT CCTGTATGTG ATAAAAATAT
         990       1000       1010       1020       1030       1040       1050
   GCAAATTTTC TTCTTTATTG CTCACTATAC ATATACATTT GGAATATGGG GTTGAACCAT TTGATAAGCT
        1060       1070
   TAATTGTTAT GAGAAAAATA CTTAAAT 3'
```

Figure 11. DNA sequences of B5-60 and B5-61.
Both are part of the clone B5-2F1.

4. Discussion

Both the chicken gene as a probe and the oligonucleotides as probes have been been tried, neither succeeded. The use of heterologous probes entails the risk that the approach may not work. Two clones isolated by using the chicken gene as a heterologous probe are not cytochrome b_5 genes as judged by computer analysis of homology to known cytochrome b_5 genes at the DNA and deduced amino acid sequence levels.

The clone AB5-13-0 has three open reading frames coding for proteins with MW of 11 KD, 14 KD, and 16 KD respectively. The open reading frame 2 encodes a protein of 11 KD which has 24.6% identity over 130 amino acid residues to Notch protein of Drosophila (10). Sequence similarity is as high as 66% if we include substitutions of similar amino acids. If the mistakes in the sequence were corrected, the open reading frame 2 could possibly encode a larger protein without decreasing the percentage of the similarity to the Notch protein. Since another slightly higher identity (25.7% over 101 amin acid residues) was found in reading frame 3 which overlaps a little bit with the end of open reading frame 2. I estimated the significance of these two matches according to Doolittle's method (12). The result is that they are "probably" related, but not "certainly". The product of the Notch locus in Drosophila melanogaster is involved in a cell interaction which controls the accurate differentiation of certain tissues during development, including the embryonic nervous system (10). Since Notch protein is a much larger protein with 2703 amino acid residues, it is difficult

	1				5					10					15					20					25					30				
Ab5-13-0	I	N	A	T	D	V	G	G	L	T	V	L	H	R	A	I	I	G	K	K	Q	A	I	T	N	Y	L	L	R	E	S	A	N
lin-12 CONSENSUS					D			G	R	T	A	L	H M		A	A		N						V		Y	L					A	
Notch, cdc 10, SW 16 CONSENSUS		N			D			G		T		L	H		A	A	R							V			L					A	N D

Figure 12. The sequence of the 33-amino acid motif from Arabidopsis clone AB5-13-0 and comparison to the consensus sequences from lin-12 of nematode, Notch of Drosophila, cdc 10 and SW 16 of yeast (13, 14).

to see how clone AB5-13-0 can encode for a protein with a similar function. However, a 33-amino acid segment within the open reading frame 2 was found to be similar to segments not just in Notch gene, but also in cell-cycle genes cdc 10, SW 16 of yeast and lin-12 gene of nematode Caenorhabditis elegans. The 33-amino acid motif repeats 6 times in lin-12 of C elegans (13), 5 times in Notch and 2 times in both cdc 10 and SW 16 of yeast (14). Figure 12 shows the sequence of the 33-amino acid motif of the clone AB5-13-0 and the consensus sequences of the motifs from Notch protein, cdc 10 gene, SW 16 gene, and lin-12 gene. As Breeden and Nasmyth predicted (14), this motif is conserved in a set of evolutionarily distant organisms and it will be found in other organisms. The presence of such motifs in genes of animal, plant, yeast and nematode suggests that it plays an important role in the cell.

The clone B5-2F1 was not completely sequenced. The gap which separates the sequenced parts, B5-60 and B5-61, is about 200 nucleotides long. But unlike AB5-10-0, this is a cDNA clone, and was isolated by oligonucleotide b_5-2. The sequence which recognized the oligonucleotide is located on B5-61. The coding sequence of B5-61 is completely different from animal cytochrome b_5 except for a region of five amino acid residues of identity from which we derived our oligonucleotide. The clone AB5-2F1 is not likely to be a plant cytochrome b_5 gene.

None of the clones sequenced (or semi-sequenced) were assigned a function. This indicates that the sequence of the plant cytochrome b_5 gene has diverged from that of the animal cytochrome b_5 gene to the point that other sequences have greater homology to the animal

gene at low stringency. After the rabbit and bovine liver cytochrome b_5 cDNA sequences were published recently (9, 11), I analyzed the nucleotide homology between the rabbit and bovine and between rabbit and the chicken liver cytochrome b_5. Even though the homology between rabbit and bovine is about 77%, the homology between rabbit and chicken is only 63%, whereas the amino acid sequence homology among animal cytochrome b_5 proteins is at least above 70%. We can predict that the homology between plant and animal cytochrome b_5 will be lower. This is probably why it is so difficult to clone the plant cytochrome b_5 by this approach. It would be better to select the "right" clone to sequence, instead of sequencing every possible clone. This can be done by using the candidate clones to hybrid-select mRNA, in vitro translating the hybrid-selected mRNA, using anti-cytochrome b_5 antibodies to precipitate the translated protein. Those clones which hybrid-select the cytochrome b_5 mRNA will be the right clones to sequence. Jollie et al. (1987) showed that the antibody raised againt the rat cytochrome b_5 can cross react with the plant cytochrome b_5 (3). This raised the possibility that the conservation of cytochrome b_5 might be extended to plant, and might well be at the nucleotide level. Unfortunately, this is a "Horse Behind Artillery" (a Chinese Saying which means It Is Too Late, And It's Over).

The use of mixed oligonucleotides overcomes the problems of codon usage and should have been a usable approach, if the plant cytochrome b_5 amino acid sequence is homologous to the animal proteins. This approach failed because of the failure of the b_5-3 probe to hybridize to anything in both genomic and cDNA libraries. The lack of recovery

of clones which hybridized to b_5-3 could be due to a bias in the library, or some other problems. Additional work would be needed to determine the causes. The 2.0 kb fragment detected on genomic Southern blots by oligonucleotide mixtures b_5-3 and b_5-5 is the best candidate for plant cytochrome b_5 gene.

5. References

1). C. Bonnerot, A. Galle, A. Jolliot, and J. C. Kader (1985). Biochem. J. 226:331-334

2). K. Madyastha, N. Krishnamachary (1986). Biochem. Biophys. Res. Commun. 136:570-576

3). D. R. Jollie, S. G. Sligar, and M. Schuler (1987). Plant Physiol. 85:457-462

4). K. Abe, S. Kimura, R. Kizawa, F. K. Anan, and Y. Sugita (1985). J. Biochem. (Tokyo) 91:1659-1668

5). T. Maniatis, E. F. Fritsch, and J. Sambrook (1982). Molecular Cloning, a Laboratory Manual. Cold Spring Harbor Laboratory. Cold Spring, New York

6). S. Henikoff (1984). Gene 28:351-359

7). H. Zhang, R. Scholl, J. Browse, and C. Somerville (1988). Nucleic Acids Res. 16:1220

8). H. Zhang, and C. Somerville (1988). Arch. Biochem. Biophys. 264:343-347

9). R.J. Cristiano, and A.W. Steggles (1989). Nucleic Acids Res. 17:799

10. S. Artavanis-Tsakonas (1988). Trends In Genetics 4:95-100

11). N. Dariush, C.W. Fisher, and A.W. Steggles (1988) Prot. Seq. Data Anal. 1:351-353

12). R.F. Doolittle (1986). Of Urfs And Orfs. ---A Primer on How to Analyze Derived Amino Acid Sequences. Univerity Science Books. Mill Valley, CA, USA

13). J. Yochem, K. Weston, and I. Greenwald (1988). Nature 335:547-550

14). L. Breeden, and K. Nasmyth (1987). Nature 329:651-654

CHAPTER V

SUMMARY AND SUGGESTIONS

The goals of this research initially were:

1. Clone the cytochrome b_5 genes from chicken and study its structure, expression, and the relationship of erythrocyte cytochrome b_5 mRNA and liver cytochrome b_5 mRNA;

2. Use the cloned chicken gene as a probe to clone the plant cytochrome b_5 genes, then study its regulation and expression in plant and possibly explore the relationship of its structure and function.

