Molecular Biology of G Protein Alpha Subunits From Bovine Photoreceptors and the Nematode *Caenorhabditis elegans*

Thesis by

Michael Alan Lochrie

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1991

(Submitted August 29, 1990)

Dedicated to my parents and my sister Angie

The years of anxious searching in the dark...

and the final emergence into the light—

only those who have experienced it can understand it.

—Albert Einstein

Just do it!

—Mel Simon

ACKNOWLEDGEMENTS

There are more people than there are G proteins that I must thank for making this thesis possible. I'll attempt to mention some of them here.

Most of the credit goes to my thesis advisor Mel Simon. Working in his lab and being exposed to his way of doing science has been a stimulating experience. It was great to be associated with a prolific scientist in a lab with so many diverse things going on. It was also a pleasure to start graduate school at the ground level in an emerging field of research, to work in a lab that set the pace in a field where there seemed to be never ending excitement, and to take advantage of Caltech's rich resources (I fear I've been spoiled!).

My thesis committee members (Barbara Wold, Henry Lester, Paul Sternberg, Mark Tanouye, and Mary Kennedy) also deserve special praise. I am very grateful to them for their fairness, the excessive amount of time devoted to committee meetings, and their interest in my work. Frankly it was an unexpected pleasure. Paul's lab was especially courteous during the *C. elegans* work.

In the early days it was a special treat to work with Jim Hurley one ofthe world's experts on transducin biochemistry, and Henry Fong who provided critical technical tips during the cloning of the cone transducin alpha subunit cDNA. Numerous people in the lab have made the work enjoyable throughout the years. My next door neighbor Bruce Birren has enriched my life in numerous ways. Tom Amatruda, Kathy Borkovich, Janis Lem, Ryn Miake-Lye, Fred Hess, and many others also bring back special memories. I'll never forget the time they had mirrors installed on the corners of our walls to help us navigate around tricky corners with precious experiments. Emil Bogenmann has been both a good collegue and friend. Mike Strathmann's technical wizardry was utilized throughout the *C. elegans* work.

I thank various friends outside the lab in the Sierra Club, the University of Colorado Young Alumni Association, and so on for interesting diversions outside the lab. In particular I thank Melanie for her love and company.

I feel compelled to thank certain people from my undergraduate career particularly Scott Power and Robert Poyton who gave me my first taste of what research in molecular biology would be like. I realize now I was very fortunate to have been associated with them at an early stage of my scientific development.

Looking to the future, as I go to Harold Varmus' lab to study oncogenes and retroviruses as a postdoc, I thank my candidacy exam study group for giving me the topic of oncogenes to research.

My work was supported by funds from a National Science Foundation Graduate Fellowship, the Markey Foundation, the National Institutes of Health, and DARPA.

ABSTRACT

This thesis examines the molecular biology of G protein alpha subunits from bovine photoreceptors and the nematode *Caenorhabditis elegans*.

Chapter one describes the nucleotide sequence of the bovine cone photoreceptor transducin alpha subunit ($T_c\alpha$). Analysis of the sequence defined regions homologous to other GTP binding proteins which may be involved in guanine nucleotide binding, allowed the positions of amino acids which are ADP-ribosylated by pertussis toxin and cholera toxin to be determined, and led to the prediction that G proteins are posttranslationally modified with lipids, which serve to anchor G protein alpha subunits to membranes. Comparison of the $T_c\alpha$ amino acid sequence with other alpha subunit sequences as they became available provided the first indications that G proteins would be more numerous and diverse than previously thought. The diversity observed among G protein subunits and its structural and functional implications are reviewed in the introduction.

In Chapter 2 the characterization of G protein alpha subunits in the nematode *C. elegans* is described. Two genes were isolated and their DNA sequences were determined. The protein products of these genes appear to be unique to *C. elegans*. A cDNA encoding a homolog of $G_o\alpha$ was also isolated and sequenced. Thus, *C. elegans* has identifiable homologs of mammalian G proteins as well as G proteins that may be unique to it. The chromosomal positions of the genes were determined. Each maps to a unique location near mutations that could be in G protein alpha subunits.

In Appendix 1 the characterization of photoreceptor specific gene expression in human retinoblastoma cultures is described. These cells express cone photoreceptor-specific genes, but not rod photoreceptor specific genes. Therefore they may provide a system for studying the DNA elements required for the expression of genes specifically in cone cells.

Appendix 2 surveys systems for the heterologous expression of transducin alpha subunits in *E. coli*, yeast, and insect cells. The *E. coli* expression system offers the most promise for obtaining adequate amounts of active, pure protein.

TABLE OF CONTENTS

Dedication.....ii

Acknowledgements.....iv

Abstract.....vi

List of Tables.....viii

List of Figures.....ix

Introduction: G Protein Multiplicity in Eukaryotic Signal Transduction Systems.....1

Chapter 1: Sequence of the Alpha Subunit of Photoreceptor G Protein: Homologies Between Transducin, *ras*, and Elongation Factors.....11

Chapter 2: G Protein Alpha Subunits in the Nematode *Caenorhabditis elegans*....16

Appendix 1: Cone Cell-Specific Genes Expressed in Retinoblastoma.....73

Appendix 2: Heterologous Expression of Transducin Alpha Subunits in *Escherichia coli*, Yeast, and Insect Cells.....77

Conclusion.....297

LIST OF TABLES

Introduction

Table 1. G-protein gene and cDNA characterization.

Table 2. Percent amino sequence identity between G protein α subunits.

Appendix 1

Table 1. Summary of transducin gene and photopigment gene expression in RB cells *in vitro*.

Appendix 2

Table 1. General characteristics of heterologous expression systems.

Table 2. Strain list.

Table 3. Features of transducin alpha subunit expression vectors.

Table 4. Saturation density of yeast strains.

Table 5. Plasmid recovery from yeast strains.

Table 6. Efficiency of yeast transformation.

Table 7. Characteristics of strains containing pML 18.

Table 8. Expression of $rT_c\alpha$ in CAG1139 pML 18(+).

Table 9. Expression of $rT_r\alpha$ in CAG1139 pML 46 at different temperatures.

Table 10. Chromatographic behavior of $rT_r\alpha$.

LIST OF FIGURES

Introduction

Figure 1. Amino acid sequence homologies among G protein alpha subunits.

Chapter 1

Figure 1A. Synthetic oligonucleotides used to isolate and confirm the identity of a cDNA clone for the alpha subunit of transducin.

Figure 1B. Organization and nucleotide sequencing strategy of λTα1 cDNA insert.

Figure 2. Nucleotide sequence of λTα1 cDNA and deduced amino acid sequence of Tα protein.

Figure 3. Hybridization of bovine genomic DNA to Tα cDNA.

Figure 4. Homologies of Tα to the protein sequences of translation elongation factors and *ras* oncogenes.

Chapter 2

Figure 1. Restriction map and DNA sequencing strategy of *C. elegans* $G_o\alpha$ cDNA.

Figure 2. Sequence of the *C. elegans* $G_o\alpha$ cDNA.

Figure 3. Hybridization of *C. elegans* total mRNA to *C. elegans* $G_o\alpha$ cDNA.

Figure 4. Comparison of $G_o\alpha$ amino acid sequences.

Figure 5. Immunoblot of *C. elegans* $G_o\alpha$ protein.

Figure 6. Restriction map, DNA sequencing strategy, and gene organization of *C. elegans gpa-1* gene.

Figure 7. DNA sequence of the *C. elegans gpa-1* gene.

Figure 8. Restriction map, DNA sequencing strategy, and gene organization of *C. elegans gpa-3* gene.

Figure 9. DNA sequence of the *C. elegans gpa-3* gene.

Figure 10. Intron sizes and locations in *gpa-1*, *gpa-2*, and *gpa-3* coding sequences.

Figure 11. Comparison of amino acid sequences of *C. elegans gpa-1*, *gpa-2*, and *gpa-3*.

Appendix 1

Figure 1. Expression of the $T_c\alpha$ subunit in cultured RB cells.

Figure 2. Expression of retinal $T\gamma$ in RB cells.

Figure 3. Expression of red or green photopigment gene in cultured RB cells.

Appendix 2

Figure 1. Diagram of pML 1.

Figure 2. Sequence of the promoter region of pML 1.

Figure 3. Diagram of pML 4.

Figure 4. Sequence of the promoter region of pML 4.

Figure 5. Diagram of pML 7.

Figure 6. Diagram of pML 8.

Figure 7. Diagram of pML 10.

Figure 8. Diagram of pML 11.

Figure 9. Sequence of the promoter region of pML 11.

Figure 10. Diagram of pML 12.

Figure 11. Sequence of the promoter region of pML 12.

Figure 12. Diagram of pML 15.

Figure 13. Sequence of the region upstream of the $rT_r\alpha$ coding region in pML 15 and pML 25.

Figure 14. Diagram of pML 17.

Figure 15. Sequence of the promoter region of pML 17.

Figure 16. Diagram of pML 18.

Figure 17. Sequence of the promoter region of pML 18.

Figure 18. Diagram of pML 25.

Figure 19. Diagram of pML 34.

Figure 20. Sequence of the promoter region of pML 34.

Figure 21. Diagram of pML 35.

Figure 22. Sequence of the promoter region of pML 35.

Figure 23. Diagram of pML 44.

Figure 24. Diagram of pML 45.

Figure 25. Sequence of the promoter region of pML 45.

Figure 26. Diagram of pML 46.

Figure 27. Sequence of the promoter region of pML 46 and pML 50.

Figure 28. Diagram of pML 49.

Figure 29. Sequence of the promoter region of pML 49.

Figure 30. Diagram of pML 50.

Figure 31. Diagram of pML 52.

Figure 32. Sequence of the promoter region of pML 52.

Figure 33. Diagram of pML 53.

Figure 34. Sequence of the promoter region of pML 53.

Figure 35. Diagram of pML 55.

Figure 36. Sequence of the polylinker region of pML 55.

Figure 37. Diagram of pML 56.

Figure 38. Sequence of the protein fusion junction region of pML 53.

Figure 39. Diagram of pGD 108.

Figure 40. Sequence of the promoter region of pGD 108.

Figure 41. Diagram of M13 ml 1.

Figure 42. Diagram of M13 ml 4.

Figure 43. Diagram of M13 ml 1-Nco I.

Figure 44. Diagram of M13 ml 91.

Figure 45. Derivation history of plasmids containing the $rT_r\alpha$ cDNA.

Figure 46. Derivation history of plasmids containing the $rT_c\alpha$ cDNA.

Figure 47. Dot blots using $T_r\alpha$ and $T_c\alpha$ antipeptide antisera.

Figure 48. Specificity of antipeptide antisera.

Figure 49. Western blots of yeast strains 20B-12 pML 35 and NNY105 pML 35.

Figure 50. Western blots of yeast strains HR125 pML 35, EJ102 pML 35, and DSR865-11α pML 35.

Figure 51. Western blots of yeast strains containing pML 11.

Figure 52. SDS gel analysis of proteins in Sf 9 insect cells infected with baculoviruses.

Figure 53. Plaques formed by AcNPV-Ac360βgal.

Figure 54. Time course of $rT_r\alpha$ expression in Sf 9 cells infected with AcNPV-ML34(+).

Figure 55. Uninfected Sf 9 cells.

Figure 56. Sf 9 cells infected with AcNPV-ML34(+).

Figure 57. Sf 9 cells infected with AcNPV-ML45 at 24 hours post-infection.

Figure 58. Sf 9 cells infected with AcNPV-ML45 at 36 hours post-infection.

Figure 59. Sf 9 cells infected with AcNPV-ML45 at 48 hours post-infection.

Figure 60. Sf 9 cells infected with AcNPV-ML45 at 72 hours post-infection.

Figure 61. Sf 9 cells infected with AcNPV-ML45 at 120 hours post-infection.

Figure 62. Distribution of polyhedrin-$rT_r\alpha$ in soluble and insoluble fractions of AcNPV-ML45 infected Sf 9 cells.

Figure 63. Expression of $rT_r\alpha$ in strains containing pML 12, pML 15, and pML 25.

Figure 64. Analysis of $rT_r\alpha$ expression in strains containing pML 25.

Figure 65. Analysis of $rT_c\alpha$ expression in CAG1139 pML 4 and CAG1139 pML 18.

Figure 66. Diagram of Glutagene protein expression system.

Figure 67. Analysis of $gT_r\alpha$ produced using the Glutagene system.

Figure 68. Western blot analysis of $rT_r\alpha$ in extracts from CAG1139 pML 46 and CAG1139 pML 50 cells.

Figure 69. Mobility of $rT_r\alpha$ in an SDS gel.

Figure 70. Partial purification of $rT_r\alpha$ and $rT_c\alpha$.

Introduction

G Protein Multiplicity in Eukaryotic Signal Transduction Systems

Michael A. Lochrie and Melvin I. Simon

Division of Biology (147-75),
Califonia Institute of Technology,
Pasadena, California 91125

(Published in Biochemistry 27, 4957-4965)

G proteins comprise a specific family of guanine nucleotide binding regulatory proteins that serve as intermediaries in a variety of transmembrane signaling processes in eukaryotic cells (Gilman, 1987). They are located on the cytoplasmic surface of membranes, where they physically couple ligand-bound receptors to the regulation of effector proteins which produce changes in intracellular metabolism. G proteins are heterotrimers composed of α (M_r 39000–54000), β (M_r 35000 or 36000), and γ (M_r 8000–10000) subunits. The α subunit interacts with guanine nucleotides. The β and γ subunits form a tightly associated $\beta\gamma$ complex. Different G proteins contain different α and probably different γ subunits but only one or two types of β subunits.

G proteins function by utilizing a guanine nucleotide binding and hydrolysis cycle. In this process ligand-bound receptors induce a nucleotide exchange reaction on the α subunit in which GDP can be exchanged for GTP. In the best characterized G protein coupled systems α-GTP directly regulates effector proteins.[1] The regulatory activity of α-GTP is terminated by an intrinsic GTPase activity of the α subunit, and the α-GDP that results can then initiate another GTP binding and hydrolysis cycle. Through multiple cycling events, a large amplification of the original stimulus can be obtained. Two bacterial toxins can interrupt the signal transduction cycle at specific stages through covalent ADP-ribosylation of the α subunit. Pertussis toxin uncouples receptors from G proteins to abolish signaling, while cholera toxin inhibits GTPase activity to enhance signaling.

G_α Protein Diversity. Over the past few years it has become evident that the number of G proteins is larger than previously suspected. The best studied G proteins are transducin, G_s and G_i. Transducin (T_r) is the retinal *rod* photoreceptor G protein that couples photolyzed rhodopsin to the activation of a cGMP phosphodiesterase. G_s and G_i are found in many cell types and function to couple various receptors to a *stimulation* or *inhibition* of adenylyl cyclase activity. Less well characterized proteins that have been studied more recently include G_o, the *other* G protein, that is abundant in brain and heart (Sternweis & Robishaw, 1984), $T_c\alpha$, which is found in *cone* photoreceptors (Lerea et al., 1986), various novel G_i-like pertussis toxin substrates such as those isolated from brain (Katada et al., 1987), neutrophils (Dickey et al., 1987), and erythrocytes (Iyengar et al., 1987) and a protein called G_p (Waldo et al., 1987). The existence of several more G proteins from yeast [SCG1 (Dietzel & Kurjan, 1987); GP1 (Nakafuku et al., 1987); GP2 (Nakafuku et al., 1988)], the fruit fly *Drosophila* (J. Hurley, personal communication), the slime mold *Dictyostelium* [G1, G2 (R. Firtel and P. Devreotes, personal communication)], and humans [G_z (Fong et al., 1988)] has been inferred from isolation of the corresponding α-subunit gene or cDNA.

Diversity in signal transducing systems is not just restricted to G proteins but is being uncovered at all levels. For example, the mas oncogene product, the substance K receptor (Masu et al., 1987), the *a* and α factor yeast pheromone receptors, and a protein called G-21 (Kobilka et al., 1987) all exhibit homology to receptors that are known to interact with G proteins (Dohlman et al., 1987). The list of effectors that may be regulated by G proteins has expanded from enzymes involved in cyclic nucleotide metabolism to include calcium and potassium channels (Dunlap et al., 1987) and phospholipases A_2 and C (Jelsema & Axelrod, 1987; Taylor & Merritt, 1986). Protein kinase C, which is activated by breakdown products of phospholipase C metabolism, consists of at least seven distinct species (Ono et al., 1987). Even a "simple" single cell eukaryote such as yeast may have at least three G proteins (Dietzel & Kurjan, 1987).

This review focuses on information that has been obtained as a result of the recent isolation and characterization of several cDNAs and genes encoding different G protein subunits. The emphasis will be on discussing α subunits since much of the knowledge that has emerged about G protein similarity and diversity is a result of analyzing their sequences. To date over

[1] Recent data indicate that $\beta\gamma$ may also regulate certain effectors (Katada et al., 1986; Jelsema & Axelrod, 1987; Logothetis et al., 1987).

Table I: G-Protein Gene and cDNA Characterization

gene[a] (protein)	chromosome location[b]	organism	mRNA size[c]	tissue distribution[d]	references
GNAI1 ($G_{i-1}\alpha$)	7	human	2.2, 3.9	Ad, Fi, He, Ki, Lu, Pa, Sp; not HL-60, Ly	Kim et al., 1988
		human	2.2	Br ≫ Li	Bray et al., 1987
		bovine		Pi	Michel et al., 1986
		rat	3.9	Br	Nukada et al., 1986a
		rat	3.5	Br, Ki > Lu, In > 01 > Li, He	Jones & Reed, 1987
GNAI2 ($G_{i-2}\alpha$)	3, 12	human	2.6	Br = Ki; Ly, Me, EBV; not SKN	Beals et al., 1987
		human		Mo	Didsbury et al., 1987
		human	2.7	Ad, Fi, He, HL-60, Ht, Ki, Li, Lu, Ly, Pa, Sp	Kim et al., 1988
		bovine		Pi	Michel et al., 1986
		rat		C6	Itoh et al., 1986
		rat	2.35	Lu > He, In, Ki > Br > 01, Li	Jones & Reed, 1987
	9	mouse	2.7	Ma, S49	Sullivan et al., 1986
GNAI3 ($G_{i-3}\alpha$)	1	human	2.6, 2.4	Ki ≫ Br; Ly, SKN; not Me, EBV	Beals et al., 1987
		human		HL-60	Didsbury & Synderman, 1987
		human		Li	Suki et al., 1987
		human	2.8	Ad, Fi, HL-60, Ht, Ki, Li, Lu, He, Pa, Sp	Kim et al., 1988
		rat	3.5	He, In, Ki, Li, Lu, 01 > Br	Jones & Reed, 1987
GNAS1 ($G_{s-L}\alpha$) ($G_{s-S}\alpha$)	20	human	1.9	Fe, Br	Bray et al., 1986
		human		Ad, Ht, Ki, Li, Ov, Te	Mattera et al., 1986
		bovine		Br	Nukada et al., 1986b
		bovine	1.9	Ad, Br, Li, COS	Robishaw et al., 1986a,b
		rat		C6	Itoh et al., 1986
		rat	1.85	He, Ki > Br, In, Li, Lu, 01	Jones & Reed, 1987
		mouse		S49	Rall & Harris, 1987
	2	mouse	1.9	Ma, S49; not S49 cyc-	Sullivan et al., 1986
GNAO1 ($G_o\alpha$)		bovine		Br	Ovchinnikov et al., 1987b
		bovine	4.0	Br, Re > He; Li, Lu	Van Meurs et al., 1987; Price et al., 1987
		rat		C6	Itoh et al., 1986
		rat	4.5, 4.1	Br, Ki, In > He, Li, Lu, 01	Jones et al., 1987
GNAZ1 ($G_z\alpha$)	22	human			Fong et al., 1988
		bovine	3.0, 2.5, 2.2	Br, Re > Ad, Ki, Li > Lu, Sp > He, Te	Fong et al., 1988
GNAT1 ($T_r\alpha$)	3	human			
		bovine	2.6	Re; not Br, He, Li	Medynski et al., 1985
		bovine	2.9	Re	Tanabe et al., 1985
		bovine	2.45	Re	Yatsunami & Khorana, 1985
	9	mouse			
GNAT2 ($T_c\alpha$)	1	human			
		bovine		Re	Lochrie et al., 1985
		bovine	6.0, 8.0	Re; not Ad, He, Br	R. Miake-Lye, personal communication
	17	mouse			
SCG1 (SCG1)		yeast		haploid; not diploid	Dietzel & Kurjan, 1987
GPA1 (GP1α)	8	yeast	1.7	haploid; not diploid	Nakafuku et al., 1987; Miyajima et al., 1987
GPA2 (GP2α)	5	yeast	1.9	haploid and diploid	Nakafuku et al., 1988
GNB1 (β36)	1	human	3.0	Li	Codina et al., 1986
		human	3.4, 1.9	ubiquitous	Fong et al., 1987; unpublished results
		bovine	3.0–3.3, 1.5–1.7	ubiquitous	Fong et al., 1986; unpublished results
		bovine	3.2, 1.7	Br, Re, Li	Sugimoto et al., 1985
	19	mouse			
GNB2 (β35)	7	human	1.9	ubiquitous	Fong et al., 1987; unpublished results
		bovine	1.7	ubiquitous	Fong et al., 1987; unpublished results
		bovine	1.8	Ad	Gao et al., 1987a
GNGT1 ($T_r\gamma$)		bovine		Re	Hurley et al., 1984
		bovine	0.6	Re	Yatsunami et al., 1985

[a] Gene names are according to McAlpine et al. (1987). The name of the corresponding protein is given in parentheses. The yeast SCG1 and GP1α genes are the same, but the two reported protein sequences differ by five amino acids possibly because of strain polymorphisms. As many as four $G_s\alpha$ proteins, two long (L) and two short (S), may be encoded by a single gene (Bray et al., 1986). Also there are three sequences that have been named G_i solely on the basis of their close sequence homology. It remains to be determined which cDNA corresponds to which purified protein and which of these inhibits adenylyl cyclase. [b] Mammalian chromosome locations are taken from Ashley et al. (1987), Blatt et al. (1988), and Neer et al. (1987). [c] Messenger RNA sizes are given in kilobases for only the most prevalent species. [d] Tissue distribution of mRNAs was taken from Northern blot data, the cDNA library source, or other reported experiments. Relative levels of expression are indicated (>, greater than; =, equal to; not, low or undetectable). Abbreviations used are as follows: Ad, adrenal; Br, brain; C6, C6 rat glioma cell line; COS, monkey kidney cell line; EBV, Epstein–Barr transformed polyclonal B cell; Fi, fibroblast; He, heart; HL-60, human myeloid leukemia cell line; Ht, hepatoma; In, intestine; Ki, kidney; Li, liver; Lu, lung; Ly, lymphocytes; Ma, PU-5 murine macrophage cell line; Me, melanoma; Mo, human U937 monocyte cell line; 01, olfactory epithelium; Ov, ovary; Pa, pancreas; Pi, pituitary; Sp, spleen; Re, retina; S49 or S49 cyc-, S49 or S49 cyc- murine lymphoma cell line; SKN, neuroblastoma cell line; Te, testes.

30 amino acid sequences corresponding to 10 different classes of α subunits are available (Table I). In contrast to the diversity of α subunits there are probably only two β subunits. The sequences of two cDNA clones, β1 and β2, have been

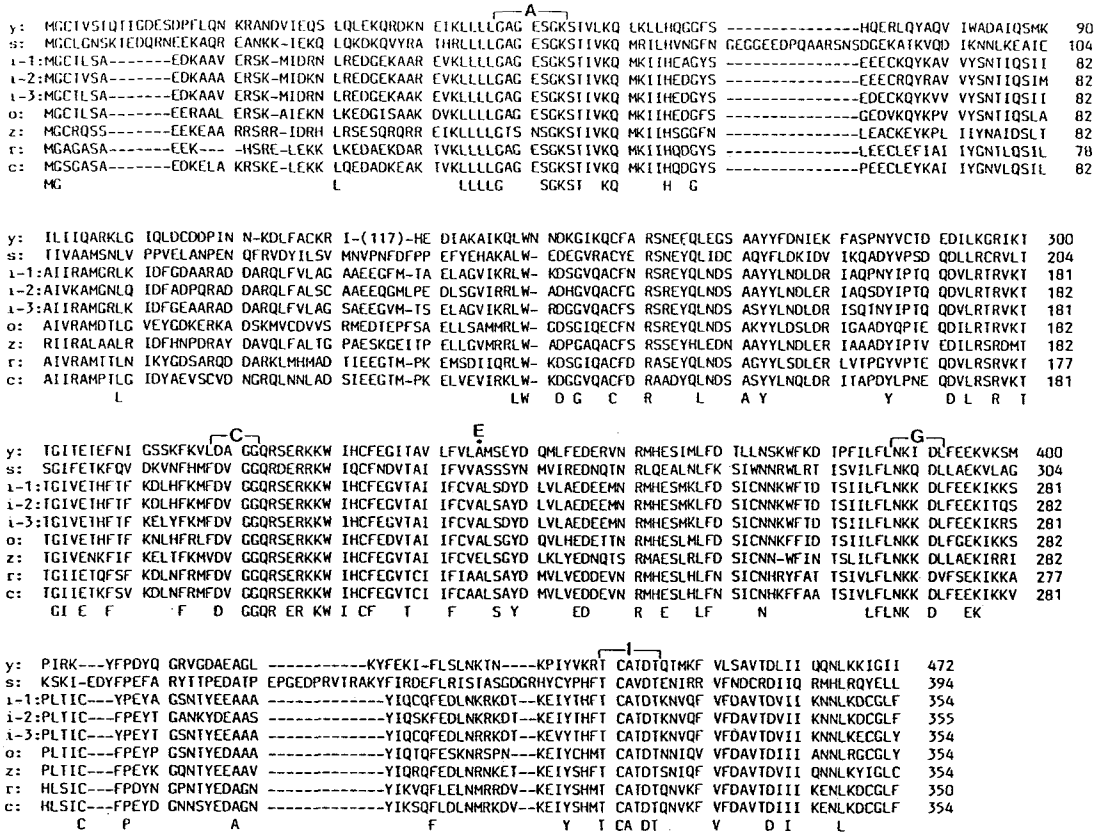


FIGURE 1: Amino acid sequence homologies among G protein α subunits. Sequences were aligned by using the FASTP program of Lipman and Pearson (1985). Hyphens indicate positions where gaps were introduced to achieve an optimal alignment. The α-subunit sequences compared are yeast GP1 (y), rat $G_{s\text{-}L}$ (s), $G_{i\text{-}1,2,3}$ (i-1, -2, -3), and G_o (o), human G_z (z), and bovine rod (r) and cone (c) transducin. A sequence of 117 nonhomologous amino acids from GP1α is not shown. The sources of the sequences are given in Table I. Letters beneath the sequences indicate amino acids that are invariant at that position. Regions marked A, C, E, G, and I are discussed in the text. The standard (IUPAC) one letter amino acid code is used.

determined. They have been shown to correspond to two β proteins, β36 and β35, respectively, which are 10% different in amino acid sequence (Gao et al., 1987b; Amatruda et al., 1988). However, the functional significance of these differences is unknown. With regard to γ-subunit diversity, only one γ-subunit sequence (T_γ) is available (Hurley et al., 1984; Yatsunami et al., 1985), but it is clear that there are other γ subunits to be characterized (Hildebrandt et al., 1985).

A variety of other proteins exist that are known or postulated to specifically bind guanine nucleotides. Certain initiation and elongation cofactors of the protein synthesis machinery such as Ef-Tu clearly utilize a molecular switch similar to that of G proteins in which alternating GDP and GTP forms interact with distinct macromolecules. Members of the ras-like oncogene product family (Barbacid, 1987) are also thought to use a similar device since they bind GDP and GTP specifically and possess GTPase activity. However, the details of a putative nucleotide exchange reaction are lacking. Tubulins bind and hydrolyze GTP but appear to function by mechanisms that are fundamentally different from G proteins or protein synthesis cofactors. The LepA (March & Inouye, 1985) and ARF (Kahn & Gilman, 1986) proteins can bind GTP. The yeast *SEC4* protein (Salminen & Novick, 1987) and others are predicted to bind GTP on the basis of strong amino acid sequence homology to ras. However, little is known about the biochemical mechanisms involved in the function of these proteins. All of these proteins can be distinguished from G_α subunits, however, because (a) they are not susceptible to pertussis or cholera toxin modification, (b) they do not appear to interact with $\beta\gamma$ subunits, and (c) the G_α subunits have more extensive amino acid sequence similarity to each other. They can also be distinguished by size. In general, G_α proteins have 350–400 amino acids, while ras-like proteins have about 200 and proteins similar to Ef-Tu often have from about 400 to 900 amino acids. Therefore, while we will compare the other GTP binding proteins to G_α subunits particularly with regard to their shared structural features that presumably function in guanine nucleotide binding and hydrolysis, we also distinguish them from the G_α's. Except for the tubulins, all known GTP binding proteins share limited sequence homology motifs but the subset that is represented by G proteins shares a number of additional features that are unique to G proteins.

Guanine Nucleotide Interactions. The percent sequence identity between α subunits varies widely from 35 to 94% (Table II), but there are 82 invariant amino acids among them (Figure 1). A majority of the most highly conserved sequences are near regions predicted to participate in guanine nucleotide interactions. The basis for these predictions comes from biophysical studies on the crystal structures of the GDP form of Ef-Tu (Jurnak, 1985; la Cour et al., 1985) and genetic studies of mutations that affect the guanine nucleotide interactions of ras proteins. These two independent approaches

have yielded complementary data and identified five homologous regions that are critical for guanine nucleotide interactions. Analogous regions can be found in G proteins. These are indicated in Figure 1 as A, C, E, G, and I from the nomenclature of Halliday (1983–1984). Halliday regions have properties that allow them to be readily identified. Most importantly, each is characterized by an almost invariant consensus sequence. In the Ef-Tu crystal structure, these consensus sequences are preceded by short hydrophobic sequences that form β strands and are followed by longer, more polar sequences that form α helices. It is at the junction of such $\beta\alpha$ secondary structures, near the consensus sequences, that the guanine nucleotide binds. The same motif has been found in the crystal structures of other nucleotide binding proteins. Halliday regions are also found in a characteristic order and spacing in proteins that contain them. For example, they are sequentially arrayed along the primary protein structure as ACEGI from the amino toward the carboxyl terminus. The C, E, G, and I regions are separated from each other by about 20–50 amino acids. The distance of the A region from this CEGI group is more variable, ranging from 50 residues in ras to 270 residues in GP1α, but it is usually found within 10–40 residues of the amino terminus. Several discussions on the precise sequence homologies of ras and Ef-Tu to the α subunits have been published (Masters et al., 1986; references in Table I). Here only a brief updated review is provided.

The A-region consensus sequence is Gly-X-X-X-X-Gly-Lys. In Ef-Tu this region is near the α-phosphate of GDP. The lysine is thought to neutralize the negative charge on the phosphate. Mutations in the analogous region of ras reduce GTP binding and GTPase activity. Similar sequences are found in other purine nucleotide binding proteins.

The C-region consensus sequence is Asp-X-X-Gly. In the Ef-Tu structure, aspartic acid chelates a magnesium ion which is closest to the β-phosphate of GDP. A mutation of Asp-X-Ala-Gly to Asp-X-Thr-Gly in ras results in autophosphorylation of the threonine if GTP is used as a substrate. Mutation of other residues near this sequence in ras results in reduced GTPase activity (Der et al., 1986).

The E region is characterized as a very hydrophobic region between the C and G regions. It often contains an alanine that is located 26 amino acids distal to aspartic acid in the C-region consensus sequence. In Ef-Tu this alanine, along with other hydrophobic amino acids from the G and I regions, forms a hydrophobic pocket near the ribose and guanine rings. In ras, mutation of the analogous alanine to threonine results in a 30-fold reduction in GTP affinity (Feig et al., 1986).

The G region, which has the consensus sequence Asn-Lys-X-Asp, is appropriately named because this region determines the nucleotide binding specificity. In Ef-Tu asparagine is hydrogen bonded to the C-6 keto group of the guanine ring and aspartic acid is hydrogen bonded to the C-2 amino group. As might be expected, a mutation of aspartic acid to asparagine in Ef-Tu alters the nucleotide specificity of the protein to favor xanthine diphosphate over guanosine diphosphate (Hwang & Miller, 1987). Mutations in this region of ras reduce, alter, or abolish nucleotide binding (Feig et al., 1986; Sigal et al., 1986b; Walter et al., 1986; Clanton et al., 1986). The conserved lysine may hydrogen bond to backbone carbonyls or side chains of the two amino acids preceding Gly-Lys in the A region (le Cour et al., 1985). In ras mutation of this lysine to a glutamine results in a 75% reduction in nucleotide binding (Clanton et al., 1986). Such a change may distort the architecture of nucleotide binding but would conserve the hydrogen-bonding amino group.

A fifth region, the I region, may also be involved in nucleotide binding. Currently this region is not well defined. In Ef-Tu it may be near Gly-172, Ser-173, Ala-174 (la Cour et al., 1985), but the precise function of each amino acid awaits a more refined crystal structure. The analogous region of ras may be Thr-144, Ser-145, Ala-146. In a genetic screen designed to isolate GTP binding mutants of ras, one was obtained in which Thr-144 was changed to isoleucine (Feig et al., 1986). This change results in a 25-fold reduction in GTP affinity. The existence of an I region in α subunits has not been widely discussed. One possible I region with the sequence Thr-Cys-Ala is shown in Figure 1. It is found after the G region and, as with other Halliday regions, contains a cluster of invariant amino acids. The homology of this region to the Ef-Tu family of proteins is low, but a more striking and extended homology can be found when it is compared to the ras family, particularly the rab proteins (Touchot et al., 1987).

Mutations in the A, C, E, G, and I regions of ras have dominant biological effects in a variety of systems (references above) including yeast (Kataoka et al., 1985; Schmitt et al., 1986), *Xenopus* oocytes (Birchmeier et al., 1985), PC12 cells (Bar-Sagi & Feramisco, 1985), NIH3T3 cells (Seeburg et al., 1984), and transgenic mice (Quaife et al., 1987). Specific effects of recessive mutations have also been observed in yeast. The phenotypic and biochemical effects of analogous mutations made in G proteins is under investigation.

In guanine nucleotide binding proteins that have Halliday regions the immediate area near each consensus sequence is similar within a family but is different from that of other families. Therefore, these regions form a "fingerprint" by which G proteins can be identified and distinguished from other guanine nucleotide binding proteins. The sequence of $G_z\alpha$ is interesting in this regard because it differs slightly from other α sequences in the A and E regions. Thus, $G_z\alpha$ may define a new subclass of G proteins. In fact, there may exist a wide continuum of α subunits, each slightly different from the other. Such a situation is emerging for the ras-like proteins, which now includes at least five related but distinct families.

Amino acid sequences outside of the Halliday regions are thought to be necessary for α-subunit specific functions such as receptor, effector, and $\beta\gamma$-subunit interactions. Structure/function studies on G proteins, particularly transducin, have provided a tentative functional assignment for some of those regions.

$\alpha/\beta\gamma$-Subunit Interactions. There is some evidence to suggest that the amino terminus of α is necessary for $\beta\gamma$ binding but the data does not indicate if it is also sufficient. Protease treatments that remove about the first 20 amino acids of $T_r\alpha$ abolish several reactions that depend on $T_r\beta\gamma$ such as ADP-ribosylation of α by pertussis toxin, rhodopsin binding, Gpp(NH)p binding, and immunoprecipitation of $T_r\beta\gamma$ by a monoclonal antibody against $T_r\alpha$ (Navon & Fung, 1987). The amino terminus is one of the more variable regions among α subunits. It is unusually hydrophilic and displays an amphipathic characteristic in which hydrophobic or uncharged amino acids are found at about every fourth position. It is not clear how these properties may contribute to $\beta\gamma$ binding. The γ subunits may provide most of the binding specificity since they are also known to be variable, but it has not been possible to separate the $\beta\gamma$ complex into pure subunits, without using denaturating conditions, to assess their individual binding properties.

α-Subunit/Receptor Interactions. Receptor/G protein interactions may be of two types. Conserved G protein sequences

may recognize conserved receptor sequences and function in processes basic to all receptor/G protein coupling systems such as the guanine nucleotide exchange reaction. Variable receptor and G protein sequences may interact to allow specific discriminations between closely related proteins.

There is evidence that receptor interactions are influenced by the extreme carboxyl terminus of the α subunit. First, pertussis toxin, which uncouples receptors from G proteins, ADP-ribosylates $T_r\alpha$ on Cys-347. Second, the UNC mutation, which uncouples $G_s\alpha$ from the β-adrenergic receptor, results in an arginine to proline change at amino acid 389 of $G_s\alpha$ (Rall & Harris, 1987; Sullivan et al., 1987).[2] Finally, there is a correlation between sequence homology to the carboxyl terminus of $T_r\alpha$ and the ability to interact with rhodopsin (Dohlman et al., 1987; Wistow et al., 1986).

However, a comparison of the interactions of rhodopsin and the α_2-adrenergic receptor with G_o, G_i, and T_r (Cerione et al., 1986) indicates another region may be important for providing receptor binding specificity. Rhodopsin can interact equally with G_o, G_i, and T_r whereas the α_2 receptor shows a distinct preference for G_o and G_i over T_r. Most of the positions where $G_o\alpha$ and $G_i\alpha$ sequences are identical but differ from $T_r\alpha$ are concentrated in the extreme amino terminus and in the carboxyl terminus on both sides of the putative I region. The variable amino terminus of α could indirectly be a determinant of receptor specificity since the amino terminus of α is required for binding to $\beta\gamma$ and $\beta\gamma$ is required for binding to receptor (Fung, 1983). A further indication that the extreme carboxyl terminus may not be a sole determinant of receptor interactions is the observation that the last 40 amino acids of $T_r\alpha$ and $T_c\alpha$, which are likely to bind to different receptors, are identical.

α-Subunit/Effector Interactions. Sequences involved in regulating effector proteins are thought to lie between the Halliday A and C regions where there is the most sequence variability among α subunits. The analogous region of Ef-Tu binds aminoacyl-tRNA in a GTP-dependent reaction, and mutations in ras in a limited region between A and C abolish its biological activity but do not interfere with nucleotide or membrane interactions (Sigal et al., 1986a; Stein et al., 1986). In fact a second site, allele specific suppressor of such a yeast ras mutation maps to adenylyl cyclase, an effector for ras in yeast (Marshall et al., 1988).

A mutation in $G_s\alpha$ (H21a) has been isolated that prevents it from activating adenylyl cyclase. It is a single base pair change which results in a glycine to alanine change at position 228 of $G_s\alpha$ in the C-region consensus sequence. However, there is evidence that the inability of the H21a protein to activate adenylyl cyclase may be due to its inability to undergo the conformational change associated with GTP binding that is a prerequisite for effector interaction rather than a direct disruption of effector binding per se (R. T. Miller and H. R. Bourne, personal communication).

Recently a mechanism of cGMP phosphodiesterase (PDE) activation has been proposed in which $T_r\alpha$-GTP binds to PDE_γ, a known inhibitor of $PDE_{\alpha\beta}$, the catalytic subunits (Wensel & Stryer, 1986; Deterre et al., 1986). The amino acid sequence of PDE_γ has been determined (Ovchinnikov et al., 1986). In its 89 amino acid long sequence there is a region of 35 amino acids with 12 basic but no acidic residues. This probably explains why PDE can be activated by trypsin treatment. However, the molecular basis of the proposed interaction between PDE_γ and $T_r\alpha$ is not apparent from an examination of their amino acid sequences alone.

ADP-Ribosylation. The amino acid substrates of cholera and pertussis toxin in $T_r\alpha$ have been identified as Arg-174 (Van Dop et al., 1984) and Cys-347 (West et al., 1985), respectively. The cholera toxin substrate is conserved in all α-subunit sequences and the pertussis toxin substrate is found in $G_{i\text{-}1,2,3}\alpha$, $G_o\alpha$, $T_r\alpha$, and $T_c\alpha$. α subunits that have been purified can be ADP-ribosylated as predicted from this sequence similarity. However, since the precise toxin recognition elements have not been defined, it is not known if the presence of the substrate amino acid at a position analogous to that found in $T_r\alpha$ is necessary to allow ADP-ribosylation. Thus, for example, it remains to be determined if $G_z\alpha$ will be a pertussis toxin substrate since it has a cysteine offset by three amino acids from the other pertussis toxin substrates.

The amino acids that are ADP-ribosylated by cholera toxin and pertussis toxin are remarkably conserved. The cholera toxin substrate is invariant, and the region around it is highly conserved even in the yeast G proteins. This degree of conservation suggests these amino acids may perform an essential function. Since all G proteins contain similar β subunits and the region near the cholera toxin substrate is well conserved, it has been suggested to play a role in $\beta\gamma$ interactions (Stryer & Bourne, 1986). It has been reported that cholera toxin ADP-ribosylation promotes dissociation of the α and $\beta\gamma$ subunits (Kahn & Gilman, 1984).

Lipid Acylation. Most G proteins can be extracted from membranes only by using detergents, a property that is inconsistent with the hydrophilic nature of their amino acid sequences. One possible mechanism for membrane localization of G proteins that involves acylation with lipids can be formulated on the basis of subtle sequence homologies (Lochrie et al., 1985) to two other hydrophilic proteins, src and ras, that are also associated with the cytoplasmic surface of the plasma membrane.

The src oncogene product has been shown to be N-myristylated on a glycine at the second position (Gly-2) that is presumably exposed after removal of the initiating methionine by an aminopeptidase. Several observations indicate myristylation of src is involved in its membrane binding. Mutations that remove Gly-2 from src result in a protein that does not associate with the membrane, is not myristylated, and is incapable of neoplastic transformation (Kamps et al., 1985). Fusion of the first 15 amino acids of src to the amino terminus of globin results in membrane binding (Pellman et al., 1985). Numerous other proteins are known to be N-myristylated, in each case on a Gly-2 residue. In general, the function of fatty acid acylation is not well understood. Some myristylated proteins are soluble, indicating that myristylation can be, but is not always, a sole determinant of membrane association. The yeast and rat *N*-myristyltransferases (NMT) have been purified and their substrate specificities examined by using over 80 synthetic peptide substrates (Towler et al., 1988). On the basis of these studies and sequence comparisons to known N-myristylated proteins one consensus sequence that can be derived is Met-Gly-X-X-X-Ser. All G protein subunits listed in Table I except $G_s\alpha$, β35, β36, and $T_r\gamma$ fit this consensus. The β35, β36, and $T_r\gamma$ subunits do not have Gly-2, which is an indispensible substrate requirement. $G_s\alpha$ has a serine at position 7 but asparagine at 6. According to Towler et al. (1988) this difference could reduce the K_m of $G_s\alpha$ for NMT by a factor of 6000. Recently Buss et al. (1987) have published the first study to determine if G proteins are modified with lipids. It was found that $G_i\alpha$ and $G_o\alpha$ are myristylated but that $G_s\alpha$, β35, β36, and $T_r\gamma$ are not. It was also reported that

[2] Since $G_s\alpha$ is longer than $T_r\alpha$, Arg-389 of $G_s\alpha$ and Cys-347 of $T_r\alpha$ are actually at adjacent positions when the homologous sequences in them are aligned.

$T_r\alpha$ is not myristylated. However, the analysis was done on transducin, which was purified without using detergent. Since there are indications that two forms of transducin exist, one that is easily solubilized and one that is more tightly membrane bound, it is possible that the difference between these forms is that one is acylated with lipid and one is not. It should be noted, though, that while NMT substrate consensus sequences may define amino acids that are critical for the myristylation reaction, other nearby residues can also have a substantial influence on the ability of a peptide to be an efficient NMT substrate.

Ras proteins are acylated with various lipids, primarily palmitate, on a cysteine that is the fourth amino acid from the carboxyl terminus (−4 Cys). Mutants that lack −4 Cys have normal nucleotide binding and hydrolysis activities but are not palmitylated and possess a greatly reduced biological activity. This deficiency can be compensated for by overexpression of the protein, indicating that the purpose of membrane association may be to increase the efficiency of interaction between ras and other membrane proteins. $G_o\alpha$, $G_{i\text{-}1,2,3}\alpha$, $T_c\alpha$, and $T_r\alpha$ have a −4 Cys. In fact, it is this amino acid that can be ADP-ribosylated by pertussis toxin, but there are no data to indicate whether it is posttranslationally modified by eukaryotic enzymes with ADP-ribose or any other adduct. Buss et al. (1987) found no evidence for palmitate or thioester-linked myristate in the G protein subunits they examined.

$T_r\gamma$ also has a −4 Cys. In this case, it is notable that two independent research groups failed to determine the amino acid sequence of $T_r\gamma$ past this cysteine as well as its presence by amino acid analysis (Ovchinnikov et al., 1985; McConnell et al., 1984), indicating that $T_r\gamma$ may be modified at this position. Pure $\beta\gamma$ will associate with phospholipid vesicles whereas α subunits do not unless $\beta\gamma$ is also present (Sternweis, 1986). Since the β and γ protein sequences are hydrophilic, a γ subunit modified with a hydrophobic group could mediate this interaction. Although Buss et al. found no myristate or palmitate linked to $T_r\gamma$, the carboxyl-terminal cysteines of other proteins have been found to be modified with other hydrophobic adducts besides fatty acids. A cysteine at the carboxyl terminus of the Thy-1 antigen is linked to phosphatidylinositol–glycan (Tse et al., 1985), and one in the yeast *Tremella* mating pheromone is linked to a farnesyl group (Sakagami et al., 1981). Perhaps the α and γ subunits are linked to one of these or another hydrophobic moiety. Of final note is that PDEα has a −4 Cys and a Gly-2 (Ovchinnikov et al., 1987a).

Phosphorylation. Since the membrane signal transduction system is the first step in a complicated pathway, it might be subject to feedback regulation. Regulation of receptors by phosphorylation has been well documented (Sibley et al., 1987). G proteins ($T_r\alpha$ and $G_i\alpha$ but possibly not $G_s\alpha$) have been reported to be substoichiometric substrates for phosphorylation by protein kinase C both in vitro and in vivo (Katada et al., 1985; Zick et al., 1987). The preferred protein substrate is α-GDP and not $\alpha\beta\gamma$ or α-GTP-γ-S. The amino acid substrate is a serine. In addition, $\beta\gamma$ inhibits phosphorylation of α-GDP. The precise substrate requirements for protein kinase C phosphorylation are not well understood, but for many kinase C substrates a basic amino acid is located two or three amino acids proximal to the substrate amino acid, which is a serine or threonine. In some cases the substrate is within about 10 amino acids of membrane attachment sites (Hunter et al., 1984; Gould et al., 1985; Ballester et al., 1987), consistent with the known membrane translocation properties of kinase C when it is activated. Given these guidelines Ser-12 (in $T_r\alpha$) is a good candidate for a kinase C substrate. This serine is conserved in all α subunits except $G_s\alpha$, 11 amino acids from putative membrane attachment site, and near several basic amino acids. Since it is located at the amino terminus, it is also in a region that could be shielded from phosphorylation in the $\alpha\beta\gamma$ trimer by $\beta\gamma$. Furthermore, since $\beta\gamma$ dissociates from α-GTP, this region may undergo a conformational change when α binds GTP such that Ser-12 would be unavailable for phosphorylation in the GTP-γ-S form of the α subunit. The functional consequences of transducin phosphorylation have not been reported, but it might regulate the interactions of α with $\beta\gamma$.

Table II: Percent Amino Acid Sequence Identity between G Protein α Subunits[a]

s	(100, M/H; 99.8, R/H; 99.8, B/M; 99.8, B/M; 99.8, M/R; 99.5, R/B; 99.2, B/H)							
o	46	(98.3, R/B)						
i-1	44	72	(100, H/B; 99.7, H/R; 99.7, R/B)					
i-2	43	69	88	(99.2, R/M; 98.6, H/R; 97.7, H/M)				
i-3	43	70	94	85	(98.6, R/H)			
c	44	62	70	69	69			
r	43	63	69	68	66	81		
z	41	60	67	67	67	57	55	
y	35	44	48	47	46	45	44	44
	s	o	i-1	i-2	i-3	c	r	z

[a] Abbreviations, sequences used, and computer program for calculating percent identity are as in Figure 1. For these calcualtions, the sequences of $G_s\alpha$ and GP1α were deleted in nonhomologous regions since they are significantly longer than the others. The percent identity between the same sequence from pairs of different species and the species compared (R, rat; B, bovine; H, human; M, mouse) is indicated on the diagonal in parentheses.

Evolution of G Protein Sequences. The interspecies divergence of any given mammalian α-subunit amino acid sequence ranges from 0 to 3% (Table II), but nucleotide sequences are typically about 10% different. G proteins from one organism can also have similar amino acid but different nucleotide sequences. For example, the amino acid sequences of $G_{i\text{-}1}\alpha$ and $G_{i\text{-}3}\alpha$ are 94% identical, yet their nucleotide sequences are only about 70% identical in the coding region. This suggests there is a very strong selective pressure to maintain protein structure. In fact, it may be that little if any of the amino acid sequence of an α subunit is nonfunctional. This may explain their high degree of evolutionary conservation. A similar situation is found for the two β proteins (Fong et al., 1987).

Differences in nucleotide sequence are also an indication that similar gene products are not encoded by one gene. In fact, the $G_{i\text{-}1}\alpha$ and $G_{i\text{-}3}\alpha$ genes do map to different human chromosomes (Table I). Other G protein genes also map to dispersed chromosomal positions in mice, humans, and yeast. Of those genes that have been mapped, only the $G_{i\text{-}2}\alpha$ and $T_r\alpha$ genes map near each other, possibly within 1 centimorgan (approximately one million base pairs), in a syntenic region that is found on both human and mouse chromosomes. Since the sequences, function, and tissue distribution of $G_{i\text{-}2}\alpha$ and $T_r\alpha$ are different, the reason for the proximity of their genes is not evident. It may be that the region as a whole has been conserved in organization for other reasons not related to G proteins. In any case, it would be interesting to determine if there is a cluster of G protein genes at this locus.

Tissue Distribution. By using gene specific probes, one can study the expression of mRNA encoding G proteins in different tissues by Northern blot analysis (Table I). Although G protein mRNA abundance does not always quantitatively agree with protein abundance (Brann et al., 1987), certain

trends emerge. First, very similar G proteins such as $G_{i-1}\alpha$ and $G_{i-3}\alpha$ can be expressed in different tissues. Second, there are "housekeeping" G proteins such as G_s that are expressed in almost every cell, and there are specialized G proteins such as the transducins that are expressed in one or a few cell types. Third, the sizes and relative expression levels for a G protein mRNA can vary between different tissues in the same organism and in the same tissue in different organisms. Finally, one tissue or clonal cell line can express multiple G proteins.

Implications of G Protein Multiplicity. The realization that multiple G proteins exist should alter experimental procedures and may provoke a reexamination of previous data. For example, the old view that pertussis toxin affects one protein, G_i, was derived from the observation that after ADP-ribosylation of membrane proteins using radioactive NAD^+ as a substrate a single band could be seen after analysis of the labeled proteins by sodium dodceyl sulfate gel electrophoresis. This is clearly not valid since multiple pertussis toxin substrates with nearly identical molecular weights can exist in one cell type. In many other types of studies such as those of in vitro biochemical reconstitution and in situ localization it now becomes important to use highly specific probes to characterize the preparations being used. Synthetic oligonucleotides and antibodies against synthetic peptides that correspond in sequence to the most divergent regions found in G proteins are being used successfully to uniquely identify the numerous related G protein subunits.

Another aspect of G protein diversity is the isolation of cDNA clones encoding G proteins that have not been purified and thus have an unknown function. Several approaches can be taken to identify the function of such G proteins. Hopefully, as new G proteins of known function are purified and sequenced, they will be found to correspond to a previously isolated cDNA. In cases where a predicted G protein is rare or has not yet been purified one could express cDNAs in a heterologous system such as bacteria and assay for a particular function such as potassium channel activation. The expression of G proteins for such purposes is just beginning. For example, Graziano et al. (1987) have shown that the long and the short forms of $G_s\alpha$ expressed in *Escherichia coli* are capable of activating adenylyl cyclase. Such in vitro approaches may not be sufficient to determine the true in vivo function of a new G protein given the moderate promiscuity observed in receptor/G protein interactions, but effector/G protein interactions may be more monogamous. An examination of the precise cellular expression pattern of a novel G protein by in situ hybridization may aid in providing a clue to its function since a G protein and the receptors and effectors it interacts with should be expressed in the same cells. Such studies are beginning [cf. Worley et al. (1986)].

The discovery of G proteins in lower eukaryotes should facilitate a genetic analysis of G proteins and their transduction pathways. In yeast the ability to isolate, express, mutate, delete, and replace genes will provide powerful tools that are currently unavailable in other systems. By isolating suppressor mutations, it should be possible to identify components of the G protein signaling systems of yeast as has been elegantly done for the yeast ras/adenylyl cyclase pathway. In fact, one yeast α-subunit gene was isolated fortuitously as a suppressor of a mating hormone supersensitive mutant (Dietzel & Kurjan, 1987). Similar techniques can be used in *Drosophila* and *Dictyostelium*, which have the added advantage of being multicellular organisms with more observable phenotypes. It may be possible to analyze the genetics of G proteins in mammals by introducing genes with dominant mutations into transgenic mice or cultured cells. Ultimately, recessive phenotypes could be generated in such systems by methods, such as the use of antisense RNA, which are currently under development.

Future Goals. One topic of interest that will continue to be of importance is that of structure/function relationships. How do G proteins interact with other proteins at a molecular level? What is the basis for changes in molecular recognition that are regulated by the presence or absence of a single phosphate group? How might posttranslational modifications alter function? Ultimately, it will be desirable to obtain X-ray crystal structures. As more information becomes available on receptor/effector interactions with G proteins, it may be possible to predict the type of signal transduction system a new G protein operates in, design G proteins with novel interaction specificities, and engineer drugs which affect the function of specific coupling systems.

Another interesting subject that has been largely unexplored is the temporal nature of G protein expression and how it changes. Are there G proteins that are only expressed during embryogenesis? What signals regulate the expression of the signaling proteins themselves? Why are G proteins that are almost identical in structure expressed in different cells? Such questions can be approached by the isolation of G protein genes and the examination of promoter function.

Given the number of receptors reported to interact with G proteins, the variable responses of different cells to the same signal, and the number of highly specialized cell types, it is not surprising that multiple species of G proteins with diverse structures and functions exist. It is possible that there are still many more than those mentioned here. Therefore, a continuing goal will be to determine how many G protein coupled transduction systems exist and to understand the significance of their diversity.

ACKNOWLEDGMENTS

We thank our colleagues for the communication of their results prior to publication, the members of our laboratory for valuable comments, and Cathy Elkins for efficient preparation of the manuscript.

REFERENCES

Amatruda, T. T., III, Gautam, N., Fong, H. K. W., Northup, J. K., & Simon, M. I. (1988) *J. Biol. Chem. 263*, 5008–5011.

Ashley, P. L., Ellison, J., Sullivan, K. A., Bourne, H. R., & Cox, D. R. (1987) *J. Biol. Chem. 262*, 15299–15301.

Ballester, R., Furth, M. E., & Rosen, O. M. (1987) *J. Biol. Chem. 262*, 2688–2695.

Barbacid, M. (1987) *Annu. Rev. Biochem. 56*, 779–827.

Bar-Sagi, D., & Feramisco, J. R. (1985) *Cell (Cambridge, Mass.) 42*, 841–848.

Beals, C. R., Wilson, C. B., & Perlmutter, R. M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7886–7890.

Birchmeier, C., Broek, D., & Wigler, M. (1985) *Cell (Cambridge, Mass.) 43*, 615–621.

Blatt, C., Eversole-Cire, P., Cohn, V. H., Zollman, S., Fournier, R. E. K., Mohandes, L. T., Nesbitt, M., Lugo, T., Jones, D. T., Reed, R. R., Weiner, L. P., Sparkes, R. S., & Simon, M. I. (1988) *Proc. Natl. Acad. Sci. U.S.A.* (in press).

Brann, M. R., Collins, R. M., & Spiegel, A. (1987) *FEBS Lett. 222*, 191–1987.

Bray, P., Carter, A., Simons, C., Guo, V., Puckett, C., Kamholz, J., Spiegel, A., & Nirenberg, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8893–8897.

Bray, P., Carter, A., Guo, V., Puckett, C., Kamholz, J., Spiegel, A., & Nirenberg, M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5115–5119.

Buss, J. E., Mumby, S. E., Casey, P. J., Gilman, A. G., & Sefton, B. M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7493–7497.

Cerione, R. A., Regan, J. W., Nakata, H., Codina, J., Benovic, J. L., Gierschik, P., Somers, R. L., Spiegel, A. M., Birnbaumer, L., Lefkowitz, R. J., & Caron, M. G. (1986) *J. Biol. Chem. 261*, 3901–3909.

Clanton, D. J., Hattori, S., & Shih, T. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5076–5080.

Codina, J., Stengel, D., Woo, S. L. C., & Birnbaumer, L. (1986) *FEBS Lett. 207*, 187–192.

Der, C. J., Finkel, T., & Cooper, G. (1986) *Cell (Cambridge, Mass.) 44*, 167–176.

Deterre, P., Bigay, J., Robert, M., Pfister, C., Kühn, M., & Chabre, M. (1986) *Proteins: Struct., Funct., Genet. 1*, 188–193.

Dickey, B. F., Pyun, H. Y., Williamson, K. C., & Navarro, J. (1987) *FEBS Lett. 219*, 289–292.

Didsbury, J. R., & Snyderman, R. (1987) *FEBS Lett. 219*, 259–263.

Didsbury, J. R., Ho, Y.-S., & Snyderman, R. (1987) *FEBS Lett. 211*, 160–164.

Dietzel, C., & Kurjan, J. (1987) *Cell (Cambridge, Mass.) 50*, 1001–1010.

Dohlman, H. G., Caron, M., & Lefkowitz, R. J. (1987) *Biochemistry 26*, 2657–2664.

Dunlap, K., Holz, G. G., & Rane, S. G. (1987) *Trends NeuroSci. (Pers. Ed.) 10*, 241–244.

Feig, L. A., Pan, B.-T., Roberts, T. M., & Cooper, G. M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4607–4611.

Fong, H. K. W., Hurley, J. B., Hopkins, R. S., Miake-Lye, R., Johnson, M. S., Doolittle, R. F., & Simon, M. I. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2162–2166.

Fong, H. K. W., Amatruda, T. T., III, Birren, B. W., & Simon, M. I. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3792–3796.

Fong, H. K. W., Yoshimoto, K. K., Eversole-Cire, P., & Simon, M. I. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3066–3070.

Fung, B. K.-K. (1983) *J. Biol. Chem. 258*, 10495–10502.

Gao, B., Gilman, A. G., & Robishaw, J. (1987a) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6122–6125.

Gao, B., Mumby, S., & Gilman, A. G. (1987b) *J. Biol. Chem. 262*, 17254–17257.

Gilman, A. G. (1987) *Annu. Rev. Biochem. 56*, 615–649.

Gould, K. L., Woodgett, J. R., Cooper, J. A., Buss, J. E., Shalloway, D., & Hunter, T. (1985) *Cell (Cambridge, Mass.) 42*, 849–857.

Graziano, M. P., Casey, P. J., & Gilman, A. G. (1987) *J. Biol. Chem. 262*, 11375–11381.

Halliday, K. R. (1983–1984) *J. Cyclic Nucleotide Protein Phosphorylation Res. 9*, 435–448.

Hildebrandt, J. D., Codina, J., Rosenthal, W., Birnbaumer, L., Neer, E., Yamazaki, A., & Bitensky, M. W. (1985) *J. Biol. Chem. 260*, 14867–14872.

Hunter, T., Ling, N., & Cooper, J. A. (1984) *Nature (London) 311*, 480–483.

Hurley, J. B., Fong, H. K. W., Teplow, D. B., Dreyer, W. J., & Simon, M. I. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6948–6952.

Hwang, Y.-W., & Miller, D. L. (1987) *J. Biol. Chem. 262*, 13081–13085.

Itoh, H., Kozasa, T., Nagata, S., Nakamura, S., Katada, T., Ui, M., Iwai, S., Ohtsuka, E., Suzuki, K., & Kaziro, Y. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3776–3780.

Iyengar, R., Rich, K. A., Herberg, J. T., Grenet, D., Mumby, S., & Codina, J. (1987) *J. Biol. Chem. 262*, 9239–9245.

Jelsema, C., & Axelrod, J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3623–3927.

Jones, D. T., & Reed, R. R. (1987) *J. Biol. Chem. 262*, 14241–14249.

Jurnak, F. (1985) *Science (Washington, D.C.) 230*, 32–36.

Kahn, R. A., & Gilman, A. G. (1984) *J. Biol. Chem. 259*, 6235–6240.

Kahn, R. A., & Gilman, A. G. (1986) *J. Biol. Chem. 261*, 7906–7911.

Kamps, M. P., Buss, J., & Sefton, B. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4625–4628.

Katada, T., Gilman, A. G., Watanabe, Y., Bauer, S., & Jakobs, K. H. (1985) *Eur. J. Biochem. 151*, 431–437.

Katada, T., Oinuma, M., & Ui, M. (1986) *J. Biol. Chem. 261*, 5215–5221.

Katada, T., Oinuma, M., Kusakabe, K., & Ui, M. (1987) *FEBS Lett. 213*, 353–358.

Kataoka, T., Powers, S., McGill, C., Fasano, O., Strathern, J., Broach, J., & Wigler, M. (1984) *Cell (Cambridge, Mass.) 37*, 437–445.

Kim, S., Arg, S.-L., Bloch, D. B., Bloch, K. D., Kawahara, Y., Tolman, C., Lee, R., Seidman, J. G., & Neer, E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* (in press).

Kobilka, B. K., Frielle, T., Collins, S., Yang-Feng, T., Kobilka, T. S., Francke, U., Lefkowitz, R. J., & Caron, M. G. (1987) *Nature (London) 329*, 75–79.

la Cour, T. F. M., Nyborg, J., Thirup, S., & Clark, B. F. C. (1985) *EMBO J. 4*, 2385–2388.

Lerea, C., Somers, D. E., Hurley, J. B., Klock, I. B., & Bunt-Milam, A. H. (1986) *Science (Washington, D.C.) 234*, 77–80.

Lipman, D. J., & Pearson, W. R. (1985) *Science (Washington, D.C.) 227*, 1435–1441.

Lochrie, M. A., Hurley, J. B., & Simon, M. I. (1985) *Science (Washington, D.C.) 228*, 96–99.

Logothetis, D. E., Kurachi, Y., Galper, J., Neer, E., & Clapham, D. E. (1987) *Nature (London) 325*, 321–326.

March, P. E., & Inouye, M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7500–7504.

Marshall, M. S., Gibbs, J. B., Scolnick, E. M., & Sigal, I. S. (1988) *Mol. Cell. Biol. 8*, 52–61.

Masters, S. B., Stroud, R. M., & Bourne, H. R. (1986) *Protein Eng. 1*, 47–54.

Masu, Y., Nakayama, K., Tamaki, H., Harada, Y., Kuno, M., & Nakanishi, S. (1987) *Nature (London) 329*, 836–838.

Mattera, R., Codina, J., Crozat, A., Kidd, V., Woo, S. L. C., & Birnbaumer, L. (1986) *FEBS Lett. 206*, 36–41.

McAlpine, P. J., Van Cong, N., Boucheix, C., Pakstis, A. J., Doute, R. C., & Shows, T. B. (1987) *Cytogenet. Cell Genet. 46*, 29–101.

McConnell, D. G., Kohnken, R. E., & Smith, A. J. (1984) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 43*, 1585.

Medynski, D. C., Sullivan, K., Smith, D., Van Dop, C., Chang, F.-H., Fung, B., K.-K., Seeburg, P. H., & Bourne, H. R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4311–4315.

Michel, T., Winslow, J. W., Smith, J. A., Seidman, J. G., & Neer, E. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7663–7667.

Miyajima, I., Nakafuku, M., Nakayama, N., Brenner, C., Miyajima, A., Kaibuchi, K., Arai, K., Kaziro, Y., & Matsumoto, K. (1987) *Cell (Cambridge, Mass.) 50*, 1011–1019.

Nakafuku, M., Itoh, H., Nakamura, S., & Kaziro, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2140–2144.

Nakafuku, M., Obara, T., Kaibuchi, K., Miyajima, I., Miyajima, A., Itoh, H., Nakamura, S., Arai, K.-I., Matsumoto, K., & Kaziro, Y. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1374–1378.

Navon, S. E., & Fung, B. K.-K. (1987) *J. Biol. Chem. 262*, 15746–15751.

Neer, E. J., Michel, T., Eddy, R., Shows, T., & Seidman, J. G. (1987) *Hum. Genet. 77*, 259–262.

Nukada, T., Tanabe, T., Takahashi, H., Noda, M., Haga, K., Haga, T., Ichiyama, A., Kangawa, K., Hiranaga, M., Matsuo, H., & Numa, S. (1986a) *FEBS Lett. 197*, 305–310.

Nukada, T., Tanabe, T., Takahashi, H., Noda, M., Hirose, T., Inayama, S., & Numa, S. (1986b) *FEBS Lett. 195*, 220–224.

Ono, Y., Fujii, T., Ogita, K., Kikkawa, U., Igarashi, K., & Nishizuka, Y. (1987) *FEBS Lett. 226*, 125–128.

Ovchinnikov, Y. A., Lipkin, V. M., Shuvaeva, T., M., Bogachuk, A. P., & Shemyakin, V. V. (1985) *FEBS Lett. 179*, 107–110.

Ovchinnikov, Y. A., Lipkin, V. M., Kumarev, V. P., Gubanov, V. V., Khramtsov, N. V., Akhmedov, N. B., Zagranichny, V. E., & Muradov, K. G. (1986) *FEBS Lett. 204*, 288–292.

Ovchinnikov, Y. A., Gubanov, V. V., Khramtsov, N. V., Ischenko, K. A., Zagranichny, V. E., Muradov, K. G., Shuvaeva, T. M., & Lipkin, V. M. (1987a) *FEBS Lett. 223*, 169–173.

Ovchinnikov, Y. A., Slepak, V. Z., Pronin, A. N., Shlensky, A. B., Levina, N. B., Vocikov, V. L., & Lipkin, V. M. (1987b) *FEBS Lett. 226*, 91–95.

Pellman, D., Garber, E. A., Cross, F. R., & Hanaf, H. (1985) *Nature (London) 314*, 374–377.

Price, S. R., Tsai, S.-C., Adamik, R., Angus, C. W., Van Meurs, K. P., Czarnecki, S., Bruckwick, E. C., Moss, J., & Vaughan, M. (1987) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 42*, 2230.

Quaife, C. J., Pinkert, C. A., Ornitz, D. M., Palmiter, R. D., & Brinster, R. L. (1987) *Cell (Cambridge, Mass.) 48*, 1023–1034.

Rall, T., & Harris, B. A. (1987) *FEBS Lett. 224*, 365–371.

Robishaw, J. D., Russel, D. W., Harris, B. A., Smigel, M. D., & Gilman, A. G. (1986a) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1251–1255.

Robishaw, J. D., Smigel, M. D., & Gilman, A. G. (1986b) *J. Biol. Chem. 261*, 9587–9590.

Sakagami, Y., Yoshida, M., Isogai, M., & Suzuki, A. (1981) *Science (Washington, D.C.) 212*, 1525–1527.

Salminen, A., & Novick, P. J. (1987) *Cell (Cambridge, Mass.) 49*, 527–538.

Schmitt, H. D., Wagner, P., Pfaff, E., & Gallwitz, D., (1986) *Cell (Cambridge, Mass.) 47*, 401–412.

Seeburg, P. H., Colby, W. W., Capon, D., Goeddel, D. V., & Levinson, A. D. (1984) *Nature (London) 312*, 71–75.

Sibley, D. R., Benovic, J. L., Caron, M. G., & Lefkowitz, R. J. (1987) *Cell (Cambridge, Mass.) 48*, 913–922.

Sigal, I. S., Gibbs, J. B., D'Alonzo, J. S., & Scolnick, E. M. (1986a) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4725–4729.

Sigal, I. S., Gibbs, J. B., D'Alonzo, J. S., Temeles, G., Wolanski, B. S., Socher, S. H., & Scolnick, E. M. (1986b) *Proc. Natl. Acad. Sci. U.S.A. 83*, 952–956.

Stein, R. B., Tai, J. Y., & Scolnick, E. M. (1986) *J. Virol. 60*, 782–786.

Sternweis, P. C. (1986) *J. Biol. Chem. 261*, 631–637.

Sternweis, P. C., & Robishaw, J. D. (1984) *J. Biol. Chem. 259*, 13806–13813.

Stryer, L., & Bourne, H. R., (1986) *Annu. Rev. Cell Biol. 2*, 391–419.

Sugimoto, K., Nukada, T., Tanabe, T., Takahashi, H., Noda, M., Minamino, N., Kangawa, K., Matsuo, H., Hirose, T., Inayama, S., & Numa, S. (1985) *FEBS Lett. 191*, 235–240.

Suki, W. N., Abramowitz, J., Mattera, R., Codina, J., & Birnbaumer, L. (1987) *FEBS Lett. 220*, 187–192.

Sullivan, K. A., Liao, Y.-C., Alborzi, A., Beiderman, B., Chang, F.-H., Masters, S. B., Levinson, A. D., & Bourne, H. R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6687–6691.

Sullivan, K. A., Miller, R. T., Masters, S. B., Beiderman, B., Heideman, W., & Bourne, H. R. (1987) *Nature (London) 330*, 758–760.

Tanabe, T., Nukada, T., Nishikawa, Y., Sugimoto, K., Suzuki, H., Takahashi, H., Noda, M., Haga, T., Ichiyama, A., Kangawa, K., Minamino, N., Matsuo, H., & Numa, S. (1985) *Nature (London) 351*, 242–245.

Taylor, C. W., & Merritt, J. E. (1986) *Trends Pharmacol. Sci. 7*, 238–242.

Touchot, N., Chardin, P., & Tavitian, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8210–8214.

Towler, D. A., Adams, S. P., Eubanks, S. R., Towery, D. S., Jackson-Machelski, E., Glaser, L., & Gordon, J. I. (1988) *J. Biol. Chem. 263*, 1784–1790.

Tse, A. G. D., Barclay, A. N., Watts, A., & Williams, A. F. (1985) *Science (Washington, D.C.) 230*, 1003–1008.

Van Dop, C., Tsubokawa, M., Bourne, H. R., & Ramachandran, J. (1984) *J. Biol. Chem. 259*, 696–698.

Van Meurs, K. P., Angus, C. W., Lavu, S., Kung, H.-F., Czarnecki, S. K., Moss, J., & Vaughan, M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3107–3111.

Waldo, G. L., Evans, T., Fraser, E. D., Northup, J. K., Martin, M. W., & Harden, T. K. (1987) *Biochem. J. 246*, 431–439.

Walter, M., Clark, S. G., & Levinson, A. D. (1986) *Science (Washington, D.C.) 233*, 649–652.

Wensel, T. G., & Stryer, L. (1986) *Proteins: Struct., Funct., Genet. 1*, 90–99.

West, R. E., Jr., Moss, J., Vaughan, M., Liu, T., & Liu, T.-Y. (1985) *J. Biol. Chem. 260*, 14428–14430.

Wistow, G., Katial, A., Craft, C., & Shinohara, T. (1986) *FEBS Lett. 196*, 23–28.

Worley, P. F., Baraban, J. M., Van Dop, C., Neer, E., & Snyder, S. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4561–4565.

Yatsunami, K., & Khorana, H. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4316–4320.

Yatsunami, K., Pandya, B. V., Oprian, D. D., & Khorana, H. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1936–1940.

Zick, Y., Sagi-Eisenberg, R., Pines, M., Gierschik, P., & Spiegel, A. M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9294–9297.

Chapter 1

Sequence of the Alpha Subunit of Photoreceptor G Protein: Homologies Between Transducin, *ras*, and Elongation Factors

Michael A. Lochrie, James B. Hurley, and Melvin I. Simon

Division of Biology (147-75),
California Institute of Technology,
Pasadena, California 91125

(Published in Science 228, 96-99)

Abstract. *A bovine retinal complementary DNA clone encoding the α subunit of transducin (T_α) was isolated with the use of synthetic oligodeoxynucleotides as probes, and the complete nucleotide sequence of the insert was determined. The predicted protein sequence of 354 amino acids includes the known sequences of four tryptic peptides and sequences adjacent to the residues that undergo adenosine diphosphate ribosylation by cholera toxin and pertussis toxin. On the basis of homologies to other proteins, such as the elongation factors of protein synthesis and the* ras *oncogene proteins, regions are identified that are predicted to be acylated and involved in guanine nucleotide binding and hydrolysis. Amino acid sequence similarity between T_α and* ras *is confined to these regions of the molecules.*

An early step in the processing of visual information involves transducin, a protein found in the rod outer segments of photoreceptors. It functions to couple the photolysis of rhodopsin to changes in intracellular levels of cyclic guanosine monophosphate (GMP) (*1*). When photoreceptors are illuminated, transducin interacts with rhodopsin, and, as a result of this interaction, guanosine diphosphate (GDP) bound to the α subunit of transducin (T_α) is exchanged for guanosine triphosphate (GTP). In this form, T_α is able to activate a cyclic GMP phosphodiesterase. The activation is terminated when the intrinsic guanosine triphosphatase activity of T_α hydrolyzes the bound GTP, restoring the transducin-GDP complex and completing the cycle. Transducin is a member of a family of proteins collectively called G proteins (*2*). Other G proteins interact with various hormone and neurotransmitter receptors and couple receptor-ligand binding events to changes in amounts of intracellular "second messenger," such as cyclic adenosine monophosphate and perhaps calcium ion (*3*). The G proteins are also substrates for adenosine diphosphate (ADP)–ribosylation by cholera toxin, pertussis toxin, or both. The ADP-ribosylation of transducin by cholera toxin inhibits GTP hydrolysis and fixes the α subunit in the GTP-bound state, whereas pertussis toxin acts to stabilize the GDP-bound form (*4*).

We have shown that the α subunits of transducin and G_o, an abundant GTP-binding protein in brain tissue, have extensive amino acid sequence homology (*5*). Furthermore, the amino acid sequence of an NH_2-terminal tryptic peptide derived from T_α was shown to have 59 percent homology with a corresponding region of the *ras* gene products. The *ras* gene family encodes guanine nucleotide–binding proteins that may be important in regulating cell growth and oncogenesis. On the basis of the sequence homology and other similarities with G proteins, we suggested that *ras* might also function as a coupling protein to transduce signals from receptors that interact with factors that regulate cell growth. In order to study G protein structure and function and more clearly define similarities and differences between the *ras* and G proteins, we isolated a complementary DNA (cDNA) corresponding to the α subunit of transducin and determined its nucleic acid sequence.

The sequences of the mixtures of oligonucleotides that were used to screen the λgt10 library prepared from bovine retinal RNA (*6*) are shown in Fig. 1A. Approximately 125,000 plaques were screened with the T_αA mixed probe and 54 were found to hybridize. These were again screened with the T_αB probe and one clone ($\lambda T_\alpha 1$) hybridized specifically with both probes. By subcloning three fragments of the insert and using synthetic oligonucleotides as primers, we obtained the complete nucleic acid sequence of the insert. The sequencing strategy is shown in Fig. 1B and the nucleotide sequence and predicted amino acid sequence are shown in Fig. 2. The nucleic acid sequence extends 429 base pairs past the TAG codon signaling the end of the coding region, and no extensive stretch of poly(A) (polyadenylate) sequence was found. There is also a 174-base-pair 5′ untranslated segment.

The open reading frame predicts an amino acid sequence including 354 amino acids. The predicted molecular weight of the protein (40.1 kD), the amino acid composition, and the size of the tryptic fragments agree with recorded values (*5*, *7*). The predicted amino acid sequences of the amino terminus of the 9-kD tryptic fragment (Lys^{209}-Tyr^{230}) and of the tetrapeptide that includes the

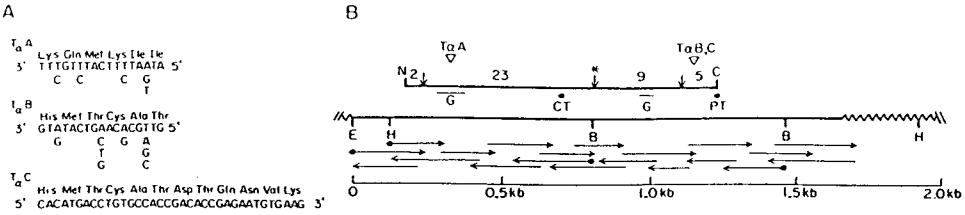


Fig. 1. (A) Synthetic oligonucleotides used to isolate and confirm the identity of a cDNA clone for the α subunit of transducin. Amino acid sequences of tryptic fragments of T_α (*5*) were selected to synthesize the corresponding oligonucleotides ($T_\alpha A$ and $T_\alpha B$) representing all possible codon combinations. Specifically for $T_\alpha A$, this included 24 molecules, each 17 nucleotides long; for $T_\alpha B$, 64 molecules each 17 nucleotides long; and for $T_\alpha C$, 1 molecule, 36 nucleotides long. Oligonucleotides were synthesized with an Applied Biosystems model 380A automated DNA synthesizer and purified by gel electrophoresis before use. The bovine retinal cDNA library was screened as described (*15*). $T_\alpha A$ and $T_\alpha B$ were hybridized at 35° and 43°C, respectively. The construction of oligonucleotide $T_\alpha C$ was based on codons most commonly found in bovine rhodopsin and other eukaryotic proteins. $T_\alpha C$ was used as a nucleotide sequencing primer (Fig. 1B) to confirm the identity of a clone ($\lambda T_\alpha 1$) which hybridizes to $T_\alpha A$ and $T_\alpha B$. (B) Organization and nucleotide sequencing strategy of $\lambda T_\alpha 1$ cDNA insert. The 0.65-kilobase-pair (kbp) Bam HI, 0.8-kbp Eco RI–Bam HI, and 1.8-kbp Hind III restriction endonuclease fragments of $\lambda T_\alpha 1$ were subcloned into M13mp10 or 11 and sequencing templates were prepared by standard methods (*19*). Nucleotide sequences were determined by the dideoxynucleotide chain-termination method (*20*). Sequences determined with $T_\alpha C$ (Fig. 1A) and a universal sequencing primer (Bethesda Research Laboratories) initially served as the basis for the construction of other synthetic oligonucleotide sequencing primers. Symbols used are as follows: thick top line, extent of protein coding region; vertical arrows, sites of tryptic proteolysis of native T_α (*5*); asterisk, tryptic site accessible in T_α bound to GDP but not in T_α bound to guanosine 5′-(β,γ-imido) triphosphate; numbers, molecular size in kilodaltons of tryptic fragments as determined by gel electrophoresis; triangles, location of amino acid sequences that served as a basis for construction of oligonucleotides shown in Fig. 1A; N and C, amino and carboxyl terminus of T_α, respectively; CT and PT, location of amino acids ADP-ribosylated by cholera toxin and pertussis toxin; G, regions predicted to be involved in guanine nucleotide binding and hydrolysis; thick bottom line, extent of cDNA insert; jagged line, λgt10 DNA; horizontal arrows, direction and extent of nucleotide sequences determined using synthetic oligonucleotides as sequencing primers; and dotted horizontal arrows, direction and extent of nucleotide sequences determined with the use of a universal sequencing primer. A scale of nucleotide lengths in kilobase pairs is shown at the bottom. Relevant restriction endonuclease sites are abbreviated as follows: B, Bam HI; E, Eco RI; and H, Hind III.

Fig. 2. Nucleotide sequence of $\lambda T_\alpha 1$ cDNA and deduced amino acid sequence of T_α protein. Numbering of the nucleic acid sequence is shown in parentheses. The predicted amino acid sequence is numbered every ten residues. Amino acid sequences from tryptic peptides of T_α (*5*, *8*) are underlined. Other symbols and abbreviations are as in Fig. 1B. The insert of $\lambda T_\alpha 1$ cannot be excised with Eco RI even though the cDNA library was constructed with Eco RI linkers. The nucleotide sequence from bases 1666 to 1672, which was GAAATTC instead of GAATTC, explains this anomaly, but the origin of the extra adenine is unknown.

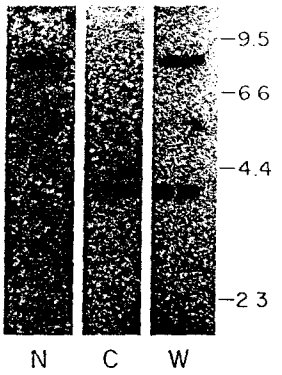


Fig. 3. Hybridization of bovine genomic DNA to T_α cDNA. Twenty micrograms of high molecular weight bovine liver DNA was cleaved with Bam HI, fractionated by agarose gel electrophoresis, transferred to nitrocellulose, and hybridized to primer extended probes (*19*) derived from M13 subclones containing the EcoRI–Bam HI fragment corresponding to the amino-terminal two-thirds of T_α (N), the Bam HI fragment corresponding to the carboxyl-terminal third of T_α (C), or the Hind III fragment that contains the entire coding region (W) (Fig. 1B). The amount of [α^{32}P]dATP (deoxyadenylate) used in the reactions was limited so that a probe of an average length of 1000 bases was made. Therefore, probe W, which begins at base 129, does not extend beyond the last base of probe C at base 1458. Hybridization was at 65°C in 6× SSC (saline sodium citrate), 0.1 percent sodium dodecyl sulfate (SDS), 2× Denhardt's solution, 50 μg per milliliter sheared salmon testes DNA, and 10 percent dextran sulfate. The filter was washed at 42°C in 0.2× SSC and 0.1 percent SDS.

cholera toxin ADP-ribosylation site (Ser^{177}-Lys^{180}) are identical with the amino acid sequences determined by Edman degradation. The amino acid reported to be ADP-ribosylated by cholera toxin is an arginine in the sequence Ser-Arg-Val-Lys (*8*). This sequence is found only once, is preceded by an arginine as expected for a tryptic peptide, and is located on the 23-kD tryptic fragment (*9*). The predicted sequence of the 5-kD tryptic fragment (Asp^{315}-Phe^{354}) is also identical with the amino acid sequence as reported (*5*), except for Cys^{351} and Leu^{353}, which were not identified. It has been suggested that the target for ADP-ribosylation by pertussis toxin is an asparagine residue corresponding to position 350 (*10*). However, the cDNA sequence (Fig. 2) predicts an aspartic acid residue at position 350 in the primary translation product. Thus, a modified form of aspartic acid could be the substrate for pertussis toxin. Alternatively, Cys^{351}, which was not previously detected (*10*), could also be the site of ADP-ribosylation.

There are four amino acid differences between the predicted sequence and the reported sequence of the 23-kD tryptic fragment. They are conservative differences and cluster at the amino terminus. Amino acid sequence analysis resulted in identification of Lys^{24}, Glu^{28}, Asp^{30}, and Arg^{32}. The nucleic acid sequence predicts Gin^{24}, Asp^{28}, Glu^{30}, and Lys^{32}. It is possible that these differences do not arise from multiple transducin genes within an organism but instead represent polymorphisms in the different animals whose tissues were used to prepare transducin protein for sequence determination and poly(A)$^+$ RNA (polyadenylated RNA) for cloning. This argument is supported by the hybridization experiment shown in Fig. 3. Digestion of genomic DNA with Bam HI and hybridization with probes made from the NH_2-terminal, COOH-terminal, and complete coding regions of the cDNA give a pattern that is only consistent with the presence of a single gene corresponding to T_α.

The complete sequence of transducin allows us to compare it with other GTP-binding proteins. Similarities were found in the sequences of mammalian and yeast *ras*, the T_α subunit, and the elongation factors (EF-Tu and EF-1α). In each of these proteins GTP binding and hydrolysis are central to their function. Homologies between EF-Tu and *ras* have been previously noted (*11*). Sequence similarities between T_α and the other proteins are primarily clustered in two regions (Fig. 4). One region is at the NH_2-terminal end starting at Lys^{29} in T_α and extending through Lys^{70}. There is evidence to suggest that this region is involved in GTP hydrolysis in *ras*. Mutagenesis resulting in changes in amino acids in this region affect guanosine triphosphatase and oncogenic activity (*12*). A second region of homology is found toward the COOH-terminal end of the T_α subunit and includes the residues from Ala^{260} to Lys^{276}. This region may participate in GTP binding. The homologous region in all elongation factors includes an asparagine residue (Asn^{135} in EF-Tu) which has been shown by x-ray crystallographic analysis to form a hydrogen bond with the guanosine moiety of GDP (*13*). Furthermore, modification of Cys^{137} in EF-Tu eliminates nucleotide binding. In EF-Tu, this region forms part of a six-stranded β sheet structure common to many nucleotide-binding proteins (*14*).

There is evidence to suggest that the guanyl nucleotide binding site in T_α interacts with a region of the protein that is located between amino acids 177 to 208. Thus, for example, bound analogs of GTP block trypsin cleavage at Arg^{208} (*7*). Furthermore, the ADP ribosylation site at Arg^{178} is only recognized by cholera toxin when T_α binds nonhydrolyzable GTP analogs (*4*). Finally, ADP ribosylation at this site blocks the GTP hydrolysis activity of transducin.

A striking feature that is conserved between T_α and *ras* is the presence of a

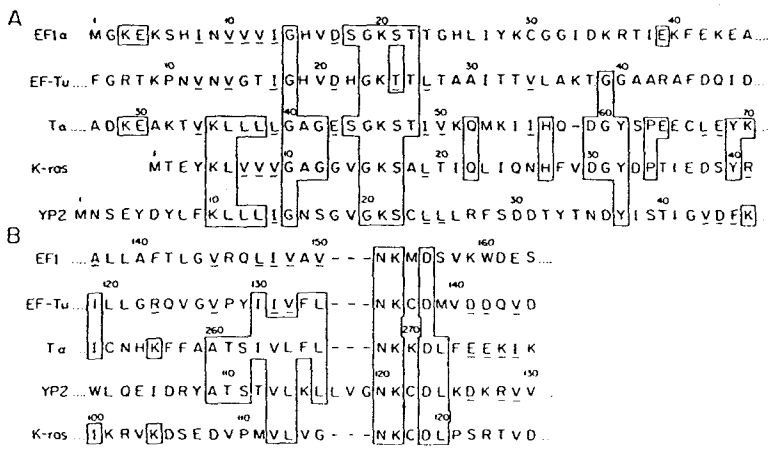


Fig. 4. Homologies of T_α to the protein sequences of translation elongation factors and *ras* oncogenes (*21*). Homologies were identified with the use of a computer homology matrix program (*22*) which determines the degree of identity found over an adjustable number of amino acids. (A) Homologies among regions predicted to play a role in guanine nucleotide hydrolysis. (B) Homologies among regions predicted to be involved in guanine nucleotide binding. Sequence identities to T_α are boxed and close functional homologies are underlined. Single letter abbreviations used for amino acids are: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; and Y, tyrosine.

cysteine residue as the fourth amino acid from the carboxyl terminus. Both the α and the γ (*15*) subunits of transducin as well as 11 *ras* gene products that have been examined (*16*) have a cysteine in this position. This residue is acylated with lipid in *ras* and is important for *ras* function (*17*). Acylation could occur at Cys^{351} in T_α. Another possible site for acylation in T_α is on Gly^2. The amino terminus of T_α is blocked, and all proteins known to have NH_2-terminal myristic acid blocking groups have glycine as the NH_2-terminal amino acid (*18*).

The homologies found between *ras*, elongation factors, and T_α reflect the guanine nucleotide binding and hydrolytic activities that are necessary for an alternation between GDP- and GTP-bound conformations that determine the nature of the reversible association of these proteins with other macromolecules. The portions of *ras* and T_α that are not directly involved in nucleotide binding or hydrolysis may govern the specificity for interaction with different subcellular components. The availability of the intact cDNA clones makes it possible to overproduce T_α and to modify the gene product by mutagenesis. Reconstruction of the G protein–coupling system in vitro and in vivo will provide an experimental system that may allow us to understand in detail how these proteins function.

MICHAEL A. LOCHRIE
JAMES B. HURLEY
MELVIN I. SIMON

Division of Biology,
California Institute of Technology,
Pasadena 91125

References and Notes

1. L. Stryer, J. B. Hurley, B. K.-K. Fung, *Curr. Top. Membr. Transp.* **15**, 93 (1981).
2. A. Gilman, *Cell* **36**, 577 (1984); E. Ross and A. Gilman, *Annu. Rev. Biochem.* **49**, 553 (1980); D. R. Manning and A. Gilman, *J. Biol. Chem.* **258**, 7059 (1983); P. Greengard *et al.*, Eds., *Advances in Cyclic Nucleotide and Protein Phosphorylation Research* (Raven, New York, 1984), vol. 17.
3. M. J. Berridge and R. F. Irvine, *Nature (London)* **312**, 315 (1984); F. Okajima and M. Ui, *J. Biol. Chem.* **259**, 13863 (1984); G. Bokoch and A. Gilman, *Cell* **39**, 301 (1984); B. Gomperts, *Nature (London)* **306**, 64 (1983).
4. M. Abood, J. B. Hurley, M.-C. Pappone, H. Bourne, L. Stryer, *J. Biol. Chem.* **257**, 10540 (1982); C. Van Dop *et al.*, *ibid.* **259**, 23 (1984).
5. J. B. Hurley, M. I. Simon, D. B. Teplow, J. D. Robishaw, A. G. Gilman, *Science* **226**, 860 (1984).
6. J. Nathans and D. Hogness, *Cell* **34**, 807 (1983).
7. Y.-K. Ho and B. K.-K. Fung, *J. Biol. Chem.* **259**, 6694 (1984); D. R. Manning and A. Gilman, *ibid.* **258**, 7059 (1983); B. K.-K. Fung and C. R. Nash, *ibid.*, p. 10503.
8. C. Van Dop, M. Tsubokawa, H. Bourne, J. Ramachandran, *J. Biol. Chem.* **259**, 696 (1984).
9. S. Navon and B. K.-K. Fung, *ibid.*, p. 6686.
10. D. Manning, B. Fraser, R. Kahn, A. Gilman, *ibid.*, p. 749.
11. R. Leberman and U. Egner, *EMBO J.* **3**, 339 (1984); K. R. Halliday, *J. Cyclic Nucleotides Prot. Phosph. Res.* **9**, 435 (1984).
12. J. McGrath, D. Capon, D. Goeddel, A. Levinson, *Nature (London)* **310**, 644 (1984); R. Sweet *et al.*, *ibid.* **311**, 273 (1984); J. Gibbs, I. Sigal, M. Poe, E. Scolnick, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5704 (1984).
13. R. Leberman and U. Egner, *EMBO J.* **3**, 339 (1984); F. Duisterwinkel *et al.*, *ibid.*, p. 113.
14. J. S. Richardson, *Adv. Prot. Chem.* **34**, 167 (1981); M. Rossman, D. Moras, K. Olsen, *Nature (London)* **250**, 194 (1974).
15. J. B. Hurley *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6948 (1984).
16. C. D. Reymond, R. H. Comer, M. C. Mehdy, R. A. Firtel, *Cell* **39**, 141 (1984); F. Neuman-Silberberg, E. Schejter, F. Hoffman, B.-Z. Shilo, *ibid.* **37**, 1027 (1984); E. Taprowsky, K. Shimizu, M. Goldfarb, M. Wigler, *ibid.* **34**, 581 (1983).
17. B. Willumsen, K. Norris, A. Papageorge, N. Hubbert, D. Lowy, *EMBO J.* **3**, 2581 (1984).
18. A. Aitken and P. Cohen, *Methods Enzymol.* **106**, 205 (1984); A. Schultz and S. Oroszlan, *Virology* **133**, 431 (1984); J. Ozols, S. Carr, P. Strittmatter, *J. Biol. Chem.* **259**, 13349 (1984); A. M. Schultz, L. E. Henderson, S. Oroszlan, E. A. Garber, H. Hanafusa, *Science* **227**, 427 (1985).
19. J. Messing, *Methods Enzymol.* **101**, 20 (1984).
20. F. Sanger, S. Nicklen, A. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977); R. Sanchez-Pescador and M. Urdea, *DNA* **3**, 339 (1984).
21. D. Gallwitz, C. Donath, C. Sander, *Nature (London)* **306**, 704 (1983); S. Powers *et al.*, *Cell* **36**, 607 (1984); S. Nagata *et al.*, *EMBO J.* **3**, 1825 (1984); K. Arai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1326 (1980).
22. J. Pustell and F. Kafatos, *Nucleic Acids Res.* **12**, 643 (1984).
23. We thank Jeremy Nathans for providing the retinal cDNA library, Suzanna Horvath for synthesis of oligodeoxynucleotides, and Ryn Miake-Lye for primer-extended M13 probes. Supported by NSF grant PCM 8410039 (M.S.), a Helen Hay Whitney fellowship (J.H.), and an NSF graduate fellowship (M.L.).