I have successfully accomplished the first goal, but failed in the second. The reason for this failure was discussed in chapter IV.

The findings of this research are summarized as follows:

1. The cloning of the chicken liver cytochrome b_5 cDNA was the first cloned cytochrome b_5 gene. The availability of the cDNA sequence revealed several new features about cytochrome b_5 structure. The chicken cytochrome b_5 gene encodes a protein which is 138-amino acid long, not 133-amino acid residues found for all the sequenced

cytochrome b_5 proteins. More heterogeneities were found at the N- and C-terminal ends. The chicken polypeptide lacks one amino acid at the C-terminus which is present on all the other known sequences. Several errors in the protein sequences were identified by comparison of the sequence deduced from the cDNA sequences and the published amino acid sequences.

2. Chicken erythrocyte cytochrome b_5 cDNA was also isolated and found to be identical in sequence to the cDNA from liver, except it was much longer. The cDNA from liver lacked a 3' poly(A) sequence, possibly due to premature termination during the second strand synthesis from mRNA. Northern Analysis of the size of the liver cytochrome b_5 mRNA indicated it was of the same size as the cDNA sequence of erythrocyte cytochrome b_5 mRNA.

3. The cloning and mapping of the genomic cytochrome b_5 gene suggested that there is only one cytochrome b_5 gene in chicken. This implys that all forms of cytochrome b_5 proteins are produced from the same gene.

4. Cytochrome b_5 mRNA is much more abundant in the liver than in erythrocytes, possibly because the liver is a very active tissue biosynthetically. A large quantity of cytochrome b_5 may have to be present since cytochrome b_5 plays a vital role for many biological reactions. The turnover of cytochrome b_5 in erythrocytes might be much slower than that in liver. In this way, the amount of cytochrome b_5 can be maintained at levels adequate to carry out all biological reactions therein.

Further research may be targetted at:

1. Using site-specific mutagenesis to alter the amino acid residues within hydrophobic region, study the interaction between the mutated protein and the membrane. This might tell us about the functions of amino acid residues within the hydrophobic domain and their contributions toward the two kinds of membrane binding of cytochrome b_5. Since our results indicate that it is posttranslationally modified, it might be interesting to attempt to identify the protease. This may represent a new class of protease since it must have substantial sequence specificity and, more importantly, it appears to insert a new amino acid residue to the C-terminal of the proteolytically processed microsomal cytochrome b_5 protein.

2. Further characterize the cytochrome b_5 genomic clones by completely sequencing them to locate all the introns, exons, and regulatory sequences.

3. Consider additional strategies to clone the plant cytochrome b_5 genes. If the cytochrome b_5 gene was cloned, it would be possible to study its expression and regulation a in plant systems. It might then be possible to use reverse genetics methods to explore its functions in plant. These methods could include the use of antisense mRNA to explore the mechanisms of organ-specific regulation or developmental regulation, and the use of the cloned gene to over express cytochome b_5 so that antibodies can be raised against it. Since cytochrome b_5 is about 100-fold less abundant in plants than in animals, it would be useful to have a more abundant

source for biochemical studies. Expression in yeast or E.coli might be suitable in this regard. Alternatively, it might be possible to overexpress the gene in transgenic plants.

APPENDIX I

TRANSFER OF THE MAIZE TRANSPOSABLE ELEMENT Mu1 INTO ARABIDOPSIS THALIANA

-----Reprint from PLANT SCIENCE,
48(1987), pp 165-173

Plant Science, 48 (1987) 165—173
Elsevier Scientific Publishers Ireland Ltd.

TRANSFER OF THE MAIZE TRANSPOSABLE ELEMENT *MU1* INTO *ARABIDOPSIS THALIANA*

HONG ZHANG and CHRIS R. SOMERVILLE*

MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824 (U.S.A.)

(Received June 25th, 1986)
(Revision received October 27th, 1986)
(Accepted October 29th, 1986)

The maize transposable element *Mu1* was transferred to *Arabidopsis thaliana* by Ti plasmid-mediated transformation and a fertile line containing *Mu1* was regenerated. Southern analysis of transformed tissue indicated that the *Mu1* DNA remained entirely within a segment of T-DNA during three sexual generations. The results of a search for spontaneous mutations in a large number of *Mu1*-containing seedlings suggests that the presence of *Mu1* did not cause a major increase in the spontaneous mutation frequency.

Key words:* *Arabidopsis thaliana*; transposable element; transformation; Ti plasmid; plant regeneration; Robertson's *Mutator

Introduction

The isolation of genes by transposon tagging has recently been accomplished in several plant species [1—3]. Unfortunately, the application of this potentially powerful approach is currently limited to those few species in which active transposable elements have been characterized at the genetic and molecular levels [4,5]. In order to extend this approach to other species it would be necessary to either identify endogenous transposable elements or to introduce a suitable element by transformation. Although the latter approach has recently become feasible for many plant species, there are many uncertainties concerning the autonomy of the plant transposable elements and the mechanisms involved in transposition [5].

The maize transposable element *Mu1* is thought to represent a member of a family of transposable elements which are present only in certain maize lines which are said to carry Robertson's *Mutator* [6,7]. These lines produce spontaneous mutations at a variety of loci at rates which are 30—50-fold higher than non-*Mutator* lines [7]. The *Mu1* element was identified as the causal agent of spontaneous unstable mutants at the *Adh-1* locus in a Robertson's *Mutator* line of maize [8]. Determination of the DNA sequence of *Mu1* [9] revealed an overall structure which is consistent with it being a transposable element. In particular, *Mu1* has inverted terminal repeats and creates direct repeats of target DNA at the site of insertion. Each strand has two open reading frames but no transcription product has been identified.

The *Mu1* element has recently been introduced into tobacco and tomato by Ti plasmid-mediated transformation [10]. The preliminary results indicated that *Mu1* DNA was present in the tobacco transformants in higher copy number than the T-DNA, and the pattern of restriction fragments which hybridized to *Mu1* were different than those

*To whom correspondence should be sent.
Abbreviations: 2,4-D, 2,4-dichlorophenoxyacetic acid; kbp, kilobase pair: SDS, sodium dodecyl sulfate.

0618-9452/87/$03.50 Elsevier Scientific Publishers Ireland Ltd.
Printed and Published in Ireland

of the transforming vector. However an unambiguous interpretation of the results was not possible. We report here the results of similar experiments in which *Mu1* was introduced into *Arabidopsis thaliana* by Ti plasmid-mediated transformation. *A. thaliana* was considered advantageous for such studies because the small size and rapid generation time permitted a facile analysis of the effect of *Mu1* on the spontaneous mutation rate.

Materials and methods

Plant material and growth conditions

The Columbia wild-type of *A. thaliana* (L.) Heynh. used for these studies was originally provided by G.P. Redei. Plants were generally grown on an artificial potting mixture irrigated with mineral nutrients at 23°C in natural light.

For some purposes sterile plants were grown from surface-sterilized seed in unsealed Petri plates (90 × 23 mm) containing mineral salts supplemented with 10 g/l sucrose and solidified with 7 g/l agar.

Bacterial strains and plasmids

Agrobacterium tumefaciens C58C1 rif^r carrying the nononcogenic Ti plasmid pGV3850 [11] was obtained from J. Schell. The plasmid pMJ9 [9], which carries *Mu1*, was obtained from M. Freeling. The plasmid pRK2013 and the intermediate vector pMON200 [12] which carries a bacterial spectinomycin resistance gene, a nopaline synthase gene and a neomycin phosphotransferase (NPTII) II gene was obtained from S. Rogers (Monsanto).

The growth of plasmid-bearing bacterial strains and the procedures for triparental matings were as described [12]. The *Escherichia coli* strain LE392 (F^- *hsdR supE supF lac gal metB trpR*) was used as the host for all *E. coli* plasmids described here.

Tissue culture media

The media used for tissue culture were modifications of those described by Negrutiu et al. [13] for *A. thaliana*. The basal medium contained the mineral salts of Murashige and Skoog [14] supplemented with 20 g/l sucrose, 0.4 mg/l glycine, 0.1 mg/l nicotinic acid, 1 mg/l thiamine-HCl, 0.1 mg/l pyridoxine-HCl, 7 g/l agarose at pH 5.8. Phytohormones were added to the media after autoclaving in the following amounts: MSa1 contained 1 mg/l 2,4-D, 0.05 mg/l kinetin; MSa2 contained 0.5 mg/l 2,4-D, 0.1 mg/l kinetin; PR29 contained 5.0 mg/l naphthaleneacetic acid, 0.5 mg/l benzyladenine; PR33i contained 1 mg/l isopentenyl adenine, 0.1 mg/l indole acetic acid; MSa4 contained one fifth the concentration of salts, 5 g/l sucrose, normal levels of vitamins but no phytohormones.

Transformation of callus tissue

Callus was established by placing sterile leaves on MSa1 for 3 weeks in continuous low light. The callus was then subcultured on MSa2 for at least 2 weeks before use. Approximately 0.5 g of callus was chopped into pieces of about 2 mm diameter and covered with 10 ml of liquid MSa2. Two drops of a fresh saturated culture of *A. tumefaciens* was added to the cell suspension which was left shaking at 40 rev./min for 20 h at 20°C in low light. At this time the liquid was discarded and replaced with fresh MSa2 containing 250 μg/ml Cefotaxime (Calbiochem) and the incubation continued for 5 days with slow shaking. The cell slurry was then transferred to solid MSa2 medium containing 250 μg/ml Cefotaxime and 25 μg/ml G418 (Gibco) and incubated at 23°C in continuous low light. After several subcultures on the same medium over a period of 4—6 weeks, the G418-resistant callus was transferred to PR29 for 4 weeks without antibiotic selection then transferred to PR33i to induce shoot formation. The transfer to PR29 before attempting shoot regeneration appears to enhance the frequency of shoots in long-term callus cultures. Shoots were rooted by dipping the base of the shoots in 0.1 mg/ml NAA then transferring the shoots

to MSa4. After roots had developed the plants were transferred to potting mixture.

Opine assays

The presence of nopaline in transformed tissue was determined by the method of Rogers et al. [12]. Nopaline was visualized by grinding single leaves of approximately 0.5 cm^2 surface area in 3 μl of water and applying the entire extract to the paper.

DNA Manipulations

Arabidopsis DNA was prepared as described previously [15]. Southern hybridizations were performed at 68°C for 36 h in 6 × SSC (SSC is 150 mM NaCl, 15 mM sodium citrate, pH 7.0), 0.5% (w/v) SDS, 0.1% (w/v) Ficoll, 0.1% (w/v) polyvinyl pyrrolidone, 0.1% (w/v) bovine serum albumin 0.1 mg/ml denatured salmon sperm DNA and 10 mM EDTA. The filters were washed at 68°C as follows: 5 min in 2 × SSC, 0.5% SDS; 15 min in 2 × SSC, 0.1% SDS; 2 h in 0.1 × SSC, 0.5% SDS; 30 min in 0.1 × SSC, 0.5% SDS. All other nucleic acid manipulations were performed by standard methods [16].

Results

Plasmid constructions

Mu1 was introduced into the non-oncogenic Ti plasmid pGV3850 in two steps. First, the 2.9 kbp BamHI-HindIII fragment from pMJ9 was ligated into the BglII-HindIII sites of pMON200 and transformed into *E. coli* strain LE392. The resulting plasmid, designated pHZ1 (Fig. 1), was then introduced into the *A. tumefaciens* strain C58C1(pGV3850) by a triparental mating in which *E. coli* strain LE392(pRK2013) provided the mobilization functions [12]. Since pMON200 derivatives do not replicate in *A. tumefaciens*, a cointegrate between pHZ1 and pGV3850 was obtained by selecting for stable expression of the spectinomycin resistance gene carried on pHZ1 in *A. tumefaciens*.

There are two regions of homology between pMON200 (or pHZ1) and the modified

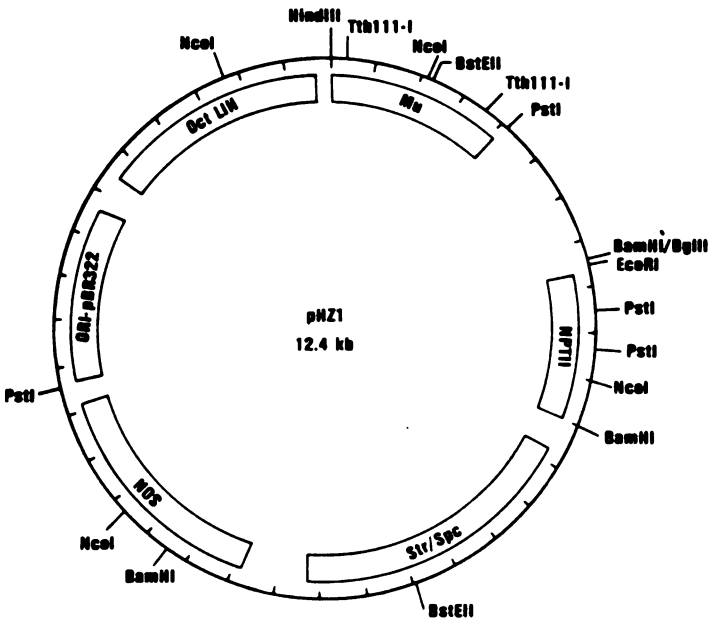

Fig. 1. Map of the *Mu1*-containing intermediate vector pHZi.

168

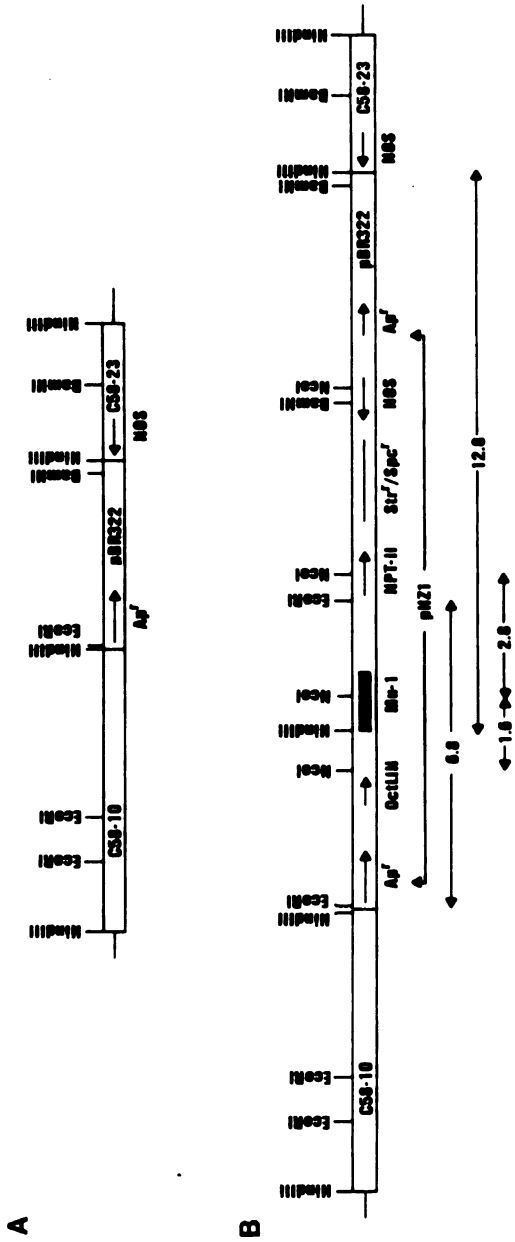

Fig. 2. The arrangement of *Mul* in the cointegrate plasmid pGV3850::pHZ1. (A) The modified T-DNA region of pGV3850 [11] in which most of the T-DNA has been replaced by pBR322 which is joined to pTiC58 HindIII fragments 10 and 23. (B) Integration of pHZ1 into pGV3850. The arrows below the figure indicate the size of the corresponding restriction fragments in kbp.

T-DNA region of pGV3850. Both plasmids contain the nopaline synthase gene and the origin or replication of pBR322. Thus, if cointegration occurs by a single crossover between the two circular plasmids, there are two possible configurations for the organization of DNA sequences in the cointegrate pGV3850::pHZ1. Southern hybridization analysis of DNA from C58C1(pGV3850::pHZ1) using the internal Tth111I fragment from *Mu1* as a probe was consistent with the orientation shown in Fig. 2B (results not presented) in which *Mu1* is positioned between the left and right border sequences of pTiC58 carried on HindIII fragments 23 and 10, respectively [11]. Thus, the configuration is that expected if cointegrate formation occurred by a single recombinational event between the regions of pBR322 homology on pHZ1 and pGV3850.

Transformation of Arabidopsis

Previous studies had shown that *A. thaliana* could be transformed with tumorigenic strains of *A. tumefaciens* [17]. However, as there were no previous reports of transformation of *A. thaliana* with a disarmed Ti plasmid we explored several approaches to transformation. We initially observed a very low frequency of transformation of *A. thaliana* by the leaf disc procedure used with other species [12] but obtained a satisfactory frequency of transformants with minor modifications of the methods used by Muller et al. [18] to transform tobacco callus. By this method, 5–7% of the calli exposed to *A. tumefaciens* C58C1(pGV3850::pHZ1) developed outgrowths which were resistant to at least 25 μg/ml G418.

We obtained a very low frequency of shoot regeneration from the G418-resistant calli using published procedures [13]. We are uncertain whether this is due to the properties of the accession of *A. thaliana* we have used or some other reason. However, it has previously been noted that the regeneration potential of *A. thaliana* callus declines sharply after several months in culture [13]. Whatever the case, from several dozen G418-resistant callus lines we obtained only five shoots and from these we recovered only one fertile plant which accumulated nopaline (results not presented). This line, which was very vigorous, was designated MSU252 and the original regenerant is designated the R1 generation.

Analysis of Mu1 DNA in transformants

Southern Analysis of the restriction pattern of DNA from G418-resistant callus tissue using the internal Tth111I fragment from *Mu1* as a hybridization probe indicated that the G418-resistant calli contained sequences homologous to *Mu1* This is evident from the results in Fig. 3 in which the combined DNA from eight independently transformed G418-resistant calli was cleaved with EcoRI to