22 January 1985; accepted 9 February 1985

Chapter 2

G Protein Alpha Subunits in the Nematode *Caenorhabditis elegans*

Michael A. Lochrie, Jane E. Mendel,
Paul W. Sternberg, and Melvin I. Simon

Division of Biology (147-75),
California Institute of Technology,
Pasadena, California 91125

(Prepared for submission to Cell Regulation)

ABSTRACT

The genomic fragments of three *C. elegans* G protein alpha subunit genes (*gpa-1*, *gpa-2*, and *gpa-3*) have been isolated by using the polymerase chain reaction (PCR). The corresponding cosmid clones were isolated and mapped to disperse locations on chromosome five. Another cosmid that weakly hybridizes to the *gpa-1* PCR-derived gene fragment maps to the X chromosome. Each maps near known genes that could correspond to G protein alpha subunit genes. The sequences of two of the genes, *gpa-1* and *gpa-3*, were determined. Four of the intron positions in the *gpa-1* and *gpa-3* genes are conserved relative to the mammalian $G_i\alpha$, $G_o\alpha$, and $T_r\alpha$ genes. The predicted amino acid sequences of *gpa-1* and *gpa-3* are 48% identical to each other and they show no obvious similarity to other G protein alpha subunit sequences that would allow them to be classified. Therefore, they are likely to have distinct functions. In contrast, a cDNA corresponding to a known G protein α subunit, the alpha subunit of G_o ($G_o\alpha$), was also isolated and sequenced. The predicted amino acid sequence of *C. elegans* $G_o\alpha$ is 80-87% identical to other $G_o\alpha$ sequences. An mRNA that hybridizes to the *C. elegans* $G_o\alpha$ cDNA can be detected on Northern blots. A protein that may be *C. elegans* $G_o\alpha$ can be detected on immunoblots. A cosmid clone containing the *C. elegans* $G_o\alpha$ gene (*goa-1*) was isolated and mapped to chromosome 1. Thus *C. elegans* has G proteins that are identifiable homologs of mammalian G proteins and G proteins that are unique to *C. elegans*. Study of the identifiable homologs may result in a further understanding of their function in other organisms. The novel G proteins may aid in understanding unique aspects of *C. elegans* physiology.

ABBREVIATIONS

ADP, adenosine 5'-diphosphate

BCIP, 5-bromo-4-chloro-3-indolyl phosphate

BSA, bovine serum albumin

cDNA, complementary DNA

DNA, deoxyribonucleic acid

EDTA, ethylenediamine tetraacetic acid

G protein, guanine nucleotide-binding regulatory protein

$G_i\alpha$, alpha subunit of the inhibitory G protein

$G_o\alpha$, alpha subunit of the "other" G protein

$G_q\alpha$, alpha subunit of the "quacky" G protein

$G_s\alpha$, alpha subunit of the stimulatory G protein

$G_z\alpha$, alpha subunit of the "zany" G protein

GTP, guanosine 5'-triphosphate

goa, *C. elegans* $G_o\alpha$

gpa, *C. elegans* G protein alpha subunit

kbp, kilobase pairs

mRNA, messenger RNA

NBT, 4-nitro blue tetrazolium chloride

PCR, polymerase chain reaction

RNA, ribonucleic acid

rRNA, ribosomal RNA

$T_r\alpha$, alpha subunit of the rod photoreceptor G protein

INTRODUCTION

G proteins mediate transmembrane signal transduction by physically coupling cell surface receptors to "effector" proteins that influence intracellular metabolism (Gilman, 1987). A large family of G protein-linked receptors has been identified (Caron, 1989). All of these receptors have seven putative transmembrane segments and bind small organic molecules or peptide hormones. The effectors regulated by G proteins that have been identified include adenylate cyclase, cGMP phosphodiesterase, phospholipase A_2, phospholipase C, and various ion channels.

G proteins are composed of three subunits: α, β, and γ. The α subunit undergoes a guanine nucleotide exchange and hydrolysis cycle. In most cells it is the α subunit that directly interacts with effector proteins. G protein α subunits are substrates for several covalent modifications. They are ADP-ribosylated by pertussis toxin, cholera toxin, or both (Moss and Vaughan, 1988). These toxins have opposite effects. Pertussis toxin uncouples receptors from G proteins and causes a deactivation of the signal transduction system. Cholera toxin inhibits GTPase activity and causes a persistant activation of the signal transduction system. Some G protein α subunits are also myristoylated (Buss *et al.*, 1987). Myristoylation occurs on glycine-2 and is thought to be important for localization of G protein α subunits to the inner face of the cytoplasmic membrane (Mumby *et al.*, 1989; Jones *et al.*, 1990). G protein α subunits are also phosphorylated (Sagi-Eisenberg, 1989; Gunderson and Devreotes, 1990)

G_O, the "other" G protein (Neer *et al.*, 1984; Sternweis and Robishaw, 1984), is abundant in brain and is found only in organisms with nervous systems. G_O has been found in *Drosophila* (deSousa *et al.*, 1989; Schmidt *et al.*, 1989; Thambi *et al.*, 1989; Yoon *et al.*, 1989), rat (Jones and Reed, 1987), cow (Van Meurs *et al.*, 1987), hamster (Hsu *et al.*, 1990), and *Xenopus* (Olate *et al.*, 1989) but not in yeast or the slime mold *Dictyostelium*. In vertebrates it is concentrated in neuropil (Gabrion *et al.*, 1989; Worley *et al.*, 1986) and growth cones (Strittmatter *et al.*, 1990). In *Drosophila* G_O has been found

in brain and ovaries (deSousa *et al.*, 1989). G_o may function in several signal transduction pathways. In many cases it seems to be able to interact with the same receptors as the closely related G_i proteins, although there are some exceptions (Ueda *et al.*, 1990; Senogles *et al.*, 1990). However G_o may regulate a set of effectors distinct from those regulated by the G_i proteins. The G_i proteins regulate adenylate cyclase and atrial potassium channels while there is evidence that G_o regulates calcium channels (Hescheler *et al.*, 1987; Harris-Warrick *et al.*, 1988; McFadzean *et al.*, 1989), neuronal potassium channels (Van Dongen *et al.*, 1988), and phospholipase C (Moriarty *et al.*, 1989). G_o has no effect on adenylate cyclase activity.

The nematode *C. elegans* offers distinct advantages for studying developmental biology and the function of the nervous system (Wood, 1988). It is the only multicellular organism for which a complete cell lineage is known (Sulston, 1988). The hermaphrodite contains 959 somatic nuclei and of these 302 are neuronal. The entire neuroanatomy and neuronal wiring diagram is known (Chalfie, 1988; White *et al.*, 1988). *C. elegans* exhibits several behaviors including chemotaxis (Ward, 1973) and touch sensitivity (Chalfie and Sulston, 1985). It has many of the neurotransmitters and signaling systems found in higher eukaryotes (Willet, 1980; Chalfie, 1988). Numerous mutants that affect developmental and neuronal processes have been isolated (Hodgkin *et al.*, 1988). Finally, the molecular genetics of *C. elegans* has advanced to the stage where gene transformation is possible and a cosmid library covering most of the genome is available to facilitate gene cloning and mapping.

To understand more about the role of G proteins in the process of development and the function of the nervous system and to study G proteins in an organism where an extensive genetic analysis could also be undertaken, we sought to characterize G protein genes in *C. elegans*. Our first goal was to determine how many G protein genes exist in *C. elegans*. The characterization of a β subunit gene (*gpb-1*; Van der Voorn, 1990) and an α subunit gene (*gpa-2*; Silva and Plasterk, 1990) has been reported. This paper reports the molecular

cloning and sequencing of a *C. elegans* $G_o\alpha$ cDNA and the genes encoding two other apparently novel *C. elegans* G protein α subunits (*gpa-1* and *gpa-3*). With the availability of these sequences one can begin to develop tools to study their role in *C. elegans* biology.

MATERIALS AND METHODS

Polymerase chain reactions. Twenty-five nanograms of *C. elegans* genomic DNA was amplified in a 10 µl reaction containing 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl, pH 8.3, 0.01% gelatin, 200 µm deoxyribonucleotides (dATP, dCTP, dGTP, and dTTP), and 0.3 µg of each primer. The primer pairs that were used are: oMP19, oMP20/oMP21 and oMP19, Tα29. The primers, oMP19, oMP20, and oMP21, have been described (Strathmann *et al.*, 1989). The primer oMP19 corresponds to the sense strand that encodes the amino acid sequence Lys-Trp-Ile-His-Cys-Phe/Leu. The primers, oMP20 and oMP21, correspond to the antisense strand that encodes the amino acid sequence Phe-Leu-Asn-Lys-Lys-Asp. The sequence of Tα29 is GAATTC(GATC)GT(AG)T-C(GATC)GT(GATC)GC(AG)CA(GATC)GT. This sequence contains an Eco RI restriction site at one end to facilitate subcloning and corresponds to the anitsense strand that encodes the amino acid sequence Thr-Cys-Ala-Thr-Asp-Thr. Thirty amplification cycles consisting of 1 minute at 92 °C, 30 seconds at 37 °C, and 1 minute at 72 °C were performed. At the end of the ampification cycles the reactions were incubated another 10 minutes at 72 °C to complete partially extended chains. The products of the reaction were analyzed on 3% NuSieve (FMC) agarose gels. The PCR products were excised from the gel, reamplified in a 100 µl reaction using the same primers, made blunt with Klenow, phophorylated with T4 polynucleotide kinase, ligated to linkers, and subcloned into pBluescript KS/- (Stratagene). Double stranded miniprep DNA for use as a sequencing template was prepared by a modification of the boiling lysis method of Holmes and Quigley (1981). In this procedure 1.5 ml cultures were grown in Luria broth with ampicillin (100 µg/ml) for 6-8 hours or just until the cultures were saturated. The cells were pelleted in 1.5

ml microfuge tubes and resuspended in 300 µl STET (8% sucrose, 5% Triton X-100, 50 mM EDTA, 50 mM Tris-HCl, pH 8.0). Twenty microliters of lysozyme solution (10 mg/ml in 50 mM Tris-HCL, pH 8.0) was added. After incubation at room temperature for 1 minute the mixture was boiled for 2 minutes and immediately microfuged at room temperature for 5 minutes. The pellet was removed with a sterile toothpick and 300 µl 75% isopropanol, 2.5 M ammonium acetate was added. After mixing, the tubes were microfuged at room temperature for 5 minutes. The supernatant was aspirated and the pellet was washed in 1 ml 70% ethanol, 1 ml 100% ethanol, and then dried in a Spin-Vac. The pellet was resuspended in 50 µl H_2O. The plasmid DNA was denatured with alkali by adding 5 µl 2 M NaOH and 2 mM EDTA and incubated at room temperature for 5 minutes. The solution was neutralized by adding 25 µl 0.9 M sodium acetate, pH 5.3. Plasmid DNA was precipitated by adding 200 µl ethanol, mixing, incubating at -70°C for 5 minutes, and microfuging for 10 minutes at 4 °C. The supernatant was aspirated and the pellet was washed in 70% ethanol and dried in a Spin-Vac. The denatured DNA was resuspended in 20 µl H_2O. For DNA sequencing reactions 7 µl denatured template was annealed to 1 µl (10 ng) primer with 2 µl 5 X Sequenase buffer at 37°C for 20 minutes. DNA sequencing reactions were performed using the Sequenase, version 2.0 kit according to the instructions provided by the manufacturer (U. S. Biochemicals).

Isolation of $G_o\alpha$ cDNA. A *C. elegans* cDNA library containing about 250,000 clones was constructed in λgt10 and kindly supplied by Stuart Kim (Stanford). The mRNA used to construct this library represented all developmental stages of the hermaphrodite wild-type N2 (Bristol) strain. The library was plated on 15 cm LB agar plates and replicated to nitrocellulose filters (Maniatis *et al.*, 1982). The filters were prehybridized in 6 X SSC (1 X SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0), 0.1% sodium dodecyl sulfate, 30% formamide, 100 µg/ml sonicated salmon sperm DNA, 5 X Denhardt's solution (1 X Denhardt's solution is 0.02% Ficoll 400, 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone) at 37°C for 2 hours. Probes were added to the prehybridization

solution and hybridized under the same conditions for 24 hours. The probes consisted of the PCR-derived genomic fragments of the *gpa-1*, *gpa-2*, and *gpa-3* genes. These were isotopically labeled with a Multiprime kit (Amersham) using α-^{32}P-dATP. The probes were denatured by boiling for five minutes before use and 1 X 10^6 cpm/ml of each probe was added to the hybridization solution. After hybridization the filters were washed with 0.1 X SSC, 0.1% SDS at 23°C and exposed to x-ray film for 48 hours.

DNA sequence analysis of $G_o\alpha$ cDNA. Restriction fragments of the Eco RI insert of λCe6-2 were subcloned into pBluescript KS/-. The DNA sequences of the inserts were determined using the Sequenase, version 2.0 kit (U. S. Biochemicals) as described above. Most of the sequence was determined in this manner using sequencing primers complimentary to pBluescript KS/-. The remaining sequence was determined using synthetic oligonucleotides as insert-specific sequencing primers. A T7 DNA polymerase stop obscurring the sequence of positions 171-181 was resolved by using the TaqTrack DNA sequencing system (Promega) at a reaction temperature of 80°C. DNA sequences were analyzed using Pustell programs (Pustell and Kafatos, 1984) run on an IBM PC or the University of Wisconsin Genetics Computer Group sequence analysis software package (Devereux *et al.*, 1984) run on a VAXstation 2000.

Northern blot analysis. Total mRNA was isolated by the guanidine isothiocyanate procedure described by Maniatis (1982). The mRNA was denatured and fractionated on a 1% formaldehyde gel. After electrophoresis the RNA was blotted to a nitrocellulose filter. The insert from the $G_o\alpha$ cDNA was labeled by the method of Feinberg and Vogelstein (1983). The filter was hybridized to the probe in 50% formamide, 5 X SSC, 5 X Denhardt's solution, 25 mM $NaPO_4$, 0.1% SDS, 0.25 mg/ml salmon sperm DNA at 42°C. The filter was washed in 0.2 X SSC, 0.1% SDS at 60°C and exposed to x-ray film.

Immunoblot analysis. Mixed populations of nematodes at various developmental stages were washed from 10 cm NGM plates (Sulston and Hodgkin, 1988) in M9 media. The

nematodes were pelleted and 5 X concentrated gel sample buffer (Laemmli, 1970) was added to a final concentration of 1 X. The nematodes were boiled for 5 minutes and the dissolved proteins were fractionated by SDS electorphoresis on a 10% acrylamide/0.27% bis-acrylamide gel (Laemmli, 1970). The proteins were electroblotted onto a nitrocellulose filter by the method of Towbin *et al.* (1979) in 25 mM Tris-HCl, 192 mM glycine buffer, pH 8.3, 20% methanol at 200 mA for 2 hours. The filter was blocked in TBS (20 mM Tris-HCl, pH 7.5, 500 mM NaCl)/3% BSA for 30 minutes and incubated in the primary antibody for 1 hour at room temperature in TBS/1% bovine serum albumin. The filter was washed with TBS for 30 minutes and then incubated in the secondary antibody for 1 hour at room temperature. Goat anti-rabbit IgG (Fc) conjugated to alkaline phosphatase (Promega) was used as the secondary antibody at a dilution of 1:7,500 in TBS/1% bovine serum albumin. After washing in TBS for 10 minutes the blot was developed with NBT and BCIP according to the manufacturers instructions.

Isolation and mapping of cosmid clones. A cosmid library was constructed from N2 genomic DNA that was partially digested with Xho II. The DNA was size fractionated by field inversion gel electrophoresis and subcloned into the Bam HI site of pWEKan. The resulting cosmid library contains about 3 or 4 genome equivalents and the average insert size is 35-50 kilobase pairs. The library was gridded in microtiter plates and replicated to nitrocellulose filters. PCR fragments from the *gpa-1*, *gpa-2*, and *gpa-3* genes were labeled by random primed-labeling (Feinberg and Vogelstein, 1983) and hybridized to the nitrocellulose filter replicas in 1% BSA, 7% SDS, 1 mM EDTA, 0.5% formamide, 0.2M $NaPO_4$, pH 7.2 at 65°C for 12 hours using 4 X 10^5 cpm per ml of probe. The filters were washed in 2 X SSC at 50°C for 30 minutes, 0.2 X SSC at 50°C for 30 minutes, 0.2 X SSC at 65°C for 30 minutes and then exposed to x-ray film. Positives clones were picked, DNA was prepared from them, and their identities confirmed by PCR analysis. The cosmid clones were mapped by Alan Coulson and John Sulston (MRC, Cambridge, England) using a fingerprinting technique that compares the Hind III restriction map of a

cosmid to that of an overlapping set of cosmids (contigs) that now represents over 95% of the *C. elegans* genome (Coulson *et al.*, 1986).

DNA sequence analysis of gpa-1 and gpa-3 genes. Hind III restriction fragments of the *gpa-1* and *gpa-3* cosmid clones that hybridize to the *gpa-1* and *gpa-3* PCR-derived probes were identified by Southern blotting and subcloned into pBluescript KS/- (see Figure 6A and 8A). To obtain the complete sequence of the *gpa-3* gene it was also necessary to subclone a Nsi I restriction fragment that overlaps the Hind III fragment. The DNA sequence of the *gpa-1* and *gpa-3* genes was determined by a combination of three methods. First, specific restriction fragments of the original subclones were subcloned further and sequenced as described above. Second, γδ transposons were inserted at random sites in subclones constructed in pMOB (M. Strathmann, personal communication). The location of the γδ transposons was determined by PCR and DNA sequences were determined by using unique transposon-specific primers corresponding to each of the different ends of γδ. Since γδ generates a five base pair duplication at the site of insertion it is possible to obtain sequences that overlap at the site of γδ insertion. Third, any remaining sequences that were not obtained by the first two methods were determined by constructing synthetic primers based on available sequence and using these as sequencing primers. Sequence compressions were resolved using dITP instead of dGTP in the sequencing reactions and T7 polymerase stops were resolved by using the TaqTrack system at a reaction temperature of 80°C.

RESULTS

Analysis of PCR products.

The primers that were used in the polymerase chain reactions are degenerate and correspond to three closely spaced regions that are highly conserved in G protein α subunits. They represent three of the five regions that are thought to be involved in guanine nucleotide binding and therefore would be expected to be present in most G protein α subunit sequences. The distances between the oMP19, oMP20/oMP21 and the oMP19,

Tα29 primer pairs in a cDNA would be about 200 and 375 base pairs, respectively. Since genomic DNA and not cDNA was amplified in this study the PCR products could be larger as they might contain introns. PCR amplification of *C. elegans* genomic DNA with the oMP19 and oMP20 primer pair produced PCR products of 270, 525, and 1,000 base pairs. Amplification using oMP19 and oMP21 produced PCR products of 260, 300, and 550 base pairs. Amplification using oMP19 and Tα29 produced PCR products of 500, 600, and 800 base pairs. The 550 base pair PCR product from the oMP19 and oMP21 reaction, and the 500, 600, and 800 base pair PCR products from the oMP19 and Tα29 reaction were subcloned. The sequences of several subclones of each PCR product were analyzed. The sequences fell into three classes designated *gpa-1*, *gpa-2*, and *gpa-3* that had amino acids that are found in all G protein α subunits (Lochrie and Simon, 1988). The 550 base pair PCR product corresponds to *gpa-2*, the 500 base pair PCR product to *gpa-3*, and the 600 base pair PCR product to *gpa-1*. The 800 base pair PCR product derived by amplification with oMP19 and Tα29 is an artifact of the PCR reaction. Most of the subclones of this PCR product had truncated *gpa-1* or *gpa-3* sequences fused to unknown sequences. Two had *gpa-3/gpa-1* or *gpa-2/gpa-1* hybrid sequences. Out of a total 85 subclones that were sequenced, 86% were G protein α subunit sequences. All three sequences contained introns. The positions of the introns were identical to the sixth and seventh introns found in the mammalian $G_i\alpha$, $G_o\alpha$, and $T_r\alpha$ genes (Kaziro *et al.*, 1988; Raport *et al.*, 1989).

$G_o\alpha$ cDNA isolation.

A mixture of the three PCR-derived gene fragments was used as a probe to screen a *C. elegans* cDNA library at low stingency. Fifty-two out of 45,000 plaques in the cDNA library hybridized to the PCR-derived probes. Seven of these tested positive by PCR analysis for the presence of G protein α subunit-related sequences using the primer pairs oMP19, oMP20/21 or oMP19, Tα29. The insert of one of these clones (λCe6-2) was subcloned. The restriction map and sequencing strategy for this cDNA are shown in

Figure 1 and the DNA sequence is shown in Figure 2. An open reading frame was found in this sequence which encodes a protein that is clearly homologous to $G_o\alpha$ sequences identified in other organisms. The $G_o\alpha$ cDNA has a 5' leader that is at least 90 base pairs in length which is about average for a eukaryotic mRNA (Kozak, 1984). Based on amino acid sequence homologies to other $G_o\alpha$ proteins the first ATG in the cDNA sequence is probably the site where translation begins. The sequence found at this position (GCCATGG) fits the optimal eukaryote translation initiation site (ACCATGG; Kozak, 1986) well. No sequences resembling the SL1- or SL2- spliced leader sequences which are posttranscriptionally spliced on certain *C. elegans* mRNAs (Krause and Hirsh, 1987; Huang and Hirsh, 1989) were found at the 5' end of the $G_o\alpha$ cDNA. However the $G_o\alpha$ cDNA sequence in Figure 2 is probably not a full length sequence since the size of the mRNA detected on a Northern blot is about 2.2 kilobase pairs (Figure 3). Thus, excluding the poly-A^+ tail, the sequence may be missing about 400 base pairs. The results of Southern blot analysis indicate that the *C. elegans* $G_o\alpha$ gene is present in a single copy in the *C. elegans* genome (not shown).

The predicted amino acid sequence of *C. elegans* $G_o\alpha$ is 87% identical to *Drosophila* $G_o\alpha$, 80% identical to *Xenopus* $G_o\alpha$, and 82% identical to rat, mouse, hamster, or bovine $G_o\alpha$ (Figure 4). The observation that *C. elegans* $G_o\alpha$ is more similar to *Drosophila* $G_o\alpha$ than to mammalian $G_o\alpha$ is in accord with the idea that nematodes are slightly more related to insects than to vertebrates in terms of evolution (Wood, 1988). Two different forms of $G_o\alpha$, called $G_o\alpha$-A and $G_o\alpha$-B, have been described from hamster (Hsu *et al.*, 1990) and mouse (M. Strathmann, personal communiciation). They differ in their sixth and seventh exons as a result of alternative splicing patterns. *C. elegans* $G_o\alpha$ is as different from $G_o\alpha$-A as from $G_o\alpha$-B. All of the $G_o\alpha$ proteins are 354 amino acids long. All of the motifs involved in GTP binding and hydrolysis (Lochrie and Simon, 1988) are highly conserved in the *C. elegans* $G_o\alpha$ sequence. Glycine-2, which has been identified as a site for myristylation on rat $G_o\alpha$ (Mumby *et al.*, 1990), is found. Amino acids which are

substrates for cholera toxin (Arg-179) and pertussis toxin (Cys-351) are also found in *C.elegans* $G_o\alpha$ at the same position as they are found in other $G_o\alpha$ proteins. The region of *C. elegans* $G_o\alpha$ that is most different from other $G_o\alpha$ sequences is in the region of amino acids 90-140 and 290-320. This would be expected since this region is also the most variable between any other pair of G protein α subunits.

A protein with a molecular weight of 40 kDal can be detected on immunoblots of total *C. elegans* extracts with an affinity-purified antibody to bovine $G_o\alpha$ (Figure 5). The intensity of the signal is the same in extracts from the wild-type N2 hermaphrodite strain as in extracts from a *him-5* strain which produces about 30% males. A protein of the same molecular weight can also be detected with an antipeptide antibody (OC1; McFadzean *et al.*, 1989) that was made against the sequence Ala-Asn-Asn-Leu-Arg-Gly-Cys-Gly- Leu-Phe (not shown). This sequence is found as the last 10 amino acids of all $G_o\alpha$ proteins except *Xenopus* $G_o\alpha$.

Based on the sequence of the $G_o\alpha$ cDNA insert two PCR primers were made (Tα33: TCGCTATTTTGCGCGCCA, position 337-354 and Tα34:AATATCTTGCTCAGTTGG, position 601-618) that correspond to a region that varies among G protein alpha subunits and thus would be expected to be $G_o\alpha$ specific. These were used to screen the six other bacteriophage isolates that were amplified by the PCR primers oMP19, oMP20/21 and oMP19, Tα29. All six were amplified by Tα33 and Tα34 implying that they are also $G_o\alpha$ cDNAs. One of these was subcloned and its 5' and 3' sequences were found in the sequence in Figure 2. Five other clones were isolated that were not amplified by Tα33, Tα34 or by oMP19, oMP20/21 and oMP19, Tα29. These hybridize strongly to the *gpa-1* probe and all five appeared to contain the same 2.2 kilobase pair Eco RI insert. The sequence of the insert of one of these clones was analyzed in several different regions. In total about 40% of the insert sequence was determined. However no sequences with homology to the *gpa-1* probe or to other G protein alpha subunits were found. No cDNAs encoding *gpa-1*, *gpa-2*, or *gpa-3* were isolated even after a high density screening of ten

million plaques.

Isolation of cosmid clones.

The PCR-derived fragments of *gpa-1*, *gpa-2*, and *gpa-3* and the $G_o\alpha$ cDNA insert were used to isolate cosmid clones. The number of cosmids that hybridizied to the *gpa-1*, *gpa-2*, *gpa-3* , and $G_o\alpha$ probes was three, one, three, and one, respectively. In addition one cosmid was isolated that hydridizes weakly to the *gpa-1* probe, but not to the *gpa-2* or *gpa-3* probes. The putative gene on this cosmid is referred to here as *gpa-1**. The $G_o\alpha$ gene is called *goa-1*. The identity of the *gpa-1*, *gpa-2*, and *gpa-3* cosmids was confirmed by PCR analysis using oMP19, oMP20/21 and oMP19, Tα29 as primers. Because of the presence of introns of variable lengths each genomic PCR product has a unique fragment size when amplified with these sets of primers. When the candidate *gpa-1*, *gpa-2*, and *gpa-3* cosmid clones were amplified by PCR using these primers, DNA fragments of the same size as the corresponding PCR-derived fragments were obtained. The cosmid that contains the putative *gpa-1** gene was not amplified by either set of primers. The candidate $G_o\alpha$ cosmid was screened by PCR using the oMP19, oMP20 and Tα33, Tα34 primer pairs. The size of the PCR product obtained by amplification of the *C. elegans* $G_o\alpha$ cDNA with the Tα33 and Tα34 primers is 282 base pairs. Since the mammalian $G_o\alpha$ gene has two introns between the amino acid sequences represented by these primers, the genomic fragment from the *C. elegans goa-1* gene would be expected to be larger than 282 base pairs. When the candidate $G_o\alpha$ cosmid clone was amplified in polymerase chain reactions using the Tα33, Tα34 and the oMP19, oMP20 primer pairs, the sizes of the DNA fragments that were obtained were about 1,400 and 500 base pairs, respectively. Southern blot analysis of the candidate $G_o\alpha$ cosmid and genomic DNA using the $G_o\alpha$ cDNA as a probe showed that both had several restriction fragments in common. The sizes of the PCR fragments and the restriction fragments observed on the Southern blot indicate the $G_o\alpha$ gene may be at least 2-3 kilobase pairs long.

Chromosome mapping.

The *gpa-1*, *gpa-2*, *gpa-3*, and *goa-1* cosmids mapped to unique positions on the *C. elegans* contig map. Each cosmid also appeared to contain a unique G protein α subunit gene as determined by the PCR analysis described above. Therefore it will be assumed that the map location of the cosmid defines the map location of the G protein α subunit gene it contains. The *gpa-1*, *gpa-2*, and *gpa-3* genes map to chromosome five. The *gpa-1* gene maps between the actin (*act-1*, *act-2*, and *act-3*) gene cluster and *myo-3* and is about 100 kilobases from the actin gene cluster. The *gpa-2* gene maps between *mec-1* and *his-2* in agreement with Silva and Plasterk (1990). The *gpa-3* gene maps between *hsp-16* and *her-1*. The *goa-1* gene maps to chromosome one between *unc-13* and *lin-10* and is about 50-100 kilobases pairs from *unc-13*. The *gpa-1** gene maps to the X chromosome between *clb-1* and *sup-10*. Each gene maps near a number of mutations. Although the positions of many of these mutations on the genetic and physical maps are ambiguous, some of these mutations have phenotypes which suggest that they could be in G protein α subunit genes.

For example the *gpa-1* gene may correspond to *daf-11*. *Daf-11* mutants are pleiotropically defective in chemotaxis and are also defective in dauer larvae formation (Riddle, 1988). The dauer larvae is a special developmental stage that arises when early stage larvae are starved. The process of dauer larvae formation is regulated by a hydrophobic pheromone. Epistasis analysis of dauer larvae constitutive and defective mutations has led to the conclusion that the *daf-11* gene product is required for a process early in the sensory processing pathway which results in dauer larvae formation. The *gpa-1* gene may also correspond to *egl-3*, *egl-9*, or *egl-47* genes that are required for egg laying. Egg laying in *C. elegans* is stimulated by serotonin (Horvitz *et al.*, 1982). In mammals serotonin receptors are members of a large family of proteins that are coupled to G proteins (Lubbert *et al.*, 1987; Fargin *et al.*, 1988; Julius *et al.*, 1988). The *che-11* and *che-12* genes which are required for chemotaxis may also map in the region of the *gpa-1* gene. *C. elegans* exhibits chemotaxis toward a number of compounds including cAMP

(Ward, 1973). In the slime mold *Dictyostelium* chemotaxis toward cAMP is regulated by a G protein-mediated signal transduction system (Kumagai *et al.*, 1989; Firtel *et al.*, 1989).

The *gpa-2* and *gpa-3* genes map near several recessive lethal mutations. Since some G proteins are involved in signaling pathways that are found in almost every cell, it is conceivable that their inactivity would be a lethal event.

The *goa-1* gene maps near *unc-55*, a mutation that affects the developmental fate of a specific class of motorneurons (Chalfie and Au, 1989), and *mec-6* and *mec-8*, genes that are required for mechanosensation. However these three genes appear to map on the side of *unc-13* which is positioned opposite t*goa-1*. In addition it is unlikely that G proteins mediate mechanosensation. The positions of the *egl-31, che-1, che-3*, and *che-13* genes are not very precise, but one of these might correspond to *goa-1*.

The *gpa-1** cosmid maps near the *lin-15* gene. *Lin-15* is involved in vulval development. Vulval development is known to involve an inductive signal that is sent from the anchor cell to vulval precursor cells (Sternberg, 1990). Mutations in *lin-15* cause the appearence of extra vulvas. The *gpa-1** cosmid rescues the multivulval defect of *lin-15* mutant animals (Min Hahn, personal communication). However subsequent analysis indicates the putative *gpa-1** gene and the DNA that rescues *lin-15* are distinct. *Gpa-1** also maps near several lethal mutations.

Gene sequences of *gpa-1* and *gpa-3*.

The restriction map, gene organization, and sequencing strategy of the *gpa-1* and *gpa-3* genes is shown in Figures 6 and 8. The DNA sequences of the *gpa-1* and *gpa-3* genes are shown in Figures 7 and 9. The *gpa-1* coding region is divided by 7 introns and the *gpa-3* coding region is divided by 5 introns. More introns may exist 5' or 3' to the coding regions especially in the 5' region of the *gpa-3* gene where several 3' consensus splice junctions are found (see below). Intron boundaries were unambiguously established by several criteria: homology to the amino acid sequences of known G protein α subunit proteins, the spacing between amino acids found in all G protein α subunits, the presence

of consensus splice junction sequences which are more extensive in *C. elegans* than in other eukaryotes (Emmons, 1988), the absence of stop codons in open reading frames, the absence of any sequence homology to G protein α subunits in putative introns, and conservation of intron position relative to mammalian G protein α subunit genes. In almost every case the 5' GTAAGTT and 3' TTTCAG consensus *C. elegans* mRNA splice sites (Emmons, 1988) or closely related variants are found at each intron/exon junction. The sequence GTAAG is found at all of the putative 5' splice junctions in the *gpa-1* gene and three out of five 5' junctions in the *gpa-3* gene. The sequence TTXCAG is found at most of the 3' junctions. An A at position -16 to -17 from the 3' splice junction border, that is conserved and is believed to be the site where lariat formation occurs, is also found in most of the putative introns in the *gpa-1* and *gpa-3* genes except for some of the smaller ones. Most of the introns are also AT-rich as observed for other *C. elegans* introns. All of the introns in the *gpa-1* and *gpa-3* genes are more than 65% AT except for introns 1, 5, and 7 in the *gpa-1* gene.

The sizes and positions of the introns in the *gpa-1*, *gpa-2*, and *gpa-3* genes are shown in Figure 10. The positions of four introns are conserved in the *gpa-1*, *gpa-2*, and *gpa-3* genes as well as in the mammalian $G_i\alpha$, $G_o\alpha$, and $T_r\alpha$ genes (Kaziro *et al.*, 1988; Raport *et al.*, 1989). These introns in the *gpa-1*, *gpa-2*, and *gpa-3* genes, that occur at the same position as an intron in a mammalian gene, not only occur at the same position within the amino acid sequence, but also at the same position within the codon. In addition these introns occur in amino acid sequences that are well conserved (i.e., those involved in guanine nucleotide binding). The introns in the *gpa-1*, *gpa-2*, and *gpa-3* genes, that are extra or missing relative to the mammalian $G_i\alpha$, $G_o\alpha$, and $T_r\alpha$ genes occur within the region of G protein α subunits which varies the most from one to the other.

A remarkable feature of the intron/exon organization of both genes is that the intron at the first position within the coding region is unusually large. The first intron of the *gpa-1* gene is 718 base pairs and the first intron of the *gpa-3* gene is 745 base pairs. The first

introns of the *gpb-1* gene (1,447 base pairs) and the *gpa-2* gene (1,020 base pairs) are also large. This is unusual because in *C. elegans* most of the introns are about 50 base pairs in length (Emmons, 1988). In fact the average length of the other introns in the *gpa-1*, *gpa-2*, *gpa-3*, and *gpb-1* genes is about 130 base pairs. The only extended region of sequence homology that has been detected between any pair of these four large introns is between the first introns of *gpa-2* and *gpa-3*. In this case a stretch of 85 base pairs with 63% nucleotide sequence identity was found. No open reading frames resembling G protein α subunit sequences have been detected within them, indicating that they do not contain alternatively spliced exons. At the other extreme the third intron in the *gpa-1* gene, which is 43 base pairs long, is among the smallest introns known.

Another interesting feature of the *gpa-1* gene is the presence of an almost perfect inverted repeat (see Figure 6C). One repeat element is located at positions 583-918 within the first intron and the other is at positions 2,012-2,350 within the fifth intron. The elements of the repeat are about 95% identical. The repeats are about 330 base pairs long and are about 1100 base pairs apart. This corresponds well to the average dimensions of repeats in *C. elegans* DNA reported by Emmons *et al.* (1980). Palindromic sequences are found at the boundaries of the repeats.

The nucleotide sequences 5' to the initiation codons of the *gpa-1* and *gpa-3* genes were examined for the presence of CAAT-box and TATA-box sequences, and the consensus transcription initiation site sequence (Breathnach and Chambon, 1981). The *gpa-1* gene has sequences at positions 50-55 (GCTAAT), 86-92 (TATATA), and 113 that fit the spacing constraints and consensus sequences for the CAAT-box, TATA-box, and transcription initiation site well. In addition, a consensus 3' splice junction acceptor site (TTTCAG) is found immediately preceeding the initiation codon. This may be a site where trans-splicing occurs. A very similar pattern of consensus CAAT-box, TATA-box, transcription initiation site, and 3' splice junction sequence is found in the 5' region of the *gpa-2* gene (Silva and Plasterk, 1990). The consensus site for translation initiation in the

gpa-1 gene (Kozak, 1986) is poorly conserved. The highly conserved A at position -3 from the initiation codon is a C.

The 760 base pairs of sequence that is available 5' to the *gpa-3* coding region has only one TATA sequence located at position 721-724 which is 35 base pairs from the translation initiation codon. No sequences resembling the CAAT-box sequence are readily found. However, the region from positions 6-52 has 68% nucleotide sequence identity to a region in the *gpa-2* gene that is between its putative CAAT-box and TATA-box. Based on this homology, positions 3-7 (GCATT) and 38-42 (TTTAT) may be the CAAT-box and TATA-box sequences of the *gpa-3* gene, respectively. Several potential 3' splice junctions are found at positions 83-90 (TTTTCCAG), 253-257 (TTTAG), 320-326 (TTTTCAG), 694-700 (TTTTTAG), and 730-736 (TTACAG). However no sequences resembling the consensus 5' splice junction sequence are found within this region. An A at position 757 that fits the Kozak consensus for eukaryotic translation initiation sites is found.

The sequences 5' to the initiation codon were compared to each other to detect sequence homologies that might, for example, be indicative of similarities in gene regulation. Other than the homology between *gpa-2* and *gpa-3* mentioned above, pairwise sequence comparisons of available sequences from the 5' regions of the *gpa-1*, *gpa-2*, *gpa-3*, and *gpb-1* genes did not reveal any striking homologies. The open reading frames on both strands that are found 5' and 3' to the coding regions of the *gpa-1* and *gpa-3* genes were compared to the GenBank database, but no significant homologies were detected.

The available sequences 3' to the termination codons of the *gpa-1* and *gpa-3* genes were examined for an AATAAA consensus polyadenylation signal sequence (Breathnach and Chambon, 1981). This sequence was found in the 3' region of the *gpa-1* gene at positions 3,340-3,345. It was not found in the *gpa-3* 3' region, but minor variants which might be functional are found at positions 2,825-2,830 (AATACA) and 2,891-2,896 (AATATA).

The predicted amino acid sequences of the *gpa-1*, *gpa-2*, and *gpa-3* proteins are compared in Figure 11. The *gpa-1* sequence is 41% identical to the *gpa-2* sequence. The

gpa-1 sequence is 48% identical to the *gpa-3* sequence. The *gpa-2* sequence is 58% identical to the*gpa-3* sequence. The *gpa-1*, *gpa-2*, and *gpa-3* proteins are 49%, 46%, and 53% identical to the *C. elegans* $G_o\alpha$ sequence, respectively. All three *gpa* proteins have the five regions implicated in binding guanine nucleotides (Lochrie and Simon, 1988). All three have the consensus myristylation signal Met-Gly-X-X-X-Ser at the amino terminus (Towler *et al.*, 1988). All three have the arginine (Arg-179 in *gpa-1*, Arg-181 in *gpa-2*, and Arg-179 in *gpa-3*) that is the substrate amino acid for cholera toxin in other G protein α subunits. The *gpa-1* and *gpa-3* proteins have the cysteine (Cys-354 in *gpa-1* and Cys-351 in *gpa-3*) four amino acids from the carboxyl terminus that is the substrate amino acid for pertussis toxin in other G protein α subunits. However the *gpa-2* protein has a serine at this position (Silva and Plasterk, 1990). The presence of cholera and pertussis toxin substrates in *C. elegans* has been reported (Van der Voorn *et al.*, 1990), but the effects of these toxins on *C. elegans* physiology has not been well studied. The *gpa-3* protein has a cysteine instead of a serine in the highly conserved sequence GAGESGK found at the amino terminus. The *gpa-2* protein also has a cysteine at this position. These are the only two known G protein α subunits that have a cysteine at this position (Silva and Plasterk, 1990). The sequence of $G_z\alpha$ also difers in this region from other G protein α subunits. It has the sequence GTSNSGK. These differences found in $G_z\alpha$ are thought to be responsible for its very low GTPase activity (Casey *et al,* 1990), since this region is close to the phosphates of GDP in the crystalline structure of the analogous *ras* proteins (Milburn *et al.*, 1990), and since mutations in this region of $G_s\alpha$ reduce its GTPase activity (Graziano and Gilman, 1989). Thus *gpa-2* and *gpa-3* may have a GTPase activity that differs from *gpa-1*.

DISCUSSION

We have determined the sequence of a *C. elegans* $G_o\alpha$ protein and have found that it is very similar to $G_o\alpha$ proteins found in other species. Therefore it is likely that *C. elegans* $G_o\alpha$ is similar in function to other $G_o\alpha$ proteins. $G_o\alpha$ is found predominantly in neurons.

However the signal transduction pathways it functions in are not well delineated. The presence of a $G_o\alpha$ homolog in *C. elegans* may provide new approaches for understanding the role of $G_o\alpha$ in the function of the nervous system.

The results of our Southern blot analysis indicates that the *C. elegans goa-1* gene is single copy and only one size class of mRNA is observed on a Northern blot using the *C. elegans* $G_o\alpha$ cDNA as a probe. Therefore it is likely that the protein detected on the immunoblot is the same as that encoded by the $G_o\alpha$ cDNA described here and not some other closely related G protein α subunit. Thus the anti-bovine $G_o\alpha$ antibody may be a useful tool for localization of the *C. elegans* $G_o\alpha$ protein by immunofluorescence. However the possibility that there may be multiple $G_o\alpha$ proteins can not be excluded. Multiple species of $G_o\alpha$ which arise as a result of alternative mRNA splicing patterns have been found in *Drosophila* (deSousa *et al.*, 1989; Thambi *et al.*, 1989; Yoon *et al.*, 1989) and rodents (Hsu *et al.*, 1990; M. Strathmann, personal communication). The sequence of the *goa-1* gene may shed light on whether there are also multiple $G_o\alpha$ proteins in *C. elegans*.

In contrast to *C. elegans* $G_o\alpha$, the *gpa-1*, *gpa-2*, and *gpa-3* proteins can not be classified according to their relatedness to other G protein α subunits. They are 40-50% identical to any other non-*C. elegans* G protein α subunit sequence. G protein α subunits that are over 80% identical are generally considered to be in the same class (Lochrie and Simon, 1989) and can activate the same effector proteins (Gillespie and Beavo, 1988; Yatani *et al.*, 1988). G protein α subunits that are as little as 60-70% identical (e.g., $G_o\alpha$, $G_i\alpha$ and $T_r\alpha$) can interact with some of the same receptors, although heterologous interactions are somewhat less efficient than homologous ones (Kahano *et al.*, 1984; Cerione *et al.*, 1986). However G proteins in this class have not been found to regulate the same effector proteins. G protein α subunits that are 40-60% identical have not been observed to interact with the same receptors or effectors and are generally considered to be distinct in function. Therefore *gpa-1*, *gpa-2*, and *gpa-3* are likely to have separate

functions. This does not exclude the possibility that they may operate in signaling systems similar to those in mammals since the receptors and effectors that *gpa-1*, *gpa-2* , and *gpa-3* interact with may also be different in sequence from their mammalian counterparts.

Several interesting features were observed in the noncoding regions of the *gpa-1* and *gpa-3* genes. One is that the first introns within the coding regions are unusually large compared to other introns in *C. elegans*. Curiously two other *C. elegans* G protein genes, *gpa-2* and *gpb-1*, also have large introns at this position. These introns may serve an important function. However the lack of significant linear sequence homology between them indicates that their length or structural features may be more important than their sequence. Another interesting feature is the inverted repeats found in the *gpa-1* gene. These may also have a function although it is not known at the present what that may be. About 2,400 inverted repeats exist in the *C. elegans* genome. These are organized into hundreds of families which usually contain less than ten members each. The inverted repeats found in the *gpa-1* gene are unusual in that they are found in introns rather than between genes. They may influence *gpa-1* gene expression or RNA processing. Because of their position within the gene, if the inverted repeats in *gpa-1* were to invert *in vivo* they would disrupt the coding region of the *gpa-1* protein. The sequence organization of the 5' region of the *gpa-1* gene indicates that the *gpa-1* mRNA may undergo trans-splicing. The significance of trans-splicing is not known. In some cases only one of two almost identical mRNAs will undergo trans-splicing. It would be interesting to determine if *gpa-1* is trans-spliced and if so whether it is splicedwith one of the two known trans-spliced leaders or a different one. Finally the sequence homology between the 5' regions of *gpa-2* and *gpa-3* genes suggests that some aspect of their regulation may be similar. The 5' region of *gpa-3* is similar to mammalian 5' regions which also lack easily detected TATA-box and CAAT-box sequences. Instead the mammalian genes have numerous GC-boxes which are hypothesized to be binding sites for the Sp1 transcription factor (Itoh *et al.*, 1988; Kozasa *et al.*, 1988; Weinstein *et al.*, 1988; Matsuoka *et al.*, 1990).