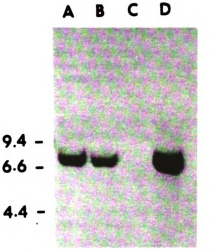

Fig. 3. Southern blot of DNA from wild-type and G418-resistant tissue probed with *Mu1* DNA. The DNA from all sources was restricted with EcoRI. The position and size (kbp) of molecular weight markers is indicated at the left side of the figure. Lanes: (A) DNA from eight G418-resistant calli; (B) DNA from leaf tissue of an R3 plant descended from a G418 resistant callus; (C) wild-type *A. thaliana;* (D) A mixture of equivalent amounts of DNA from *A. tumefaciens* C58C1 (pGV3850::*Mu1*) and the DNA preparation in lane B.

release the *Mu1* DNA and surrounding sequences on a 6.8 kbp fragment (Fig. 2). DNA from leaf tissue of the wild-type showed no homology to *Mu1* under the conditions used. The only band of homology in DNA from the callus tissue was the same size as the major band of homology in DNA from *A. tumefaciens* carrying the Ti-plasmid pGV3850::pHZ1, and from leaf tissue of a single R3 plant descended from a transformed regenerant. Thus, it is inferred that the *Mu1* element did not rearrange or transpose from the surrounding vector DNA during the early cell divisions following the initial transformation events that gave rise to the eight callus lines.

A more detailed analysis was conducted on DNA from plants of the R2 and R4 generations of the *Mu1*-containing line MSU252. DNA from individual plants or the combined DNA from 30 plants was cleaved with NcoI, EcoRI and HindIII transferred to nitrocellulose and probed with the Tth111I fragment from *Mu1* (Figs. 4 and 5). The apparent molecular weights of the fragments which hybridized with the probe are consistent with those expected from the analysis of the orientation of *Mu1* sequences in pGV3850::pHZ1 (Fig. 2). Thus, when cleaved with EcoRI, a single 6.8 kbp fragment showed homology to *Mu1* When the DNA was cleaved with NcoI, which cuts within *Mu1*, two fragments of 1.6 and 2.8 kbp were the most prominent bands, and when cleaved with HindIII the major signal was a 12.8 kbp band. A less intense signal corresponding to a HindIII fragment of about 4.0 kbp in the HindIII digests of the transformants was also found in digests of *A. tumefaciens* C58C1-(pGV3850) (results not presented). On the basis of this observation, and considering the size and intensity of this band, it is probably due to contamination of the gel-purified probe with vector sequences which hybridize with the pBR322 sequences present in the T-DNA region of pGV3850. Thus, there was no evidence that *Mu1* DNA has rearranged or moved from the adjoining vector sequences during the four generations following transformation.

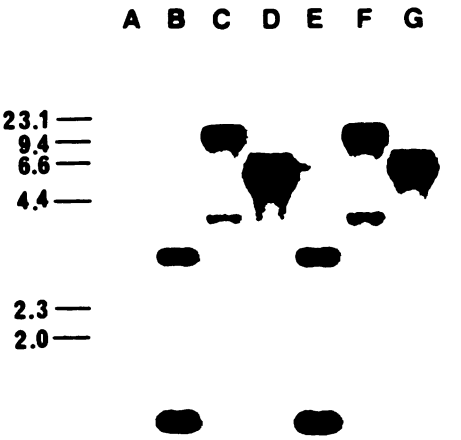

Fig. 4. Southern blot of DNA from transformed plants of *A. thaliana* probed with *Mu1* DNA. The position and size (kbp) of molecular weight markers are indicated at the left side of the figure. Lanes: (A) wild-type *A. thaliana* digested with EcoRI; (B, C and D) DNA from 30 R2 plants digested with NcoI, HindIII and EcoRI, respectively; (E, F and G) DNA from one R2 plant digested with NcoI, Hind III and EcoRI, respectively.

Inheritance of introduced DNA

On the basis of the intensity of the hybridization signal obtained with *Mu1* DNA as the probe, the *Mu1*-containing line MSU252 carried approximately 2—5 copies of *Mu1* (results not presented). The inheritance of the introduced DNA in line MSU252 was followed for four generations by scoring for nopaline production or antibiotic resistance. Twenty-eight randomly chosen R2 progeny resulting from the self-fertilization of the R1 plant were examined for the presence of nopaline and were all found to be positive. Since a single insertion of T-DNA would be expected to behave as a single dominant mendelian locus, this observation suggested

Fig. 5. Southern blot of DNA from a variegated *Mu1*-containing plant. Lanes A to E were digested with EcoRI. Lanes F to I were digested with HindIII. (A) pMJ9; (B and F) DNA from a single variegated R4 plant; (C, D, G and H) DNA from single non-variegated R4 plants; (E and I) wild-type *A. thaliana*.

that the original transformation event was more complex. One hundred R2 progeny were also scored for G418-resistance by germinating surface-sterilized seed on agar-solidified mineral medium to ensure that spurious problems with seed viability did not bias the results, then transferring the seedlings to minimal medium containing 25 μg/ml G418. Ninety-seven R2 progeny were G418-resistant and three were sensitive. This pattern of inheritance is an acceptable fit to the hypothesis that resistance is due to two unliked dominant nuclear genes which were present in heterozygous condition in the original regenerant ($\chi^2 = 1.80; P > 0.1$).

The R3 progeny resulting from self-fertilization of nine R2 individuals were also scored for G418-resistance in the same way. Two lines had all G418-resistant progeny. The progeny of two other lines had segregation ratios which were excellent fits to a 3:1 (resistant/susceptible) pattern, and the other five lines had patterns of resistance which were an acceptable fit to a 16:1 (resistant/susceptible) pattern of inheritance (results not presented). These results are also consistent with the proposal that the original regenerant was heterozygous for two unliked copies of the NPTII gene each of which could confer G418-resistance.

In order to obtain a line of *A. thaliana* with simple inheritance of G418-resistance a number of the R3 individuals from the two lines showing 3:1 segregation of the resistance marker were self-fertilized and the resulting R4 progeny scored for G418-resistance. This resulted in the recovery of two lines, designated MSU253 and MSU254, which produced only G418-resistant R5 progeny and are, therefore, presumed to be homozygous for a simply-inherited G418-resistance determinant.

Spontaneous mutation rate

In order to examine the possibility that the *Mu1* element might enhance the spontaneous mutation rate we examined approximately 30 000 R3 seedlings descended from self-fertilization of 227 phenotypically wild-type R2 individuals for mutations affecting chlorophyll content. Ten variegated individuals were observed in the R3 progeny whereas none were observed among a similar number of wild-type seedlings. No other obvious mutant phenotypes were observed among the 30 000 R3 seedlings. Thus, aside from the variegated phenotype, no increase in mutation frequency was observed in this experiment.

A recessive nuclear mutation which causes

variegation has previously been recovered from a plant regenerated from untransformed callus tissue of *A. thaliana* [19]. Similarly, in the R2 generation of line MSU252, four of 231 individuals were variegated. Most and possibly all of the progeny resulting from self-fertilization of the variegated individuals appeared to be either variegated or completely chlorotic. Several progeny from these plants did not appear to be variegated, but because the amount of chlorotic tissue varied widely from individual to individual, small variegated sectors may have been missed on these individuals.

The possibility that the variegation was directly related to *Mu1* was investigated by Southern analysis of DNA from leaf tissue of a heavily variegated plant. However, no difference was observed in the restriction pattern of the sequences with homology to the Tth111I fragment from *Mu1* (Fig. 5). Thus, the variegation does not appear to be due to a rearrangement of *Mu1* DNA in the *A. thaliana* genome.

Discussion

Bennetzen [6] has estimated that in order to account for the observed increase in spontaneous mutation rate in Robertson's *Mutator* lines of maize, each of the large number of *Mu1* sequences present in these lines would have to transpose, on the average, once every plant generation. Since we did not observe any evidence for a transposition event among more than 30 individuals separated by two or three generations from the original transformant it appears that, if *Mu1* is an autonomous element, the frequency of transposition is markedly reduced in *A. thaliana* as compared with maize. In this regard our results substantiate similar experiments in which *Mu1* did not appear to transpose in tomato [10].

In order to examine the possibility of *Mu1* transposition by genetic criteria we examined a relatively large number of seedlings for spontaneous mutations. There are more than 100 loci in both barley and maize at which mutations giving rise to chlorophyll-deficient phenotypes are known to occur. Thus, the frequency of chlorotic mutations is a relatively sensitive indicator of mutation rate. Since no chlorotic mutations were observed in approximately 30 000 seedlings it is apparent that the presence of *Mu1* DNA did not cause a major increase in the spontaneous mutation rate. The only genetic anomaly observed was the occurrence of variegated plants in the *Mu1*-containing line. However, as we did not observe any rearrangement of the *Mu1* DNA in variegated leaf tissue, the variegation appears to be an unrelated phenomenon which may, nevertheless, be relevant in another context to the search for a transposable element in *A. thaliana*.

As with most negative results this experiment does not provide any information as to the reason that *Mu1* does not appear to transpose. As noted by others [20] it may be that *Mu1* is not the autonomous element in Robertson's *Mutator* lines or host-specific factors may be required. Alternatively, since the frequency of transposition of *Mu1* is correlated with copy number [9] it may be that the copy number of *Mu1* in the transformants we obtained is too low. Whatever the case, on the basis of the results presented here we do not consider *Mu1* a promising candidate for a broad-host-range plant transposon.

As there have not been previous reports of transformation of *A. thaliana* with non-oncogenic Ti plasmids, a few comments on the methodology seem appropriate. First, it should be noted that the protocol described here for obtaining transformed plants is relatively slow and the conditions we have employed may not be optimal. The methods recently developed by others appear to substantially more efficient [21]. Second, although the NPTII gene from pMON200 does confer resistance to G418 in *A. thaliana*, there is a relatively narrow range of antibiotic concentration at which the wild-type tissue does not grow and the transformed

tissue will grow. Thus we believe that alternate selectable markers which provide greater discrimination between transformed and non-transformed tissue would be very useful.

Acknowledgments

We thank R. Horsch for advice concerning the transformation of *Arabidopsis* and J. Schell, M. Freeling, W. Taylor, J. Bennetzen and S. Rogers for the generous gifts of plasmids. This work was supported in part by grants from the National Science Foundation (No. PCM8351595) and the U.S. Department of Energy (AC02-76ER01338).

References

1 N. Fedoroff, D.B. Furtek and O.E. Nelson Jr., Proc. Natl. Acad. Sci. U.S.A., 81 (1984) 3825.
2 C. Martin, R. Carpenter, H. Sommer, H. Saedler and E.S. Coen, EMBO J., 4 (1985) 1625.
3 C. O'Reilly, N.S. Shepherd, A. Pereira, Z. Schwarz-Sommer, I. Bertram, D.S. Robertson, P.A. Peterson and H. Saedler, EMBO J., 4 (1985) 877.
4 M. Freeling, Ann. Rev. Plant. Physiol. 35, (1984) 277.
5 P. Nevers, N.S. Shepherd and H. Saedler, Adv. Bot. Res., 12 (1986) 103.
6 J.L. Bennetzen, J. Mol. Appl. Genet., 2 (1984) 519.
7 D.S. Robertson, Mutat. Res., 51 (1978) 21.
8 J.N. Strommer, S. Hake, J.L. Bennetzen, W.C. Taylor and M. Freeling, Nature (London), 300 (1982) 542.
9 R.F. Barker, D.V. Thompson, D.R. Talbot, J. Swanson and J.R. Bennetzen. Nucleic Acids Res., 12 (1984) 5955.
10 M. Lillis, A. Spielmann and R.B. Simpson, in: M. Freeling (Ed.), Plant Genetics, Alan R. Liss, New York, 1985, p. 213.
11 P. Zambryski, H. Joos, C. Genetello, J. Leemans, M. Van Montagu and J. Schell, EMBO J., 2 (1983) 2143.
12 S.G. Rogers, R.B. Horsch and R.T. Fraley, Methods Enzymol., 118 (1986) 627.
13 I. Negrutiu, F. Beeftink and M. Jacobs, Plant Sci. Lett., 5 (1975) 293.
14 T. Murashige and F. Skoog, Physiol. Plant., 15 (1962) 473.
15 L.S. Leutwiler, B.R. Hough-Evans and E.M. Meyerowitz, Mol. Gen. Genet., 194 (1984) 15.
16 T. Maniatis, E.F. Fritsch and J. Sambrook, Molecular Cloning, Cold Spring Harbor, 1982.
17 M. Aerts, M. Jacobs, J.P. Hernalsteens, M. Van Montagu and J. Schell, Plant Sci. Lett., 17 (1979) 43.
18 A. Muller, T. Manzara and P.F. Lurquin, Biochem. Biophys. Res. Commun., 123 (1984) 458.
19 J. Martinez-Zapater, S.C. Somerville and C.R. Somerville, in: M. Freeling (Ed.), Plant Genetics, Alan R. Liss Inc., New York, 1985, p. 828.
20 L.P. Taylor and V. Walbot, EMBO J., 4 (1985) 869.
21 A.M. Lloyd, A.R. Barnason, S.G. Rogers, M. Byrne, R.T. Fraley and R. Horsch, Science, 234 (1986) 464.

APPENDIX II

TRANSFER OF THE Ac ELEMENT TO ARABIDOPSIS BY SEED TRANSFORMATION

1. Introduction

The Ac transposable element is an autonomous element in Maize (1, 2), and has been successfully used to clone bz locus by transposon tagging (3). Furthermore Ac was recently shown to excise from chromosomal DNA in a heterologous transgenic plant system (4). Thus, Ac is a good candidate to be put into Arabidopsis where there is no endogenous element available for transposon tagging.

Although it was posssible to obtain transgenic plants (Appendix I), Arabidopsis transformation had been very difficult until recently. In order to try and improve the frequency, Feldmann developed the seed transformation protocol (5). With this simple protocol, transformation seems extremely easy. Germinating seeds of Arabidopsis were cocultivated with an Agrobacterium for 24 hrs before being planted in pot. Seedlings resulting from this treatment were called T1 generation. The selection is applied in the T2 generation on a selective petri dishes. Transformants will grow with suitable antibiotics selection.

The following experiments were carried out in an attempt to transfer Ac to Arabidopsis by this transformation protocol.

2. Materials and methods

Plasmid pAc9 contained intact Ac element, and was from N. Fedoroff (Carnegie Institution of Washington). Vector pBIN19 (6) was from M. Bevan (Plant Preeding Institute, Cambridge). Agrobacterium strain LBA 4404 was from M.D. Chilton (Ciba-Geigy Biotecnology). Arabidopsis seeds (WS race) were from Feldmann (Sandoz Crop Protection).

Agrobacterium transformation protocol:

Grow 2 ml of LBA 4404 in YEP[a] plus 50ug/ml rifampicin overnight at 28°C

Innoculate 50 ml YEP plus rifampicin with entire 2 ml culture

Continue to shake vigorously at 28°C for 4 hrs (OD_{600}>1.0)

Pellet in sterile Oak Ridge tubes at 7000 rpm 4°C for 10 minutes

Wash each tube in 2.5 ml of 150 mM NaCl (ice cold)

Pool in one tube and pellet again

Resuspend in 0.5 ml of ice cold 75 mM CaCl

Add 0.2 ml cells to 20 ul pBIN::Ac DNA (10-20 ug) in Eppendorf tube

Leave on ice 30 min, then freeze in dry ice/ethanol 5 min

Thaw at 37°C for 3 min, then innoculate into 5 ml of YEP

Incubate at 28°C 1 hour, then pellet, resuspend in 1 ml YEP

Plate entire transformation mixture over 4-8 selective plates[b]

[a] YEP (per liter):

10 g peptone

10 g yeast extract

5 g NaCl

[b] Selective Plate (per ml)

50 ug Kanamycin

50 ug Rifampicin

200 ug Streptomycin

Arabidopsis seed transformation protocol:

Surface sterilize seeds with 30% bleach plus 1ul/ml of 20% Triton-X100

Imbibe 3000-5000 seeds in 50ml of Murashige and Skoog (MS) salts (8), 4% sucrose, vitamin[c] for 12 hrs, Shake imbibition mix, maintain temp at 22^{o}C in dim light (about 1000 lux)