The sequences of the *gpa-1*, *gpa-2* , and *gpa-3* proteins provide no clues about their function, but several approaches are available for investigating their function. One approach described in this study is the mapping of *gpa* genes near known mutations with visible phenotypes. However further work will need to be done to determine which of these mutations, if any, correspond to mutations in G protein α subunit genes. This can be done by transformation of mutant strains with the cosmid clones isolated in the work described here. However, it is possible that none of the mutations that map near a given *gpa* gene will be in that gene since the genetic map is not saturated with mutations. If this is the case then recessive mutations can be produced by Tc1 transposon insertion and dominant mutations can be generated by transformation with mutants of the *gpa* genes constructed *in vitro*. Based on studies of mutant G protein α subunits found in human tumor cells (Landis, *et al* ., 1989) and the biochemical properties of G protein α subunits synthesized in *E. coli* (Graziano and Gilman, 1989; Freissmuth and Gilman, 1989) and mammalian cells (Masters *et al.*, 1989) several amino acid changes in the *gpa* sequences might be expected to result in a dominant phenotype. These include the glutamine in the sequence DVGGQ and the arginine that is a cholera toxin substrate. One dominant phenotype that might be observed after transformation with such mutant *gpa* genes is suggested by studies of the *C. elegans tpa-1* gene (Tabuse *et al.*, 1989). The *tpa-1* gene encodes protein kinase C. The protein kinase C pathway is regulated by G proteins in vertebrates. G proteins activate phospholipase C which hydrolyzes phosphatidylinositol diphosphate to inositol triphosphate and diacylglycerol. Inositol triphosphate releases calcium from the endoplasmic reticulum and diacylglycerol activates protein kinase C. Phorbol esters mimic diacylglycerol and persistantly activate protein kinase C. Phorbol esters cause *C. elegans* to become uncoordinated and stunt development. Dominant activation of the G protein that regulates the protein kinase C pathway in *C. elegans* could have a similar effect. Tabuse *et al.* (1989) have shown that the target of phobol esters in *C. elegans* is the protein kinase C encoded by the *tpa-1* gene. The location and timing of

gpa-1, *gpa-2*, and *gpa-3* gene or protein expression may also provide a clue about their function. Our attempts to detect mRNA on Northern blots or isolate cDNA clones corresponding to *gpa-1*, *gpa-2*, or *gpa-3* have not been successful. The reason for this is unknown, but there are several possibilities. 1) The mRNAs encoding these proteins could be too large to be efficiently cloned in λgt10. 2) These G proteins could be male specific. The frequency of male animals in a hermaphrodite population is about 1 in 700. Thus male-specific mRNA would be underrepresented in the library we screened. 3) Another possibility is that the mRNAs are rare because they are unstable or expressed in one or a few cells or are expressed only during a narrow period of development.

There may be G protein genes in *C. elegans* other than those reported to date. A *C. elegans* homolog of the mammalian $G_S\alpha$ gene (*gsa-1*) has been isolated (Yasumi Ohshima, personal communication). This gene could not have been isolated by the methods used in this paper since $G_S\alpha$ is too divergent from other G proteins. From the PCR analysis reported here there is evidence that other G protein α subunit genes may exist, since not all of the PCR products can be accounted for by the clones that are available. The 1,000 base pair oMP19, oMP20 PCR product and the 300 base pair oMP19, oMP21 PCR product may represent other G protein α subunit genes. The 525 base pair oMP19, oMP20 PCR product is probably from the *goa-1* gene since the cosmid has a similar sized fragment. Similarly the 270 base pair oMP19, oMP20 PCR product and the 260 base pair oMP19, oMP21 PCR product probably contain fragments of the *gpa-1* and *gpa-3* genes, respectively. However it is also possible that a given size class of PCR product may contain more than one G protein α subunit sequence as observed by Strathmann *et al.* (1989). The identity of the putative *gpa-1** gene is not known at this time. However the restriction map of the *gpa-1** gene is different from the *gsa-1* gene (Yasumi Ohshima, personal communication). In addition degenerate oligonucleotide primers specific for mammalian $G_S\alpha$ do not amplify DNA from the *gpa-1** cosmid in PCR reactions. This rules out the possibility that *gpa-1** might encode $G_S\alpha$. *C. elegans* may have other G proteins

that have so far been found only in mammals, such as $G_z\alpha$ and Gα11–16 (M. Strathmann, T. Wilke, and T. Amatruda, personal communication), or G proteins that have been found in both *Drosophila* and mammals, such as $G_i\alpha$ (Provost *et al.*, 1988) and $G_q\alpha$ (M. Strathmann, personal communication). Conversely other organisms may have homologs of the *C. elegans gpa-1*, *gpa-2*, and *gpa-3* proteins. Since G proteins are always found as heterotrimers it is likely that G protein γ subunit genes also exist in *C. elegans*, but there have been no reports of their existence to date.

CONCLUSION

C. elegans has at least five G proteins. Two of these (G_o and G_s) are identifiable homologs of mammalian G proteins whereas three (*gpa-1*, *gpa-2*, and *gpa-3*) appear to be unique to *C. elegans*. All multicellular organisms that have been examined have G_o and G_s. However unicellular organisms such as yeast and *Dictyostelium* do not have G_o and G_s. Instead they have G proteins with novel sequences. Therefore it may be that G_o and G_s arose during the development of multicellular organisms and have been highly conserved across phylogeny because they have fundamental roles in intercellular signalling. On the other hand the G proteins that are not well conserved may have functions specific to the organism in which they are found.

ACKNOWLEDGEMENTS

We thank Mike Strathmann for advice on PCR reactions and providing materials for the $\gamma\delta$ transposon-based sequencing method, the Caltech Microchemical Facility for oligonucleotide synthesis, Andrea Holbocke for *C. elegans* genomic DNA, Alan Coulson and John Sulston for gene mapping data, Eva Neer and Graeme Milligan for $G_o\alpha$ antibodies, Stuart Kim and Russel Hill for the *C. elegans* cDNA library, and Yasumi Ohshima and Ron Plasterk for generous communication of results prior to publication.

REFERENCES

Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349-383.

Buss, J. E., Mumby, S., Casey, P. J., Gilman, A. G. and Sefton, B. (1987). Myristoylated α subunits of guanine nucleotide-binding regulatory proteins. Proc. Natl. Acad. Sci. U. S. A. 84, 7493-7497.

Caron, M. G. (1989). The guanine nucleotide regulatory protein-coupled receptors for nucleotidsides, nucleotides, amino acids, and amine neurotransmitters. Curr. Opinion in Cell Biol. 1, 159-166.

Casey, P., Fong, H. K. W., Simon, M. I. and Gilman, A. G. (1990). G_z, a guanine nucleotide-binding protein with unique biochemical properties. J. Biol. Chem. 265, 2383-2390.

Cerione, R. A., Regan, J. W., Nakata, H., Codina, J., Benovic, J. L., Gierschik, P., Somers, R., Spiegel, A. M., Birnbaumer, L., Lefkowitz, R. J. and Caron, M. G. (1986). Functional reconstitution of the α_2-adrenergic receptor with guanine nucleotide regulatory proteins in phospholipid vesicles. J. Biol. Chem. 261, 3901-3909.

Chalfie, M. and Au, M. (1989). Genetic control of differentiation of the *Caenorhabditis elegans* touch receptor neurons. Science 243, 1027-1033.

Chalfie, M., Sulston, J., White, J. G., Southgate, E., Thompson, J. N. and Brenner, S. (1985). The neural circuit for touch sensitivity in *Caenorhabitis elegans*. J. Neurosci. 5, 956-964.

Chalfie, M. and White, J. (1988). The nervous system. In: The Nematode *Caenorhabditis elegans*. ed., W. B. Wood. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp. 337-391.

Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986). Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U. S. A. 83, 7821-7825.

deSousa, S. M., Hoveland, L. L., Yarfitz, S. and Hurley, J. B. (1989). The *Drosophila* $G_o\alpha$-like G protein gene produces multiple transcripts and is expressed in the nervous system and in ovaries. J. Biol. Chem. 264, 18544-18551.

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

Emmons, S.W. (1988). The genome. In: The Nematode *Caenorhabditis elegans*. ed., W. B. Wood. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp. 47-80.

Emmons, S. W., Rosenzweig, B. and Hirsh, D. (1980). Arrangement of repeated sequences in the DNA of the nematode *Caenorhabditis elegans*. J. Mol. Biol. 144, 481-500.

Fargin, A., Raymond, J. R., Lohse, M. J., Kobilka, B. K., Caron, M. G. and Lefkowitz, R. J. (1988). The genomic clone G-21 which resembles a β-adrenergic receptor sequence encodes the 5-HT_{1A} receptor. Nature 335, 358-360.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction fragments to high specific activity. Anal. Biochem. 132, 6-13.

Fino Silva, I. F. and Plasterk, R. H. A. (1990). Characterization of a G-protein α subunit gene from the nematode *Caenorhabditis elegans*. submitted to J. Mol. Biol.

Firtel, R., van Haastert, P. J. M., Kimmel, A. R. and Devreotes, P. (1989). G protein linked signal transduction pathways in development: *Dictyostelium* as an experimental system. Cell 58, 235-239.

Freissmuth, M. and Gilman, A. G. (1989). Mutations of $G_{s\alpha}$ designed to alter the reactivity of the protein with bacterial toxins. Substitutions at arg^{187} result in loss of GTPase activity. J. Biol. Chem. 264, 21907-21914.

Gabrion, J., Brabet, P., Dao, B. N. T., Homburger, V., Dumuis, A., Sebben, M., Rouot, B. and Bockaert, J. (1989). Ultrastructural localization of the GTP-binding protein G_O in neurons. Cell Signaling 1, 107-123.

Gillespie, P. G. and Beavo, J. A. (1988). Characterization of a bovine cone photoreceptor phosphodiesterase purified by cyclic GMP-Sepharose chromatography. J. Biol. Chem. 263, 8133-8141.

Gilman, A. G. (1987) G proteins: transducers of receptor-generated signals. Ann. Rev. Bioch. 56, 615-650.

Graziano, M. P. and Gilman, A. G. (1989). Synthesis in *Escherichia coli* of GTPase-deficient mutants of $G_{s\alpha}$. J. Biol. Chem. 264, 15475-15482.

Gunderson, R. E. and Devreotes, P. N. (1990). In vivo receptor-mediated phosphorylation of a G protein in *Dictyostelium*. Science 248, 591-593.

Harris-Warrick, R. M., Hammond, C., Paupardin-Tritsch, D., Homburger, V., Rouot, B., Bockaert, J. and Gerschenfeld, H. M. (1988). An α_{40} subunit of a GTP-binding protein immunologically related to G_O mediates a dopamine-induced decrease of Ca^{2+} current in snail neurons. Neuron 1, 27-32.

Hescheler, J., Rosenthal, W., Trautwein, W. and Schultz, G. (1987). The GTP-binding protein, G_O, regulates neuronal calcium channels. Nature 325, 445-447.

Hodgkin, J., Edgley, M., Riddle, D. L. and Albertson, D. G. (1988). Genetics. In: The Nematode *Caenorhabditis elegans*. ed., W. B. Wood. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp. 491-586.

Holmes, D. S. and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114, 193-197.

Horvitz, H. R., Chalfie, M., Sulston, J. and Evans, P. (1982). Serotonin and octopamine in *Caenorhabditis elegans*. Science 216, 1012-1014.

Hsu, W. H., Rudolph, U., Sanford, J., Bertrand, P., Olate, J., Nelson, C., Moss, L. G., Boyd, III, A. E., Codina, J. and Birnbaumer, L. (1990). Molecular cloning of a novel splice variant of the α subunit of the mammalian G_O protein. J. Biol. Chem. 265, 11220-11226.

Huang, X.-Y. and Hirsh, D. (1989). A second trans-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U. S. A. 86, 8640-8644.

Itoh, H., Toyama, R., Kozasa, T., Tsukamoto, T., Matsuoka, M. and Kaziro, Y. (1988). Presence of three distinct molecular species of G_i protein α subunit. Structure of rat cDNAs and human genomic DNAs. J. Biol. Chem. 263, 6656-6664.

Jones, D.T. and Reed, R.R. (1987). Molecular cloning of five GTP-binding protein cDNA species from rat olfactory neuroepithelium. J. Biol. Chem. 262, 14241-14249.

Jones, T. L. Z., Simonds, W. F., Merendino, Jr., J. J., Brann, M. R. and Spiegel, A. M. (1990). Myristoylation of an inhibitiory GTP-binding protein α subunit is essential for its membrane attachment. Proc. Natl. Acad. Sci. U. S. A. 87, 568-572.

Julius, D., MacDermott, A. B., Axel, R. and Jessell, T. M. (1988). Molecular cloning of a functional cDNA encoding the serotonin 1c receptor. Science 241, 558-564.

Kahano, Y., Tsai, S.-C., Adamik, R., Hewlett, E. L., Moss, J. and Vaughan, M. (1984). Rhodopsin-enhanced GTPase activity of the inhibitory GTP-binding protien of adenylate cyclase. J. Biol. Chem. 259, 7378-7381.

Kaziro, Y., Itoh, H., Kozasa, T., Toyama, R., Tsukamoto, T., Matsuoka, M., Nakafuku, M., Obara, T., Takagi, T. and Hernandez, R. (1988). Structures of the genes coding for G-protein α subunits from mammalian and yeast cells. Cold Spring Harbor Sym. Quant. Biol. 53, 209-220.

Kozasa, T., Itoh, H., Tsukamoto, T. and Kaziro, Y. (1988). Isolation and characterization of the human $G_s\alpha$ gene. Proc. Natl. Acad. Sci. U. S. A. 85, 2081-2085.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucl. Acid Res. 12, 857-872.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44, 283-292.

Krause, M. and Hirsh, D. (1987). A trans-spliced leader sequence on actin mRNA in *C. elegans*. Cell 49, 753-761.

Kumagai, A., Pupillo, M., Gunderson, R., Miake-Lye, R., Devreotes, P. and Firtel, R. A. (1989). Regulation and function of Gα protein subunits in *Dictyostelium*. Cell 57, 265-275.

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

Landis, C. A., Masters, S. B., Spada, A. M., Bourne, H. R. and Vallar, L. (1989). GTPase inhibiting mutations activate the α chain of G_S and stimulate adenylate cyclase in human pituitary tumours. Nature 340, 692-696.

Lipman, D. J. and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. Science 227, 1435-1441.

Lochrie, M. A. and Simon, M. I. (1988). G protein multiplicity in eukaryotic organisms. Biochemistry 27, 4957-4965.

Lubbert, H., Hoffman, B. J., Snutch, T. P., Van Dyke, T., Levine, A. J., Hartig, P. R., Lester, H. A. and Davidson, N. (1987). cDNA cloning of a serotonin 5-HT_{1C} receptor by electrophysiological assays of mRNA-injected *Xenopus* oocytes. Proc. Natl. Acad. Sci. U. S. A. 84, 4332-4336.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Masters, S. B., Miller, R. T., Chi, M.-H., Chang, F.-H., Beiderman, B., Lopez, N. G. and Bourne, H. R. (1989). Mutations in the GTP-binding site of $G_{s\alpha}$ alter stimulation of adenylyl cyclase. J. Biol. Chem. 264, 15467-15474.

Matsuoka, M., Itoh, H. and Kaziro, Y. (1990). Characterization of the human gene for $G_X\alpha$, a pertussis toxin-insensitive regulatory GTP-binding protein. J. Biol. Chem. 265, 13215-13220.

McFadzean, I., Mullaney, I., Brown, D. A. and Milligan, G. (1989). Antibodies to the GTP binding protein, G_O, antagonize noradrenaline-induced calcium current inhibition in NG108-15 hybrid cells. Neuron 3, 177-182.

Milburn, M. V., Tong, L., deVos, A. M., Brunger, A., Yamaizumi, Z., Nishimura, S. and Kim, S.-H. (1990). Molecular switch for signal transduction: Structural differences between active and inactive forms of the protooncogenic *ras* proteins. Science 247, 939-945.

Moriarty, T. M., Padrell, E., Carty, D. J., Omri, G., Landau, E. M. and Iyengar, R. (1989). G_O protein as signal transducer in the pertussis toxin-sensitive phosphatidylinositiol pathway. Nature 343, 79-82.

Moss, J. and Vaughan, M. (1988). ADP-ribosylation of guanyl nucleotide-binding regulatory proteins by bacterial toxins. Adv. Enzymol. 61, 303-379.

Mumby, S. M., Heukeroth, R. O., Gordon, J. I. and Gilman, A. G. (1990). G-protein α-subunit expression, myristoylation, and membrane association in COS cells. Proc. Natl. Acad. Sci. U. S. A. 87, 728-732.

Neer, E. J., Lok, J. M. and Wolf, L. G. (1984). Purification and properties of the inhibitory guanine nucleotide regulatory unit of brain adenylate cyclase. J. Biol. Chem. 259, 14222-14229.

Olate, J., Jorquera, H., Purcell, P., Codina, J., Birnbaumer, L. and Allende, J. E. (1989). Molecular cloning and sequence determination of a cDNA coding for the α–subunit of a G_O-type protein of *Xenopus laevis* oocytes. FEBS Lett. 244, 188-192.

Proudfoot, N. J. and Brownlee, G. G. (1976). 3' Non-coding region sequences in eukaryotic messenger RNA. Nature 263, 211-214.

Provost, N. M., Somers, D. E. and Hurley, J. B. (1988). A *Drosophila melanogaster* G protein α subunit gene is expressed primarily in embryos and pupae. J. Biol. Chem. 263, 12070-12076.

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

Raport, C. J., Dere, B. and Hurley, J. B. (1989). Characterization of the mouse rod transducin α subunit gene. J. Biol. Chem. 264, 7122-7128.

Riddle, D. (1988). Dauer larvae. In: The Nematode *Caenorhabditis elegans*. ed., W. B. Wood. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp. 393-412.

Sagi-Eisenberg, R. (1989). GTP-binding proteins as possible targets for protein kinase C action. Trends Biochem. Sci. 14, 355-357.

Schmidt, C. J., Garen-Fazio, S., Chow, Y.-K. and Neer, E. (1989). Neuronal expression of a newly identified *Drosophila melanogaster* G protein α_o subunit. Cell Regulation 1, 125-134.

Senogles, S.E., Spiegel, A. M., Padrell, E., Iyengar, R. and Caron, M. (1990). Specificity of receptor G-protein interactions. Discrimination of G_i subtypes by the D_2 dopamine receptor in a reconstituted system. J. Biol. Chem. 265, 4507-4514.

Serventi, I. M., Price, S. R., Murtagh, J. J. and Moss, J. (1990). Structure and function of $G_{o\alpha}$. In: ADP-Ribosylating Toxins and G Proteins, eds. J. Moss and M. Vaughan, American Society for Microbiology, Washington, D. C. pp. 325-348.

Sternberg, P.W. (1990). Genetic control of cell type and pattern formation in *Caenorhabditis elegans*. Adv. Genetics 27, 63-116.

Sternweis, P. C. and Robishaw, J. D. (1984). Isolation of two proteins with high affinity for guanine nucleotides from membranes of bovine brain. J. Biol. Chem. 259, 13806-13813.

Strathmann, M. P., Wilke, T. M. and Simon, M. I. (1989). Diversity of the G-protein family: sequences from five additional α subunits in mouse. Proc. Natl. Acad. Sci.. U. S. A. 86, 7407-7409.

Strittmatter, S. M., Valenzuela, D., Kennedy, T. E., Neer, E. and Fishman, M. (1990). G_o is a major protein of growth cone protein subject to regulation by GAP-43. Nature 344, 836-841.

Sulston, J. (1988) Cell lineage. In: The Nematode *Caenorhabditis elegans*. ed., W. B. Wood. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp. 123-156.

Sulston J. and Hodgkin, J. (1988). Methods. In: The Nematode *Caenorhabditis elegans*. ed., W. B. Wood. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp. 587-606.

Sulston, J., Horvitz, H. R. and Kimble, J. (1988). Cell lineage. In: The Nematode *Caenorhabditis elegans*. ed., W. B. Wood. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp.457-490.

Tabuse, Y., Nishiwaki, K. and Miwa, J. (1989). Mutations in a protein kinase C homolog confer phorbol ester resistance on *Caenorhabditis elegans*. Science 243, 1713-1716.

Thambi, N. C., Quan, F., Wolfgang, W. J., Spiegel, A. and Forte, M. (1989). Immunological and molecular characterization of $G_{o\alpha}$-like proteins in the *Drosophila* central nervous system. J. Biol. Chem. 264, 18552-18560.

Towbin, H., Staehlin, T. and Gordon, J. (1979). Electorphoretic transfer of proteins from polyacrylamide gels to nitrocellulsoe sheets: procedure and some applications. Proc. Natl. Acad. Sci. U. S. A. 76, 4350-4354.

Towler, D. A., Gordon, J. I., Adams, S. P., Glaser, L. (1988). The biology and enzymology of eukaryotic protein acylation. Ann. Rev. Biochem. 57, 69-99.

Ueda, H., Uno, S., Harada, J., Kobayashi, I., Katada, T., Ui, M. and Satoh, M. (1990). Evidence for receptor-mediated inhibition of intrinsic activity of GTP-binding proteins, G_i1 and G_i2, but not G_o in reconstitution experiments. FEBS Lett. 266, 178-182.

Van der Voorn, L., Gebbink, M., Plasterk, R. H. A. and Ploegh, H. L. (1990). Characterization of a G-protein β-subunit gene from the nematode *Caenorhabditis elegans*. J. Mol. Biol. 213, 17-26.

Van Dongen, A. M. J., Codina, J., Olate, J., Mattera, R., Joho, R., Birnbaumer, L. and Brown, A. M. (1988). Newly identified brain potassium channels gated by the guanine nucleotide binding protein G_o. Science 242, 1433-1437.

Van Meurs, K. P., Angus, C. W., Lavu, S., Kung, H.-F., Czarnecki, S. K., Moss, J. and Vaughan, M. (1987). Deduced amino acid sequence of bovine retinal $G_{o\alpha}$: homologies to other guanine nucleotide-binding proteins. Proc. Natl. Acad. Sci. U. S. A. 84, 3107-3111.

Yatani, A., Mattera, R., Codina, J., Graf, R., Okabe, K., Padrell, E., Iyengar, R., Brown, A. and Birnbaumer, L. (1988). The G protein-gated atrial K^+ channel is stimulated by three distinct $G_i\alpha$-subunits. Nature 336, 680-682.

Ward, S. (1973). Chemotaxis by the nematode *Caenorhabditis elegans*: Identification of attractants and analysis of the response by the use of mutants. Proc. Natl. Acad. Sci. U. S. A. 70, 817-821.

Weinstein, L. S., Spiegel, A. M. and Carter, A. D. (1988). Cloning and characterization of the human gene for the α-subunit of G_{i2}, a GTP-binding signal transduction protein. FEBS Lett. 232, 333-340.

White, J., Southgate, E. and Durbin, R. (1988). Neuroanatomy. In: The Nematode *Caenorhabditis elegans*. ed., W. B. Wood. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp.433-456.

White, J., Southgate, E., Thomson, J. N. and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. Phil. Trans. R. Soc. Lond. B 314, 1-340.

Wickens, M. and Stephenson, P. (1984). Role of the conserved AAUAAA sequence: Four AAUAAA point mutations prevent messenger RNA 3' end formation. Science 226, 1045-1051.

Willet, J. D. (1980). Control mechanisms in nematodes. In: Nematodes as biological models. Volume 1: Behavioral and developmental models. ed. B. M. Zuckerman. Academic Press, New York. pp. 197-225.

Wood, W. B. (1988). In: The Nematode *Caenorhabditis elegans*. ed., W. B. Wood. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. pp. 1-16.

Worley, P. F., Baraban, J. M., Van Dop, C., Neer, E. J. and Snyder, S. H. (1986). G_O, a guanine nucleotide-binding protein: immunohistochemical localization in rat brain resembles distribution of second messenger systems. Proc. Natl. Acad. Sci. U. S. A. 83, 4561-4565.

Yoon, J., Shortridge, R. D., Bloomquist, B. T., Schneuwly, S., Perdew, M.H. and Pak, W. L. (1989). Molecular characterization of *Drosophila* gene encoding $G_O\alpha$ subunit homolog. J. Biol. Chem. 264, 18536-18543.

Figure 1. Restriction map and DNA sequencing strategy of *C. elegans* $G_o\alpha$ cDNA. The *C. elegans* $G_o\alpha$ cDNA is represented as a horizontal line. The $G_o\alpha$ coding region is represented as a thicker horizontal line. The arrows indicate the direction and extent of DNA sequences determined. Abbreviations for restriction enzyme recognition sites are: Ba, Bam HI; Bc, Bcl I; Cl, Cla I; E, Eco RI; H2, Hinc II; Nc, Nco I; Ns, Nsi I; S, Sph I; T, Taq I. Other abbreviations used are: N, amino terminus; C, carboxyl terminus.

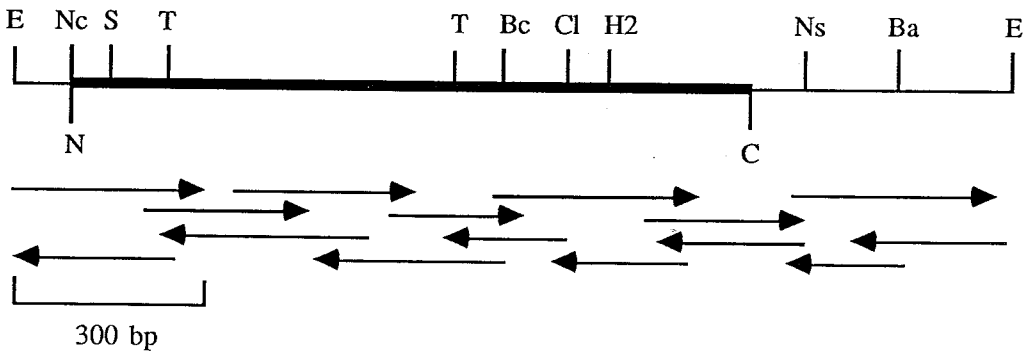

E
Nc
S
T
T
Bc
Cl
H2
Ns
Ba
E
N
C
300 bp

Figure 2. Sequence of the *C. elegans* $G_o\alpha$ cDNA. Amino acid abbreviations used are: Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine.

```
CGAGCTGCACCACATTACAGTGAGTGAGTAGAGGATATCAAGTGGAGACCGCTACCGGGG     60

CATAGGTCCACCGTTCATCAACTCTAGGTGCCATGGGTTGTACCATGTCACAGGAAGAGC    120
                                    MetGlyCysThrMetSerGlnGluGluA     10

GTGCCGCTCTTGAAAGATCACGAATGATTGAGAAAAATCTTAAAGAAGACGGCATGCAAG    180
rgAlaAlaLeuGluArgSerArgMetIleGluLysAsnLeuLysGluAspGlyMetGlnA     30

CGGCAAAAGATATCAAACTGCTGCTACTTGGTGCAGGAGAATCAGGAAAATCGACTATTG    240
laAlaLysAspIleLysLeuLeuLeuLeuGlyAlaGlyGluSerGlyLysSerThrIleV     50

TAAAACAGATGAAAATTATTCACGAATCAGGATTCACAGCAGAAGACTACAAACAGTACA    300
alLysGlnMetLysIleIleHisGluSerGlyPheThrAlaGluAspTyrLysGlnTyrL     70

AGCCGGTTGTCTACAGTAACACGGTTCAATCATTGGTCGCTATTTTGCGCGCCATGAGCA    360
ysProValValTyrSerAsnThrValGlnSerLeuValAlaIleLeuArgAlaMetSerA     90

ACTTAGGCGTTTCATTTGGTTCGGCTGACAGAGAGGTAGATGCAAAATTAGTGATGGATG    420
snLeuGlyValSerPheGlySerAlaAspArgGluValAspAlaLysLeuValMetAspV    110

TGGTGGCACGAATGGAGGACACAGAGCCATTCTCAGAAGAATTGCTCAGTTCAATGAAAC    480
alValAlaArgMetGluAspThrGluProPheSerGluGluLeuLeuSerSerMetLysA    130

GGTTGTGGGGAGACGCAGGTGTACAGGATTGTTTCTCAAGGAGTAACGAGTATCAATTGA    540
rgLeuTrpGlyAspAlaGlyValGlnAspCysPheSerArgSerAsnGluTyrGlnLeuA    150

ATGATTCAGCCAAATATTTCCTTGACGACCTGGAAAGGTTAGGAGAGGCAATATATCAAC    600
snAspSerAlaLysTyrPheLeuAspAspLeuGluArgLeuGlyGluAlaIleTyrGlnP    170

CAACTGAGCAACATATTCTCCGAACTCGTGTCAAAACAACTGGTATTGTTGAAGTTCACT    660
roThrGluGlnHisIleLeuArgThrArgValLysThrThrGlyIleValGluValHisP    190

TCACATTCAAAAATCTCAATTTCAAATTGTTCGATGTGGGAGGTCAAAGATCAGAAAGGA    720
heThrPheLysAsnLeuAsnPheLysLeuPheAspValGlyGlyGlnArgSerGluArgL    210

AGAAGTGGATTCATTGTTTCGAAGATGTTACTGCTATTATTTTCTGTGTTGCCATGTCAG    780
ysLysTrpIleHisCysPheGluAspValThrAlaIleIlePheCysValAlaMetSerG    230

AGTATGATCAACTGTTGCACGAAGATGAGACAACAAACCGAATGCACGAATCGCTGAAGC    840
luTyrAspGlnLeuLeuHisGluAspGluThrThrAsnArgMetHisGluSerLeuLysL    250

TGTTCGATTCGATCTGTAATAACAAATGGTTCACAGATACATCGATTATTCTGTTCCTGA    900
euPheAspSerIleCysAsnAsnLysTrpPheThrAspThrSerIleIleLeuPheLeuA    270
```

ACAAGAAGGATCTGTTTGAAGAGAAAATCAAGAAAAGCCCGTTAACGATCTGCTTCCCAG 960
snLysLysAspLeuPheGluGluLysIleLysLysSerProLeuThrIleCysPheProG 290

AATATTCAGGACGACAAGACTACCACGAGGCATCTGCGTATATTCAAGCACAATTTGAGG 1020
luTyrSerGlyArgGlnAspTyrHisGluAlaSerAlaTyrIleGlnAlaGlnPheGluA 310

CTAAAAACAAATCAGCGAATAAAGAAATCTATTGCCACATGACATGTGCCACAGACACAA 1080
laLysAsnLysSerAlaAsnLysGluIleTyrCysHisMetThrCysAlaThrAspThrT 330

CTAACATTCAATTTGTGTTTGACGCTGTCACCGATGTGATTATTGCCAATAATCTTCGTG 1140
hrAsnIleGlnPheValPheAspAlaValThrAspValIleIleAlaAsnAsnLeuArgG 350

GATGCGGCTTGTATTAAGCTGTCGTCTTCGCCGCCCTCTTCTACCATTCTGTGTGTATCT 1200
lyCysGlyLeuTyrEnd 354

TTGCTTACTTTTCCCAAATTTTTAAAGATTCGTTATTTTTCTCATCCCGCGAGTATGCAT 1260

CTATAATTTGAGAGCTTTACATGTACATTATGGTTGAACTGTTTTTATTTTTTGGAAAAA 1320

GTTGCAATTTCAGATGTATATGGGCTTTTTTTTAACCTCTGATTCATTCATAATAGTCTC 1380

ACCGCCCTTCTTTTTCTCAACTAGGATCCCTAATTTTTCTGTCACAAAAATAACACATAA 1440

AACTCACAATTATTTATTTTTTCCTTTTTTCTTATTTATTATATATTGTCATTATTCACG 1500

CCACATCCCCCGCCCTTCTCACTCCCGTGTAGCCTTATCGCTATTTGAAAACAAGAAAAC 1560

TTTAAAAATCTAAATTCAGTTG 1582

Figure 3. Hybridization of *C. elegans* total mRNA to *C. elegans* $G_o\alpha$ cDNA. Total mRNA (12.5 μg) from mixed populations of the N2 wild-type hermaphrodite strain was fractionated on a 1% formaldehyde gel and hybridized to the *C. elegans* $G_o\alpha$ cDNA. The positions of the 18S (1,750 bp) and 28S (3,500 bp) rRNAs are indicated.

28S rRNA

18S rRNA

Figure 4. Comparison of $G_o\alpha$ amino acid sequences from four species. $G_o\alpha$ protein sequences from C, *C. elegans* (this study), D, *Drosophila melanogaster* (deSousa *et al.*, 1989), R, rat (Jones and Reed, 1987), and X, *Xenopus laevis* (Olate *et al.*, 1989) were aligned. Bold-faced letters indicate amino acids that are not found in all four sequences at that position. The mouse (M. Strathmann, personal communication), hamster (Hsu *et al.*, 1990), and bovine (Van Meurs *et al.*, 1987) $G_o\alpha$ amino acid sequences are not shown since they are over 95% identical to the rat $G_o\alpha$ sequence. Similarly the locust $G_o\alpha$ sequence (Hsu *et al.*, 1990) is 93% identical to the *Drosophila* sequence. Amino acid abbreviations used are: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

```
C: MGCTMSQEER AALERSRMIE KNLKEDGMQA AKDIKLLLLG AGESGKSTIV  50
D: MGCTTSAEER AAIQRSKQIE KNLKEDGIQA AKDIKLLLLG AGESGKSTIV  50
R: MGCTLSAEER AALERSKAIE KNLKEDGISA AKDVKLLLLG AGESGKSTIV  50
X: MGCTLSAEER AALERSKQIE KNLKEDGVTA AKDVKLLLLG AGESGKSTIV  50

C: KQMKIIHESG FTAEDYKQYK PVVYSNTVQS LVAILRAMSN LGVSFGSADR 100
D: KQMKIIHESG FTAEDFKQYR PVVYSNTIQS LVAILRAMPT LSIQYSNNER 100
R: KQMKIIHEDG FSGEDVKQYK PVVYSNTIQS LAAIVRAMDT LGVEYGDKER 100
X: KQMKIIHEDG FSGEDVKQYK PVVYSNTIQS LAAIVRAMDT LGIEYGDKER 100

C: EVDAKLVMDV VARMEDTEPF SEELLSSMKR LWGDAGVQDC FSRSNEYQLN 150
D: ESDAKMVFDV CQRMHDTEPF SEELLAAMKR LWQDAGVQEC FSRSNEYQLN 150
R: KADSKMVCDV VSRMEDTEPF SAELLSAMMR LWGDSGIQEC FNRSREYQLN 150
X: RADAKMVCDV VSRMEDTEPY SPELLSAMVR LWADSGIQEC FNRSREYQLN 150

C: DSAKYFLDDL ERLGEAIYQP TEQDILRTRV KTTGIVEVHF TFKNLNFKLF 200
D: DSAKYFLDDL DRLGAKDYQP TEQDILRTRV KTTGIVEVHF SFKNLNFKLF 200
R: DSAKYYLDSL DRIGAADYQP TEQDILRTRV KTTGIVETHF TFKNLHFRLF 200
X: DSAKYYLDSL DRIGAADYQP TEQDILRTRV KTTGIVETHF TFKNLHFRLF 200

C: DVGGQRSERK KWIHCFEDVT AIIFCVAMSE YDQLLHEDET TNRMHESLKL 250
D: DVGGQRSERK KWIHCFEDVT AIIFCVAMSE YDQVLHEDET TNRMQESLKL 250
R: DVGGQRSERK KWIHCFEAVT AIIFCVALSG YDQVLHEDET TNRMHESLML 250
X: DVGGQRSERK KWIHCFEAVT AIIFCVALSG YDQVLHEDET TNRMHESLKL 250

C: FDSICNNKWF TDTSIILFLN KKDLFEEKIK KSPLTICFPE YSGRQDYHEA 300
D: FDSICNNKWF TDTSIILFLN KKDLFEEKIR KSPLTICFPE YTGGQEYGEA 300
R: FDSICNNKFF IDTSIILFLN KKDLFGEKIK KSPLTICFPE YPGSNTYEDA 300
X: FDSICNNKWF TDTSIILFLN KKDIFQEKIK SSPLTICFPE YTGPNSFTEA 300

C: SAYIQAQFEA KNKSANKEIY CHMTCATDTT NIQFVFDAVT DVIIANNLRG 350
D: AAYIQAQFEA KNKSTSKEIY CHMTCATDTN NIQFVFDAVT DVIIANNLRG 350
R: AAYIQTQFES KNRSPNKEIY CHMTCATDTN NIQVVFDAVT DIIIANNLRG 350
X: VAHTQHQYES RNKSENKEIY THITCATDTQ NIQFVFDAVT DVIIAYNLRG 350

C: CGLF 354
D: CGLF 354
R: CGLF 354
X: CGLF 354
```

Figure 5. Immunoblot analysis of *C. elegans* $G_o\alpha$ protein. An affinity-purified rabbit anti-bovine $G_o\alpha$ antisera was used at a dilution of 1:1,000 to detect $G_o\alpha$ proteins as described in Materials and Methods. Lane 1, 1 μg purified bovine $G_o\alpha$; Lane 2, 50 μg total protein from *C. elegans* N2; Lane 3, 50 μg total protein from *C. elegans him-5* (e1490).


kDal
40
1
2
3

Figure 6. Restriction map, DNA sequencing strategy, and gene organization of *C. elegans gpa-1* gene. A) Restriction map of 10 kbp Hind III fragment which contains the *C. elegans gpa-1* gene. The arrows indicate the direction and extent of DNA sequences determined. B) Restriction map of 3.1 kbp Xho I/Bgl II subclone from 10 kbp Hind III fragment. The arrows indicate the direction and extent of DNA sequences determined. Abbreviations for restriction enzyme recognition sites are: Bg, Bgl II; H3, Hind III; Nc, Nco I; Ns, Nsi I; K, Kpn I; P, Pst I; Sa, Sal I; X, Xba I; Xh, Xho I. C) Intron/exon organization of *C. elegans gpa-1* gene. Exons are represented as horizontal lines. Introns are represented as boxes. Arrows indicate the location of the inverted repeats. Abbreviations used are: N, amino terminus; C, carboxyl terminus.

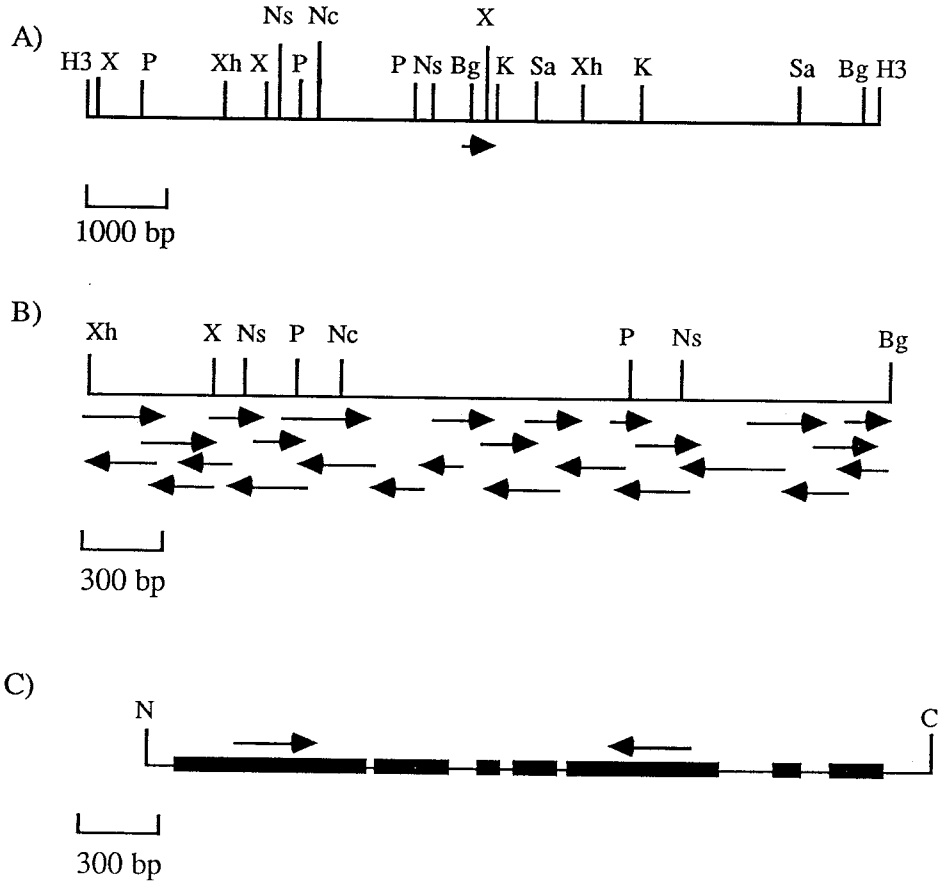

A)
Ns Nc
X
H3 X P Xh X P P Ns Bg K Sa Xh K Sa Bg H3
1000 bp
B)
Xh X Ns P Nc P Ns Bg
300 bp
C)
N
C
300 bp

Figure 7. DNA sequence of *C. elegans gpa-1* gene. The predicted amino acid sequence of the *gpa-1* protein is shown below the DNA sequence. The standard three letter amino acid abbreviations are given in the Figure 2 legend. Coding sequences are shown in bold-faced lettering.

```
CTCGAGTTGTTTATGCGGCGAAAAAAAATCTGTAGAGGTATACGAAAAAGCTAATTAGGT   60
GGTTTGCCGTCTAGTAATACTCGATTATATATTTCTTGTCGTCATCGTTACCACTGCTCT  120
CCATTTTCTCTCCTCACTCACAAAAACTTGAATATTCTTCGCTTCCTGTTGTGATCAGTG  180
AAAATTCGTCTGATTTCTTCAATATTATTCAATGTCATTTATTCGAAGACTTCAGATGGG  240
                                                        MetGl

AAACTGCGAAAGTCGAGAACTTGTAGCTCAAGCCAAACAGAATAAAATAATCAACACGGA  300
yAsnCysGluSerArgGluLeuValAlaGlnAlaLysGlnAsnLysIleIleAsnThrGl

ACTTGATAAGGCCAAAAAAACCGATGAAAATATAATCAAATTATTGCTGTTAGGTAAGTT  360
uLeuAspLysAlaLysLysThrAspGluAsnIleIleLysLeuLeuLeuLeuG

AGATTCTGAACCAAAAACAGCTTTCTGCAATACTTGTTGGTACATTCAAATCGCGCTCAT  420
TTTTCATTGAAAATTAAACGCGGCGAACATTTTTCTTACAAATAGGTGGTGACAACCTCA  480
TTTCATTCTGGTATCGGCACGAAAACATGTTTCTAGAAAGGTCTAGTTACCCAACCAAAA  540
AGTTGAAAGCACCACAGCTGTAGCGAAACCCAAAATAGAAGGAATTAAAATCTACCCACT  600
TGAGTTCACTAACACCCGCAAATGCATTAAAGAAACGTTTAAAAAACATCAGGTTGAAAT  660
CTTGTTCAATTATATCCGCTCAGAAGAGCAAGCCACTACGGCTCCAAAATTCATCGGCTC  720
CGTGCGGTCATCAAGTTCCGAAGAACTGTTTCACGCTCGCGGCTCGCCCGTCTCCCACTC  780
ATCGCGGATGGCATCTCTACCAATTTGATGGGAATTACGAGATACATACTGCAGAAATGA  840
TTCTGTATAGTATGGTCTCGATGTAAGCAGAGTGCTGACGGGGTTAGTGTGTACACGACA  900
GCCGACACCTCGCGGGTTTAGCCTGCCACCCTGTTTTGTGTGTCTGCCTGACCACCTGCC  960
GCCGATTGCATTGATAGCATTGATGCTCCATGGAAATGCGATTACTAATCAAACATCATT 1020
CGAATGTTTGACTTTTCTATCGTTTAAACATCTGAAATTAACTATTTCCAGGAGCTGGAG 1080
                                                    lyAlaGlyG

AGAGTGGAAAAAGTACCGTGCTGAAACAAATGAAGTAAGTTCTTAAAGATGAGTTCGAAA 1140
luSerGlyLysSerThrValLeuLysGlnMetLy

AAGAGTGTTTGCTATAAAATGCTCGGAATAATGCAATATGGGGCAAATGACGGATTAATC 1200
TTTCAGCTTTTATTTTAAGAGGAAGTTCAAAAATTTACAATTATTATTAGATTTTTTTTC 1260
AAAATTTCTTTTTTGTGGCATTTCATCATATTGGAATTTTCCAAGGTCTTTAAAGCAAAA 1320
GCTAAGTTAACCGAACTCTACTTTCAATCAAATGAAGAAATGTACCAATAAAATGATTTA 1380
CAGAATCATCCATAACAGTGGATTCTCTCAAGAAGAAATATCAAATAAACGAAATGTTGT 1440
   sIleIleHisAsnSerGlyPheSerGlnGluGluIleSerAsnLysArgAsnValVa

CTGTGCGAATACTGTGCAAGCAATGGGAGCATTGTTAGATGGAATGAAACAACTTCAATT 1500
lCysAlaAsnThrValGlnAlaMetGlyAlaLeuLeuAspGlyMetLysGlnLeuGlnPh

CGATTTTTCAACCCGAGTTTGTAATGTAAGAAATGTTAATTTTTGTGTTACATAGTACTT 1560
eAspPheSerThrArgValCysAsn

ATTTATAGGCACATGAGAAGTTGATACGTGAAACATTGAATGATAAAGCTGAGTATGGAC 1620
        AlaHisGluLysLeuIleArgGluThrLeuAsnAspLysAlaGluTyrGlyP

CATTCAGTGATGCAATGTTCAAGTAAGAACCTTATGTCATTTTGTGAATTTATTCAAAAA 1680
roPheSerAspAlaMetPheAs

ATATTTTGATAGCAATTCTCCGATTTGACAAAACTTTCATAACAATGGAAGTTCACAATA 1740
AAATTTTAATAAATCTAAACTATATTCAGCGCACTTACTGAGTTGTGGGCGGACAAAGGA 1800
                                nAlaLeuThrGluLeuTrpAlaAspLysGly

GTTCAGTGTGCATACGATAAGCGAGAGTTTTTTTACCTTCATGATTCTGCAAAATAGTAA 1860
ValGlnCysAlaTyrAspLysArgGluPhePheTyrLeuHisAspSerAlaLysTy
```

```
GTTAAACTTTTTTTTAAATATCAGAACTTTGCGATATTTTAAGTATTTAAAAACATAGAG 1920
CGTTAGAGGCTCAATATTAGTTCAACAAAACAGGTTATTCCATTTCCTAATTGCAATCAA 1980
GAGGCGTGTCTGTGCCTACACTTAGGATGCTGAATGACTGAACCCGCGAGGTGTCGGCTG 2040
TCGTGTACACACTAGCCCCGCCAGCACTCTGCGTACATCGAGACCATACTATACAGAATC 2100
ATTTCTGCAGTATGTATCTCGTAATTCCCATCAAATTGGTAGAGATGCCATCCGCGATGA 2160
GTGGGAGACGGGCGAGCCGCGAGCGTGAAACAGTTCTTCGGAACTTGATGACCCGCACGG 2220
AGCCGATGAATTTTGGAGCCGTAGTGGCTTGCTCTTCTGAGATCGGATATAATCGAACAA 2280
GGTTTCAATCTGATGTTTTTTTAAACGTTTTTTTAATGCATTTGCGGGTGTCTAGTGCGA 2340
TTTTATAATTGAATTACATTAAAAATTGCGAGTAGGTAAATAAGATTTAATTTTCAGTTT 2400
                                                          rPh

TCTTGATCGAATTGCAAGGGTTCACACTCCCAATTACGTACCAACTGAAAATGATATTCT 2460
eLeuAspArgIleAlaArgValHisThrProAsnTyrValProThrGluAsnAspIleLe

GCATACACGGGTTCCAACAATGGGTGTTATAGAAGTTAATTTCACAATAAAGGGCAAATT 2520
uHisThrArgValProThrMetGlyValIleGluValAsnPheThrIleLysGlyLysPh

TTTTCGAGTATTTGATGTGGGCGGACAGCGATCGCAAAGGAAGAAATGGATTCACTGTTT 2580
ePheArgValPheAspValGlyGlyGlnArgSerGlnArgLysLysTrpIleHisCysPh

TGATGATGCAAAAGCTATGATATATGTTGCTTCGTTGTCAGAGTATGATCAAGTTTTATT 2640
eAspAspAlaLysAlaMetIleTyrValAlaSerLeuSerGluTyrAspGlnValLeuLe

AGAAGACAATACTACTGTAAGTTTTGATATTTTCAAAAACAAATAATTCAATAGTAATTT 2700
uGluAspAsnThrThr

CTTGTTTTTCAGAATCGAATGCACGAATCAATACAGTTATTTAAGCAAGTAATCAACAAT 2760
            AsnArgMetHisGluSerIleGlnLeuPheLysGlnValIleAsnAsn

AAATATTTTGTAAACACTTCAGTTATATTATTCCTGAACAAAATTGATTTATTCGAGGAA 2820
LysTyrPheValAsnThrSerValIleLeuPheLeuAsnLysIleAspLeuPheGluGlu

AAAATTGTTACCAAAAAACGAAGTCTTGGTATTGCATTTGAGTCATTTAGTGGTAAGTGC 2880
LysIleValThrLysLysArgSerLeuGlyIleAlaPheGluSerPheSerG

AATTTTTGAAGGCGAATGTGGAGGCACTTGAACCCCGTACACCTGTCCGCCGTTCTGGCC 2940
GAGTGGTCTAAGGCGCTGCGTTCAGGTCGCAGTCCTCTCCGGAGGGCGCAGGTTCGAATC 3000
CTGCGGACGGCATTTCTTTTTGTGCATTCTCTTTTTTTTTGCAGGACCGAGCCAAGATCT 3060
                                                    lyProSerGlnAspLe

CAATGCAGCTGTAGCGTTCGTTGAAAAAAAGTATAGAAGTATGGCAGAGAACAAAGAGAA 3120
uAsnAlaAlaValAlaPheValGluLysLysTyrArgSerMetAlaGluAsnLysGluLy

GAACATTTATTGTCATCACACTTGTGCTACAGACACACAACAGGTGCAGTACGTACTCGA 3180
sAsnIleTyrCysHisHisThrCysAlaThrAspThrGlnGlnValGlnTyrValLeuAs

TGCGGTTCTAGATACGATACTATCTACCAAACTGAAGGGATGTGGATTGTATTGAGAGAT 3240
pAlaValLeuAspThrIleLeuSerThrLysLeuLysGlyCysGlyLeuTyrEnd

TTTAAATTATTGTTACCTTTGGTACAAACCTTGTAAACCTGAGTCATAATTATCGGGTAC 3300
CGTTCGGTATTATCTTGATTGGTGTAATATATAGAATTTAATAAATAATGAATCAAATTT 3360
```

Figure 8. Restriction map, DNA sequencing strategy, and gene organization of *C. elegans gpa-3* gene. A) Restriction map of 3 kbp Hind III fragment from *C. elegans gpa-3* gene. The arrows indicate the direction and extent of DNA sequences determined. B) Restriction map of 2.7 kbp Nsi I fragment from *C. elegans gpa-3* gene. The arrows indicate the direction and extent of DNA sequences determined. Abbreviations for restriction enzyme recognition sites are: Ns, Nsi I; H2, Hinc II; H3, Hind III; P, Pst I; X, Xba I. C) Gene organization of *C. elegans gpa-3* gene. Exons are represented as horizontal lines. Introns are represented as boxes. Abbreviations used are: N, amino terminus; C, carboxyl terminus.