Add 3-5ml overnight culture of Agrobacterium (3-5 X 10^{9} cells of LBA 4404::Ac) into the imbibition mix

Shake resultant culture for 24 hrs at 28^{o}C, in dim light

Wash seeds on a Buchner funnel with water, then leave seeds on 3MM papers

Allow them to dry in a hood until just dry enough to plant

Plant about 200 seeds in one pot, place all pots in 4^{o}C for 24 hrs

Grow plants in greenhouse to harvest and collect seeds in bulk, usually get about 150-200 seeds per plant

Plate T2 progeny seeds on medium containing 100 ug/ml of Kanamycin

[c]Vitamin (per liter):

10.0 mg Thiamine

0.5 mg Pyridoxine

0.5 mg Nicotinic Acid

100.0 mg Inositol

All DNA manipulations were performed according to the methods in molecular cloning (7).

3. Results:

1). Transforming vector construction

pAc9 plasmid (pKP32::Ac9) was cut with SalI, and ligated into pBIN19 at SalI site. See figure 1 below.

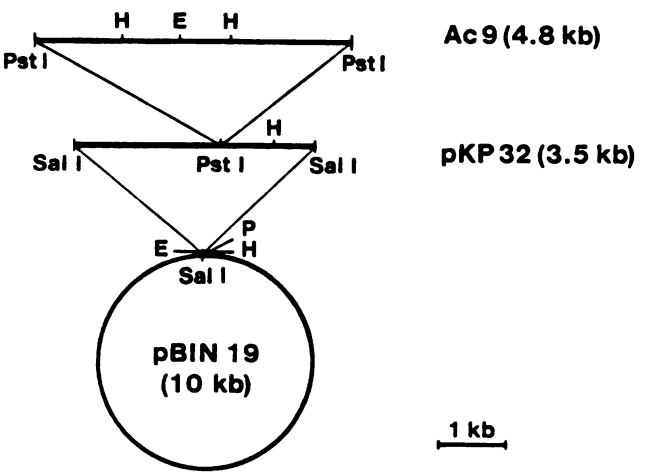

Figure 1. The construction of the transforming vector pBIN19::Ac.

2). Agrobacterium transformation

The transforming vector pBIN19::Ac made from above construction was directly transformed into Agrobacterium LBA4404. Three Kam^R transformants were obtained. However only one showed to have the unrearranged pBIN19::Ac plasmid. See the expected restriction hybridization pattern in figure 2 below.

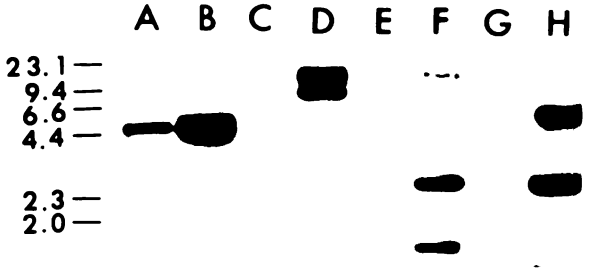

Figure 2. Restriction pattern of two selected transforming Agrobacteria. Strain 1 (A, C, E and G) and strain 2 (B, D, F and H) were digested with enzymes Pst I (A and B), Sal I (C and D), Hind III (E and F), and Bgl I(G and H). Only strain 2 gave the expected pattern.

3). Arabidopsis Transformation

The Agrobacterium which contained the correct pBIN19::Ac vector was used to infect approximately 20,000 germinating seeds of Arabidopsis Ws race according to the protocol developed by Feldmann. From approximately 300,000 'T2' progeny of treated seedings, about 30 appeared to be more resistant than wild type to grow on selection plates. Figure 3 showed some resistant 'T2' seedlings on Kanamycin

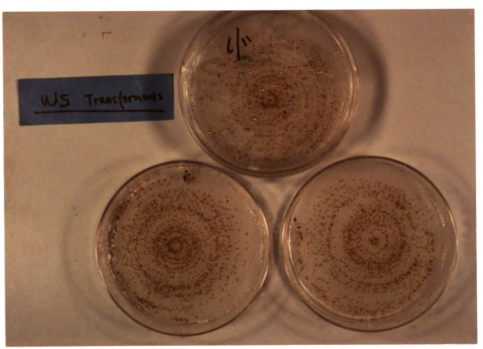

Figure 3. An example of some resistant T2 seedlings on Kanamycin plates.

plates. However, only 5 seedings have survived after transplanting from the selective medium to potting mixture. I isolated the DNA from the most vigorously growing putative transformant, and cut it with EcoRI and HindIII, then hybridized it to Ac element. Unfortunately, the hybridization pattern did not match the expected pattern. See figure 4 below. The result simply can not be explained.

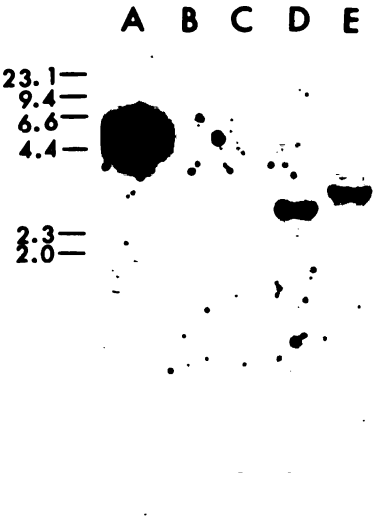

Figure 4. Restriction pattern of one putative transformant. Lane A, Ac element alone. Lanes B and C, wild-type Arabidopsis WS race and Columbia race cut with Eco RI. Lanes D and E, the putative transformant cut with Eco RI and Hind III respectively.

4. Discussion

The result in figure 4 can not be explained by premature breakage during single stranded T-DNA transfer, because within 4 kb there is no other EcoRI site from either side of the EcoRI site in the Ac construct (see figure 1). Where did the 2.6 kb EcoRI band from? We can only assume that there was a strange rearrangement which happened

after T-DNA transfer (or just before). I have to believe that this putative transformant must have some DNA sequences which hybridize to Ac element, because the control DNAs (both Columbia and Ws wild type of Arabidopsis) did not show any signals at all! If they are true transforants, the transformation frequency would be 1/10,000 or 1/60,000 (from 30/300,000 or 5/300,000), which is much lower than 4/1,000 Feldmann originally reported (5). Nevertheless, the technical simplicity of this seed transformation method seems very atractive, even though there are tremendous amount of work involved.

5. References

1). H.-P. Doring and P. Starling, Ann. Rev. Genet. (1986). 20:175-200

2). N. V. Fedoroff, Cell (1989). 56:181-191

3). N. Fedoroff, D. B. Furtek, and O. E. Nelson, Jr (1984). Proc. Natl. Acad. Sci. USA 81:3825-3829

4). B. Baker, J. Schell, H. Lorz, and N. Fedoroff (1986). Proc. Natl. Acad. Sci. USA 83:4844-4848

5). K. A. Feldmann, and M. D. Marks (1987). Mol. Gen. Genet. 208:1-9

6). M. Bevan (1984). Nucleic acids Res. 12:8711-8721

7). T. Maniatis, E. F. Fritsch, and J. Sambrook (1982). Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Sporing Harbor, New York

8). T. Murashige, and F. Skoog (1962). Physil. Plant. 15:473-497

APPENDIX III

SEARCHING FOR THE DESATURASE GENE IN ARABIDOPSIS AND AGMENELLUM

1. Introduction

A cDNA clone coding for rat liver stearoyl-CoA desaturase gene had been isolated and sequenced by Thiede et al (1). Unfortunately, this clone did not recognize any sequences in Arabidopsis by genomic Southern analysis (Hugly, personal comm.). Martin et al cloned and sequenced the yeast stearoyl-CoA desaturase gene by complementing the Ole 1 mutation in yeast. (personal comm.). By comparing the coding sequence of the yeast gene with that of the rat gene, regions of homology can been found (figure 1).

We believe that these homologies represent the conservation of the functional domains of stearoyl-CoA desaturase, and may extended to plants and cyanobacteria. We designed mixed oligonucleotides from these conserved regions and examined their use as probes to clone the plant and blue-green desaturase genes. In addition, we also used the yeast gene as a heterologous probe, because, from preliminary Southern blotting, the yeast gene seemed to recognize a few sequences from Arabidpsis (figure 2).

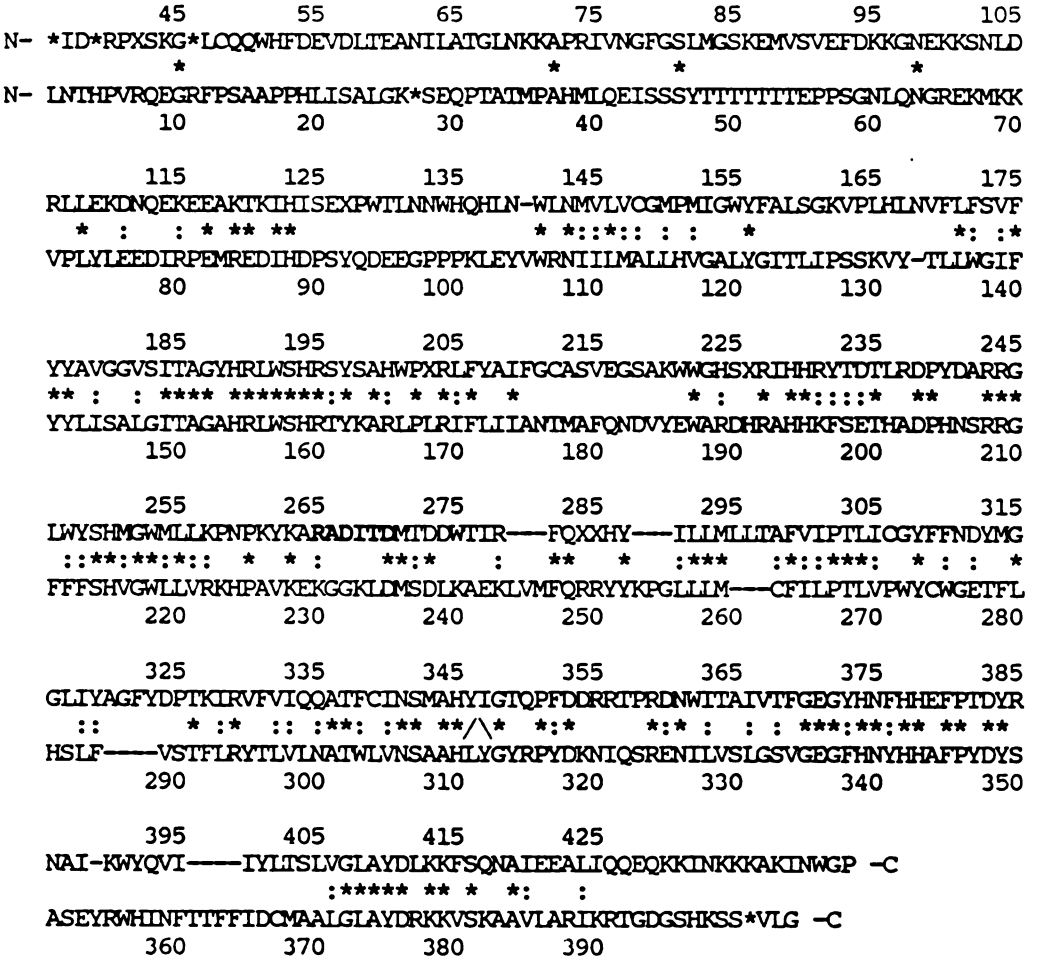

Figure 1. Comparison of the deduced amino acid sequence of yeast stearoyl-CoA desaturase gene with that of the rat stearoyl-CoA desaturase gene. Upper strand is yeast sequence.

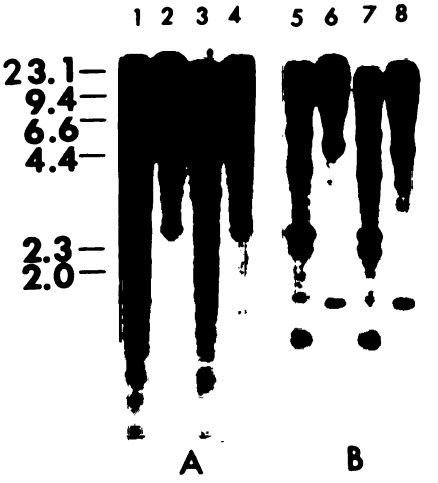

Figure 2. Arabidopsis genomic DNA digests were probed with p433 (A) and p403 (B) both of which contain parts of the yeast stearoyl-CoA gene. Lanes 1, 2, 5 and 6 are Columbia race; lanes 3, 4, 7 and 8 are WS race. Lanes 1, 3, 5 and 7 are Eco RI digests,; Lanes 2, 4, 6 and 8 are Bam HI digests.

	His	Arg	Leu	Trp	Ser	His	Arg
D1T	CAC	AGI	CTI	TGG	TGI	CAC	AG
	T	C	T		C	T	C
	His	Arg	Leu	Trp	Ser	His	Arg
D1A	CAC	AGI	CTI	TGG	AGI	CAC	AG
	T	C	T		C	T	C
	Glu	Gly	Phe	His	Asn	Tyr	His
D2	GAA	GGI	TAC	CAC	AAC	TTC	CA
	G		TT	T	T	AT	
	Leu	Ala	Tyr	Asp	Arg	Lys	Lys Val
D3	CTI	GCI	TAC	GAC	IGI	AAA	AAA TT

Figure 3. Oligonucleotides designed from the conserved regions of the stearoyl-CoA desaturase.

2. Materials and methods:

The plasmids p433 and p403 both containing part of the yeast desaturase gene were from D.C. Martin (Rutgers, New Jersey). The conditions for oligonucleotide hybridization, southern blotting, and sequencing were essentially the same as in chapter II. The heterologous probing using yeast gene to screen Arabidopsis libraries was carried out under conditions exactly like that used in oligonucleotide hybridization.

The Arabidopsis lambda gt10 cDNA library was from N. Crawford (stanford, CA). The Arabidopsis lambda gt10 A1 library was made by digesting Arabidopsis genomic DNA with EcoRI, and collecting fragments from 0.4-3.5kb, then ligating them into lambda gt10 arms. I made an Agmenellum lambda gt10 Ag1 library the same way as A1 libyary, except I collected fragments from 2.7-5kb. An Agmenellum lambda 2000 library was made by completely digesting Agmenellum DNA with EcoRI, then directly packaging them into lambda 2000 arms. The original titers for all these libraries were at least above 10^5 pfu. The oligonucleotides designed from the homologies between the yeast and rat desaturase genes are shown in figure 3.

3. Results

1). Yeast desaturase gene as a probe

We screened both an Arabidopsis cDNA library and the A1 genomic library by using the yeast desaturase gene as a heterologous probe.

```
         10         20         30         40         50         60         70         80
5' ACATCGATCC AAGAGATGTT TAGGCGGGTG AGCGAGCAGT TCACTGCTAT GTTCAGGAGG AAAGCTTTCT TGCATTGGTA

         90        100        110        120        130        140        150        160
   CACAGGTGAA GGAATGGACG AGATGGAGTT TACTGAAGCT GAGAGCAACA TGAACGATCT AGTCTCAGAG TACCAGCAAT

        170        180        190        200        210        220        230        240
   ACCAAGACGC AACTGCAGAT GACGAAGGCG AGTATGAAGA AGACGAGGAT GAAGAAGAGA TATTGGATCA TGAGTGAGTG

        250        260        270        280        290        300        310        320
   AAAAGAGCTG ATATTACCGA TTTTTAAATA CCTCTCTTAT CTTCTTTTCG TTTGGTCGGT ATATGTTTAT TGAGTTTCAT

        330        340        350        360        370        380        390        400
   GTATTGTTTG TGATGGTCTG TGTGTAATAG TGAGGTCGGA TCTAAACTTT TATGCGTGGT TTTTATATGA AATTGTCCAT


   TGTGGTT 3'
```

Figure 4. DNA sequence of the clone D2-62.

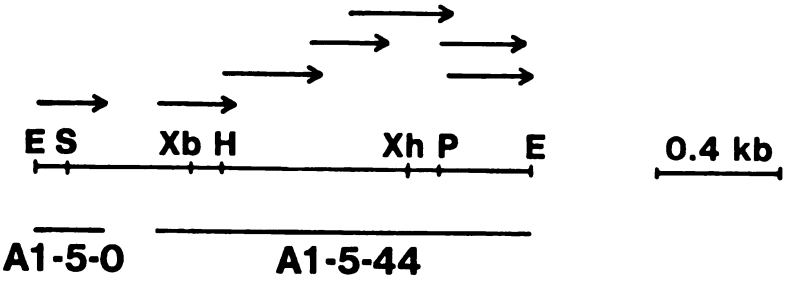

Figure 5. Sequencing strategies for clone A1-5.

I isolated 1 clone from the cDNA library and 7 clones from the A1 library. Among the 7 clones from A1 library, only three are independent fragments because they migrated differently during electrophoresis (results not shown).

The clone from the cDNA library was named D2-65, because it was recognized by D2 oligo also. The insert was excised by digestion with EcoRI and cloned into the EcoRI site of pBluescript to produce plasmid pD2-65. This clone is a very short sequence, only 410 bp. The deletion and sequencing were done as described in Methods and Materials section of previous chapters. The DNA sequence of this clone is shown in figure 4.

After we searched the protein bank (PIR R18.0 September 1988) with all three reading frames of this clone, to our surprise, reading 3 was found to encode beta-Tubulin. The homology between the coding sequence of D2-65 and sea urchin beta-Tubulin is as high as 85% in 73 amino acids overlap. The beta-tubulin gene family of Arabidopsis has been characterized by Oppenheimer et al (2). However, the clone D2-65 appears to represent a new member of the gene family, since the published sequences do not completely match the sequence of D2-65. When I compared the sequence of clone D2-65 with that of the yeast desaturase gene, I found that there is a 18 nucleotides identity continuously between the two sequences. The oligonucleotides D2 has 77% identity (14/18) to the sequence from nucleotides 149 to 167 of clone D2-65. That is why the clone D2-62 was picked up by yeast desaturase gene and oligonucleotides D2.

A clone designated A1-5 among the 7 isolated from Arabidopsis A1 library has an insert of 1.6 kb, and was chosen for sequence analysis

A1-5-0

```
          10         20         30         40         50         60         70         80
5' CTCAGTTTGG TGAAGTCGAC ATTGATGGAT CACTTGTGGC AGCTCAAACT GCGGGACGGA GGATATAATG ATGCTTAACA
          90        100        110        120        130        140        150        160
   ATGGCTGTCT GTGTTGTACT GTTAGGGGTG ATCTTGTGAG GATGATTTCT GAAATGGTCC AGACCAAGAA AGGAAGGTTC
         170        180        190        200        210        220
   GACCATATCG TTATTGAGAC GACAGGTTCA ATATTTCTAT CACTCTGGAG CTACATGATT CGTTAAGAA 3'
```

A1-5-44

```
          10         20         30         40         50         60         70         80
5' TTCTATGCTG AGGATGAAAT TTTCAATGAT GTCAAGCTGG ATGGGGTTGT CACTCTGGTT GATGCTAAAC ATGCTCGTTT
          90        100        110        120        130        140        150        160
   GCATCTAGAT GAGGTCAAAC CTGAAGGCTA TGTCAATGAG GCGGTTGAAC AAATAGCTTA CGCGGATCGT ATCATTGTTA
         170        180        190        200        210        220        230        240
   ACAAGGTATT GTGAAGTTCT ATTTCGACCT AATTTTCATG ATTAACATCA CCATAACTTG TATTTGAAAT GTGTGGTGGA
         250        260        270        280        290        300        310        320
   TGCTTATAGT AGAAGCTTTT CAATGGTCTG CTATGTATGT TCTCCTCGAC ATATCATACA AGCTGAAAAT TGTTTGTACT
         330        340        350        360        370        380        390        400
   TTGTTGTAGA CTGATCTTGT TGGTGAGCCA GAACTAGCTT CAGTGATGCA GCGGATAAAG GTAGGTTCTG CTTCTTCTGA
         410        420        430        440        450        460        470        480
   TTCTTTCTTT TTTCTCTTTG AACATCTATT TTTCTCTGTG CACTGCGTTC AGTTTATGAT GTGGTTATTT CCAGACCATA
         490        500        510        520        530        540        550        560
   AACCAGCATG GCTCACATGA AGCGGACAAA GTATGGGAAG GTTGACTTGG ATTATGTTCT TGGAATTGGA GGGTTTGATC
         570        580        590        600        610        620        630        640
   TAGAAAGGTC CGGTTTTATT TTACGTCTTA CTAAAAATCT TATTACGAGT CACTCCGAAA ACTTCTGTCC AAAGTGAATT
         650        660        670        680        690        700        710        720
   GTAGCTCCTT ACATTTGTTA TGGTTTATTG GTTGTGGATT TCCTGGTTTA AGAGGGCTGA TTTTGTTACT CTGATTTATT
         730        740        750        760        770        780        790        800
   GCTGGTGGCC AACTATTTTA TCTTTCCAGA ATTGAAAGCT CTGTGAATGA AGAAGAGAAA GAAGATCGCG AGGGTCATGA
         810        820        830        840        850        860        870        880
   TGATCATCAC CATGGTCATG ACTGCCATGA TCACCACAAT GAGCATGAGC ATGAGCATGA ACACGGTATG TTATAGTGTT
         890        900        910        920        930        940        950        960
   ACCAAATAAT GGCCTTATTA CTTAATAACG ATCGCCTACA TGCATTGATT AGTTTGCTTC TGAAACTGCA GAGATCACCA
         970        980        990       1000       1010       1020       1030       1040
   TTCTCATGAT CACACCCATG ACCCTGGTGT TGGTTCAGTC AGTATAGTTT GCGAAGGAGA CTTAGACCTC GAGAAGGTAT
        1050       1060       1070       1080       1090       1100       1110       1120
   TAACCCAATA TGGCTCAAGT TGTGTCACAT GTGATGCATA AAACTGTAGC ACCTGACGTG TTACTTATAA TTAAAACGAT
        1130       1140       1150       1160       1170       1180       1190       1200
   TGTTGAGATG TGCGATTGAT TACTTGCATC ATCTATAACC TTAGTTTTGA AAATGGATCA GGCTAACATG TGGCTTGGGG
        1210
   CGCTATTGTA CCAACGT 3'
```

Figure 6. DNA sequences of A1-5-0 and A1-5-44. Both are parts of the clone A1-5.

because of its slightly higher hybridization signal. The insert was released from the recombinant phage with EcoRI and then cloned into the EcoRI site of pBluescript to give rise to the plasmid pA1-5. The A1-5 was sequenced in one orientation and was not completely sequenced, figure 5 shows the sequencing strategy. Two sets of sequences generated, A1-5-0 and A1-5-44, were separated by a 200 bp gap. The two DNA sequences and all the possible reading frames of these two sequences were analyzed in the same way as the clone Ab5-10 (Chapter IV), no informative similarities with any known DNAs and proteins could be found. The best match for the DNA data bank search is below 50%, and the best match for the protein data bank search is 26% identity over 47 amino acid residues. The 200 bp gap which has not been sequenced does not seem to cover the conserved sequences, because both in rat and yeast genes, the sequences corresponding to D1 and D3 oligonucleotides are approximately 500 bp apart. The figure 6 shows part sequences of the clone A1-5.

2). Oligonucleotides as probes

Genomic Southern analysis indicated that in Agmenellum there are two EcoRI fragments of approximately 3.8 and 10 kb recognized by both D1 (D1A and D1T) and D3 (figure 7).

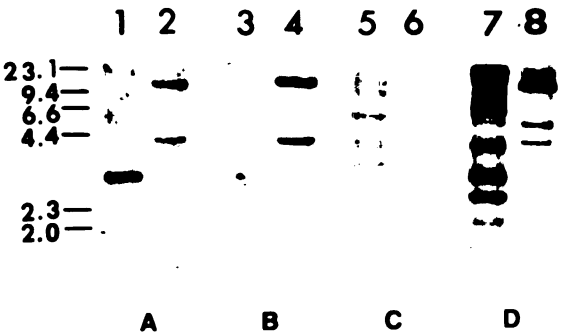

Figure 7. Genomic blotting of Agmenellum and yeast by oligonucleotide mixtures D1A, D1T, D2, and D3 (A, B, C and D). Lanes 1, 3, 5 and 7 are yeast DNA; Lanes 2, 4, 6 and 8 are Agmenellum DNA. All DNA were digested with Eco RI to completion (2 ug per lane).

Another 5 kb fragment was recognized by the D3 oligo also. After extensive screening of 4 independent libraries (2 Agmenellum lambdo gt10 libraries and 2 Agmenellum lambdo 2000 libraries), not a single clone could be isolated by using oligonucleotides D1A and D1T as probes. By contrast, more than 100 clones were isolated by using oligo D3 as a probe, none of these hybridized to oligo D1A or D1T at all. DNA was made from 25 clones isolated on the basis of homology to probe D3, and all of these clones contained a 5 kb EcoRI fragment with homology to D3 probe. This is presumably the fragment evident on the Southern blot, which did not hybridize to probe D1T or D1A. A 1 kb fragment called B4-43 from this 5 kb clone was shown to hybridize to probe D3, and was subsequently sequenced. Figure 8 is the sequencing strategy and figure 9 is the sequence for the fragment

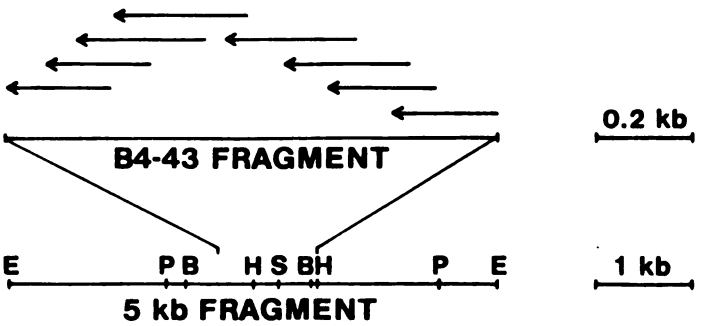

Figure 8. Sequencing strategy for fragment B4-43.