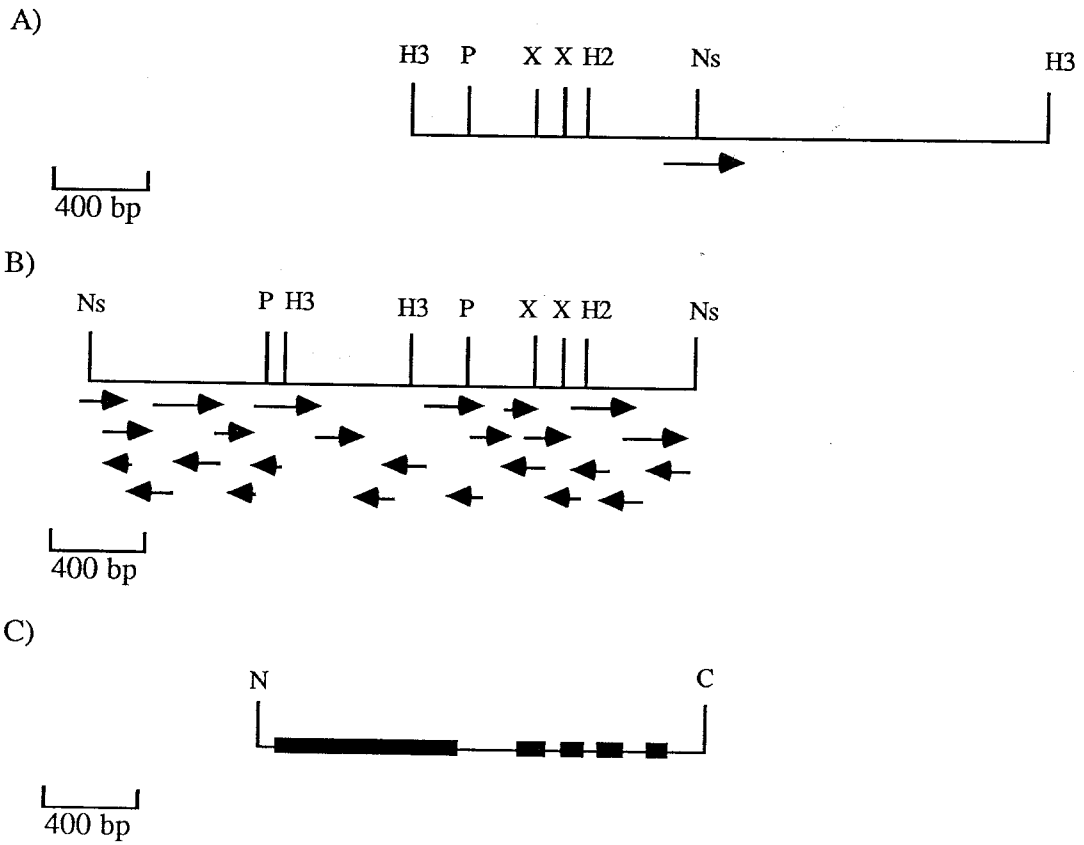

A)
H3
P
X
X
H2
Ns
H3
400 bp
B)
Ns
P
H3
H3
P
X
X
H2
Ns
400 bp
C)
N
C
400 bp

Figure 9. DNA sequence of *C. elegans gpa-3* gene. The predicted amino acid sequence of the *gpa-3* protein is shown below the DNA sequence. The standard three letter amino acid abbreviations are given in the Figure 2 legend. Coding sequences are shown in bold-faced lettering.

```
ATGCATTCATTTTGAGTCTATTTTTGTATGCTTCTTTTTTATGTCTAAGCATTCGTCGAC   60
ATTTGAAATATGCGTGAAATAATTTTCCAGTGGAACTAGAGCTGGGGGGCTTCCTATTGA  120
TTTTTGATTGATTTTTGAAAAAGACTTCGACCACGAATATCAATAAAATAATCTAAGTTT  180
CCGGGAAGACTCGAAACATCAAAAAGAAACGACTGGAACTATTCCATCGAACAATAGTCT  240
ATCTTTCGTTCATTTAGATGTGAAATTCATGATGATGTACAAATTGATTCTATTCGAAAG  300
ATTAATGTATTATTGAACATTTTCAGCTGAATCGAACATCTTTTATCACTCATTATTATG  360
ATATTTGTTCATTTTCAATGCGTCACAGCTTGGTGTCGTTTTGTTGAATTGGCAGCACCA  420
TTTCTTGTGAAAAAATGCTTGCTGTCGTCTTTCATTATTGTCGAGGTGAAACAGAAAAAG  480
AAAGAGACACACGAGATATCAACAACACAAAATTCTGAATATCTGACTTTGGCTCTGACC  540
GCCGTCGCCGCCCGCCCGCTAACGAGCATTTTCTCATTTTCTTCTCATTTACCCCGTGTC  600
TACTTCTTTTCTTATTTTTCGTTTCGTTAAACTTCGGTTGTCGCATAATCAGGCCTGCTT  660
TTCATTTTTCACTTTTTGTTTGAACTTTTGGGCTTTTTAGAGTTGTGCTTCATTCAAATC  720

TATATTCAATTACAGAAAATTATTAGCCAAAGAACAATTATGGGATTATGCCAATCTGCA  780
                                        MetGlyLeuCysGlnSerAla

GAGGACAAAGAACTGACGTTGAAATCAAAAGCGATCGATAAGGAAATGATGCAGAACCAT  840
GluAspLysGluLeuThrLeuLysSerLysAlaIleAspLysGluMetMetGlnAsnHis

ATGTCACAGCAAAAAGTTGTGAAGCTTCTGCTTTTGGGTAAAATTTTAAAACTTATTTTT  900
MetSerGlnGlnLysValValLysLeuLeuLeuG

TGTATTCAGCTTCGCGCAGTCAGAAATTGTTTTACGGTAATCATATGACTCATTGAGAAC  960
TTGTTTGACTTTTTGATAATACGCCGGAATCAAATACAAGCAAAACAAAAGTGTCGTGTG 1020
AGAAATTGTTACCCGCTTTTGTTGAAAAGTCCAAAAAATTTAATTTATGAGAACGTTGGG 1080
ATAAAAAATATTATCGTTTCAGAAGAAATGTTCCCTTCTGTTCTCGAACACTTTTTCAAA 1140
AAATTACTTCAAATTTGTAATAACTGTAATTAATAATTCAGAGAATGAAAGAGCTGTGAT 1200
TTTTAAATACTATATTTACAGACAAAAAGGATAAGTCAGAGAAAATGAGAAATGATCAGT 1260
AATAAGAATAAAGAACATCAACTCAATTTTCTCGAATCATATAAAATAATTATTAAAGTT 1320
TTAAGTTTCTTCTCTCACCCGGCTTCTTGCAAACACTACTTTTTTAACTGCCCGCAAAAA 1380
TTTTCATGTGTTCTAAATATTAATAATCAAATTATTTATCAAGTTATAATTCTCTTTTGA 1440
AGCTTATCAAAGAACTACAAACTACGTAGAACCCGTTGATGTCAAATTTTCAGATATCAA 1500
GTGTGCGTACGCTTAGGTTTTCAGCTGTTTCTTTCATTTGATTCTGAGAATTTAGAAGTT 1560
GAAAAATAAATTAGCTGGGGTTTGTGACAAGTTGGGCAAAATTAAACTCTTACTAATTTC 1620

AGGTGCTGGAGAATGTGGAAAAAGTACGGTTCTGAAACAAATGAGGATTCTTCACGATCA 1680
  lyAlaGlyGluCysGlyLysSerThrValLeuLysGlnMetArgIleLeuHisAspHi

CGGATTTACTGCAGAAGAGGCTGAACAACAGAAAAGTGTCGTTTTCAATAATACCCTTCA 1740
sGlyPheThrAlaGluGluAlaGluGlnGlnLysSerValValPheAsnAsnThrLeuGl

AGCAATGACTGCAATTCTGAAAGGAATGGAAGCACTTCGAATGACCTTTGATAAGCCAAT 1800
nAlaMetThrAlaIleLeuLysGlyMetGluAlaLeuArgMetThrPheAspLysProIl

TCGAGAAAATGATGCAAAATTTGTGATGGAGTCTCATAAAATGCTCCAAGAAGCGAAAGT 1860
eArgGluAsnAspAlaLysPheValMetGluSerHisLysMetLeuGlnGluAlaLysVa

TTTCCCAGAAGAATTAGCAAATGCCATTCAAGCATTATGGAATGATAAAGCCGTTCAGCA 1920
lPheProGluGluLeuAlaAsnAlaIleGlnAlaLeuTrpAsnAspLysAlaValGlnGl

AGTTATTGCAAAAGGAAATGAGTTCCAAATGCCTGAAAGTGCACCACAGTAAGATTTTAC 1980
nValIleAlaLysGlyAsnGluPheGlnMetProGluSerAlaProHi
```

TGTACTTCTAGAACTTCAAGATTCACATTTATATTTTTTTAG**TTTTCTATCAAGTCTTGA** 2040
sPheLeuSerSerLeuAs

TCGTATCAAGTTACCCGATTATAATCCAACTGAACAGGATATTCTTCTGTCTAGAATCAA 2100
pArgIleLysLeuProAspTyrAsnProThrGluGlnAspIleLeuLeuSerArgIleLy

GACAACTGGAATTGTTGAAGTTAAATTTCAAATGAAAAGTGTAGATTTTAGGTGTGAATA 2160
sThrThrGlyIleValGluValLysPheGlnMetLysSerValAspPheAr

GTTAATGATTCATGAGTTTGCAAGACATGAGTTTCAG**GGTATTGGACGTTGGAGGTCAAC** 2220
gValPheAspValGlyGlyGlnA

GATCAGAGAGAAAGAAATGGATTCATTGTTTTGAAGATGTTAATGCAATTATTTTCATCG 2280
rgSerGluArgLysLysTrpIleHisCysPheGluAspValAsnAlaIleIlePheIleA

CTGCAATTTCAGAATATGATCAAGTTCTGTTTGAAGATGAGACGACGGTTAGAACAAGGA 2340
laAlaIleSerGluTyrAspGlnValLeuPheGluAspGluThrThr

AAACTTTATGGTTTATTTGAGAATTGTTTTTTTTTTTGCAG**AATCGAATGATTGAATCAA** 2400
AsnArgMetIleGluSerM

TGAGGCTGTTTGAATCAATTTGTAATTCACGATGGTTCATCAATACTTCAATGATTCTTT 2460
etArgLeuPheGluSerIleCysAsnSerArgTrpPheIleAsnThrSerMetIleLeuP

TCTTAAATAAGAAGGATTTGTTTGCCGAGAAGATTAAAAGAACTTCAATTAAATCCGCAT 2520
heLeuAsnLysLysAspLeuPheAlaGluLysIleLysArgThrSerIleLysSerAlaP

TCCCAGACTATAAAGGTAAGATTTAATTGAACTTTTTCACACTCTGCCCATATATTTTCA 2580
heProAspTyrLysG

GG**TGCACAAACCTACGACGAATCGTGTCGATATATTGAGGAAAAGTTTGATGGATTGAAT** 2640
lyAlaGlnThrTyrAspGluSerCysArgTyrIleGluGluLysPheAspGlyLeuAsn

GCAAATCCAGAAAAGACGATTTACATGCATCAGACGTGTGCAACAGATACAGATCAAGTA 2700
AlaAsnProGluLysThrIleTyrMetHisGlnThrCysAlaThrAspThrAspGlnVal

CAAATGATTTTGGACTCAGTGATTGATATGATTATCCAGGCCAACTTGCAAGGATGCGGT 2760
GlnMetIleLeuAspSerValIleAspMetIleIleGlnAlaAsnLeuGlnGlyCysGly

TTGTACTGAAATGTTTGTGCCTTTCTCTCTCTCTCTCTTTCTCTCGCTATTTCTATCACT 2820
LeuTyrEnd

TGGAAATACACATAGGGAAAGCATTTTCAGAAAAAAAAACTGGATTACTTTCTGTTAAGAA 2880
GTAACCTGAAAATATACTTAATCGAAATTTAAGACAAACATGGTAAAGAAGGAAACATGG 2940
GTAATGCCAGGTCAATTCTTTGTTATGA 2970

Figure 10. Intron sizes and locations in *gpa-1*, *gpa-2*, and *gpa-3* coding sequences. The solid lines represent the coding regions of *gpa-1*, *gpa-2*, and *gpa-3*. The arrows indicate the location of introns and the numbers indicate their size in base pairs. Abbreviations used are: N, amino terminus; C, carboxyl terminus.

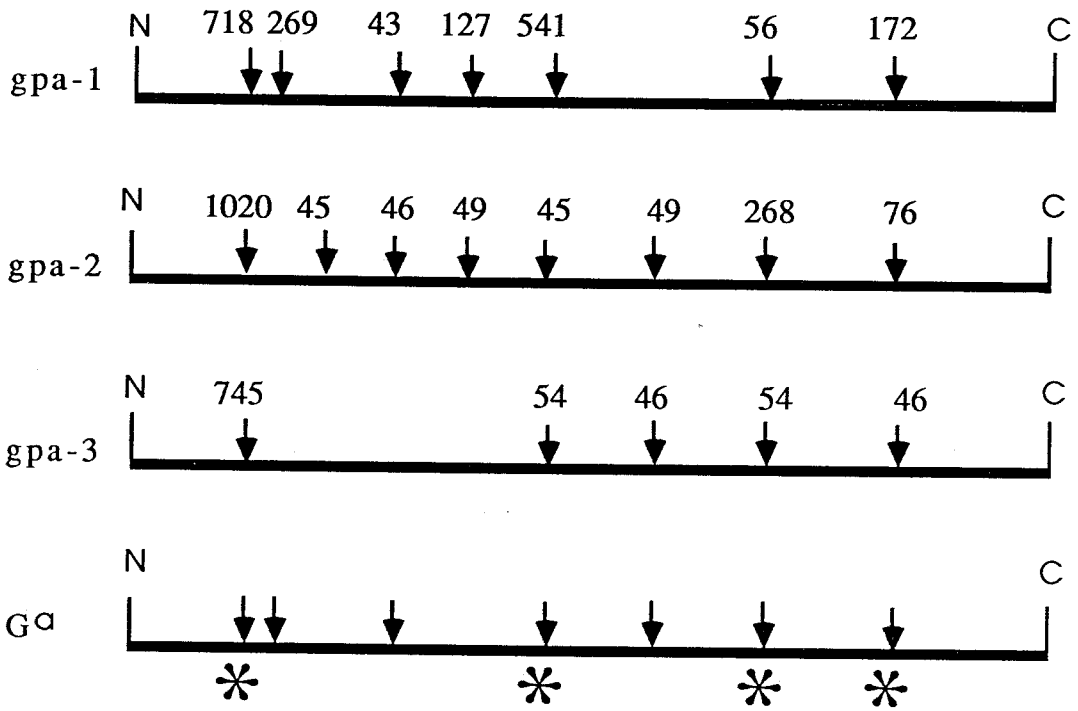

N
718 269
43
127
541
56
172
C
gpa-1
N
1020
45
46
49
45
49
268
76
C
gpa-2
N
745
54
46
54
46
C
gpa-3
N
C
Gα

Figure 11. Comparison of the amino acid sequences of *C. elegans gpa-1*, *gpa-2*, and *gpa-3*. Sequences were aligned using the FASTP program of Lipman and Pearson (1985). Dashes indicate gaps introduced to obtain an optimal alignment. One letter amino acid abbreviations are given in the legend to Figure 4. Bold-faced lettering indicates amino acids that are found in all three sequences at that position.

```
1: MGNCESRELV AQAKQNKIIN TELDKAKKTD ENIIKLLLLG AGESGKSTVL  50
2: MGLCQSEEEK VGTLKSRAID KEIKQLQTSE ERTVKLLLLG AGECGKSTVL  50
3: MGLCQSAEDK ELTLKSKAID KEMMQNHMSQ QKVVKLLLLG AGECGKSTVL  50

1: KQMKIIHNSG FSQEEISNKR NVVCANTVQA MGALLDGMKQ LQFDFSTRVC 100
2: KQMRLLTSKQ YTDEELLTQA KLVYTNIVIE MDHLVKAMPA AGLNFSDPMR 100
3: KQMRILHDHG FTAEEAEQQK SVVFNNTLQA MTAILKGMEA LRMTFDKPIR 100

1: NAHEKLIRET LNDKAEYGPF SDAMFNALTE LWADKGVQCA YDKR-EFFYL 149
2: EHDVHMLTLY IKDMQH-KNF QQDAADHVEK LWKDPVVKRL YAERKELNIR 149
3: ENDAKFVMES HKMLQEAKVF PEELANAIQA LWNDKAVQQV IAKGNEFQMP 150

1: H--DSAKYFL DRIARVHTPN YVPTENDILH TRVPTMGVIE VNFTIKGKFF 197
2: DIGDNTEYFF ENLPRISKED YHPNATDTLL LRTKTTGIVE VGFEIKKVKF 199
3: ---ESAPHFL SSLDRIKLPD YNPTEQDILL SRIKTTGIVE VKFQMKSVDF 197

1: RVFDVGGQRS QRKKWIHCFD DAKAMIYVAS LSEYDQVLLE DNTTNRMHES 247
2: RVFDVGGQRS ERKKWIHCFE DVNAIIFIAA LSEYNEVLFE DETTNRMIES 249
3: RVFDVGGQRS ERKKWIHCFE DVNAIIFIAA ISEYDQVLFE DETTNRMIES 247

1: IQLFKQVINN KYFVNTSVIL FLNKIDLFEE KIVTKKRSLG IAFESFSGPS 297
2: MRLFESICNS RWFHNTNIIL FLNKKDLFEE KI--KKENIH KAFPEYRGE- 297
3: MRLFESICNS RWFINTSMIL FLNKKDLFAE KI--KRTSIK SAFPDYKGA- 295

1: QDLNAAVAFV EKKYRSMAEN KEKNIYCHHT CATDTQQVQY VLDAVLDTIL 347
2: QNYAETVAFI KTKFEALSNN PKKTFYVHET CATDTNQVQK ILDSVISMII 346
3: QTYDESCRYI EEKFDGLNAN PEKTIYMHQT CATDTDQVQM ILDSVIDMII 344

1: STKLKGCGLY 357
2: QSNLHKSGLY 356
3: QANLQGCGLY 354
```

Appendix 1

Cone Cell-Specific Genes Expressed in Retinoblastoma

E. Bogenmann,[+] M. A. Lochrie,[*] and M. I. Simon[*]

[+] Division of Hematology-Oncology,
Children's Hospital of Los Angeles,
Los Angeles, CA 90027

[*]Division of Biology (147-75),
California Institute of Technology,
Pasadena, CA 91125

(Published in Science 240, 76-78)

Retinoblastoma, an intraocular tumor that occurs in children, has long been regarded, on the basis of morphological criteria, as a malignancy of the photoreceptor cell lineage. Here it is shown that when this tumor is grown in vitro, the cells express highly specialized photoreceptor cell genes. Transcripts for the transducin alpha subunit, $T_{C\alpha}$, which is specific to the cone cell, as well as transcripts for the red or green cone cell photopigment, were found in seven out of seven low-passage retinoblastoma cell lines. No marker genes specific to rod cells were expressed, suggesting that retinoblastoma has a cone cell lineage.

RETINOBLASTOMA (RB) IS AN INTRAOCULAR tumor that occurs in children. The neoplasm develops early in life on a sporadic or hereditary basis and may be unilateral or bilateral, with the hereditary form being mainly bilateral (*1*). A chromosomal abnormality has been described for both forms of RB (*2*), and a gene in the RB locus was recently cloned (*3*).

The tumor is derived from neuroectodermal cells, but its cell of origin has not been unequivocally identified. Primary tumors are composed of anaplastic cells with little cytoplasm, and morphologically differentiated structures (fleurettes and Flexner-Wintersteiner rosettes) are found that express structural characteristics of mature photoreceptor cells (*4*). A bidirectional differentiation potential to photoreceptor cells or glial cells, or both, has been postulated on the basis of studies with the Y79 cell line (*5*).

Rod and cone cells are the photoreceptors of the mammalian retina. They express the photopigment genes (rhodopsin and color photopigments) (*6*) that have been isolated from the human genome (*7, 8*). Photopigment photolysis is coupled to cyclic guanosine 5′-phosphate metabolism through a membrane-associated retinal G protein, transducin (*9–11*), and results in a hyperpolarization of the photoreceptor cell membrane (*12, 13*). Bovine complementary DNAs (cDNAs) specific for the rod cell ($T_{R\alpha}$) and cone cell ($T_{C\alpha}$) alpha subunits of transducin have been isolated (*10, 14, 15*).

Here we show that cultured RB cell lines established from individual patients express the red or green photopigment gene, or both genes, but not the gene coding for rhodopsin. Concomitantly, the cells express the cone cell $T_{C\alpha}$ subunit but not the rod cell $T_{R\alpha}$ subunit of transducin. We therefore postulate that RB may be a neoplasm of the cone cell lineage.

Previously isolated RB cell lines that show

various degrees of morphological differentiation in vitro were grown in mass cultures (*16*). These cell lines represent early-passage tumor populations, which represent a total culture period of less than 12 months. We investigated the expression of the various transducin genes by means of Northern blot analysis. Bovine cDNA probes for the $T_{R\alpha}$ and $T_{C\alpha}$ (*15*) were used since human probes were not available. The RNAs from seven RB lines were hybridized with the $T_{C\alpha}$ subunit cDNA, and two transcripts were detected in the RNAs from RB cell lines but not in RNA from an osteosarcoma cell line (HTLA145) (*17*) (Fig. 1). In RB cells, as in bovine retina, the $T_{C\alpha}$ probe detected a high and a low molecular weight transcript; the low molecular weight message was generally expressed at a higher level. The intensity and electrophoretic mobility of the $T_{C\alpha}$ transcripts in different RB cell lines varied, although equal amounts of total RNA were present when the RNAs were hybridized with the β subunit cDNA probe, which is expressed in all tissues. The extent of expression varied also with culture conditions but could not be correlated with the morphology of the cells, suggesting that expression of $T_{C\alpha}$ is independent of Flexner-Wintersteiner rosette differentiation in vitro.

The Northern blots hybridized with the $T_{C\alpha}$ probe were then used to study the expression of the gene for the $T_{R\alpha}$ subunit. In none of the seven low-passage RB cell lines could we detect transcripts for the rod cell–specific $T_{R\alpha}$. Since Southern blot analysis revealed the presence of the $T_{R\alpha}$ gene in all cell lines tested, we conclude that these cultured RB cell lines express only $T_{C\alpha}$.

The β subunit (T_β) was expressed as two distinct messages in all RB cell lines tested, as has been found in all tissues thus far examined (*18*). A transcript (0.7 kb) for the gamma subunit (T_γ) (Fig. 2) was found in four out of seven RB lines (Table 1) when a bovine cDNA probe was used (*19*); however, no expression was found in the HTLA145 cell line. The intensities and electrophoretic mobilities of the transcripts for T_γ also varied in different cell lines. However, the message size was similar to that found in the mature bovine retina (*20*).

We also investigated the presence of transcripts for the different photopigments in RB cell lines since transducin and pigment proteins are components of the same signal transduction system. In Northern blots used for the analysis of $T_{C\alpha}$ gene expression we did not detect a message for the opsin gene with the bovine opsin cDNA probe (*21*), although strong hybridization with rabbit retina RNA occurred. Southern blots revealed the presence of the opsin gene in all RB lines tested. In seven out of seven early-passage RB lines a transcript for the red or green or both color pigment genes but not for the blue pigment gene was found (Fig. 3). The message size was similar to that found in the mature retina (*8*), and expression levels varied between individual cell lines but could not be correlated with the morphology of the cells. A summary of the analysis is given in Table 1.

The presence of transcripts for gene products involved in cone cell signal transduction is a marker for a lineage analysis of RB and may provide a molecular clue to its developmental origin. The expression of such cone cell specific marker genes therefore suggests that RB is committed within the cone cell lineage, but morphological observations indicate that RB does not represent a terminally differentiated cell (*5, 16*). Transcripts of rod cell–specific genes, if present, were below the limit of detection relative to cone cell markers. In vitro selection for cone cell precursors in RB seems unlikely since these tumors represent early cell passages; however, it is known that chicken cone cells do survive in culture whereas rod cells do not (*22*). Absence of any RB with blue pigment expression may be due to the combination of the small sample size and the low number of blue cones in the human retina (*23*).

This study demonstrates the expression of cone-specific genes in cultured RB. These cells may therefore be an attractive system in which to study photoreceptor development.

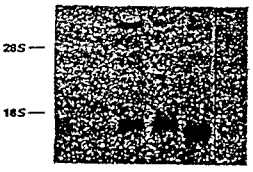


Fig. 1. Expression of $T_{C\alpha}$ subunit in cultured RB cells. RB cell lines of passage 15 through 25 that were in culture less than 12 months were grown as described (*16*). Cells were pipetted off the feeder layers, centrifuged, and lysed in lysis buffer (0.14*M* NaCl, 0.01*M* tris-HCl, *p*H 7.4, 1.5 m*M* $MgCl_2$, and 0.5% NP-40). Nuclei were centrifuged (1500*g*), and the supernatant was extracted three times with phenol and then once with a mixture of chloroform and isoamyl alcohol (24:1 by volume). Total cellular RNA (30 μg per lane) was electrophoresed on a 1% agarose-formaldehyde gel and transferred to Biodyne (ICN), and filters were baked at 80°C for 2 hours. Filters were prehybridized and hybridized in 50% formamide, 5× standard saline citrate (SSC), 1× Denhardt's (0.02% each of Ficoll, polyvinylpyrrolidone, and bovine serum albumin), 20 m*M* sodium phosphate, *p*H 6.5, 0.1% SDS, and denatured salmon sperm DNA (250 μg/ml) at 42°C. Filters were washed at a final stringency of 0.3× SSC at 65°C. The $T_{C\alpha}$ cDNA probe consisting of a 1.7–kilobase pair (kbp) Hind III fragment (*15*) was radiolabeled with [^{32}P]deoxycytidine triphosphate (ICN), according to Feinberg and Vogelstein [(*24*), see addendum]. Cell lines: Y79, RB cell line; HTLA230, neuroblastoma; RBLA22, RBLA18, and RBLA12, low-passage RB cell lines; and HTLA145, osteosarcoma.

Table 1. Summary of transducin gene and photopigment gene expression in RB cells in vitro. ND, not determined; +, detectable expression; −, no detectable expression.

Cell line	$T_{R\alpha}$	$T_{C\alpha}$	T_γ	T_β	Red or green photopigment	Blue photopigment	Rhodopsin
RBLA10	−	+	+	+	+	−	−
RBLA12	−	+	+	+	+	ND	−
RBLA13	−	+	ND	+	+	ND	−
RBLA18	−	+	−	+	+	−	−
RBLA19	−	+	−	+	+	−	−
RBLA20	−	+	+	+	+	−	−
RBLA22	−	+	+	+	+	ND	−
Y79	ND	+	−	+	ND	ND	−

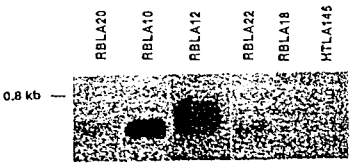


Fig. 2. Expression of retinal transducin T_γ in RB cells. The two 0.2-kbp Eco RI fragments from the bovine retinal transducin T_γ cDNA (*19*) were used and hybridized to RB RNA previously analyzed for the $T_{C\alpha}$ subunit, as described in the legend to Fig. 1.

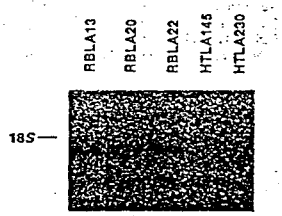


Fig. 3. Expression of red or green color photopigment gene in cultured RB cells. Total cellular RB RNA was hybridized with the human cDNA probe for the red photopigment clone hs7 (*8*), as described in the legend to Fig. 1.

REFERENCES AND NOTES

1. A. G. Knudson, *Cancer Res.* **45**, 1437 (1985); A. L. Murphree and W. F. Benedict, *Science* **223**, 1028 (1984).
2. R. S. Sparkes *et al.*, *Science* **208**, 1042 (1980); W. K. Cavenee *et al.*, *Nature (London)* **305**, 779 (1983).
3. S. H. Friend *et al.*, *Nature (London)* **323**, 643 (1986); W.-H. Lee *et al.*, *Science* **235**, 1394 (1987); Y-K. T. Fung *et al.*, *ibid.* **236**, 1657 (1987).
4. M. G. M. Ts'o, L. E. Zimmerman, B. S. Fine, *Am. J. Ophthalmol.* **69**, 339 (1970).
5. A. P. Kyritsis, M. Tsokos, T. J. Triche, G. J. Chader, *Nature (London)* **307**, 471 (1984).
6. D. Bok, J. F. Basinger, M. O. Hall, *Exp. Eye Res.* **18**, 225 (1974); D. S. Papermaster, B. G. Schneider, M. A. Zorn, J. P. Kraehenbuhl, *J. Cell Biol.* **77**, 196 (1978); R. Boynton, *Human Color Vision* (Holt, Rinehart and Winston, New York, 1979).
7. J. Nathans and D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4851 (1984).
8. J. Nathans *et al.*, *Science* **232**, 193 (1986).
9. G. B. Grunwald, P. Gierschik, M. Nirenberg, A. Spiegel, *ibid.* **231**, 856 (1986).
10. C. L. Lerea, D. E. Somers, J. B. Hurley, I. B. Klock, A. H. Bunt-Milam, *ibid.* **234**, 77 (1986).
11. A. G. Gilman, *Annu. Rev. Biochem.* **56**, 615 (1987).
12. B. K-K. Fung, J. B. Hurley, L. Stryer, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 152 (1981).
13. E. E. Fesenko, S. S. Kolesnikov, A. L. Lyubarsky, *Nature (London)* **313**, 310 (1985); K.-W. Yau and K. Nakatani, *ibid.* **317**, 252 (1985).
14. D. C. Medvinski *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4311 (1985); T. Tanabe *et al.*, *Nature (London)* **315**, 242 (1985); K. Yatsunami and H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4316 (1985).
15. M. A. Lochrie, J. B. Hurley, M. I. Simon, *Science* **228**, 96 (1985).
16. E. Bogenmann and C. Mark, *J. Natl. Cancer Inst.* **70**, 95 (1983); E. Bogenmann, *Int. J. Cancer* **38**, 883 (1986).
17. E. Bogenmann, H. Moghadam, Y. A. DeClerck, A. Mock, *Cancer Res.* **47**, 3808 (1987).
18. H. K. W. Fong, T. T. Amatruda III, B. W. Birren, M. I. Simon, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3792 (1987).
19. J. B. Hurley, H. K. W. Fong, D. B. Teplow, W. J. Dreyer, M. I. Simon, *ibid.* **81**, 6948 (1984).
20. C. Van Dop *et al.*, *Biochem. Biophys. Res. Commun.* **124**, 250 (1984).
21. J. Nathans and D. S. Hogness, *Cell* **34**, 807 (1983).
22. R. Adler, J. D. Lindsey, C. L. Elsner, *J. Cell Biol.* **99**, 1173 (1984).
23. H. J. A. Dartnall, J. K. Bowmaker, J. D. Mollon, in *Color Vision*, J. D. Mollon and L. T. Sharpe, Eds. (Academic Press, New York, 1983), p. 69.
24. A. P. Feinberg and B. Vogelstein, *Anal. Biochem.* **132**, 6 (1983); addendum, *ibid.* **137**, 266 (1984).

25. We thank J. Nathans for providing the photopigment cDNA clones, A. Mock and H. Moghadam for technical assistance, and C. Stutters for secretarial assistance. Supported by PHS grant EY04950 from the National Eye Institute (E.B.) and by a grant from the National Science Foundation (M.S.).

25 September 1987; accepted 26 January 1988

Appendix 2

Heterologous Expression of Transducin Alpha Subunits in *Escherichia coli,* Yeast, and Insect Cells

ABSTRACT

Heterologous expression of recombinant bovine rod and cone photoreceptor transducin alpha subunits ($rT_r\alpha$ and $rT_c\alpha$) has been examined in *E. coli*, yeast, and insect cells. Several combinations of expression vectors, host strains, and induction conditions were tested. Yeast expression vectors were constructed for secretion of $rT_r\alpha$ or intracellular expression of $rT_c\alpha$. In each case, no expression was observed even in a pep4 protease-deficient strain. Baculovirus vectors were constructed for the expression of $rT_r\alpha$ or a polyhedrin-$rT_r\alpha$ fusion protein in Sf9 insect cells. Upon infection of Sf9 cells with these viruses large amounts of $rT_r\alpha$ and polyhedrin-$rT_r\alpha$ were produced, but over 98% of it was insoluble. For expression of $rT_r\alpha$ and $rT_c\alpha$ in *E. coli,* several different types of expression vectors were constructed. The expression levels and physical properties of the $rT_r\alpha$ and $rT_c\alpha$ produced using these vectors varied widely. The most favorable results for producing $rT_r\alpha$ and $rT_c\alpha$ in *E. coli* were obtained by using a dual cistron expression vector in a lon100 protease deficient strain. Using this system $rT_r\alpha$ was synthesized to a level representing about 1% of total cell protein and about 50% of the $rT_r\alpha$ made was soluble. The soluble material can be partially purified by ammonium sulfate precipitation, and heptyl agarose and anion exchange chromatography. Upon further optimization, *E. coli* may serve as the most useful heterologous expression system for synthesizing transducin alpha subunits in an active form.

ABBREVIATIONS

AcNPV, *Autographa californica* nuclear polyhedrosis virus

A. S., ammonium sulfate

BCIP, 5-bromo-4-chloro-3-indolyl phosphate

BSA, bovine serum albumin

CHAPS, 3-[(3-cholamidopropyl)dimehtylammonio]-1-propane sulfonate

cGMP, 5', 3' cyclic guanosine monophosphate

DTT, dithiothreitol

E. coli, the bacterium *Escherichia coli*

EDTA, ehtylene diamine tetraacetic acid

F^+, male *E. coli* cells which carry the F factor

$G\alpha$, G protein alpha subunit

GDP, guanosine-5'-diphosphate

G protein, guanine nucleotide-binding regulatory protein

GST, glutathione S-transferase

GTP, guanosine-5'-triphosphate

GTPγS, guanosine-5'-O-(3-thiotriphosphate)

IPTG, isopropyl-β–D-thiogalactopyranoside

kbp, kilobase pairs

kDal, kilodaltons

KLH, keyhole limpet hemocyanin

KPO_4, potassium phosphate

LB, Luria broth

M_r, molecular weight

NBT, 4-nitro blue tetrazolium chloride

N. D., not determined

OD, optical density

PBS, phophate buffered saline

PDE, rod photoreceptor phosphodiesterase

PMSF, phenylmethane sulfonyl fluoride

QMA, quarternary methyl ammonium

$rG_i\alpha$, recombinant inhibitory G protein alpha subunit

$rG_o\alpha$, recombinant "other" G protein alpha subunit

$rG_s\alpha$, recombinant stimulatory G protein alpha subunit

ROS, rod outer segments

$rT_c\alpha$, recombinant bovine cone photoreceptor transducin alpha subunit

$rT_r\alpha$, recombinant bovine rod photoreceptor transducin alpha subunit

SDS, sodium dodecyl sulfate

Sf9, *Spodoptera frugiperda* clone 9 insect cell line

SPDP, N-succinimidyl 3-(2-pyridyldithio) propionate

$T\beta\gamma$, bovine rod photoreceptor beta and gamma subunit complex

$T_r\alpha$, bovine rod photoreceptor transducin alpha subunit

Tris, Tris[(hydroxylmethyl)aminomethane]

TPCK, N-tosyl-L-phenylalanine chloromethyl ketone

XGAL, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

INTRODUCTION

G proteins are a family of diverse guanine nucleotide-binding proteins that function in transmembrane signal transduction (see Gilman, 1987 or Lochrie and Simon, 1988 for review). They interact with receptors that detect extracellular signals and with effector proteins that generate intracellular signals. Each G protein has three subunits: α, β, and γ. The alpha subunit binds GDP and GTP. The GDP form interacts with ligand-bound receptors leading to a nucleotide exchange reaction in which GDP is exchanged for GTP. The GTP form interacts with effector proteins which alter intracellular metabolism. An intrinsic GTPase activity of the alpha subunit allows multiple recycling from the GTP to the GDP form. This leads to signal amplification.

Transducins are G proteins that are found specifically in retinal photoreceptor cells. Two transducin alpha subunits are known to exist. $T_r\alpha$ is found in rod cells and $T_c\alpha$ is found in cones cells (Lerea *et al.*, 1986). The biochemical properties of rod transducin have been extensively studied (Stryer, 1986; Hurley, 1987). It functions in the initial steps of visual signal transduction by coupling photolyzed rhodopsin to a cGMP phosophodiesterase. In contrast, nothing is known about the function of $T_c\alpha$. However, it may function in a visual pathway in cones similar to the process found in rods.

To learn more about how G proteins work it is desirable to know which of their amino acids are important for the various G protein alpha subunit functions. Based on sequence homologies and a comparison of G protein alpha subunits to the crystal structures of Ef-Tu (Woolley and Clark, 1989) and *ras* (Milburn *et al.*, 1990) specific amino acids have been proposed to form the guanine nucleotide binding site. Other studies have provided a clue about the location of sequences that interact with receptor, effector and the beta and gamma subunits. For a systematic study of structure/function relationships in transducin alpha subunits and other novel G protein alpha subunits of unknown function

such as $G_z\alpha$ (Fong *et al.*, 1988), a long-term strategy was conceived. First, cDNAs encoding alpha subunits would be synthesized in a heterologous expression system. Second, the proteins produced would be purified. Third, the purified proteins would be assayed and their properties compared to the natural protein, if available. Fourth, mutant recombinant proteins would be generated, assayed, and compared to the wild-type recombinant protein. Mutants having specific biochemical abnormalities would be identified and dominant mutants would be of particular interest. Fifth, genes encoding such mutant proteins would be introduced into transgenic mice or cultured cells under the control of cell type-specific promoters. In this manner it may be possible to correlate an observable phenotype with a defined biochemical alteration. The mutants could be used not only for studying the guanine nucleotide interactions of rod transducin in detail, but also the importance of other amino acids such as the cholera toxin target which is found in all G proteins. Overall this strategy provides one approach for learning more about the role of G proteins *in vivo*.

This study addresses the first stage of this long term project by surveying heterologous expression of bovine recombinant $T_r\alpha$ ($rT_r\alpha$) and $T_c\alpha$ ($rT_c\alpha$) proteins in yeast, *E. coli*, and insect cells. Expression of G protein alpha subunits in heterologous systems could have many uses other than simply assaying the effects of mutations on alpha subunit function. For example, recombinant proteins could serve as the ideal substrate for studying some of the posttranslational modifications which $G\alpha$ has been proposed to undergo (e.g., myristylation, carboxymethylation, phosphorylation, and carboxyl terminal proteolysis), since the natural proteins might already be fully modified and thus refractory to any further modification.

Recently an *in vivo* myristylation system for proteins expressed in *E. coli* has been developed (Duronio *et al.*, 1990). It would be interesting to use this system to determine

what effect myrisylation has on the properties of $rT_r\alpha$. Myristylation is necessary for the attachment of G protein alpha subunits to membranes (Jones *et al.*, 1990; Mumby *et al.* 1990). However its role in other activities of G proteins is unclear.

Heterologous expression systems can provide a more convenient and readily available source of protein than the natural source, especially in cases where the endogenous protein is expressed in a few cells, when the tissue is difficult to obtain, or when the natural protein has not been purified. For example, $T_c\alpha$ has not been purified from the retina. If large amounts of recombinant protein could be obtained it would be useful, not only for the biochemical studies mentioned above, but also for x-ray crystallographic studies.

Another way expression systems have proven to be very useful even if active protein is not obtained is in providing an antigen to stimulate the production of antibodies. An expressed protein can also provide a reagent for testing antibodies when the natural antigen is unavailable. Such a situation arises when antibodies are raised against peptide sequences derived from the DNA sequences of unidentified protein coding regions. This was the case for the $T_c\alpha$ antipeptide antibodies raised in this study.

The development of an *in vitro* reconstituted system for studying mutants "locked" in the activated GTP bound form could be useful for identifying drugs that block $G\alpha$ protein/effector interactions. Such experiments would be of special interest in view of recent results demonstrating activated forms of $G_s\alpha$ in human pituitary tumours (Landis *et al.* 1989). Drugs that block G protein activation (GTP-binding) or enhance deactivation (GTPase activity) may also have therapeutic uses. Their effects would have to be well-charaterized biochemically before clinical studies could be initiated.

The rod transducin alpha subunit is probably the best G protein candidate to choose for heterologous expression studies, at least for studying the basic properties of a G protein-mediated signal transduction system. No other G protein-mediated pathway has

been biochemically characterized in as much detail. The receptor (rhodopsin) and effector (cGMP phosphodiesterase) that transducin interacts with can be readily purified in larger amounts than those from other G protein-mediated signal trasnsduction systems. Also cDNAs encoding all of the components of the basic photoreceptor-signalling cascade (rhodopsin, transducin, and phosphodiesterase) have been cloned and sequenced. Rhodopsin (Oprian *et al.*, 1987) and the gamma subunit of phosphodiesterase (Brown and Stryer, 1989) have been expressed in heterologous hosts in an active form.

MATERIALS AND METHODS

Materials. Restriction enzymes were from New England Biolabs or Bethesda Research Labs. TPCK-treated trypsin was from Worthington. ^{35}S-GTPγS (1,200 Ci/mmol) was purchased from New England Nuclear. Other nucleotides were from Boeringer Mannheim. Frozen retinas were purchased from Hormel and stored in the dark at -70°C until use. Plasmids were obtained from Amrad (pGEX-2T), Bethesda Research Laboratories (pUC 18 and M13 mp 11), Pharmacia (pPL, pKK 223-3), Promega (pSP 64), and Stratagene (pBluescript/KS$^+$). Other plasmids and strains were obtained as indicated in the acknowledgements. Human thrombin (3200 NIH units/mg) was purchased from Calbiochem. Heptyl agarose was synthesized by the method of Shaltiel (1984). Other chromatography media was obtained from Pharmacia (Blue Sepharose CL-6B, Chelating Sepharose 6B, Sephacryl S-200 superfine, Mono Q, Phenyl superose, Superose 12, DEAE-Sepharose CL-6B), Bio-Rad (Affigel Blue, Biogel HTP), Waters Associates (QMA), Amicon (dye matrex kit: Blue A, Blue B, green, red, white, orange), Whatman (P11 phosphocellulose) and Sigma (reduced glutathione-agarose, S-linked). The strains used in this study are listed in Table 2. Luria broth (LB) was used to grow bacteria. Either minimal (SD) or rich (YPD) media was used to grow yeast. Grace's insect cell media (Gibco), containing 10% heat-inactivated fetal calf serum, 0.67% lactalbumin hydrolysate

(Difco), 0.67% TC yeastolate (Difco), penicillin (100 U/ml), streptomycin (0.1 mg/ml), and amphotericin B (0.25 μg/ml), was used to propagate Sf9 insect cells

Isolation of $T_r\alpha$ and $T_c\alpha$ cDNAs. Isolation of the $T_c\alpha$ cDNA has been described (Lochrie *et al.*, 1985). Four candidate $T_r\alpha$ cDNA clones out of 14,000 phage were isolated from a bovine retinal cDNA library (Nathans and Hogness, 1983) by the same general methods used to isolate the $T_c\alpha$ cDNA. A synthetic oligonucleotide corresponding to nucleotides 260-299 of the $T_r\alpha$ cDNA sequence (Tanabe *et al.*, 1985) was used as a probe. The four candidate clones were rescreened with TαB (Lochrie *et al.*, 1985) and TαCT oligonucleotide mixtures. The sequence of TαCT is 5'CGNAGNACPuTCPyTG-PyTC 3' (N=any nucleotide, Pu=purine, and Py=pyrimidine). This corresponds to the antisense strand encoding the amino acids Glu-Gln-Asp-Val-Leu-Arg found at amino acids 167-172 of $T_r\alpha$. One phage hybridized to all three probes. DNA was isolated from this phage and digested with Nco I. A 2,190 base pair Nco I fragment was detected as predicted from the published sequence (Tanabe *et al.*, 1985) and subcloned in pMS 631 to generate pML 7 (see below).

Construction of expression vectors. General features of the plasmids used in this study are listed in Table 3. Schematic diagrams of the plasmids and the sequences of their promoter regions are given in Figures 1-44. Details of their construction are described below. General molecular cloning methods used to construct these plasmids are outlined in Maniatis *et al.* (1982). The derivation history of the plasmids is diagrammed in Figures 45 and 46.

pML 1. pML 1 (see Figures 1 and 2) was constructed by subcloning the Hind III fragment from the $T_c\alpha$ cDNA clone (Lochrie *et al.*, 1985) into the Hind III site of the *E. coli* p_{tac} expression vector pKK 223-3 (Brosius and Holy, 1984).

pML 4. pML 4 (see Figures 3 and 4) was constructed by inserting the Eco RI

fragment of M13 ml 91 (see below) into the Eco RI site of pKK 223-3.

pML 7. pML 7 (see Figure 5) was constructed by subcloning the 2,190 base pair Nco I fragment of the $T_r\alpha$ cDNA clone into Nco I digested pMS 631 (Johnson and Simon, 1985). This Nco I fragment contains 1 bp of 5' untranslated sequence, the entire $T_r\alpha$ coding region of 1,050 bp, and 1,138 bp of 3' untranslated sequence.

pML 8. pML 8 (see Figure 6) consists of the Sal I restriction fragment from pML 7 inserted into the Sal I site of pSP 64. Sense strand RNA would be synthesized upon transcription by the SP 6 phage RNA polymerase.

pML 10. pML 10 (see Figure 7) is the same as pML 8 with the insert in the opposite orientation. Thus antisense strand RNA would be synthesized upon transcription by the SP 6 phage RNA polymerase.

pML 11. pML 11 (see Figures 8 and 9) is the Hind III fragment from the $T_c\alpha$ cDNA (Lochrie *et al.*, 1985) inserted into the Hind III site of the yeast expression vector pAAH 5 (Ammerer, 1983). pAAH 5 is a shuttle vector which contains the yeast alcohol dehydrogenase promoter and terminator elements, a yeast 2 μm plasmid replication origin, an *E. coli* Col E1 replication origin, the LEU 2 gene as a selectable gene in yeast and β-lactamase (ampicillin resistance) as a selectable gene in *E. coli.*

pML 12. pML 12 (see Figures 10 and 11) was constructed by inserting the Nco I fragment from pML 10 into the *E. coli* p_{trc} expression vector pKK233-2 (Amann and Brosius, 1985) such that the coding region is in a sense direction relative to the promoter. pML 12(-) is the same as pML 12 except that the insert is in the opposite orientation.

pML 15. pML 15 (see Figures 12 and 13) was constructed by inserting the Eco RI fragment of pGM 5 (Bruist and Simon, 1984) into the Eco RI site of pML12. The Eco RI fragment of pGM 5 was derived from λbio 256 and contains the λ p_L and λ p_R promoters and the λcI_{857} and N genes (Hedgpeth *et al.*, 1978).

pML 17. pML 17 (see Figures 14 and 15) was constructed by inserting the Hinc II fragment from pML 12 into pPL. Ligations were initially transformed into N99 to repress transcription and then into N4830 to test expression.

pML 18. pML 18 (see Figures 16 and 17) was constructed by inserting the Hinc II fragment from pML 4 into pPL. Ligations were initially transformed into N99 to repress transcription and then into N4830 to test expression.

pML 25. pML 25 (see Figure 18) was constructed by inserting the Pst I fragment from pML 15 into the Pst I site of pUC 18.

pML 34(+). pML 34(+) (see Figures 19 and 20) was constructed by inserting the Pst I fragment from pML 8 into the baculovirus transplacement vector pAc611 (Summers and Smith, 1987). The $T_r\alpha$ coding region is in a sense orientation relative to the polyhedrin promoter. pML 34(-) contains two copies of the Pst I fragment from pML 8 in the baculovirus transplacement vector pAc611 with the $T_r\alpha$ coding regions in an antisense orientation relative to the polyhedrin promoter.

pML 35. pML 35 (see Figures 21 and 22) was constructed by inserting the Nco I fragment of pML 8 into the yeast secretion vector pMFα8-Nco I. pMFα8-Nco I was made by inserting an Nco I linker (New England Biolabs #1040) in the Stu I site of pMFα8 (Miyajima *et al.*, 1985).

pML 38. pML 38 is pML 35 with a SnaB1/Hpa1 deletion which inactivates the yeast 2μm origin of replication.

pML 43. pML 43 is an Eco RI deletion derivative of pML 25.

pML 44. pML 44 (see Figure 23) was constructed by inserting the Sal I fragment from pML 8 into the Xho I site of the baculovirus transplacement vector pEV 55 (Miller, *et al.*, 1986).

pML 45. pML 45 (see Figures 24 and 25) was constructed by combining the Bgl

II/Sph I and Sph I/ Kpn I fragments of pML 44 with the Bam HI/Kpn I vector fragment of pAc360 (Summers and Smith, 1987).

pML 46. pML 46 (see Figures 26 and 27) was constructed by inserting the Nco I/Xba I fragment from pML 34 into the Nco I/Spe I site of pGD 108 (see below).

pML 49. pML 49 (see Figures 28 and 29) was constructed by inserting the Eco RI/Hind III fragment of pML 46 into the Eco RI/Hind III site of pBluescript/ KS^+.

pML 50. pML 50 (see Figure 30) was constructed by inserting the Hinc II/Hind III fragment of pML 46 into the Hinc II/Hind III site of pBluescript/ KS^+.

pML 52. pML 52 (see Figures 31 and 32) was constructed by inserting the Xba I/Hind III fragment of pML 46 into the Xba I/Hind III site of pBluescript/ KS^+.

pML 53. pML 53 (see Figures 33 and 34) was constructed by inserting the Nco I/Hind III fragment of the $T_c\alpha$ cDNA from M13 ml 1-Nco I into the Nco I/Hind III site of pGD 108.

pML 55. pML 55 (see Figure 35 and 36) was constructed by inserting an Nco I linker (New England Biolabs #1040) into the Sma I site of pGEX-2T.

pML 56. pML 56 (see Figures 37 and 38) was constructed by inserting the Nco I/Eco RI fragment of pML 34 into the Nco I/Eco RI site of pML 55.

pGD 108. pGD 108 (see Figures 39 and 40) was obtained from Gloria Dalbadie-McFarland (Caltech). It was constructed by ligating a 51 base pair double stranded synthetic DNA fragment (see figure) between the Eco R1 and Hind III site of pIN-III-ompA1 (Ghrayeb *et al.*, 1984).

M13 ml 1. M13 ml 1 (see Figure 41) was constructed by subcloning the Hind III fragment of the $T_c\alpha$ cDNA (Lochrie *et al.*, 1985) into M13 mp 11. Cells infected with M13 ml 1 secrete viral particles containing the antisense strand of the $T_c\alpha$ cDNA.

M13 ml 4. M13 ml 4 (see Figure 42) is the same as M13 ml 1 with the insert in the

opposite orientation. Cells infected with M13 ml 4 secrete viral particles containing the sense strand of the $T_c\alpha$ cDNA.

M13 ml 1-Nco I and M13 ml 91 were constructed by site-directed mutagenesis to facilitate subcloning the $T_c\alpha$ cDNA into expression vectors and to remove an ATG triplet preceeding the translation initiation codon of the $T_c\alpha$ cDNA by four base pairs.

M13 ml 1-Nco I. M13 ml 1-Nco I (see Figure 43) was constructed by oligonucleotide-directed site-specific mutagenesis (Kunkel *et al.*, 1987) of single stranded M13 ml 1 with a synthetic oligonucleotide (Tα25: 5' AGACGGATGAACCATGGGGAG TG 3') to create an Nco I site encompassing the translation initiation site. The strain RZ1032 was used to produce uracil containing DNA and candidate mutants were identified by Nco I restriction enzyme digestion. Out of six candidate mutants tested, five contained a Nco I site.

M13 ml 91. M13 ml 91 (see Figure 44) was constructed by oligonucleotide-directed site-specific mutagenesis (Smith and Zoller, 1983) of single stranded M13 ml 4 with a synthetic oligonucleotide (TαEco: 5' CACTCCCCATGAATTCATCC 3') to create an Eco RI site immediately preceeding the translation initiation codon. Mutants were initially identified by plaque hybridization to TαEco which was 5' end-labeled with ^{32}P. About one plaque per one thousand hybridized. Clones which hybridized to the probe were transformed into JM101, the replicative form was prepared, and the DNA was digested with Eco RI to confirm their identity.

Preparation of antipeptide antisera. Peptide Tα1AB has the sequence TTLNIQYGDSARQDDARKL(C) and corresponds to amino acids 85-103 of $T_r\alpha$. The peptide Tα2CD, EDKELAKRSKELE(C), corresponds to amino acids 8-20 of $T_c\alpha$. Cysteines in parentheses are not part of the natural protein sequence and were included to conjugate the peptide to immunogenic carrier proteins using SPDP (Pharmacia) as a

cross-linking agent. Each conjugated peptide (0.5 mg) was mixed with Freund's complete adjuvant and injected subcutaneously at 5-10 sites into the backs of two New Zealand White female rabbits. Booster injections of 0.5 mg conjugated peptide were given every two weeks in Freund's incomplete adjuvant for a period of 4-6 months. The rabbits were bled in the weeks between booster injections. Production of antipeptide antisera was monitored by dot blotting using the free synthetic peptides as antigens. In this procedure the free peptides were dissolved in PBS, 6 M guanidine-HCl at a concentration of 10 mg/ml and one microliter was spotted onto a strip of dry nitrocellulose. After drying, the blot was blocked, incubated with primary and secondary antibodies, and developed with BCIP and NBT as described below under "Western blotting." In addition to the free peptides purified retinal $T_r\alpha$ was dot blotted to test antisera against the Tα1 peptide. Cross reactivity to Tα1 and Tα2 peptides was first observed in sera collected four weeks after the primary injection. Antisera Tα1A and Tα1B were derived by injection of the Tα1AB peptide conjugated to keyhole limpet hemocyanin (KLH). Antisera Tα2C and Tα2D were derived by injection of the Tα2CD peptide conjugated to KLH and chicken ovalbumin, respectively.

Two other peptides were also synthesized for use as antigens: TαF: DIIKENLKD(G)GLF(C) (amino acids 338-350 in $T_r\alpha$ and 342-354 in $T_c\alpha$) and Tα2AB: PTLGIDYAEVSCVDNGRQL (amino acids 89-107 in $T_c\alpha$). In the TαF peptide, rather than using an internal cysteine found in the natural sequence, a cysteine was added to the end and the internal cysteine was replaced with glycine to prevent disulphide bond formation. In the Tα2AB peptide, a natural cysteine midway through the sequence was used for conjugation to carrier proteins. These peptides were injected into rabbits both as KLH conjugates and as free peptides. The TαF peptide conjugated to KLH, the unconjugated TαF peptide, the Tα2AB peptide conjugated to KLH, and the free Tα2AB

peptide were used to produce the TαF, TαG, TαA and TαB, and TαE sera, respectively.

Analysis of yeast cell extracts and media for $rT_r\alpha$ and $rT_c\alpha$ expression. Yeast cell extracts were prepared by the method of Lyons and Nelson (1984). Briefly, cells are lysed in 2 M NaOH, 5 mM β-mercaptoethanol. The proteins are precipitated with 10% trichloroacetic acid and then analyzed by Western blotting. Media was precipitated with 10% trichloroacetic acid and then analyzed by Western blotting or, if the volume of media was over 10 ml, it was concentrated by ultrafiltration and loaded directly on a gel for Western blot analysis. Retinal $T_r\alpha$ or *E. coli* cell extracts containing $rT_c\alpha$ were used as molecular weight standards.

Construction of recombinant baculoviruses. Basic techniques for Sf9 cell propagation, viral infection, the preparation of high titer viral stocks, and AcNPV genomic DNA purification are described in a methods manual published by the Texas Agricultural Experimental Station (Summers and Smith, 1987). AcNPV genomic DNA was prepared from the L1 strain. Transplacement vector DNA was purified by cesium chloride/ ethidium bromide density gradient centrifugation. For transfections AcNPV-L1 DNA (1μg) was mixed with transplacement vector DNA (2μg), precipitated with calcium phosphate, and transfected into Sf9 cells (method 2; Summers and Smith, 1987). After incubation at 28°C for 5 days the media was collected. Typically these "primary" viral stocks contained 0.1-1% recombinant virus. To maximize expression of recombinant proteins pure recombinant viruses free of wild-type virus must be used. In order to purify recombinant viruses containing the $rT_r\alpha$ cDNA a procedure combining limiting dilution with dot blotting was used. In each well of a 96-well microtiter plate 2 X 10^5 Sf9 cells were seeded. For each row of 12 wells 10-fold decreasing dilutions of the "primary" virus stock from transfected cells were added. These plates were incubated for 4-7 days. One half of the volume of each well (100 μl) was dot blotted onto nitrocellulose and hybridized to a ^{32}P-

labeled (Feinberg and Vogelstein, 1983) $T_r\alpha$ cDNA probe (pML10). Wells showing positive hybridization to the $rT_r\alpha$ cDNA were examined by inverted microscopy to look for viral occlusion bodies. Cells producing wild-type virus form occlusion bodies in the cell nucleus. The hydridization-positive wells with the lowest proportion of occluded cells were identified and the prodedure was repeated using viruses from those wells until wells were identified which contained no occluded cells and hybridized to the $T_r\alpha$ cDNA probe. These virus isolates were purified once more and a high titer stock prepared. Typically two to four rounds of purification were required to obtain stocks free of wild-type virus. The presence of the $T_r\alpha$ cDNA in AcNPV-ML34 was confirmed by Southern blotting (Maniatis, 1982) total nucleic acids from infected cells or by restriction endonuclease digestion of total cellular DNA. Total cellular DNA can be analyzed in this manner because viral DNA constitutes 25% of cellular DNA and insert-specific restriction fragments can easily be detected by ethidium bromide staining. The virus (AcNPV-Ac360βgal) that expresses a polyhedrin–β-galactosidase fusion protein was purified by plating virus stocks for plaques in media containing XGAL. Blue plaques were picked and replated three times. At this point all viral plaques were blue and occlusion-free.

Immunofluorescence of $rT_r\alpha$ expressed in Sf9 cells infected with AcNPV-ML45. Glass cover slips (22 mm^2) were rinsed in ethanol, sterilized 12 or more hours under a germicidal lamp in a tissue culture hood, and then placed in sterile 35mm diameter petri dishes. These were seeded with 10^6 Sf9 cells, 10^5 AcNPV-ML45 viruses, and incubated at 28°C. At varying times the coverslips were removed and fixed in 37% formaldehyde for 5 min in a plastic weigh boat with the cell-side facing up. The coverslips were washed in PBS (10 mM sodium phophate, 150 mM sodium chloride, pH 7.2)/0.2% Triton X-100 for 5 min, PBS for 5 min and, if necessary, stored in PBS, 0.1% sodium azide at 4°C. The coverslips were then incubated in a 1 /50 dilution of Tα1B antisera in PBS/1% bovine

serum albumin (BSA) for 30-60 min at 22°C on a strip of parafilm with the cell-side facing down, washed three times (5 min each) in a weigh boat with 5 ml PBS, incubated in a 1/100 dilution of goat anti-rabbit antisera conjugated to rhodamine (Cappel), washed three times (5 min each) in a weigh boat with 5 ml PBS, and mounted on a slide in 90% glycerol after blotting off excess PBS. The coverslip was sealed to the slide with fingernail polish and viewed under a Zeiss fluorescence microscope.

Analysis of E. coli whole cell lysates for $rT_r\alpha$ and $rT_c\alpha$ expression. Overnight cultures were grown in the presence of the appropriate antibiotics. A 1/100 dilution of the cells was grown and induced to express transducin alpha subunits as described in the figure legends. At the end of the induction period the optical density of the culture was measured. The desired amount of cells was centrifuged and each 1.0 OD_{600nm} of cells was resuspended in 50 µl of gel sample buffer (Laemmli, 1970). The samples were boiled for 3 minutes, centrifuged in a microfuge for five minutes, and subjected to SDS polyacrylamide gel electrophoresis according to Laemmli (1970) using a 10% acrylamide/0.27% bis-acrylamide gel. After electrophoresis the proteins were electroblotted to nitrocellulose filters as described by Towbin (1979) in a buffer containing 25 mM Tris-HCl, pH 8.3, 192 mM glycine, 20% methanol. The blots were then subjected to Western blotting.

Western blotting. Protein blots were rinsed in TBS (20 mM Tris-HCl, pH 7.5, 500 mM NaCl) and nonspecific protein binding sites were blocked by incubation in TBS/3% BSA for 1 hour. The blots were incubated in a 1/200 dilution of the primary antipeptide antisera in TBS/1% BSA for 1 or more hours at room temperature. The blots were washed three times (5 min each) in TBS and incubated with a goat anti-rabbit secondary antibody coupled to alkaline phosphatase (Promega) at a dilution of 1/7500 in TBS/1% BSA. The blots were washed three times (5 min each) in TBS and developed with BCIP and NBT as

recommended by the supplier (Promega). Quantitation of Western blot signals was determined using an LKB 2202 Ultroscan laser densitometer interfaced with an Apple IIe computer.

Preparation of rod transducin subunits. Frozen retinas (500) were thawed on ice overnight in the dark at 4°C. The following procedures were performed under dim red light (Kodak 1A Safelight) and all buffers were ice cold. All centrifugation rotors were precooled to 4°C, operated at 4°C, and loaded in the dark. The volume of the thawed retinas was adjusted to 800 ml with transducin buffer (10 mM Tris-HCl, pH 8.0, 1 mM DTT, 1mM $MgCl_2$, 0.1 mM PMSF) containing 45% sucrose. The suspension was stirred in a 2 L beaker for 10 min at a moderate speed to break off the rod outer segments. The preparation was divided into 4-200 ml aliquots in 4-500 ml centrifuge bottles and centrifuged in a GS-3 rotor at 5,000 rpm (4,230 x g) for 2 hours. Rod outer segments (ROS) and disc membranes were found floating, pelleted on the sides of the bottles, and occasionally on the surface of the pellet. All of the ROS were scraped into the supernatant with a rubber policeman. The supernatants were poured into 500 ml centrifuge bottles and ROS buffer (10 mM MOPS, 2 mM $MgCl_2$, 30 mM NaCl, 60 mM KCl, 100 mm PMSF, 1 mM DTT, pH 7.5) was added to the top of the bottle so that the final sucrose concentration was 20-25%. After mixing, the bottles were centrifuged in a GS-3 rotor at 5,000 rpm (4,230 x g) for 2 hours. The pellets contained disc membranes (disc membranes pellet in less than about 28% sucrose). These were resuspended in 80 ml ROS buffer with a homogenizer and centrifuged in an SS-34 rotor at 18,000 rpm (40,000 x g) for 30 min. The pellets were resuspended in 80 ml transducin buffer, poured into a 250 ml beaker, and exposed to room lights on ice for 10 min with occasional agitation. During this time rhodopsin bleached from a deep orange to a light orange-yellow color. All subsequent steps were performed under room lights. After photolysis the preparation was centrifuged

in an SS-34 rotor at 18,000 rpm for 30 min. The supernatant contained the first phosphodiesterase extract. The pellet was resuspended in 80 ml transducin buffer and centrifuged in an SS-34 rotor at 18,000 rpm for 30 min three more times to obtain a second, third, and fourth extract. The final pellet was resuspended by homogenization in transducin buffer plus guanine nucleotide. Depending on experimental requirements either 10 μM GTPγS or 100 μM GTP was used. This suspension was centrifuged in an SS-34 rotor at 18,000 rpm for 30 min. The supernatant contained the first transducin extract. The membrane pellets were washed three more times with transducin buffer plus guanine nucleotide. The phosphodiesterase and transducin extracts were centrifuged in a Ti45 rotor at 40,000 rpm (200,000 x g) for 2 or more hours to remove residual membranes. After this centrifugation step the extracts were collected carefully to avoid removing any pelleted material. The extracts were concentrated by ultrafiltration using an Amicon PM-30 filter. Typically 500 mg of rod outer segment membranes, 50 mg of transducin extract, and 50 mg of phosphodiesterase extract were obtained. The transducin and rhodopsin obtained by this procedure were usually over 90% pure while the phosphodiesterase is usually about 10-20% pure. $T_r\alpha$ was separated from $T\beta\gamma$ on a Blue Sepharose column by the procedure of Yamazaki *et al.* (1988).

Urea stripped rod disc membranes. Bovine rod disc membranes at a concentration of 6 mg/ml were homogenized in 10 ml transducin buffer, centrifuged, homogenized in 10 ml transducin buffer/4 M urea, centrifuged, homogenized in 20 ml transducin buffer, centrifuged, resuspended in 600 μl transducin buffer, and stored in aliquots at -70°C.

cGMP phosphodiesterase assay. The activation of bovine retinal cGMP phosphodiesterase was measured continuously by the pH method of Yee and Liebman (1978). The pH microelectrode (Sigma model E4878) was connected to a digital pH meter (Orion model 601A), and the output was amplified on a chart recorder (Linear Instruments

model 291/MM). Using this system one inch on the chart paper corresponds to 0.2 pH unit. The sample to be assayed was in a final volume of 490 µl ROS buffer, 2 mM cGMP. Assays were incubated in the dark in a 1.5 ml Eppendorf centrifuge tube which was taped to the end of the pH electrode and immersed in a beaker of ice water. The assay was mixed constantly with a 5mm long stir bar. The electrode was preequilibrated in the sample until the pH was stabilized, usually after a few minutes. The assay was then initiated by the addition of 10 µl urea stripped rod outer segment membranes containing 140 µg protein. In some cases 10 µM GTPγS or 100 µM GTP was also added to initiate the assay.

^{35}S-GTPγS binding. Experimental assays contained 90µl of sample in transducin buffer and were initiated by adding 10 µl of 1 µM ^{35}S-GTPγS, 0.8 µg photolyzed, urea stripped rod outer segment membranes, and 0.3 µg $T_{\beta\gamma}$ in transducin buffer. Components were incubated under room lights as indicated in the text. (Some assays were incubated on ice for 5 min to minimize potential proteolysis. Others were incubated at 37 °C for 30 min to maximize binding sensitivity.) Control assays contained 90 µl of sample in transducin buffer and were initiated by adding 10µl 1µM ^{35}S-GTPγS in transducin buffer. After incubation, 900 µl of ice cold transducin buffer was added and the samples were filtered through HA filters (Millipore) using a 12-sample vacuum filtration manifold (Millipore). The filters were washed three times with 2 ml ice cold transducin buffer, dried in the air or under a heat lamp, dissolved in ACS (Beckman), and then subjected to liquid scintillation counting. The amount of ^{35}S-GTPγS bound in the control assay was subtracted from that bound in the experimental assay to obtain the amount of rhodopsin/$T_{\beta\gamma}$ stimulated-GTPγS binding. The amount of ^{35}S-GTPγS bound to rhodopsin/$T_{\beta\gamma}$ was negligible.

Preparation of $gT_r\alpha$ from E. coli pML 56. CAG1139 (Grossman *et al.*, 1983) F' lac I^{sq} lac Z_{u118} pro AB Tn 5 (Johnson *et al.*, 1982) was transformed with pML 56. A saturated culture of this strain was prepared in LB plus ampicillin (100 µg/ml). The culture

was diluted 1/10 in 100 ml LB plus ampicillin and grown for 1 hour at 30°C. The culture was induced for 2 hours with 0.1 mM IPTG. After induction the cells were pelleted and lysed in transducin buffer by passage through a French pressure cell three times at 18,000 psi. The cell lysate was centrifuged at 12,000 X *g* for 10 min. This soluble lysate was incubated with 0.3 ml of a 50% slurry of glutathione-agarose in transducin buffer for 10 min at 4°C. The beads were pelleted for 15 sec at 4°C in a clinical centrifuge at 1,000 rpm. The beads were washed three times in transducin buffer and the GST-r$T_r\alpha$ fusion was eluted from the beads in 1 ml 50 mM Tris-HCl, pH 8.0, 5 mM reduced glutathione, 1 mM $MgCl_2$, 1 mM DTT. This buffer lacks PMSF because PMSF inactivates the thrombin used in the next step. The GST-r$T_r\alpha$ fusion was cleaved into GST and r$T_r\alpha$ components by thrombin proteolysis in a buffer containing 50 mM Tris-HCl, pH 8.0, 2.5 mM $CaCl_2$, 150 mM NaCl , 1 mg/ml GST-r$T_r\alpha$, and 1 ng/ml thrombin at 25°C for 30 min.

Trypsin cleavage of T_r α and gT_r α. Samples containing 1-5 μg/ml protein were incubated on ice for one hour in 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, and 50 μg/ml TPCK-treated trypsin in a final volume of 200 ml. The samples were precipitated with trichloroacetic acid and subjected to Western blotting.

Partial purification of r$T_r\alpha$ from E. coli pML 46. CAG1139 F' lac I^{sq} lac Z_{u118} pro AB Tn 5 was transformed with pML46 and grown to saturation at 30°C in LB media containing ampicillin (50 mg/ml) and kanamycin (40 mg/ml) to select for maintenance of pML46 and the F' factor, respectively. The culture was diluted 100-fold and grown at 30°C with aeration to OD_{600}=0.7. To induce the synthesis of r$T_r\alpha$, IPTG was added to a final concentration of 100 μM and growth was continued for 30 min. All subsequent steps were conducted at 2-5°C. The cells were harvested by centrifugation at 5,000 X *g* for 10 min, resuspended in 10 ml transducin buffer (10 mM Tris-HCl, 1 mM $MgCl_2$, 1 mM DTT, 0.1 mM PMSF, pH 7.4) per liter of cells. The cells were disrupted by passage through a

one inch diameter French pressure cell three times at 18,000 psi. The cell lysate was centrifuged at 100,000 x *g* for one hour. Ammonium sulfate (242 g/L) was slowly added to the supernatant to allow gradual dissolution and after stirring for 30 min the precipitated protein was collected by centrifugation at 27,000 X *g* for 10 min. The pelleted protein was resuspended in HA buffer (50 mM Tris-HCl, 200 mM NaCl, 1 mM DTT, 1 mM $MgCl_2$, 0.1 mM PMSF)/0.5 M $(NH_4)_2SO_4$ and applied to a heptyl agarose column (≤10 mg protein/ ml of column) equilibrated in HA buffer/0.5 M $(NH_4)_2SO_4$. The column was washed in HA buffer/0.5 M $(NH_4)_2SO_4$, HA buffer/0.2 M $(NH_4)_2SO_4$, HA buffer/0.0 M $(NH_4)_2SO_4$ and eluted with transducin buffer. The eluate was loaded directly onto a DEAE-Sepharose column (1 mg protein/ml of column). The column was washed with transducin buffer and eluted with five column volumes of a linear 0-0.5 M NaCl gradient in transducin buffer. Prior to subsequent analysis the samples were concentrated by ultrafiltration using a YM-10 membrane (Amicon) or dialyzed against transducin buffer. The chromatographic behavior of $rT_r\alpha$ was monitored by Western blotting.

Miscellaneous methods. Yeast were transformed by the method of Klebe *et al.* (1983). All *E. coli* cells were transformed by the $CaCl_2$ procedure (Maniatis *et al.*, 1982) Relative protein concentrations were determined using the Bradford reagent (BioRad) with bovine serum albumin as the standard. Molecular weight standards used for SDS gel electrophoresis were bovine serum albumin (66 kDal), chicken ovalbumin (45 kDal), retinal $T_r\alpha$ (39 kDal), and carbonic anhydrase (30 kDal).

RESULTS AND DISCUSSION

Specificity of peptide sera.

Initially antisera against the Tα1AB ($T_r\alpha$) and Tα2CD ($T_c\alpha$) peptides were screened by dot blotting. The Tα1A sera recognizes the Tα1AB peptide and retinal $T_r\alpha$ (Figure

47B). The Tα2C sera recognizes the Tα2CD peptide, but does not recognize retinal $T_r\alpha$ (Figure 47C). The Tα2A, TαE, and TαF sera recognize the corresponding synthetic peptides (Figures 47C,47D, and 47E). However, none of these sera recognize retinal $T_r\alpha$, $rT_r\alpha$, or $rT_c\alpha$. Similar negative results were obtained with the Tα2B and TαG sera.

To further demonstrate the specificity of the Tα1A and Tα2C sera, they were tested on Western blots of *E. coli* extracts containing $rT_r\alpha$ and $rT_c\alpha$. It was found that the Tα1A antisera cross-reacts specifically with retinal $T_r\alpha$ and $rT_r\alpha$ expressed in *E. coli* but not with $rT_c\alpha$ (Figure 48B). The Tα2C antisera cross-reacts specifically with $rT_c\alpha$ expressed in *E. coli* but not with retinal $T_r\alpha$ or $rT_r\alpha$ (Figure 48D). Equivalent results were obtained with the Tα1B and Tα2D sera. The preimmune sera do not recognize $rT_r\alpha$, $rT_c\alpha$, or retinal $rT_r\alpha$ (Figures 48A and 48C).

The Tα1AB and Tα2CD peptide sequences are located in regions that are diverse among G protein alpha subunits. The closest match of the Tα1AB ($T_r\alpha$) peptide to known G proteins is to $G_{i\text{-}1}\alpha$ where 11 out of 20 amino acids match. The Tα1AB peptide sequence matches the corresponding region of $T_c\alpha$ in 7 out of 20 positions. The closest match of the Tα2CD ($T_c\alpha$) peptide to known G proteins is to $G_{i\text{-}2}\alpha$ where 7 out of 14 amino acids match. The Tα2CD peptide sequence matches the corresponding region of $T_r\alpha$ in 6 out of 14 positions. Therefore the Tα1A, Tα1B, Tα2C, and Tα2D antisera are likely to be suitable for specifically detecting bovine transducin rod and cone alpha subunits produced in heterologous expression systems.

Expression of $rT_r\alpha$ in yeast.

The yeast *Saccharomyces cerevisiae* was investigated as a candidate organism for transducin alpha subunit expression because its properties can be considered to be intermediate between bacteria and mammalian systems (Table 1). Like *E. coli*, it is simple to grow and can be genetically manipulated with ease, yet it has many aspects of a typical

eukaryotic cell. For example it performs most of the posttranslational modifications found in eukaryotic proteins including myristylation.

Since yeast has been shown to have G proteins (Miyajima *et al.*, 1987; Dietzel and Kurjan, 1987; Nakafuku *et al.*, 1988; Whiteway *et al.*, 1989), it is possible that mammalian G proteins expressed in yeast would be properly synthesized and localized. One class of yeast G proteins have been found to be involved in the respose to α-factor, a mating pheromone. In this response pathway the βγ subunits appear to regulate an as yet unidentified effector that leads to arrest in the G1 phase of the cell cycle. Disruption of either the β (STE 4) or the γ (STE 18) gene results in sterility because the mating pathway cannot be initiated. Disruption of the α subunit gene (SCG 1) results in constitutive activation of the mating response (arrest in the G1 phase of the cell cycle or death) ,because the βγ subunit complex is not bound to the α subunit where it is normally inactive until the mating horomone receptor binds α−factor. One concern about expressing transducin alpha subunits in yeast is whether their expression would be detrimental to yeast. For example a foreign G protein might interfere with the normal function of a yeast G protein. Retinal and yeast G protein alpha subunits are not closely related. However rat $G_s\alpha$, which is also not closely related to yeast G proteins, can complement yeast mutants deficient in SCG1 and allow them to grow normally (Dietzel and Kurjan, 1987). Rat $G_o\alpha$ complements a SCG1 mutant ,weakly, but the closely related rat $G_{i\text{-}2}\alpha$ does not complement. The ability of $T_r\alpha$ to complement a SCG 1 mutant was not investigated in these studies. Rat $G_o\alpha$ and $G_{i\text{-}2}\alpha$ are much more simliar to transducin than to $G_s\alpha$. However all of the mammalian Gα proteins are equally distant from SCG1 in terms of overall sequence homology. The aspects of $G_s\alpha$ which allow it to complement SCG1 are unclear.

The plasmid pML 35 (Figure 21) was constructed to secrete $rT_r\alpha$ from yeast. In pML 35, the prepro leader sequence of the yeast α-factor mating horomone is fused to the

entire coding region of $rT_r\alpha$. Protease cleavage sites recognized by Kex2, an enzyme involved in maturation of the α-factor precursor, are present upstream from the $rT_r\alpha$ sequence. Kex2 cleaves the alpha factor precursor after dibasic Lys-Arg sequences. It can also cleave after Arg-Lys and Arg-Arg pairs, but not well, and it cleaves after Lys-Lys poorly (Fuller *et al.*, 1989). Plasmids similar to pML 35 have been successfully used for expression of several mammalian proteins (Miyajima *et al.*, 1985; Brake, 1990). The design of pML35 is such that $rT_r\alpha$ produced after proteolysis should have one extra amino acid (proline) at the amino terminus.

One advantage of secreting $rT_r\alpha$ from yeast is that it can be easily purified. *Saccharomyces cerevisiae* secretes few proteins other than the small (1 kDal) peptide mating pheromones. On the other hand $rT_r\alpha$ may not enter the yeast secretory pathway efficiently since $rT_r\alpha$ is not normally secreted in retina. Also the molecular weight of $rT_r\alpha$ (40 kDal) may be too large to be efficiently secreted from yeast. Most foreign proteins that have been successfully secreted from yeast are less than 20 kDal. Even if $rT_r\alpha$ did enter the secretory pathway, the possiblity exists that $rT_r\alpha$ would be aberrently processed in the secretory pathway. For example, $T_r\alpha$ contains one potential site for N-linked glycosylation (Asn-Asp-Ser at position 145-147).

pML 35 was transformed into several trp^- yeast strains (EJ102, 20B-12, NNY105, DSR865-3B, HR125a, HR125α, SM1269, and SM1291). EJ102 and 20B-12 are pep4-3 strains which are deficient in proteinase A. Proteinase A is essential for the maturation of several vacuolar proteases including proteinase B, carboxypeptidase Y, aminopeptidase I and also RNase and alkaline phosphatase (Jones, 1990). Since cell lysis releases these proteases and the heat and SDS used for gel sample preparation actually activates proteinase B, the pep4-3 mutation has been useful the detecting "unstable" proteins in yeast. The strain 20B-12 has been found to be one of the best strains for producing α-factor fusion

proteins in yeast (A. Miyajima, personal communication). NNY105 is a kex2$^-$ strain. In NNY105 pML 35 cells an uncleaved prepro-α-factor-$rT_r\alpha$ fusion would be expected to be observed. This could serve as a useful control. $T_r\alpha$ has no Lys-Arg pairs, but it does have an Arg-Arg pair at position 309-310 and Arg-Lys pairs at 101-102 and 204-205. SM1269 is a strain that lacks a gene called RAM. The precise function of RAM is unclear but it appears to be involved in some type of C-terminal processing of proteins such as *ras*, **a**-factor and the alpha and gamma subunits of G proteins which have cysteines four amino acids from the C-terminal end. SM1269 was transformed to determine if any differences could be detected between the $rT_r\alpha$ produced in this strain and an isogenic strain which carries the RAM gene on a plasmid (SM1291). HR125**a** is an **a** mating-type strain which serves as a negative control. The α-factor promoter is expressed 100-fold less efficiently in **a** cells than in α mating-type cells.

The pML 35 transformants were grown for two days in the absence of tryptophan in liquid culture to provide selective pressure for maintaining pML 35 and then tested for $rT_r\alpha$ production by Western blotting. $rT_r\alpha$ was not detected in cells or in the media in any experiment. Representative results are shown in Figures 49 and 50. Expression was also not observed in SM1269 or SM1291. The detection limit was about 0.05% of total protein.

The reason for the inability to detect $rT_r\alpha$ expression is not known. One possibility is that the structure of pML 35 is incorrect. For example, the NcoI linkers used to join the α-factor leader to the $rT_r\alpha$ coding region may have had an incorrect sequence. However, this cannot be the case since DNA sequence analysis of the prepro α−factor/$rT_r\alpha$ junction in pML 35 confirmed that the correct junctional sequence was present.

Another possibility is that $rT_r\alpha$ may be rapidly degraded even in the pep4-3 strains that were tested. Pulse-chase experiments have not been performed to investigate the

kinetics of the synthesis and degradation of yeast $rT_T\alpha$. Many foreign proteins expressed in yeast are unstable (Emr, 1990). Using pep4-3 mutations does not always alleviate this instability as there are a variety of other catabolic enzymes which are not reduced in pep4-3 strains.

Another reason for the absence of $rT_T\alpha$ expression in pML 35 containing strains may be that $rT_T\alpha$ is toxic to yeast cells and that nonexpressing mutants were selected. There is some indication that this may be the case. For example, the colony size, the saturation density (Table 4), and the amount of plasmid in α–type cells carrying pML 35 (Table 5) was often, but not always, reduced compared to cells with the parent vector pMFα8. Some cells such as 20B-12 pML 35 did not grow at all in liquid YPD media even though transformants could be obtained. About half of the DSR865-3B pML 35and EJ102 pML 35 transformants grew poorly. Some of these lysed when grown in YPD liquid media. On the other hand there are arguments against the idea that pML 35 is toxic to yeast. For example there was no significant difference in the number of transformants obtained using pML 35 or pMFα8 (Table 6). In particular the number of HR125α transformants was the same as the number of HR125a transformants.

The vector pML 38 was constructed for integrating the pML 35 expression unit into the yeast genome. As would be expected if pML 38 had integrated, no plasmids were recovered in *E. coli* from candidate pML 38 integrants (Table 5). Integrants containing pML 38 were even sicker than cells carrying pML 35. The colony size of pML 38 integrants was significantly smaller than the same strain containing pML 35. In particular NNY 105 pML 38 grew very poorly (Table 4).

Expression of $rT_C\alpha$ in yeast.

To study cytoplasmic expression of $rT_C\alpha$, a plasmid (pML 11; see Figure 8) was constructed from the yeast expression vector pAAH 5 (Ammerer, 1983) which contains the

constitutive high-level alcohol dehydrogenase promoter. Strains SF838-10α and SR741-3B were transformed with pML 11(+) or pML 11(-). The SF838-10α transformants required 6 days to form a colony rather than the normal 2-3 days. The SR741-3B transformants arose in the normal time. However, no significant differences in the growth behavior of cells with pML 11(+) or pML 11(-) were observed compared to the parental strains.

When these strains were grown in a selective media (SD without leucine) no $rT_c\alpha$ was observed on Western blots of cell extracts from these strains (Figure 51, Lanes 8-13). When nonselective media (YPD) were used, the pML 11(+) and pML 11(-) strains contained a 43kDal protein cross-reactive with the $T_c\alpha$ antisera (Figure 51, Lanes 2-5) that was not observed in the parental strains (Figure 51, Lanes 6 and 7). Since both the pML 11(+) and pML 11(-) strains had the 43 kDal protein, it seemed likely that it was not $rT_c\alpha$. It is possibile that the 43 kDal protein was plasmid encoded and not related to $rT_c\alpha$. However extracts from strains with the parental plasmid pAAH5 were not examined.

One likely explanation for the lack of $rT_c\alpha$ expression from pML 11 is incorrect translation initiation. Yeast have a strong preference for beginning translation at the first ATG codon on the mRNA (Donahue and Cigan, 1990). Since the $T_c\alpha$ cDNA has an ATG codon preceeding the true translation initiation codon by four base pairs (see Figure 9) the yeast ribosome may start translation at this codon thereby precluding efficient translation at the second, correct one. In yeast the efficiency of translation initiation at a second codon has been observed to be no more than 10-20% that of the first.

Expression of $rT_r\alpha$ in Sf9 insect cells.

Since yeast did not appear to provide adequate expression of transducin alpha subunits another eukaryotic expression system was chosen. Recently an expression system has been developed that utilizes baculoviruses as expression vectors and insect cells

as hosts. This system has a number of unique features. Several proteins have been expressed in this system successsfully (Luckow and Summers, 1988). Most of these are found at their normal cellular locations and are posttranslationally modified properly. The main difference observed between mammalian proteins produced in the insect cell system and in mammalian cell systems is in their glycosylation patterns, since insect cells do not make sialic acid. However this should not present a problem in the expression of $rT_r\alpha$ in insect cells since it is not known to be glycosylated.

Baculoviruses have a very narrow host range. They productively infect about 25 *Lepidoptera* species. During the infectious cycle two types of virions are produced. One is extracellular and one is intracellular. The extracellular virus is made early in the infectious cycle (0-24 hours post-infection) and is produced by extrusion from the cell membrane. The intracellular form is made later in the infectious cycle (24-72 hours post-infection) and is found within the nucleus in bodies called occlusions. Polyhedrin is the major constituent of occlusion bodies and is expressed at levels up to 75% of the cell's protein. Occlusions protect virions from potentially adverse environmental conditions after the host cell is killed and lyses. Polyhedrin is not an essential gene, however, and can be replaced with no effect on the production of extracellular virus. These observations provided the basis for the development of the baculovirus expression system. Foreign genes are introduced onto the baculovirus genome under the control of the polyhedrin promoter through recombination of the genome with "transplacement" vectors since direct subcloning of genes onto the baculovirus genome is not possible because of its large size (about 130 kilobase pairs). These vectors contain the polyhedrin promoter followed by a restriction enzyme site polylinker where the gene of interest is subcloned. The promoter region is flanked by enough flanking baculovirus DNA to allow efficient recombination onto the viral genome after cells are transfected with a mixture of baculovirus genomic DNA and the

transplacement vector. In recombinant viruses the polyhedrin gene is disrupted. Cells infected with such viruses do not form occlusions and can easily be identified visually.

There are several advantages to the baculovirus system. The insect cell hosts perform most of the posttranslational modification that eukaryotic cells do. They are large cells that can grow to a density of 2×10^8 cells per liter so that the amount of total protein per liter of cells is about 25 times the amount from a standard "bench-top" *E. coli* culture. The polyhedrin promoter is tightly regulated and is only transcribed late in the infectious cycle. Therefore foreign gene products that are normally toxic to other cells can often be produced in the baculovirus host. The main disadvantages of the system are the long time it takes to construct recombinant viruses and the length of the expression period.

To become familiar with the procedures involved in using the baculovirus expression system a recombinant virus (AcNPV-Ac360βgal) was constructed which expresses a polyhedrin–*E. coli* β-galactosidase fusion protein. The AcNPV-Ac360βgal virus was present in the initial virus stock at a frequency of 0.1% which is the normally observed frequency. Sf9 cells infected with AcNPV-Ac360βgal produce a polyhedrin-β-galactosid -ase at a level that represents about 25% of total cell protein (Figure 52, Lane 2). The fusion is enzymatically active since XGAL was cleaved by cells infected with AcNPV-Ac360βgal to result in blue-colored plaques (Figure 53). In liquid culture media containing XGAL the cytoplasm of the cells turned blue and by 7 days post-infection, when cell lysis is almost complete, the media also turned blue.