```
         10         20         30         40         50         60         70         80
5' AGCTTTACGC TTACCACAGA GTTGATAAAT GTAGTAGAAA GCGGCCCCAG CAATATAGGT TCCCAGTAAA AAATTAAAGA

         90        100        110        120        130        140        150        160
GTCCAGGTGC GCCAATGGTG GGAACATGAG GACGACAAAA ATAACGGTGA GGATCCCCGG CACGAGATAG GCTTTCTGGA

        170        180        190        200        210        220        230        240
GAGGTCTTTC CCGGTGGCGG CGATGGGAAA AAGTTGGGTC AGGGTGAGGT TGGAGTTGGG ACTGACAATG GGGCCGGGGA

        250        260        270        280        290        300        310        320
CAGGGATATG GGAACCCGGA TCGAGGGGAG CCTGGGGATT GTGGCCTTGC TCTGGAGCTC GAAACTAAAG CTGGGGCCAT

        330        340        350        360        370        380        390        400
TGTCTCCGAG GGTGAGGCGA TCGCTGCTTG CAGATTCGTG AACCTTGTAA CCGTTGACCA TTGATAAAGG TGCCATTCGC

        410        420        430        440        450        460        470        480
ACTGCCGAGA TCAACGGCTT CCCAACCTTG ACCCTGGGGA CGAATCAGAA CGTGGCGACG GGAGACAACG GTGTAAAGAT

        490        500        510        520        530        540        550        560
TGGGGTTGAG GGCAATCTGG CAACTGGGTT CGCGGCCAAT GAGAATTTCT TGACGATGAT CAAGGGAAAA ATCGGGCAGC

        570        580        590        600        610        620        630        640
AGTGGGGCTT GAACACCGCC AGTGGAAATT TGGCGCAAAA TACCAGAAGC TTGCATCGTA ATAATTTTTT TCTAAAACGC

        650        660        670        680        690        700        710        720
AAAAGACGGA GGAGCGCCCC GCCTCAACTT TAGTTATAAC GATAGATTGC CGGATCTTAA AGATATTTTT GTAAATTGAT

        730        740        750        760        770        780        790        800
ACGGGCTGTT GAACGCCTAG TCTTGATTGG CGCGAGTATC TTGGACGGCC CACAAAAAAG AGGCAATGGT CAGGGGAATA

        810        820        830        840        850        860        870        880
CTGAGGGCCA GGATAATGGT GGTATTGGTC GCACCAAGAT CGGGTTCGCC GGAACTGAGT TCAAACACGC AACCCACAGC

        890        900        910        920        930        940        950        960
GGCGATCGAA AGACGCAGCA TAGTAATAAT AAAACACCAC TTTTGGGGGT CATCGAAACC ATGGTTTTAT GCCTCAAAGA

        970        980        990       1000       1010
TTTGGTTTAG GGGATTATAA AAAAGAGTGA GATAAGGGCA AAACAGAGGT 3'
```

Figure 9. DNA sequence of the fragment B4-43.

B4-43. Unfortunately, as we expected (because it did not hybridize to D1A or D1T), it is not desaturase gene. As before, analysis of the sequence failed to indicate any significant similarity to proteins or genes of known function. Thus we cannot assign any functions for this clone.

4. Discussion

Although we have not been able to clone desaturase genes from either plant or Agmenellum, it remains possible that, with additional efforts, the approach may work. The reason that the two Agmenellum fragments recognized by both oligos on Southern blots can not be recovered in plaque screening could be due to a variety of reasons. First, it is possible that the sequences are killing E.coli. The sequences could encode protein products which are toxic to E.coli. Second, there may be some bias in the library so that the sequences not be packaged into lambda or the recombinant phage fail to produce plaques. These problems may be circumvented by making libraries using different restriction enzymes or making plasmid libraries.

The problem with using heterologous probes to isolate genes is that there is no way to know what stringency to use. Yeast genes have been used successfully to clone the corresponding plant genes (3,4). However the probability of identifing a specific gene depends on how conserved the specific gene is between yeast and plants. The yeast desaturase gene does recognize a few Arabidopsis sequences under the condition used. The two clones I have sequenced are not desaturase genes. However, this does not mean there is no need to

sequence the rest of the clones. I do sincerely believe that it is possible to clone the desaturase genes by this approach, but I don't know how long it will take and how many clones we have to sequence before we can really get what we want.

5. Refferences

1). M. A. Thiede, J. Ozols, and P. Strittmatter (1986). J. Biol. Chem. 261:13230-13235

2). D. G. Oppenheimer, N. Haas, C. D. Silflow, and D. P. Snustad (1988). Gene. 63:87-102

3). R. L. Scholl, H. Zhang, Y. Kim, and C. Somerville (1989). Gene. Summitted

4). B. J. Mazur, C. F. Chui, and J. K. Smith (1987). Plant Physiol. 85:1110-1117

APPENDIX IV

DOUBLE STRANDED DNA SEQUENCING AS A CHOICE FOR DNA SEQUENCING

-----Reprint from NUCLEIC ACIDS
RESEARCH, Vol. 16, No. 3,
1988, p 1220

Volume 16 Number 3 1988 Nucleic Acids Research

Double stranded DNA sequencing as a choice for DNA sequencing

Hong Zhang, Randy Scholl[1], John Browse and Chris Somerville

Genetics Program, Michigan State University, E. Lansing, MI 48824 and [1]Genetics Department, Ohio State University, Columbus, OH 43210, USA
Submitted October 23, 1987

Double stranded DNA sequencing (1) is favored because of its simplicity, and convenice. However it is only recently that the quality of this method has become comparable with single stranded DNA sequencing. We believe that two factors have limited the popularity of double stranded DNA sequencing: 1, the quality of the template DNA; 2, the inherent property of the DNA polymerase. The modified T7 DNA polymerase (Sequenase™) has several properties which make it more suitable for sequencing (2). We replaced the Klenow Polymerase I with Sequenase™ in our double stranded DNA sequencing, and by paying careful attention to the conditions used to denature and recover the template DNA, we are now routinely producing sequencing results which are as good as single stranded DNA sequencing (see figure). Here, we present a step-by-step protocol for the alkaline denaturation of template DNA, and our recommendations for the sequencing reaction.

I. Template preparation:
1. use 3ug of CsCl purifed plasmid DNA, add NaOH to 0.2M and EDTA to 0.2mM (total volume 20ul), incubate at room temperature for 5min
2. neutralize by adding 2ul of 2M NH_4Ac (pH4.6), mix quickly, then add 60ul of 100% ethanol (-20°C), mix well on ice
3. precipitate the DNA in a microfuge for 20min at 4°C, wash once with 80% ethanol (-20°C), spin again for another 5min
4. carefully draw away ethanol, and dry the DNA in a vacum desicator for 10min, then use immediately

II. Sequencing Reaction:
See Sequenase™ Manual (3) for details. This manual was originally designed for single stranded DNA sequencing, for double stranded DNA sequencing, we recommend the following modifications:
1. 5ng (about 1pmol) primer will be enough, excess of primer often decreases the sequencing quality
2. anneal template and primer at 65°C for 3-5min, then cool at room temperature for about 30min
3. 1ul of 400 Ci/mmol of (^{35}S)-dATP per reaction will give satisfactory results

References:
1. Chen, E. Y., Seeburg, P. H. (1985) DNA 4, 165-170
2. Tabor, S., Richardson, C. C. (1987) Proc. Natl. Acad. Sci. 84, 4767-4771
3. Step-By-Step Protocols For DNA Sequencing With Sequenase™ (United States Biochemical Corporation, P.O.Box 22400, Cleveland, Ohio 44122)

 © IRL Press Limited, Oxford, England.

MICHIGAN STATE UNIV. LIBRARIES
31293007883972