To examine expression of $rT_r\alpha$ in insect cells three transplacement vectors (pML 34(+), Figure 19; pML 34(-); and pML 45, Figure 24) were used to construct three recombinant baculoviruses (AcNPV-ML34(+), AcNPV-ML34(-), and AcNPV-ML45). AcNPV-ML34(+) was designed to direct the synthesis of a full-length protein. AcNPV-ML34(-) is a negative control which should not synthesize $rT_r\alpha$ since the coding

region is in the antisense orientation. AcNPV-ML45 was designed to synthesize a fusion protein that consists of the first twelve amino acids of polyhedrin ,followed by seven amino acids derived from the polylinker sequence used in subcloning, and then the complete coding region of $T_r\alpha$ (350 amino acids). Each virus was constructed and purified as described in Materials and Methods.

The time course of expression of $rT_r\alpha$ in cells infected with AcNPV-ML34(+) is shown in Figure 54. Expression is first observed about 24 hours post-infection and peaks at about 48-72 hours post-infection. After about 96 hours post-infection the amount of $rT_r\alpha$ per cell is about the same as at 72 hours post-infection, but the total amount of cell-associated $rT_r\alpha$ that can be recovered per culture dish declines because of cell lysis. This pattern of expression corresponds well to the pattern reported previously for polyhedrin, indicating that the temporal pattern of expression of the polyhedrin promoter is retained in AcNPV-ML34(+)-infected cells. Cells infected with AcNPV-ML34(-) do not express $rT_r\alpha$ as detected by Western blotting. A similar expression pattern was observed in AcNPV-ML45 infected cells monitored by indirect immunofluorescence (Figures 57-60). In this case cells expressed no detectable polyhedrin-$rT_r\alpha$ at 0 or 24 hours post-infection. By 36, 48, and 72 hours post-infection about 30%, 50%, and 70% of the cells, respectively, were positive for polyhedrin-$rT_r\alpha$ expression. From about 96-120 hours after infection with wild-type AcNPV, the nuclear envelope breaks down and cell lysis occurs. This was also evident in AcNPV-ML45-infected cells. By 120 hours post-infection with AcNPV-ML45 over 90% of the cells had lysed. In the cells that did not lyse, the polyhedrin-$rT_r\alpha$ was distributed throughout the cytoplasm (Figure 61). At an earlier time point the polyhedrin-$rT_r\alpha$ is clearly excluded from the nucleus. No staining was detected with preimmune sera.

Uninfected cells have a "pancake" appearance when fixed and viewed with Nomarski

optics (Figure 55). By about 24 hours after infection the cells have enlarged and have a "fried egg" appearance where the nuclear membrane and the nucleolus are clearly distinguishable. Not all cells that where infected with AcNPV-ML45 expressed polyhedrin-r$T_r\alpha$ (compare Figure 59A with 59B). This could be because the physiology of the nonexpressing cells is different from the rest or, more likely, that expression is asynchronous even though infection is synchronous.

The levels of expression observed in AcNPV-ML34(+)-infected cells ranged from about 0.5-10 mg r$T_r\alpha$/L of cells with an average of 1 mg r$T_r\alpha$/L. The levels of polyhedrin-r$T_r\alpha$/L in AcNPV-ML45-infected cells ranged from 50-500 mg/L. The average was about 100 mg r$T_r\alpha$/L, which represents about 2% of total insect cell protein. The levels of expression were variable and are known to depend on various factors such as multiplicity of infection, cell physiology and density, cell culture procedures, and media composition. The mobility of the expressed proteins on SDS gels was 40 kDal for r$T_r\alpha$ and 42 kDal for polyhedrin- r$T_r\alpha$ as expected.

Upon lysing insect cells expressing either r$T_r\alpha$ or polyhedrin-r$T_r\alpha$, typically over 98-99% of the r$T_r\alpha$-related protein was found in the high-speed pellet after ultracentrifugation (Figure 62). The material that was soluble was only about 1% pure, which makes it no purer than r$T_r\alpha$ produced in *E. coli* (see next section). The insoluble polyhedrin-r$T_r\alpha$ is the major protein in the insoluble pelleted material. Typically 25-80% of the pelleted material was polyhedrin-r$T_r\alpha$ (see Figure 52, Lane 3 for example). This material is insoluble in the same buffers that retinal $T_r\alpha$ is soluble in, at concentrations up to 10 mg/ml. Several buffers which do not inactivate $T_r\alpha$ activity were tested for their effectiveness at solubilizing polyhedrin-r$T_r\alpha$. This includes 2 M NaCl, 10 μM GTPγS, 0.5% Lubrol, 2% cholate, 2% tetradecyl trimethyl ammonium bromide, and 1% CHAPS. The zwitterionic detergent CHAPS was found to be most effective at solubilizing

polyhedrin-r$T_r\alpha$. At a concentration of 1% CHAPS, about 1-5% of the total polyhedrin-r$T_r\alpha$ or r$T_r\alpha$ was solubilized. Only buffers containing strong protein denaturants (e.g., 1% SDS, 6 M guanidine-HCl) were significantly effective in solubilization. For example, 6 M guanidine solubilized 90% of the r$T_r\alpha$ synthesized. Attempts to "renature" the insoluble polyhedrin-r$T_r\alpha$ by solubilization in 6 M guanidine-HCL followed by slow dialysis led to reprecipitation.

It was thought that localization of the expressed protein by indirect immunofluorescence imaging might provide provide insight into its insolubiliy. The level of expression of r$T_r\alpha$ made in AcNPV-ML34(+)-infected cells was not high enough to detect (Figure 56). Therefore cells infected with AcNPV-ML45 were used. Figures 58-60 show that at least some of the insect cell polyhedrin-r$T_r\alpha$ exists in a perinuclear reticular network which contains cytoplasmic nodules reminiscent of "inclusion bodies" observed in bacterial cells that express high levels of foreign gene products. The polyhedrin-r$T_r\alpha$ appears to be in the endoplasmic reticulum/golgi area. It has recently been reported that $G_{i\text{-}2}\alpha$ is found in the golgi of epithelial cells (Ercolani *et al.*, 1990) whereas $G_{i\text{-}3}\alpha$ was found in the basolateral membrane. $G_{i\text{-}2}\alpha$ and $G_{i\text{-}3}\alpha$ are about 90% identical and yet appear to undergo different fates. Therefore subtle change in protein structure could alter localization. Thus it is not known if the localization of polyhedrin-r$T_r\alpha$ in Sf9 cells is a real reflection of the synthetic pathway taken by $T_r\alpha$ in photoreceptors or an artifact of it being a fusion protein in a heterologous cell. Very little is known about the biosynthetic pathway of $T_r\alpha$ in photoreceptor cells. It would be expected that r$T_r\alpha$ made in Sf9 cells should go to the cytoplasmic surface of the plasma membrane since this is the topological equivalent of the cytoplasmic side of the disc membranes in photoreceptor cells where transducin is found. Rhodopsin expressed in COS cells goes to the plasma membrane. In the studies reported here it was not determined how much polyhedrin-r$T_r\alpha$ was in the

plasma membrane because the cells were treated with Triton X-100 to allow antibodies access to the antigen. Thus it is possible that in these cells some of the polyhedrin-$rT_r\alpha$ made actually does go to the plasma membrane, but because of the large amount that is made, the membrane sites become saturated and the excess ends up localized around the nucleus. Nevertheless it is interesting why it ends up there rather than dispersed throughout the cytoplasm. Perhaps in retina $T_r\alpha$ normally goes to the golgi before going to the disc membranes. $G\alpha$ proteins are myrisylated. There is evidence from studying viral capsid proteins that myristylation may be a signal for targeting proteins to the plasma membrane. If this is true for G protein alpha subunits then polyhedrin-$rT_r\alpha$ may become detained in the normal course of its synthetic path because it is not myristylated. (Polyhedrin-$rT_r\alpha$ would not be myristylated because it lacks a glycine as the second amino acid.)

In any case the high levels of fusion protein in AcNPV-ML45-infected cells do not appear to be responsible for its insolubility. AcNPV-ML34(+)-infected cells make about 100-fold less $rT_r\alpha$, yet it is also over 98% insoluble. Also, polyhedrin-$rT_r\alpha$ is no more soluble 24 hours after infection when 10-fold less protein is present than at later time points when the amount of polyhedrin-$rT_r\alpha$ is at a peak.

The insoluble nature of polyhedrin-$rT_r\alpha$ and $rT_r\alpha$ in Sf9 cells was unexpected. Out of over 100 proteins that have been expressed in the baculovirus system, almost all are properly localized and behave as they do in their natural cell environment. However there are a few exceptions. Entactin is normally secreted. In Sf9 cells it remains entirely in the cytoplasm in an insoluble aggregate (Tsao *et al.*, 1990). The lymphocytic choriomeningitis virus nucleocapsid normally goes to the plasma membrane but in insect cells it forms cytoplasmic inclusion bodies (Matsuura *et al.*, 1987).

Although the levels of protein made in the baculovirus expression system are

impressive several observations made the use of another system more attractive. First, the time involved in baculvoirus vector construction (5-8 weeks per virus) is excessive. As a result it is impractical to make the number of recombinant viruses that would be required for extensive mutagenesis studies. Second, the time course of protein expression is long (2-3 days) and expression levels are not reproducible. Third, only a small amount of the expressed protein is soluble and it is no more than 1% pure. Finally Sf9 cells contain G proteins. Therefore biochemical assays would have to be carefully controlled by comparing the activity $T_r\alpha$-containing extracts to extracts from wild-type AcNPV or AcNPV-Ac360βgal-infected cells to rule out any effects endogenous Sf9 G proteins might have. Perhaps the only use of the baculovirus system in the context of this study would be for obtaining antigen for antibody production, since milligram amounts can be obtained in a nearly pure state.

Expression of $rT_r\alpha$ in *Escherichia coli* using efficient promoters.

E. coli has been used more than any other organism as a host for heterologous expression studies. It is easy to grow, has well-characterized genetics and mutagenesis procedures, and several well-regulated, highly active promoters are available. Many eukaryotic proteins have been produced in *E. coli* in an active form. Many have not.

Several different promoters that function efficiently in *E. coli* (Table 3) were tested for the ability to synthesize $rT_r\alpha$ (p_{trc} in pML 12, λp_L in pML 17, λp_R in pML 15 and pML 25, and p_{T7} in pML49). Several hosts strains were also tested. Overall the strain CAG1139 produced the most $rT_r\alpha$ from any given vector. Expression from some vectors (e.g., pML 46) was not observed in other strains. CAG1139 is deficient in the lon protease. The lon protease has been implicated as a major, but certainly not the only, protease involved in the degradation of abnormal and foreign gene products in *E. coli.* (Gottesman, 1990). Another strain, CAG629, was found to result in $rT_r\alpha$ expression

levels that were 2-fold or 3-fold higher than in CAG1139 on average. CAG629 is deficient in the lon protease and it also has a temperature-sensitive mutation in the htpR (rpoH) gene. The htpR gene is a heat shock-induced transcription factor that regulates proteases involved in the degradation of proteins in response to high temperatures or other stress such as the overproduction of foreign proteins. Others have also reported that the htpR, lon double mutant overproduces foreign proteins better than the single lon mutant. However CAG629 transforms with DNA very poorly and grows very slowly even at the permissive temperature (30°C) compared to CAG1139. CAG1139 also does not transform well, but it grows at about the same rate as wild-type *E. coli* strains. Therefore CAG1139 was used for most of the expression experiments described here.

In an initial attempt to obtain expression of $rT_r\alpha$ in *E. coli*, the vector pML 12 (see Figure 10) was constructed. pML 12 contains the p_{trc} promoter. The p_{trc} promoter differs from the p_{tac} promoter by one base pair (see Figure 11). It has been reported that the activity of the p_{trc} promoter and the p_{tac} promoter are about equal (Amman and Brosius, 1985) and both are among the strongest promoters known to function in *E. coli*. However, detectable levels of $rT_r\alpha$ expression (>0.05%) were not obtained using pML12 (Figure 63A, Lane 3). Others have also found that pKK 233-2 based vectors do not result in detectable protein expression (Biernat *et al.*, 1987; J. Banroques, personal communication). When a DNA fragment containing the λ_R promoter was placed upstream of the p_{trc} promoter in pML 12 to create pML 15 (Figure 12), expression of a 25 kDal fragment of $rT_r\alpha$ was detected (Figure 63A, Lanes 6 and 7). The simplest explanation of this effect is that the promoter of pKK 233-2 is defective. Appearently in pML 15 the λ_R promoter can function even though it is about 1,400 base pairs upstream of the $rT_r\alpha$ initiation codon. Another explanation is that the λN protein, which is encoded on pML 15, was relieving transcriptional termination. Transfer of the "expression unit" of pML 15 into

a smaller plasmid to create pML 25 (Figure 18) resulted in a further 5-fold increase in expression compared to levels obtained using pML 15 (Figure 63B).

The level of $rT_r\alpha$ expression in CAG1139 pML 25 cells induced in 0.1 mM IPTG for 20 min was about 0.1% of total cell protein (about 0.2 μg of $rT_r\alpha$ per liter of cells at OD_{600nm}=1.0). About 55% of this $rT_r\alpha$ was soluble. Compared to a 20-minute induction, about five times as much total $rT_r\alpha$ was made after 16 hours of induction, but the amount of soluble $rT_r\alpha$ that was made during the 16-hour induction period was about the same as after a 20-minute induction. Thus it appeared that only a certain amount of soluble $rT_r\alpha$ could be made and any excess would become insoluble.

Extracts of induced CAG1139 pML 25 cells contain two proteins of molecular weight 25 kDal and 40 kDal that cross-react specifically with the Tα1A antisera (Figures 63B and 64). These proteins were observed only when the promoter was induced and only in cells containing $rT_r\alpha$ expression plasmids. The 40 kDal product presumably represents $rT_r\alpha$ that is full length or missing fewer than about 10-20 amino acids, since it migrates about the same as retinal $T_r\alpha$ on SDS gels. The nature of the 25 kDal product is not known. A possible precursor/product relationship between the 25 kDal and 40 kDal proteins was not investigated directly, but the 25 kDal protein may represent a relatively stable protease cleavage product of $rT_r\alpha$. In this regard it is noteworthy that trypsin cleavage of native retinal $T_r\alpha$ (see Materials and Methods) results in the production of a 21 kDal fragment corresponding to amino acids 20-204. Perhaps the 25 kDal protein is similar to the 21 kDal product of tryptic digestion. It is reasonable that the same regions of $T_r\alpha$ that are accessible to trypsin might also be accessible to proteases in *E. coli.* The relative ratios of the 25 kDal and 40 kDal products varied in different strains and even with different expression vectors (see pML 46 below). For example, in AR58 (Shatzman and Rosenberg, 1985) pML 25 or in cells containing pML 15, primarily the 25 kDal form was

observed. The best ratio between the two forms was observed in CAG1139 pML 25 where they were present in about a 1:5 ratio.

Another explanation for the appearance of the 25 kDal form of $rT_r\alpha$ is that internal translation initiation could be occurring. This phenomenon is relatively common. In order for the Tα1 antibody to recognize the 25 kDal protein it must contain amino acids 85-103. The $T_r\alpha$ cDNA has three methionines before position 85 at positions 1, 49, and 84. If internal initiation at position 84 is the sole reason for the generation of the 25 kDal protein then the predicted molecular weight of the fragment would be about 30 kDal not 25 kDal. Therefore it seems unlikely that internal translation initiation alone generates the 25 kDal protein. Of course a combination of proteolysis and internal initiation could also explain the results. Other explanations for the appearance of the 25 kDal protein are premature transcriptional or translational termination, or specific ribonucleasease cleavage of its mRNA. These possibilities were not investigated.

The vector pML 17 (Figure 14) which contains another strong λ promoter (λ_L) was also constructed for $rT_r\alpha$ expression. However induced cells containing pML 17 did not contain detectable amounts of $rT_r\alpha$. One possible explanation for this lack of expression is that pML 17 was derived from pML 12 ,which also does not express $rT_r\alpha$. However the precise molecular reason for the lack of expression is unknown.

Another strong promoter that was tested for its ability to direct $rT_r\alpha$ expression in *E. coli* was the T7 bacteriophage ϕ10 promoter (see pML 49, Figures 28 and 29). This promoter has the advantage of having lower basal levels of expression and higher induction levels (Studier *et al.*, 1990) than other strong promoters such as the tac promoter. This property is useful when the foreign gene product is toxic to cells. In addition when the foreign gene product is expressed, most host cell protein synthesis can be inhibited with rifampicin because the T7 RNA polymerase is rifampicin resistant where the *E. coli* RNA

polymerase is not. In this study, expression of the T7 ϕ10 promoter was induced by infection with the M13 phage mGP1-2 (Tabor and Richardson, 1985). mGP1-2 carries a T7 polymerase gene that is inducible with IPTG because it is under the control of lacI. In order for cells to be infected with mGP1-2 they must be F^+. Since CAG1139 cells are F^+, they are suitable. CAG1139 pML 49 cells were infected with mGP1-2 at a multiplicity of infection of 1 in the presence of 2 mM IPTG and rifampicin (200μg/ml). The levels of expression were about the same as observed in IPTG-induced CAG1139 pML 25 cells. The amount of soluble $rT_r\alpha$ that was made was not determined.

Expression of $T_c\alpha$ in *Escherichia coli* using efficient promoters.

The p_{tac} (pML 4; Figures 3 and 4) and $p_{L/}p_{tac}$ (pML 18; Figures 16 and 17) promoters were tested for their ability to drive the synthesis of $rT_c\alpha$ in *E. coli*. In CAG1139 cells both vectors can direct the synthesis of about the same amount of $rT_c\alpha$ under optimized conditions. Almost all of the $rT_c\alpha$ produced in CAG1139 pML 4 cells is insoluble. In contrasst, about 20% of the $rT_c\alpha$ produced in CAG1139 pML 18 cells is soluble (Figure 65).

Unfortunately it was not possible to quantitate the amount of $rT_c\alpha$ synthesized because a standard does not exist since retinal $T_c\alpha$ has not been purified. However the intensity of staining observed on Western blots was about the same as for $rT_r\alpha$ produced in pML 25 cells. Thus if the antibody titer, affinity, and so on of the $rT_c\alpha$ and $rT_r\alpha$ antisera are the same, then the levels of $rT_c\alpha$ made are probably about the same as for an equivalent vector expressing $rT_r\alpha$. This might be expected since $T_c\alpha$ and $T_c\alpha$ are 82% identical in amino acid sequence.

One curious feature of pML18 was its lethality to cells (Table 7). As a consequence, cells containing pML 18 require particular growth and induction conditions for expression to be observed. The lethality was observed only when the $p_{L/}p_{tac}$ promoter was induced,

only under conditions selecting for plasmid maintainence, and only when the coding region of $rT_C\alpha$ was in the sense orientation. Furthermore there is a correlation between the amount of $rT_T\alpha$ synthesized and the degreee of lethality observed (Table 8). When the promoter is induced in nonselective media, no lethality was observed. When the coding region is in the antisense orientation, no lethality was observed. This behavior is what is typically observed when a gene product is toxic or a plasmid is being lost from cells. The data in Tables 7 and 8 suggest that $rT_C\alpha$ may be toxic to cells. Gene product toxicity can lead to plasmid loss. However plasmid loss does not prove that a gene product expressed from that plasmid is toxic. For example, an alternate possibilty is that transcription from the $p_{L/P}tac$ promoter interferes with the origin of replication. This has been observed to cause plasmid loss. Once the plasmid is lost, descendents of the cell will be sensitive to the ampicillin in the media and will die. This seems like the best explanation for why the parent plasmid of pML 18, pPL, also causes cell lethality when induced in LB media with ampicillin, but not in LB. In the case of pML 18, it may be that the coding region of $rT_C\alpha$ in the sense orientation allows transcription to read into the origin region. The antisense construct may cause transcriptional termination or reduce the amount of transcription to acceptable levels.

Expression of $rT_T\alpha$ in *Escherichia coli* as a protease-cleavable fusion protein.

Attempts to overproduce and purify $rT_T\alpha$ using vectors described in previous sections were complicated by proteolysis, difficulties in purification, and low levels of expression. To overcome these difficulties we investigated the "Glutagene" technology which is diagrammed in Figure 66. This technology was developed by Smith and Johnson (1988) and has been licensed by Pharmacia for commercial distribution. The essential components of the technology are two plasmid vectors pGEX-2T and pGEX-3X. Each expresses the

Schistosoma japonicum glutathione S-transferase (GST; EC 2.5.1.18) under the control of the p_{tac11} promoter, which is regulated by lac I^q repressor encoded on the same plasmid. Each plasmid has a restriction enzyme site polylinker at the carboxyl terminus of the coding region of GST. This allows insertion of the coding region of interest to generate a GST-protein X fusion. The polylinker also encodes cleavage sites for specific proteases. The pGEX-2T polylinker encodes a thrombin cleavage site (usually between Arg-Gly; Lundblat *et al.*, 1976) and pGEX-3X encodes a factor X_a cleavage site (after Ile-Glu-Gly-Arg; Nagai and Thorgersen, 1987). These proteases have infrequent recognition sites and therefore do not cleave most proteins. GST fusion proteins are usually soluble, represent 1-10% of total protein, and can be rapidly purified in a single step from crude cell lysates by batch absorption to glutathione agarose beads. Elution of GST fusions from these beads can be achieved in buffers with a variety of compositions as long as they contain 5 mM reduced glutathione. The eluted fusion protein can be cleaved with the appropriate protease and the contaminating GST that is released can be removed with glutathione agarose or with other chromatography columns. Potential advantages of this system over those previously described included higher expression levels, more soluble protein, and a rapid, one-step purification method.

The Glutagene vector pGEX-2T was chosen for expression of $T_r\alpha$. Thrombin was reported to be more efficient at cleaving GST fusions than factor X_a (Smith and Johnson, 1988). Also the reading frame of pGEX-2T is in frame with that of $T_r\alpha$ after insertion of an Nco I linker to facilitate subcloning.

Prior to initiating the construction of Glutagene-based expression vectors, retinal $T_r\alpha$ was tested for thrombin sensitivity. The sequence of $T_r\alpha$ does not have any predicted Arg-Gly thrombin cleavage sites. However, thrombin is a serine protease similar to trypsin and has been observed to cut after certain arginine residues that are not followed by

glycine. Therefore retinal $T_r\alpha$ was incubated with thrombin at a concentration of 10-fold greater than that used to cleave GST fusion proteins and either run on an SDS gel and stained with Coomassie Blue or subjected to Western blotting using Tα1 antisera. Control samples were untreated. Under the conditions used, no thrombin cleavage of retinal $T_r\alpha$ was detected. In addition, thrombin pretreatment of retinal $T_r\alpha$ or inclusion of thrombin in a GTPγS binding assay had no effect on the rhodopsin/Tβγ-stimulated ^{35}S-GTPγS binding activity of retinal $T_r\alpha$.

Since thrombin does not cleave retinal $T_r\alpha$ or affect its GTPγS-binding activity significantly, a vector (pML 56; see Figures 37 and 38) was constructed to express $rT_r\alpha$ as a fusion to GST. This plasmid was transformed into CAG1139. Upon induction of these cells with 0.1 mM IPTG for 30 min about 8 mg GST-$T_r\alpha$ is produced per liter of cells per OD_{600nm}=1.0. This represents about 6% of total cell protein which is within the expected range. This level of expression was easily observed when induced extracts were analyzed on a Coomassie Blue-stained SDS gel (Figure 67A, Lane 2). However quantitation of GST-$rT_r\alpha$ expression levels was judged by Western blotting using the Tα1A or B antisera rather than by examining Coomassie Blue-stained gels because GST exhibits enhanced Coomassie Blue binding. Thus quantitations based on densitometric scans of Coomassie Blue-stained gels would overestimate the amount of GST fusion made.

Since the molecular weight of GST is 26 kDal and that of $T_r\alpha$ is 40 kDal, 60% of the weight of the fusion is $rT_r\alpha$ and the theoretical maximum yield of $rT_r\alpha$ is about 5 mg per liter of cells. After disruption of IPTG-induced cells by passage through a French Press cell at 18, 000 psi and centrifugation at 100,000 X *g*, 70% of the GST-$rT_r\alpha$ fusion was found in the soluble supernatant. About 5% of the soluble fraction was GST-$rT_r\alpha$ while about 20% of the insoluble fraction was GST-$rT_r\alpha$. About 10-20% of the soluble GST-$rT_r\alpha$ bound to glutathione agarose beads. After washing the glutathione agarose

bound material with transducin buffer, about 50% of the fusion that bound to the beads could be recovered by eluting with transducin buffer containing 5 mM reduced glutathione. This material was highly purified relative to the starting crude extract (Figure 67A, Lane 3). Over 95% of the fusion was cleaved with thrombin under mild conditions (Figure 67A, Lane 4). GST, but not $rT_r\alpha$, was removed efficiently with glutathione agarose after the concentration of glutathione was reduced by dialysis in transducin buffer. It was found that $rT_r\alpha$ can also be cleaved directly from GST-$rT_r\alpha$ that is bound to the glutathione agarose beads. This would obviate the need to remove residual GST present when the soluble, eluted GST-$rT_r\alpha$ is cleaved. Such direct cleavage of $rT_r\alpha$ from the beads was found to be about 60-70% efficient. The transducin alpha subunit prepared by the procedure described above is referred to as $gT_r\alpha$ (Glutagene-derived $rT_r\alpha$). About 250 μg of $gT_r\alpha$ per liter of cells (OD_{600nm}=1.0) was obtained. This represents a final yield of about 5%. While this yield is low, the material is at least 50% pure. Such preparations should be adequate for biochemical assays, but larger (>10-100 liter) fermentations would be required to obtain amounts of $gT_r\alpha$ suitable for x-ray crystallography.

The main reason for the low yield of $gT_r\alpha$ is that only 10-20% of the GST-$rT_r\alpha$ bound to the glutathione agarose beads. The reason for this low level of binding is not clear. More than 90% of unfused GST binds to glutathione agarose beads (Smith and Johnson, 1988). One possibility is that the amount of beads used was not adequate. However the amount of beads used to bind GST-$rT_r\alpha$ was about three times more than that recommended by the manufacturer or by Smith and Johnson (1988). In addition, adding 10-fold more beads did not increase the amount of the GST-$rT_r\alpha$ fusion that bound. Another possibility is steric interference. The transducin alpha subunit may physically interfere with the binding of GST to the beads. This does not seem likely since GST protein fusions larger than GST-$rT_r\alpha$ bind efficiently to the beads. Perhaps the transducin

alpha subunit inhibits the proper folding of GST so that it can not bind to the beads. Another explanation is that binding of GST-r$T_r\alpha$ to the beads is slower than normal. For other fusions, 2 minutes appears to be adequate (Smith and Johnson, 1988). Binding of the GST-r$T_r\alpha$ fusion to the beads was usually done in 10-30 minutes to minimize proteolysis problems. The effect of longer binding periods on the amount of GST-r$T_r\alpha$ bound to the beads was not investigated. Finally the beads were added directly to an untreated crude extract. Substances in the crude extract could inhibit binding. For example, proteins which attack the glutathione moiety on the beads may be present. Also the levels of endogenous glutathione may be high enough to inhibit binding. Thus binding might be improved by dialyzing the crude extract or by passing it over a gel filtration or DEAE column before performing the glutathione agarose affinity purifiction. The material that does not bind to the glutathione agarose beads may be active. It might be useful to purify the GST-r$T_r\alpha$ fusion by standard nonaffinity methods and then cleave it with thrombin and test its activity.

About 50 % of the GST-r$T_r\alpha$ eluted from the beads. This is a reasonable yield taking into consideration the observation that some GST fusions require much stronger elution conditions for high yield (Kemp, 1989). Better elution yields might be obtained by including 0.5 M NaCl or a mild detergent such as CHAPS in the elution buffer. These conditions would not adversely affect $T_r\alpha$ activity.

Retinal $T_r\alpha$ and g$T_r\alpha$ were assayed for rhodopsin/T$\beta\gamma$-stimulated ^{35}S-GTPγS binding activity. Control assays containing retinal $T_r\alpha$ were found to be active. However an equivalent amount of g$T_r\alpha$ was inactive. The GST-$T_r\alpha$ fusion was also inactive. The amount of ^{35}S-GTPγS bound by all g$T_r\alpha$ samples tested was not significantly different from blank assays which contained no protein. At most, 0.5% of the g$T_r\alpha$ was active.

One explanation for these results is that the conditions utilized during the preparation of

$gT_r\alpha$ had an adverse effect on its activity. To test this possibility, retinal r $T_r\alpha$ was exposed to the buffer used for cell lysis, the buffer used for elution from the beads, the glutathione agarose beads, and the buffer used for thrombin cleavage. None of these treatments had a significant effect on GTPγS binding activity. As mentioned previously, residual thrombin is unlikely to affect GTPγS binding. It is also unlikely that some of the impurities remaining in the $gT_r\alpha$ preparation interfere with the GTP-binding assay. Crude *E. coli* extracts do interfere with GTPγS binding activity. However this only occurs at *E. coli* protein concentrations (>1 mg/ml) that would not have been reached in the GTPγS binding assays of $gT_r\alpha$. Of course it is possible that an inhibitor of GTP-binding selectively purified with GST-$T_r\alpha$ on the glutathione-agarose beads.

There are other more plausible explanations for $gT_r\alpha$ inactivity. The vector pML 56 was constructed such that thrombin cleavage of GST-$T_r\alpha$ would result in a protein that has six extra amino acids (Gly-Ser-Pro-Tyr-Gly-Pro) at the amino terminus (see Figure 38). $gT_r\alpha$ was not subjected to amino acid sequencing, but if these extra amino acids exist on $gT_r\alpha$, they could prevent the formation of active $gT_r\alpha$ by blocking proper folding or interfering with GTPγS binding. It is unlikely that either is the case, however. The regions of $T_r\alpha$ thought to be involved in nucleotide binding (near positions 40, 200, 220, 270, and 320) are not close to the amino terminus in the linear structure. In addition, modeling studies fitting Gα amino sequences into the crystal structures of *ras* and Ef-Tu predict that the amino terminus of $T_r\alpha$ is on a side of the protein which is opposite the guanine nucleotide binding site. Nevertheless it is still possible that extra amino terminal sequences could indirectly interfere with GTPγS binding. This is because the amino terminus is necessary for Tβγ subunit binding and Tβγ is necessary for catalyzing GTPγS binding. However, other G proteins ($G_s\alpha$, $G_{i\text{-}1}\alpha$, $G_{i\text{-}2}\alpha$, $G_{i\text{-}3}\alpha$) have been expressed as fusion proteins with nine extra amino terminal residues (Mattera *et al.*, 1989) and these are

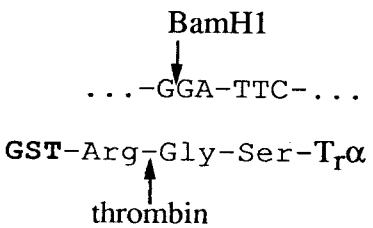
active in GTPγS binding and effector regulation assays. $gT_r\alpha$ was not assayed for cGMP phopshodiesterase activation.

A third explanation for the inactivity of $gT_r\alpha$ is that it did not fold into a native configuration. To test this notion, $gT_r\alpha$ was digested with trypsin as described in Materials and Methods. Trypsin cleavage of retinal $T_r\alpha$-GDP under native conditions produces four fragments corresponding to amino acids 1-18 (M_r=2 kDal), 19-204 (M_r=21 kDal), 205-310 (M_r=12 kDal), and 311-350 (M_r=5 kDal) (Fung *et al.*, 1984; Hurley *et al.*,1984; Tanabe *et al.*, 1985). In a control experiment, digestion of native retinal $T_r\alpha$-GDP followed by Western blotting using the Tα1B antisera detected a 21 kDal fragment (Figure 67B, Lane 2). This was expected since the Tα1B antisera was raised against a peptide corresponding to amino acids 85-103 of $T_r\alpha$. Tryptic digestion of $gT_r\alpha$ under the same conditions resulted in complete degradation (Figure 67B, Lanes 4 and 6). This result indicates that $gT_r\alpha$ is not in a native configuration, but it does not indicate what form it may be in.

Some attempts were made to subject $gT_r\alpha$ to "denaturation/renaturation" procedures that have worked for recovering the activity of other foreign proteins found in the insoluble fraction of the *E. coli* lysates. $gT_r\alpha$ was incubated in 6 M guanidine, 0.1 M β-mercaptoethanol, 0.1 M Tris-HCl, pH 8.5 followed by slow dialysis against buffers that contained progressively less guanidine and finally against transducin buffer. These procedures were done with and without GDP, GTP, Tβγ, and 0.1 mM oxidized glutathione/1.0 mM reduced glutathione. Then GTPγS binding activity was assayed. In no case was activity above a blank control observed. However there are two reasons to think that $T_r\alpha$ might not renature once it is denatured. First, α-^{32}P-GTP transfer-blot binding experiments which have been successful for the 20 kDal class of GTP binding proteins (McGrath *et al.*, 1984; Lapetina and Reep, 1987; Bokoch and Parkos, 1988;

Kikuchi *et al.*, 1988; Polakis *et al.*, 1989) have been unsuccessful for the 40 kDal G protein alpha subunits. (In these procedures protein is run on an SDS denaturing gel, electroblotted onto nitrocellulose, incubated with α-^{32}P-GTP in a "renaturation" buffer, washed, and exposed to x-ray film.) Second, $T_r\alpha$ has 8 cysteines. It is not known how these are configured. However since 3 appear to be present as sulfhydryls (Ho and Fung, 1984) and one may be lipid-modified, it is plausible that the eight are arranged as 3 sulfhydryls, 2 disulfides, and one acylated cysteine. It is somewhat unusual for intracellular proteins to have disulfide bonds. In any case, it is often difficult to recover activity from large (greater than 100 amino acids) proteins expressed in *E. coli* especially if they contain disulfide bonds or large numbers of cysteines. In the internal environment of *E. coli* such proteins tend to form "scrambled" molecules with improper pairs of disulfide bonds. In such cases it is sometimes better to secrete the protein since the exterior of bacterial cells is more oxidizing than the interior.

There are several ways to investigate the reason for $gT_r\alpha$ inactivity or attempt to recover active protein. For example, the amino ternimal region in pML 56 could be reconstructed so that the cleaved protein would begin with glycine-2 as *E. coli* $G_s\alpha$ does. This could be done by creating a BamH1 restriction site in the amino terminus of the $T_r\alpha$ coding region and then subcloning the $T_r\alpha$ coding region into the BamH1 site of pGEX-2T. The result would be:



This would change the normal amino acid at position three (alanine) to serine. However

this may be acceptable since $T_c\alpha$ has a serine at position three. Of course, it would be possible to change the serine back to an alanine once the above modified vector was constructed. Another option would be to construct a Glutagene fusion vector with the factor X_a cleavage site. However factor X_a is 10-fold more expensive than thrombin and is not as efficient at cleaving GST fusions.

It is possible that the inactivity of $gT_r\alpha$ is a property of the Glutagene system rather than of $T_r\alpha$. To test this notion, the $G_o\alpha$ cDNA could easily be subcloned into the Glutagene vector pML 55. $G_o\alpha$ has been produced as an unfused protein in *E. coli* in an active form that is virtually indistinguishable from the native protein from brain (Linder *et al.*, 1990). $G_o\alpha$ is similar to $T_r\alpha$ in many respects. However it is different from $T_r\alpha$ in that it readily binds GTPγS in the absense of receptor. This property offers a distinct advantage from the viewpoint of expression in *E. coli*. If the $G_o\alpha$ produced using the Glutagene system is inactive, this may be an indication that some property of the Glutagene system is responsible for the inactivity of $gT_r\alpha$. However if it is active, it may be that $T_r\alpha$ is inherently different from the other G protein alpha subunits that have been successfully produced in *E. coli*.

Other fusion vector systems have been developed that produce soluble proteins which can be purified in a single step (for example the FLAG technology from Immunex, Inc.) Perhaps one of these systems would work better than the Glutagene system.

Another explanation for the inactivity of $gT_r\alpha$ is that a mistake was made in constructing the vector pML 56. This was tested by sequencing the junction between GST and $rT_r\alpha$ in pML56. The junction sequence was not as expected. It was found to have an extra Nco I linker. This linker was not digested with Nco I during the construction of pML 55 because it lacks the correct recognition sequence. Nevertheless it retains the correct reading frame between GST and $rT_r\alpha$. However during the course of this

sequencing it was found that a mutation exists in the $rT_r\alpha$ protein sequence that results in changing Leu-32 (CTG) to Pro-32 (CCG). Leu-32 is a highly conserved amino acid found in all G protein alpha subnits without exception. It is in the sequence LysLeuLeu-(32)Leu LeuGlyAlaGlyGluSerGlyLys. The sequence GlyAlaGlyGluSerGlyLys is referred to as the "A region" (Lochrie and Simon, 1987). In the *ras* crystal structure the region analogous to the four leucines in $rT_r\alpha$ forms a β strand and the A region forms a random coil that is involved in GTP-binding. Thus, it is possible that the Leu-32 to Pro-32 mutation could significantly affect GTP-binding activity.

Expression of $T_r\alpha$ in *Escherichia coli* using a dual cistron vector.

Multiple steps exist in the process of protein expression from the initiation of transcription to posttranslational maturation of the protein product. As described above, using different *E. coli* promoters, different plasmid contexts, and different strains improved $rT_r\alpha$ expression levels. Another factor that often limits higher levels of expression in *E. coli* operates at the level of translational initiation. In vectors described above, the nucleotides prior to the initiation codon are ones that are frequently used in *E. coli* genes. However sequences after the initiation codon can also influence translation initiation. In particular, guanines after the initiation codon tend to be disfavored in *E. coli* genes that are highly expressed (Gold and Stormo, 1990). This may be because guanines would compete with the guanine-rich ribosome binding site for binding to the ribosome. In the $T_r\alpha$ sequence 8 out of 10 nucleotides after the initiation codon are guanines. Therefore as one approach to investigate the possibility that some aspect of translation might be preventing higher levels of $T_r\alpha$ expression, we utilized a "dual cistron" expression vector. In these vectors a short, highly expressed cistron preceeds the $T_r\alpha$ cistron. Within the carboxyl terminus of the first cistron is a second ribosome binding site that directs the translation of the second cistron. Previous reports indicate that dual cistron expression

vectors can circumvent problems related to inefficient translation initiation and can lead to higher expression levels (Schoner *et al.*, 1990). This is thought to occur because ribosomes bind efficiently at the first ribosome binding site and then those same ribosomes reinitiate translation at the second ribosome binding site.

The vector pML46 (Figure 26) was constructed to express $rT_1\alpha$ using a dual cistron system. The key features of pML46 include the omp A/$rT_1\alpha$ dual cistron, a lpp/lac promoter regulated by a plasmid-encoded, IPTG -inducible lac I repressor, and the Col E1 replication origin and β-lactamase gene for replication and selection in *E. coli.*

The expression of $rT_1\alpha$ from this type of vector has some interesting properties. 1) The level of expression is about 1% of total protein, a level ten-fold higher than that observed with any previous *E. coli* $rT_1\alpha$ expression vector. Some of the increase could be due to the fact that a dual lipoprotein/lac promoter was used. However, a comparison of expression levels from a pair of homologous T7 promoter vectors that differed only in having either a single (pML49) or a dual (pML52) cistron showed that the level of $rT_1\alpha$ made using the dual cistron vector was about 3-5 fold higher than that of the single cistron vector. Thus some of the increase may be due to the promoter, but some is also clearly due to the dual cistron expression configuration. 2) The amount of $rT_1\alpha$ that is found in the soluble fraction after cell lysis (55%) compares favorably with that observed from other bacterial vectors (Figure 68B). 3) The rate of $rT_1\alpha$ synthesis is more rapid than that observed with other systems. A time course indicates that synthesis plateaus after about 30-45 minutes (Figure 68A). With other *E. coli* $rT_1\alpha$ expression vectors described above the amount of $rT_1\alpha$ would slowly increase over a period of hours after induction. 4) Unlike pML 18, pML46 was stable in *E. coli* and not lost from cells upon induction. 5) The appearance of the 25 kDal $T_1\alpha$-related protein observed in strains with pML 15 or pML 25 was usually not observed on Western blots of induced cells containing pML 46. If it

was it was a small fraction (<10%) of the total $T_r\alpha$-related protein.

Another distinction between pML 46 and pML 25 is that a difference in mobility on SDS gels was observed between $rT_r\alpha$ produced from pML 46 and $rT_r\alpha$ produced from pML 43 (a close derivative of pML 25), even when both where produced in the same strain (Figure 69). The pML 43 form of $rT_r\alpha$ has an apparent molecular weight of 40 kDal and the pML 46 form has an apparent molecular weight of 42 kDal. One explanation for this difference is that the stop codon preceeding the start codon of $rT_r\alpha$ in pML 46 had mutated. This would result in an in-frame fusion to the ompA leader (M_r=3 kDal). However this is unlikely because DNA sequence analysis of this region confirmed that it is correct. Another possibility is that the ochre stop codon between the end of the ompA and the beginning of the $rT_r\alpha$ coding region is being suppressed. This possibility also seems unlikely because no ochre suppressors are known to exist in CAG1139. Also, since suppression of nonsense codons is not 100% efficient, both a 42 kDal and a 40 kDal form might have been observed on a Western blot. However, this was not the case. Fusion of the ompA leader to $rT_r\alpha$ could have resulted in localization of $rT_r\alpha$ to the outer membrane. However the results of crude cellular fractionation procedures suggested that this had not occurred. The most plausible explanation for the difference between the pML 43 and the pML 46 forms of $rT_r\alpha$ is in the extent of proteolysis. The rapid rate of $rT_r\alpha$ synthesis from pML 46 may afford it some protection against proteolytic degradation. In any case these observations highlight the important point that once $rT_r\alpha$ is overproduced and purified it would be wise to determine its amino and carboxyl terminal amino acid sequences especially since, in the case of retinal $T_r\alpha$, it is known that the extreme amino and carboxyl termini are important for its function.

Another vector, pML 50, was constructed by transfer of the "expression unit" (promoter and coding region of $rT_r\alpha$) of pML 46 to a smaller vector (see Figure 30). This

resulted in about a two-fold increase in the total amount of $rT_r\alpha$ synthesized. However most of the extra $rT_r\alpha$ produced appeared to be insoluble (Figure 68B).

Partial purification of $rT_r\alpha$ from CAG1139 pML 46.

Since the pML 46 form of $rT_r\alpha$ appeared to provide the best source of recombinant protein, a program was initiated to purify it. For the pupose of obtaining a maximal amount of $rT_r\alpha$ the temperature during the induction of pML 46 was chosen as 30°C. At this temperature the largest amount of soluble $rT_r\alpha$ was produced (Table 9). At lower temperatures (25°C) the percentage of soluble $rT_r\alpha$ was the same, but the doubling time of the cells is much less. At higher temperatures (37°C) the percentage of soluble $rT_r\alpha$ was significantly reduced compared to 25°C or 30°C. A correlation between higher expression temperatures and increased insolubility of foreign proteins has been observed before (Squires *et al.*, 1988; Schein and Noteborn, 1988; Piatah *et al.*, 1988; Schein, 1989). However the optimum temperature for obtaining the maximum amount of soluble protein varies for each protein.

The chromatographic behavior of $rT_r\alpha$ produced from cells containing pML 46 was tested on several columns and is summarized in Table 10. Only three procedures resulted in significant amounts of purification without substantial losses in yield. These were ammonium sulphate precipitation, heptyl agarose chromatography and anion exchange chromatography. A purification scheme using these three steps was developed (Figure 70A) that produces $rT_r\alpha$ which is about 3-5% pure at a yield of about 10%. This represents about a 10-fold purification from the soluble crude extract. The $rT_r\alpha$ precipitated at a lower ammonium sulfate concentration than most *E. coli* proteins. At 40% ammonium sulfate about 80-90% of the $rT_r\alpha$ precipitated, whereas only about 20-30% of the total protein precipitated. A similar result was obtained by heptyl agarose chromatography. The transducin buffer eluate contained about 30% of the total protein

loaded on the column, but over 90% of the $rT_r\alpha$. In most of these exploratory chromatographic studies the entire peak of $rT_r\alpha$ was collected for analysis. Therefore material of higher purity could probably be obtained by collecting only peak fractions or fractions that are better separated from peaks of other *E. coli* proteins, but a substantial reduction in yield would probably occur. Therefore in order to obtain purer material it might be necessary to use larger amounts of starting cultures than those typically used in this study (2 liters). Better purificatation levels might also be obtained by using a gradient elution of the heptyl agarose column rather than the stepwise elution method used here.

Some of the chromatographic properties of $rT_r\alpha$ are bothersome and inexplicable. Two peaks of $rT_r\alpha$ were observed on the Blue Sepharose column. One elutes at 0.1 M NaCl and the other at 0.3 M NaCl where retinal $T_r\alpha$ elutes. Very broad peaks indicative of molecular heterogeneity (aggregation?) were observed on gel filtration columns. Some of the $rT_r\alpha$ (5-10%) precipitated at low ammonium sulfate concentrations (<20%) which also suggests it may be part of a larger molecular weight or hydrophobic aggregate. On the heptyl agarose column $rT_r\alpha$ eluted at a very low salt concentration (0.0-0.1 M ammonium sulphate) indicating it may be hydrophobic or denatured. Retinal $T_r\alpha$ elutes at 0.9 M ammonium sulfate. The recovery of $rT_r\alpha$ was much better from Blue Sepharose than from Affigel Blue. These columns have the same cibracon blue dye bound to them at about the same density. However Blue Sepharose is agarose-based whereas Affigel Blue is polyacrylamide-based.

Purification schemes which attempted to utilize the specific binding and one-step elution of transducin to rhodopsin failed. When *E. coli* extracts containing $rT_r\alpha$ were mixed with retinal $T\beta\gamma$ and ROS membranes containing rhodopsin, the $rT_r\alpha$ did not bind to the ROS membranes. However retinal transducin mixed with an *E. coli* extract that did not contain $rT_r\alpha$ also failed to bind to ROS membranes indicating that something in the extract

was interfering with binding. The results of GTPγS binding assays, in which it was found that crude *E. coli* extracts partially inhibit GTP-binding, are in agreement with this conclusion. An attempt was also made to specifically elute $rT_r\alpha$ from Blue Sepharose, DEAE, and heptyl agarose columns with GTPγS. This was not successful.

During the purification of G protein alpha subunits from mammalian tissues GTPγS or AMF (AlF_4^-, $MgCl_2$) is sometimes adding during the initial stages of purification because some G protein alpha subunits are prone to denaturation. Adding GTPγS or AMF stabilizes their activity. For the purification of recombinant G protein alpha subunits in *E. coli,* GTPγS or AMF is not necessary. The effects of adding GTPγS or AMF to several column and lysis buffers on the chromatographic behavior of $rT_r\alpha$ was negligible. They also had no effect on the amount of soluble $rT_r\alpha$ that was obtained after cell lysis.

Expression of $T_c\alpha$ in *Escherichia coli* using a dual cistron vector.

The plasmid pML 53 (Figure 33) was constructed for the expression of $rT_c\alpha$. This plasmid is not toxic to CAG1139 cells like pML 18 is. Expression levels comparable to those observed using pML 4 were obtained. The $rT_c\alpha$ produced in induced CAG1139 pML 53 cells was about 45% soluble. The soluble $rT_c\alpha$ was purified about 5-fold by a procedure identical to that described for $rT_r\alpha$ in Materials and Methods. The results were similar except that 50% of the $rT_c\alpha$ precipitated at 40% ammonium sulfate and most (85%) of the $rT_c\alpha$ eluted from the heptyl agarose column in 0.0 M heptyl agarose buffer (Figure 70B).

Guanine nucleotide binding activity of $rT_r\alpha$.

In order to demonstrate the GTP-binding activity of $T_r\alpha$ two requirements must be met. First, rhodopsin and Tβγ are required in the assay to catalyze guanine nucleotide exchange. $T_r\alpha$ is not like other Gαs, such as $G_o\alpha$, which can readily bind GTPγS in the absence of receptor or Gβγ. Since rhodopsin and Tβγ will contain trace amounts of $T_r\alpha$,

careful controls must be implemented when analyzing $rT_r\alpha$ for GTP-binding activity. Second, the $T_r\alpha$ must be in a GDP-bound or be bound to no nucleotides prior to the assay because $T_r\alpha$ bound to GTP will not bind to rhodopsin/$T\beta\gamma$. Presumably in any heterologous expression host the $rT_r\alpha$ that is synthesized will be in a GDP-bound form. In most cells the concentration of GTP is greater than that of GDP. In *E. coli*, for example, the concentration of GDP and GTP are about 128 μM and 923 μM, respectively (Neuhard and Nygaard, 1987). However the affinity of $T_r\alpha$ for GDP and GTP is about equal. Therefore $rT_r\alpha$ might initially bind GTP shortly after it is synthesized, but it should hydrolyze the GTP to GDP within minutes. The majority of GTP-binding *ras* proteins expressed in *E. coli* (Poe *et al.*, 1985) or yeast (Gibbs *et al.*, 1987) are bound to GDP even though the GTP hydrolysis rate of *ras* is about 10-fold lower than that of $T_r\alpha$.

In this study guanine nucleotide binding activity was not reproducibly detected in any crude extract or chromatographic fraction containing $rT_r\alpha$. In most cases several GTPγS-binding activities were detected in fractions that did not contain $rT_r\alpha$. However in one experiment using the most highly purified fractions of $rT_r\alpha$ from a DEAE column, a low level of rhodopsin/$T\beta\gamma$ stimulated GTPγS-binding activity was observed. This fraction was about 3-5% pure. The peak fraction of the GTP-binding activity coincided with the peak of $rT_r\alpha$ protein observed on a Western blot of fractions from the DEAE column. A comparison of the amount of GTPγS bound with the amount of $rT_r\alpha$ in the fraction as judged by the Western blot indicated that 5-25% of the protein might be active.

GTPase activity of $rT_r\alpha$.

The alpha subunit of transducin is a poor GTPase (k_{cat}=1/min). For this reason the GTPase activity of any fraction of $rT_r\alpha$ was impossible to measure because of the overwhelming activity of cellular "GTPases." ATPases and other nucleotidases probably contribute to the high backround activity observed since many are able to use GTP as a

substrate. ATPγS suppressed the backround GTPase activity, but not enough to measure the GTPase activity of $rT_r\alpha$. In this study even the GTPase activity of a purified retinal $T_r\alpha$ sample was not always reproducibly detected. Therefore it will probably be necessary to obtain homogeneous $rT_r\alpha$ before its GTPase activity can be measured.

Phosphodiesterase activation by $rT_r\alpha$.

For this study the phosphodiesterase (PDE) assay of Yee and Liebman (1978) was used. In this assay GTP (or nonhydrolyzable GTP analogs such as GTPγS) or exogenous $T_r\alpha$ bound to GTP is added to ROS membranes. These membranes already contain endogenous $T_r\alpha$. Added GTP will activate the endogenous $T_r\alpha$ as long as a small fraction of the rhodopsin is bleached. Exogenously added $T_r\alpha$-GTP will directly activate the PDE. In this latter case the endogenous $T_r\alpha$ will not be activated because there is no source of GTP for it to bind.

It may not be trivial to assay the ability of $rT_r\alpha$ to activate rod PDE activity because of the nature of this PDE assay and the retinal transduction system. Only the GTP-bound form of $T_r\alpha$ can activate PDE. Rhodopsin and Tβγ are required to obtain GTP-bound $T_r\alpha$. However rhodopsin and Tβγ will contain trace amounts of retinal $T_r\alpha$. Also free nucleotide (GTPγS) can activate PDE. Therefore carefully controlled experiments must be done to distinguish activation by $rT_r\alpha$ versus activation by contaminating $T_r\alpha$ or GTPγS.

One way to circumvent the problems mentioned above relating to obtaining $rT_r\alpha$ in a GTP-bound form would be to use a mutation that permanently maintains $rT_r\alpha$ in a configuration that mimics the GTP-bound configuration but does not interfere with its ability to activate PDE. For this purpose I made a mutation in $rT_r\alpha$ that changes Gly-198 to Thr-198. In the *ras* protein, the analogous mutation causes the protein to autophosphorylate on the threonine and it behaves as if it were GTP-bound. I was hoping that the analogous mutation in $rT_r\alpha$ would result in a protein that would autophosphorylate

in the cell shortly after being synthesized. However the ability of this protein to activate PDE was not tested.

Another way to obtain $T_r\alpha$ in a conformation that mimics the GTP-bound form is to include AlF_4^- in the buffers. Aluminum fluoride appears to mimic the γ-phosphate of GTP. It has been found that as long as $T_r\alpha$ is bound to GDP, then $T_r\alpha$-GDP plus AlF_4^- behaves the same as $T_r\alpha$ bound to GTPγS. In order to take advantage of this trick, the PDE assay used in this study cannot be used because the AlF_4^- would activate the endogenous $T_r\alpha$. One would have to use purified PDE. However this would reduce the sensitivity of the assay because PDE is 10-fold less active when it is in solution.

In spite of these potential difficulties some exploratory and control experiments were done in this study. PDE was readily activated in ROS membranes by GTPγS or by exogenously added purified retinal $T_r\alpha$-GTPγS, but not by the purified Tβγ complex or by purified $T_r\alpha$-GDP. Various *E. coli* $rT_r\alpha$ extracts were tested for their ability to activate PDE. Extracts with or without $rT_r\alpha$ were inactive indicating that the backround activity is very low. The opposite situation occurred with extracts from Sf9 cells. In this case extracts containing the polyhedrin-$rT_r\alpha$ fusion protein were active. However a control cell lysate from uninfected cells containing no polyhedrin-$rT_r\alpha$ was also active. Therefore Sf9 cells may contain a component that activates PDE. This would not be too surprising. The catalytic α and β subunits of the PDE are negatively regulated by the γ subunit. The PDE γ subunit is a 73 amino acid long protein that is very basic. Trypsin can activate PDE by selectively destroying the γ subunit. In addition, basic materials like poly-lysine and histones can also activate PDE presumably by competing with the γ subunit for binding to the α and β subunits. Thus it may be best to use only homogeneous $rT_r\alpha$ for the PDE assay.

CONCLUSION

Three expression systems were surveyed for their ability to synthesize the rod and cone transducin alpha subunits. Many of the effects observed were complicated and are not well understood. In some cases subtle changes in plasmid structures could have a dramatic effect on expression. In almost every case only the end result (stable protein production) was examined. Little attempt was made to examine all of the steps that occur during the protein expression process.

In yeast no $rT_C\alpha$ or $rT_T\alpha$ expression was observed. In insect cells $rT_T\alpha$ and a polyhedrin-$rT_T\alpha$ fusion protein were expressed at high levels, but were insoluble. In *E. coli* expression levels were modest, but most of the protein made was soluble These levels would be adequate if an efficient purification procedure could be developed. However at the level of expression which was observed in *E. coli* the yield of protein during a purification procedure would be a critical factor. It would be useful to develop an affinity purification step to use as an early step in a purification procedure. In this regard the development of affinity columns that contain $\beta\gamma$ subunits for the purification of α subunits recently reported by Pang and Sternweis (1989) may be of use. The Glutagene system described in this study might also work once the Pro-32 mutation in the $rT_T\alpha$ is repaired. Even with such a purification procedure, higher expression levels or large cultures (10-100 liters) of *E. coli* might be required to obtain enough protein for x-ray crystallographic studies.

$G_S\alpha$ (Graziano *et al.*, 1989), $G_Z\alpha$ (Casey *et al.*, 1989), $G_{i\text{-}1}\alpha$, $G_{i\text{-}2}\alpha$, $G_{i\text{-}3}\alpha$ (Mattera *et al.*, 1989; Yatani *et al.*, 1988; Linder *et al.*, 1990) and $G_O\alpha$ (Linder *et al.*, 1990) have been successfully expressed in *E. coli* in an active form. Therefore it seems reasonable to assume that it should also be possible to express transducin alpha subunits in *E. coli* in an active form. The results of this study do not conclusively demonstrate that this

is possible, but do indicate that transducin alpha subunits can be produced in a soluble form at levels similar to that achieved with other Gα subunits. At this point it is possible that the reason $rT_r\alpha$ activity has not been easier to demonstrate is that the assays for $rT_r\alpha$ are more sensitive to interference than those for the other G protein alpha subunits. Once material of higher purity is obtained it may be easy to demonstrate activity. In fact, crude fractions of *E. coli* $rG_s\alpha$ were difficult to assay (Graziano *et al.*, 1987) and exhibited some unusual properties that were probably artifacts, since they were not observed in homogeneous material (Graziano *et al.*, 1989). Finally, repair of the Pro-32 mutation in $rT_r\alpha$ may result in an active protein.

A synthetic $T_r\alpha$ coding region has been expressed in mammalian COS cells (Sakmar and Khorana, 1988). That protein is insoluble and inactive (T. Sakmar, personal communication). The $rT_r\alpha$ cDNA has also been expressed in yeast at low levels using a metallothionine promoter (H. Bourne, personal communication). That protein is insoluble. In this study I found that $rT_r\alpha$ expressed in insect cells was insoluble and inactive. Therefore, surprisingly, expression of $rT_r\alpha$ in three different eukaryotic cells leads to insoluble protein. These observations contradict the dogma: it is best to express bacterial proteins in bacteria and eukaryotic proteins in eukaryotes.

There are several avenues that could be explored for improving the expression systems described in this study. It might be useful to secrete the protein. The plasmid pML 46 has the ompA leader sequence which can direct secretion (Takahara *et al.*, 1985; Takahara *et al.*, 1988). Thus pML 46 could easily be converted into a secretion vector by fusing the amino ternimus of $rT_r\alpha$ to the ompA leader. Other secretion vectors are also available. Secretion was not explored in this study and has not been tried for expressing any other Gα subunit. The bacterium *Bacilus subtilus* might be useful for secreting transducin subunits (Henner, 1990; Nagarajan, 1990).

Measuring expression levels in other strains is something that is easily done on an ongoing basis. It may be worthwhile to examine expression in *E. coli* ompT or *E. coli* type B or C strains rather than the K12 strains used in this study since such strains are thought to have fewer proteases. For the *E. coli* T7 expression system strains containing the T7 RNA polymerase gene integrated into the bacterial chromosome are now available. This alleviates the inconvenience and irreproducibility of using mGP1-2 infection to induce expression as in this study. In addition a vector (pLysS) that carries the T7 lysozyme gene has been found to be of use. T7 lysozyme inhibits the activity of T7 polymerase and thus reduces the basal level of expression even further, but it does not interfere with obtaining normal levels of induction. Using pLys vectors in conjunction with the T7 expression system has allowed production of proteins that are toxic at low levels.

Another aspect of expression that could have a significant effect is the coexpression of rTβγ subunits along with rTα. To date, such coexpression has not been reported. Early in the course of this study, rTαβγ coexpression in *E. coli* was attempted (N. Gautam and M. Lochrie, unreported experiments). Some expression of $rT_r\alpha$ and rTβ was obtained, but it was unclear at the time if rTγ was expressed. Since then rTγ has been expressed at high levels and purified to homogeneity (N. Gautam, personal communication). With the materials available now it may be possible to increase rTβ expression levels and synthesize all three subunits in the same cell. In many systems it is known that if other subunits of a multisubunit complex are not present then the unmatched proteins are often degraded. Sometimes they also do not fold properly. Also in eukaryotic cells the unmatched pairs are sometimes not localized correctly. This could also explain the appearant aberrent localization of $rT_r\alpha$ in insect cells.

Other expression hosts which were not surveyed in this study might also be useful. A variety of hosts and vector systems exist for overproduction of foreign proteins in

mammalian cells (Kaufman, 1990a). High levels of expression have been achieved by linking genes to dihydrofolate reductase (Kaufman, 1990b). The dihydrofolate reductase gene can be amplified in copy number by exposing stable transformants to several stepwise increasing levels of methotrexate, a specific inhibitor of dihydrofolate reductase. Genes linked to the dihydrofolate reductase gene will also be increased in copy number. However it takes several months to establish cell lines with high levels of expression.

To summarize, in all of the other cases in which active G protein alpha subunits have been expressed and purified the *E. coli* T7 expression system was used. Of the expression systems reported here the best appears to be either the *E. coli* dual cistron or T7 expression system. Although the levels of expression of $rT_r\alpha$ from pML 46 (dual cistron) is about 3-fold higher than from pML 52 (dual cistron and T7) pML 52 has two potentially important advantages over pML 46. First the basal level of $rT_r\alpha$ expression from pML 52 is much lower than for pML 46. Second, pML 52 carries the f1 bacteriophage origin of DNA replication which allows single stranded DNA to be generated as a template for mutagenesis procedures. Thus pML 52 may be the best vector described here for $rT_r\alpha$ expression or as a starting point for further generations of $rT_r\alpha$ expression vectors. If purification of $rT_r\alpha$ is difficult then perhaps the Glutagene-based vector pML 56 would be the best one to use.

ACKNOWLEDGEMENTS

I wish to thank Suzanna Horvath for peptide and oligonucleotide synthesis, Gloria Balbadie-McFarland for providing pGD 108, Carol Gross for CAG1139 cells, Reid Johnson for constructing CAG1139 F' lac I^{sq} Tn5 cells, R.-J. Lin for pAAH 5, Atsushi Miyajima for pMFα8 and NNY105 and 20B-12 cells, Jurgen Brosius for pKK 233-2, David Miller for pEV 55, Joan Kobori for pAc 360, pAc 610 and pAc 611, John Abelson for AR58, Scott Emr for SF838-10α, Susan Michaelis for SM1269 and SM1291,

Stephanie Ruby for SR741-3B, and Jim Urban for wild-type AcNPV and Sf9 cells.

REFERENCES

Amman, E. and Brosius, J. (1985). "ATG vectors" for regulated high-level expression of cloned genes in *Escherichia coli*. Gene 40, 183-190.

Amman, E., Brosius, J. and Ptashne, M. (1983). Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. Gene 25, 167-178.

Ammerer, G. (1983). Expression of genes in yeast using the ADC1 promoter. Meth. Enzymol. 101, 192-201.

Bennetzen, J. L. and Hall, B. D. (1982). The primary structure of the *Saccharomyces cerevisiae* gene for alcohol dehydrogenase I. J. Biol. Chem. 257, 3018-3025.

Biernat, J., Hasselmann, H., Hofer, B., Kennedy, N. and Koster, H. (1987). The construction and cloning of synthetic genes coding for artficial proteins and expression studies to obtain fusion proteins. Protein Engineer. 1, 345-351.

Bokoch, G. M. and Parkos, C. A. (1988). Identification of novel GTP-binding proteins in the human neutrophil. FEBS Lett. 227, 66-70.

Brake, A. J. (1990). α-factor leader-directed secretion of heterologous proteins from yeast. Meth. Enzymol. 185, 408-421.

Brosius, J. and Holy, A. (1984). Regulation of ribosomal RNA promoters with a synthetic *lac* operator. Proc. Natl. Acad. Sci. U. S. A. 81, 6929-6933.

Brown, R. L. and Stryer, L. (1989). Expression in bacteria of functional inhibitory subunit of retinal rod cGMP phosphodiesterase. Proc. Natl. Acad. Sci. U. S. A. 86, 4922-4926.

Bruist, M. F. and Simon, M. (1984). Phase variation and the Hin protein: *In vivo* activity measurements, protein overproduction, and purification. J. Bacteriol. 159, 71-79.

Dietzel, C. and Kurjan, J. (1987). The yeast SCG1 gene: A G_α-like protein implicated in the a- and α-factor response pathway. Cell 50, 1001-1010.

Donahue, T.F. and Cigan, A. M. (1990). Sequence and structural requirements for efficient translation in yeast. Meth. Enzymol. 185, 366-371.

Duronio, R. J., Jackson-Machelski, E., Heuckeroth, R. O., Olins, P. O., Devine, C. S., Yonemoto, W., Slice, L. W., Taylor, S. S. and Gordon, J. I. (1990). Protein N-myristoylation in *Escherichia coli*: reconstitution of a eukaryotic protein modification in bacteria. Proc. Natl. Acad. Sci. U. S. A. 87, 1506-1510.

Emr, S. D. (1990). Heterologous gene expression in yeast. Meth. Enzymol. 185, 131-133.

Ercolani, L., Stow, J. L., Boyle, J. F., Holtzman, E. J., Lin, H., Grove, J. R. and Ausiello, D. A. (1990). Membrane localization of the pertussis toxin-sensitive G-protein subunits $\alpha_{i\text{-}2}$ and $\alpha_{i\text{-}3}$ and expression of a metallothionein-$\alpha_{i\text{-}2}$ fusion gene in LLC-PK_1 cells. Proc. Natl. Acad. Sci. U. S. A. 87, 4635-4639.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction fragments to high specific activity. Anal. Biochem. 132, 6-13.

Fong, H. K. W., Yoshimoto, K., Eversole-Cire, P. and Simon, M. I. (1988). Identification of a GTP-binding protein α subunit that lacks an apparent ADP-ribosylation site for pertussis toxin. Proc. Natl. Acad. Sci. U. S. A. 85, 3066-3070.

Fuller, R. S., Brake, A. and Thorner, T. (1989). Yeast prohoromone processing enzyme (KEX2 gene product) is a Ca^{2+} -dependent serine protease. Proc. Natl. Acad. Sci. U. S. A. 86, 1434-1438.

Fung, B. K.-K. and Nash, C. (1983). Characterization of transducin from bovine rod outer segments. II. Evidence for distinct binding sites and conformational changes revealed by limited proteolysis with trypsin. J. Biol. Chem. 258, 10503-10510.

Ghrayeb, J., Kimura, H., Takahara, M., Hsiung, H., Masui, Y. and Inouye, M. (1984). Secretion cloning vectors in *Escherichia coli*. EMBO J. 3, 2437-2442.

Gibbs, J. B., Schaber, M. D., Marshall, M. S., Scolnick, E.M. and Sigal, I. S. (1987). Identification of guanine nucleotide bound to *ras*-encoded proteins in growing yeast cells. J. Biol. Chem. 262, 10426-10429.

Gilman, A.G. (1987). G proteins: transducers of receptor-generated signals. Ann. Rev. Biochem. 56, 615-649.

Goeddel, D. V. (1990). Systems for heterologous gene expression. Meth. Enzymol. 185, 3-7.

Gottesman, S. (1990). Minimizing proteolysis in *Escherichia coli:* Genetic solutions. Meth. Enzymol. 185, 119-129.

Graziano, M. P., Casey, P. J. and Gilman, A. G. (1987). Expression of cDNAs for G proteins in *Escherichia coli*. Two forms of $G_{s\alpha}$ stimulate adenyate cyclase. J. Biol. Chem. 262, 11375-11381.

Graziano, M. P., Freissmuth, M. and Gilman, A. G. (1989). Expression of $G_{s\alpha}$ in *Escherichia coli*. Purification and properties of two forms of the protein. J. Biol. Chem. 264, 409-418.

Grossman, A. D., Burgess, R. R., Walter, W. and Gross, C. A. (1983). Mutations in the *lon* gene of *E. coli* K12 phenotypically suppress a mutation in the sigma subunit of RNA polynerase. Cell 32, 151-159.

Hedgpeth, J., Balivet, M. and Eisen, H. (1978). Lambda phage promoter used to enhance expression of a plasmid-cloned gene. Molec. Gen. Genet. 163, 197-203.

Henner, D. J. (1990). Expression of heterologous genes in *Bacillus subtilis*. Meth. Enzymol. 185,199-201.

Ho, Y.-K. and Fung, B. K.-K. (1984). Characterization of transducin from bovine rod outer segmetns. The role of sulfhydryl groups. J. Biol. Chem. 259, 6694-6699.

Hurley, J. B., Simon, M. I., Teplow, D. B., Robishaw, J. D. and Gilman, A. G. (1984). Homologies between signal transducing G proteins and *ras* gene products. Science 226, 860-862.

Hurley, J. B. (1987). Molecular properties of the cGMP cascade of vertebrate phoptoreceptors. Ann. Rev. Physiol. 49, 793-812.

Johnson, R. C., Yin, J. C. P. and Reznikoff, W. S. (1982). Control of Tn5 transposition in *Escherichia coli* is mediated by protein from the right repeat. Cell 30, 873-882.

Johnson, R. C. and Simon, M. I. (1985). Hin-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp recombinational enhancer. Cell 41, 781-791.

Jones, T. L. Z., Simonds, W. F., Merendino, Jr., J. J., Brann, M. R. and Spiegel, A. M. (1990). Myristylation of an inhibitory GTP-binding protein α is essential for its membrane attachment. Proc. Natl. Acad. Sci . U. S. A. 87 568-572.

Kaufman, R. J. (1990a). Vectors used for expression in mammalian cells. Meth. Enzymol. 185, 487-511.

Kaufman, R. J. (1990b). Selection and coamplification of heterologous genes in mammalian cells. Meth. Enzymol. 185, 537-566.

Kemp, D. J., Smith, D. B., Foote, S. J., Samaras, N. and Peterson, M. G. (1989) Colorimetric detection of specific DNA segments by polymerase chain reactions. Proc. Natl. Acad. Sci. U. S. A. 86, 2423-2427.

Kikuchi, A., Yamashita, T., Kawata, M., Yamamoto, K., Ikeda, K., Tanimoto, T. and Takai, Y. (1988). Purification and characterization of a novel GTP-binding protein with a molecular weight of 24,000 from bovine brain membranes. J. Biol. Chem. 263, 2897-2904.

Klebe, R. J., Harriss, J. V., Sharp, Z. D. and Douglas, M. G. (1983). A general method for polyethylene-glycol-induced genetic transformation of bacteria and yeast. Gene 25, 333-341.

Kunkel, T. A., Roberts, J. D. and Zakour, R. A. (1987). Rapid and efficient mutagenesis without phenotypic selection. Meth. Enzymol. 154, 367-382.

Kurjan, J. and Herskowitz, I. (1982). Structure of a yeast pheromone gene (MFα): A putative α-factor precursor contains four tandem copies of mature α-factor. Cell 30, 933-943.

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

Landis, C. A., Masters, S. B., Spada, A., Pace, A. M. Bourne, H. R. and Vallar, L. (1989). GTPase inhibiting mutations activate the α subunit of G_S and stimulate adenylyl cyclase in human pituitary tumours. Nature 340, 692-696.

Lapetina, E. G. and Reep, B. R. (1987) Specific binding of [α-^{32}P]GTP to cytosolic and membrane-bound proteins of human platelets correlates with the activation of phospholipase C. Proc. Natl. Acad. Sci. U. S. A. 84, 2261-2265.

Lerea, C. L., Somers, D. E., Hurley, J. B., Klock, I. B. and Bunt-Milam, A. H. (1986). Identification of specific transdcuin α subunits in retinal rod and cone photoreceptors. Science 234, 77-80.

Linder, M. E., Ewald, D. A., Miller, R. J. and Gilman, A. G. (1990). Purification and characterization of $G_{o\alpha}$ and three types of $G_{i\alpha}$ after expression in *Escherichia coli*. J. Biol. Chem. 265, 8243-8251.

Lochrie, M. A. and Simon, M. I. (1988). G protein multiplicity in eukaryotic signal transduction systems. Biochemistry 27, 4957-4965.

Lorincz, A. (1984). Quick preparation of plasmid DNA from yeast. Focus 6, 11.

Luckow, V. A. and Summers, M. D. (1988). Trends in the development of baculovirus expression vectors. BIO/Technology 6, 47-55.

Lundblad, R. L., Kingdon, H. S. and Mann, K. G. (1976). Thrombin. Meth. Enzymol. 45, 156-176.

Lyons, S. and Nelson, N. (1984). An immunological method for detecting gene expression in yeast colonies. Proc. Natl. Acad. Sci. U. S. A. 81, 7426-7430.

Maniatis, T., Fritsch, E. F. and Sambrok, J. (1982). Molecular cloning: a laboratory manual. Cold Spring Harbor Labratory, Cold Spring Harbor, New York.

Matsuura, Y., Possee, R. D., Overton, H. A. and Bishop, D. (1987). Baculaovirus expression vectors: the requirements for high level expression of proteins, including glycoproteins. J. Gen. Virol. 68, 1233-1250.

Mattera, R., Yatani, A., Kirsch, G. E., Graf, R., Okabe, K., Olate, J., Codina, J., Brown, A. and Birnbaumer, L. (1989). Recombinant α_i-3 subunit of G protein activates G_k-gated K^+ channels. J. Biol. Chem. 264, 465-471.

McGrath, J. P., Capon, D. J., Goeddel, D. V. and Levinson, A. D. (1984). Comparative biochemical properties of normal and activated human *ras* p21 protein. Nature 310, 644-649.

Milburn, M. V., Tong,. L., deVos, A. M., Brunger, A., Yamaizumi, Z., Nishimura, S. and Kim, S.-H. (1990). Molecular switch for signal transduction: Structural differences between active and inactive forms of protooncogenic *ras* proteins. Science 247, 939-945.

Miller, D. W., Safer, P. and Miller, L. K. (1986). An insect baculovirus host-vector for high-level expression of foreign genes. In: Genetic Engineering, Vol. 8. Principles and Methods. eds., J. K. Setlow and A. Hollander. Plenum Publishing Corp., New York. pp. 277-298.

Miyajima, A., Bond, M. W., Otsu, K., Arai, K. and Arai, N. (1985). Secretion of mature mouse interleukin-2 by *Saccharomyces cerevisiae:* use of a general secretion vector containing promoter and leader sequences of the mating factor pheromone α-factor. Gene 37, 155-161.

Miyajima, I., Nakafuku, M., Nakayama, N., Brenner, C., Miyajima, A., Kaibuchi, K., Arai, K., Kaziro, Y. and Matsumoto, K. (1987). *GPA1*, a haploid-specific essential gene, encodes a yeast homolog of mammalian G protein which may be involved in mating factor signal transduction. Cell 50, 1011-1019.

Mott, J. E., Grant, R. A., Ho, Y.-S. and Platt, T. (1985). Maximizing gene expression from plasmid vectors containing the λPL promoter: strategies for overproducing transcription termination factor ρ. Proc. Natl. Acad. Sci. U. S. A. 82, 88-92.

Mumby, S. M., Heukeroth, R.O., Gordon, J. I. and Gilman, A. G. (1990). G-protein α-subunit expression, myristylation, and membrane association in COS cells. Proc. Natl. Acad. Sci. U. S. A. 87, 728-732.

Nagai, K. and Thorgersen, H. C. (1987). Synthesis and sequence-specific proteolysis of hybrid proteins produced in *Escherichia coli*. Meth. Enzymol. 153, 461-481.

Nagarajan, V. (1990). System for secretion of heterologous proteins in *Bacillus subtilis*. Meth. Enzymol. 185, 214-223.

Nakafuku, M., Obara, T., Kaibuchi, K., Miyajima, I., Miyajima, A., Itoh, H., Nakamura, S., Arai, K.-I., Matsumoto, K. and Kaziro, Y. (1988). Isolation of a second yeast *Saccharomyces cerevisiae* gene (GPA1) coding for guanine nucleotide-binding regulatory protein: Studies on its structure and possible functions. Proc. Natl. Acad. Sci. U. S. A. 85, 1374-1378.

Nathans, J. and Hogness, D. (1983). Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. Cell 34, 807-814.

Navon, S. E. and Fung, B. K.-K. (1987). Characterization of transducin from bovine retinal rod outer segments. Participation of the amino-terminal region of Tα in subunit interaction. J. Biol. Chem. 262, 15746-15751.

Neuhard J. and Nygaard, P. (1987). Purines and pyrimidines. In: *Escherichia coli* and *Salmonella typhimurium*. Cellular and Molecular Biology, Volume 1. ed., F. C. Neidhardt. American Society for Microbiology, Washington, D.C. pp. 445-473.

Oprian, D., Molday, R. S., Kaufman, R. J. and Khorana, H. G. (1987). Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. Proc. Natl. Acad. Sci. U. S. A. 84, 8874-8878.

Pang, I.-H. and Sternweis, P. C. (1989). Isolation of the α subunits of GTP-binding regulatory proteins by affinity chromatography with immobilized βγ subunits. Proc. Natl. Acad. Sci. U. S. A. 86, 7814-7818.

Piatak, M., Lane, J. A., Laird, W., Bjorn, M. J., Wang, A. and Willaims, M. (1988). Expression of soluble and fully functional ricin A chain in *Escherichia coli* is temperature-sensitive. J. Biol. Chem. 263, 4837-4843.

Poe, M., Scolnick, E. M. and Stein, R. (1985). Viral Harvey *ras* p21 expressed in *Escherichia coli* purifies as a one-to-one complex with GDP. J. Biol. Chem. 260, 3906-3909.

Polakis, P. G., Weber, R. F., Nevins, B., Didsbury, J. R., Evans, T. and Snyderman, R. (1989). Identification of the *ral* and *rac1* gene products, low molecular mass GTP-binding proteins from human platelets. J. Biol. Chem. 264, 16383-16389.

Possee, R. D. and Howard, S. C. (1987). Analysis of the polyhedrin gene promoter of the *Autogpapha californica* nuclear polyhedrosis virus. Nucl. Acid Res. 15, 10233-10248

Roberts, R. J. (1985). Restriction and modification systems and their recognition seqeunces. Nucl. Acid Res. 13, r165-r200.

Sakmar, T. P. and Khorana, H. G. (1988). Total synthesis and expression of a gene for the α-subunit of rod outer segment guanine nucleotide-binding protein (transducin). Nucl. Acid Res. 16, 6361-6372.

Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. and Petersen, G. B. (1982). Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. 162, 729-773.

Schatzman, A. R. and Rosenberg, M. (1985). Efficient expression of heterologous genes in *Escherichia coli*. The pAS system and its applications. Ann. N. Y. Acad. Sci. 478, 233-248.

Schein, C. H. (1989). Production of soluble recombinant proteins in bacteria. BIO/Technology 7, 1141-1149.

Schein, C. H. and Noteborn, M. H. M. (1988). Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperatures. BIO/Technology 6, 291-294.

Schoner, B. E., Belagaje, R. M. and Schoner, R. G. (1990). Enhanced expression efficiency with two-cistron expression system. Meth. Enzymol. 185, 94-103.

Shaltiel, S. (1984). Hydrophobic chromatography. Meth. Enzymol. 104, 69-96.

Simons, P. C. and Vander Jagt, D. L. (1981) Purification of glutathione S-transferases by glutathione-affinity chromatography. Meth. Enzymol. 77, 235-237.

Smith, D. B. and Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene 67, 31-40.

Squires, C. H., Childs, J., Eisenberg, S. P., Polverni, P. J. and Sommer, A. (1988). Production and characterization of human fibroblast growth factor from *Escerichia coli*. J. Biol. Chem. 263, 16297-16302.

St. Angelo, C., Smith, G. E., Summers, M. D. and Krug, R. M. (1987). Two of the three influenza viral polymerase proteins expressed by baculovirus vectors form a complex in insect cells. J. Virol. 61, 361-365.

Stryer, L. (1986). Cyclic GMP cascade of vision. Ann. Rev. Neurosci. 9, 87-119.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. Meth. Enzymol. 185, 60-89.

Summers, M. D. and Smith, G. E. (1987). A manual of methods for baculovirus vectors and insect cell culture procedures. Texas Agricultural Experimental Station, Texas A & M University. Bulletin No. 1555.

Tabor, S. and Richardson, C. C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci., U. S. A. 82, 1074-1078.

Takahara, M., Hibler, D. W., Barr, P. J., Gerlt, J. A. and Inouye, M. (1985). The ompA signal peptide directed secretion of Staphlococcal nuclease A by *Escherichia coli*. J. Biol. Chem. 260, 2670-2674.

Takahara, M., Sagai, H., Inouye, S. and Inouye, M. (1988). Secretion of human superoxide dismutase in *Escherichia coli*. Bio/Technology 6, 195-198.

Tanabe, T., Nukada, T., Nishikawa, Y., Sugimoto, K., Suzuki, H., Takahashi, H., Noda, M., Haga, T., Ichiyama, A., Kangawa, K., Minamino, N., Matsuo, H. and Numa, S. (1985). Primary structure of the α-subunit and its relationship to *ras* proteins. Nature 315, 242-245.

Towbin, H., Staehlin, T. and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci., U. S. A. 76, 4350-4354.

Tsao, T., Hsieh, J.-C., Durkin, M. E., Wu, C., Chakravarti, S., Dong, L.-J., Lewis, M. and Chung, A. E. (1990) Characterization of the basement membrane glycoprotein entactin synthesized in a baculovirus expression system. J. Biol. Chem. 265, 5188-5191.

Whiteway, M., Hougan, L., Dginard, D., Thomas, D. Y., Bell, L., Saari, G. C., Grant, F. J., O'Hara, P. and MacKay, V. L. (1989). The *STE4* and *STE18* genes of yeast encode potential β and γ subunits of the mating factor receptor-coupled G protein. Cell 56, 467-477.

Woolley, P. and Clark, B. F. C. (1989). Homologies in the structures of G-binding proteins: An analysis based on elongation factor Ef-Tu. BIO/Technology 7, 913-920.

Yamazaki, A., Tatsumi, M. and Bitensky, M. W. (1988). Purification of rod outer segment GTP-binding protein subunits and cGMP phosphodiesterase by single-step column chromatography. Meth. Enzymol. 159, 702-710.

Yatani, A., Mattera, R., Codina, J., Graf, R., Okabe, K., Padrell, E., Iyengar, R., Brown, A. M. and Birnbaumer, L. (1988). The G protein-gated atrial K^+ channel is stimulated by three distinct $G_i\alpha$–subunits. Nature 336, 680-682.

Yee, R. and Liebman, P. A. (1978). Light-activated phosphodiesterase of the rod segment. J. Biol. Chem. 253, 8902-8909.

Zoller, M. J. and Smith, M. (1983). Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. Meth. Enzymol. 100, 468-500.

Table 1. General characteristics of heterologous expression systems.

Characteristic	Expression system		
	E. coli	Yeast	Insect cell
1. Level of expression	high	low-high	medium-high
2. Localization of product	often insoluble	often correct	correct
3. Proteolysis	high	medium-high	low
4. Posttranslational modifications			
a. myristylation	no	yes	yes
b. phosphorylation	no	yes	yes
5. Endogenous G proteins	no	yes	yes
6. Time course of expression	minutes	hours	days
7. Correct protein folding	often incorrect	varies	usually correct
8. Cost of media	low	low	high
9. Ease of site-directed mutagenesis	easy	intermediate	difficult

Table 2. Strain list.

Yeast	Genotype*
DSR865-3B	α, leu2, his5, lys1, trp1, ura3, ade2
EJ102	α, trp1, pep4-3
HR125-5Dα	α, ura3-52, leu2-3, leu2-112, $trp1^{am}$, his3, his4
HR125-5Da	**a**, ura3-52, leu2-3, leu2-112, $trp1^{am}$, his3, his4
NNY105	α, kex2, trp1, ura3-52
SF838-10α	α, leu2-2, leu2-112, ura3-52, his4, ade6, pep4-3
SM1269	α, his3, ura3, trp1, leu2, ade8, can1, ram
SM1291	SM1269, pYep24-RAM
SR741-3B	**a**, leu2-1, leu2-112, his3-11, trp1, ura3
20B-12	α, pep4-3, trp1
Bacteria	Genotype
AR58	*E. coli* K12 [λcI^{ts}], also ts in unknown proteolytic gene(s)
CAG629	$lacZ^{am}$, trp^{am}, pho^{am}, $supC^{ts}$, mal, rpsL, phe, rel, lon, rpoH165
CAG1139	F^-, thi, thr, leu, lacY1, tonA21, supE44, galK, lon100
D1210	*E. coli* B, pro, leu, thi, lacY, str^R, r_K^-, m_K^-, endA, recA, $laci^q$
JM101	Δlac, pro, supE, thi, F' (traD36, proAB, $lacI^q$ Z ΔM15)
MC1061	F^-, araD139, ΔlacIPOZYA, Δlac X74, strA, Δara ABOIC-leu, galK, $hsdR^-$, $hsdM^+$, rpsL, galU
N4830-1	F^-, su, his, ilv, [λ N^+, cI_{857}, ΔBAM(int-cIII), ΔH1(cro-chlA)]
N99cI^+	gal K, rpsL, [λ N^+, cI^+, Δ-8(chlD-pgl), ΔH1(cro-chlA)]
Insect	Genotype
Sf9	*Spodoptera frugiperda*, clone 9 (Fall armyworm ovary cell line)
AcNPV, L1	*Autographa californica* nuclear polyhedrosis virus, wild type

* Standard genotype abbreviations appropriate for the organism listed are used.

Table 3. Features of transducin alpha subunit expression vectors.

Cell	Vector type	Plasmid	Coding region
E. COLI	TAC PROMOTER	pML 4, 6	$rT_c\alpha$
E. COLI	TRC PROMOTER	pML 12	$rT_r\alpha$
E. COLI	λ_R PROMOTER	pML15	$rT_r\alpha$
E. COLI	λ_L/TRC PROMOTER	pML17	$rT_r\alpha$
E. COLI	λ_L/TAC PROMOTER	pML18	$rT_c\alpha$
E. COLI	λ_R/TRC PROMOTER	pML 25	$rT_r\alpha$
E. COLI	T7 PROMOTER	pML 49, 52	$rT_r\alpha$
E. COLI	DUAL CISTRON	pML 46, 50	$rT_r\alpha$
E. COLI	DUAL CISTRON	pML 53	$rT_c\alpha$
E. COLI	GST FUSION	pML 56	GST-$rT_r\alpha$
YEAST	ADH PROMOTER	pML 11	$rT_c\alpha$
YEAST	SECRETION	pML 35	$rT_r\alpha$
YEAST	SECRETION/INTEGRATION	pML 38	$rT_r\alpha$
Sf9	BACULOVIRUS	pML 34	$rT_r\alpha$
Sf9	BACULOVIRUS	pML 45	polyhedrin-$rT_r\alpha$

Table 4. Saturation density of yeast strains.

Strain	Plasmid	OD*
HR125a	pMFα8	2.1
	pML 35	2.3
	pML 38	1.8
HR125α	pMFα8	3.0
	pML 35	3.1
	pML 38	2.4
20B-12	none	14.1
	pMFα8	5.9
	pML 35	0.0
	pML 38	2.1
NNY105	none	13.2
	pMFα8	4.5
	pML35	4.2
	pML 38	0.2
DSR865-3B	none	3.2
	pML35	2.2
EJ102	none	6.3
	pML35	4.5
SM1269	none	2.4
	pML 35	2.0
SM1291	none	2.6
	pML 35	1.9

* Yeast strains were grown for two days in selective media at 30°C and the optical density (OD) of the culture at 600 nm was measured.

Table 5. Plasmid recovery from yeast strains.

Strain	Plasmid	Ampicillin resistant colonies*
HR125α	pMFα8	84
	pML 35	7
	pML 38	0
HR125a	pMFα8	18
	pML 35	18
	pML 38	0

*Yeast plasmid DNA was prepared by a rapid glass bead lysis/ phenol extraction procedure (Lorincz, 1984), transformed into *E. coli* by the $CaCl_2$ procedure (Maniatis *et al.*, 1982), and plated on LB plates with ampicillin (100 μg/ml).

Table 6. Efficiency of yeast transformation.

Yeast strain:	HR125**a**	HR125α	NNY105	20B-12
	Transformants per microgram of plasmid[a]			
Plasmid				
pMFα8	2000	2000	2000	2000
pML 35	1000	1000	1000	400

*Yeast were transformed according to Klebe *et al.* (1983) and plated on the appropriate selective media. The number of colonies was counted after three days at 30°C.

Table 7. Characteristics of strains containing pML 18.

Strain	Temperature	Media	Plasmid	Growth*
N4830-1	30°C	LB, AMP	pPL	+
			pML 18(+)	+
			pML 18(-)	+
N4830-1	37°C	LB	pPL	+
			pML 18(+)	+
			pML 18(-)	+
N4830-1	37°C	LB, AMP	pPL	-
			pML 18(+)	-
			pML 18(-)	+
CAG1139	37°C	LB, AMP	pML 18(+)	+
			pML 18(-)	+
CAG1139	37°C	LB, AMP, IPTG	pML 18(+)	-
			pML 18(-)	+

*Strains were inoculated in the given media and incubated at the given temperature for 12-24 hours. A plus (+) indicates the strain grew to saturation. A minus (-) indicates the strain did not grow to saturation. In these cases the strain either did not grow at all or it grew poorly and showed clear signs of cell lysis. Ampicillin (AMP) was used at a concentration of 100 μg/ml and IPTG was used at 0.1 mM.

Table 8. Expression of $rT_c\alpha$ in CAG1139 pML 18(+).

Temperature	Media	Promoters induced	Relative amount of expression*
30°C	-IPTG	none	1
30°C	+IPTG	P_{tac}	3
37°C	-IPTG	P_L	1
37°C	+IPTG	P_{tac}/P_L	10

*CAG1139 pML 18(+) was grown in LB, AMP media at 30°C to saturation, diluted 1 to 100 in LB with or without 0.1 mM IPTG and incubated at the indicated temperature for 4 hours. Cell extracts were prepared and analyzed for $rT_c\alpha$ expression by Western blotting as described in Materials and Methods.

Table 9. Expression of $rT_r\alpha$ in CAG1139 pML 46 at different temperatures.

Temperature	$rT_r\alpha$ per liter of cells*	% of $rT_r\alpha$ that is soluble	soluble $rT_r\alpha$ per liter of cells
25°C	1.75	55	1.0
30°C	2.0	50	1.0
37°C	2.1	22	0.45

*CAG1139 pML 46 was grown to saturation in LB plus ampicillin (100 μg/ml) at the indicated temperatures. The culture was diluted 1 to 100 in LB plus ampicillin and induced with 0.1 mM IPTG at the indicated temperature for 30 minutes. Cell lysates were prepared by French pressure lysis and the soluble fraction by centrifugation of the crude lysate at 100,000 X g. The amount of $rT_r\alpha$ in the total, soluble, and insoluble fractions was determined by Western blotting using retinal $T_r\alpha$ as a standard.

Table 10. Chromatographic behavior of $rT_r\alpha$. A crude soluble extract of $rT_r\alpha$ was prepared from CAG1139 pML 46 cells as described in Materials and Methods and chromatographed as indicated. The elution profile of $rT_r\alpha$ was determined by Western blotting. The degree of purification is calculated by comparing the ratio of total protein to $rT_r\alpha$ present in the peak fraction with that of the starting material, i.e., it is not based on increases in specific activity, but on increases in the percentage of $rT_r\alpha$ protein in a sample. The yield is calculated by dividing the total $rT_r\alpha$ that eluted from the column by the total amount of $rT_r\alpha$ present in the starting material.

Method	Yield	Purification	Eluted with	Comments
QMA	30%	2	0-0.5 M NaCl	peak of $rT_r\alpha$ at 0.3 M
DEAE	50%	2	0-0.5 M NaCl	peak of $rT_r\alpha$ at 0.3 M
HEPARIN	N. D.	N. D.	0-1.0 M NaCl	might be some purification
MONO Q	N. D.	N. D.	0-0.4 M NaCl	peak of $rT_r\alpha$ at 0.35 M
METAL CHELATE	N. D.	N. D.	EDTA, GTP,	no selective elution
(Cu^{2+}, Zn^{2+})			0.5 M NaCl	
PHOPHOCELLULOSE	>50%	1	1.0 M NaCl	$rT_r\alpha$ did not bind
HEPTYL AGAROSE	30%	3	0.5-0.0 M A.S.	elutes at 0.0-0.1 M A.S.
PHENYL SUPEROSE	<10%	N. D.	0.4-0.0 M A.S.	lost $rT_r\alpha$ on column
SUPEROSE 12	>50%	2	0.1 M NaCl	2 peaks of $rT_r\alpha$
SEPHACRYL S-200	30%	2	transducin buffer	very broad peak of $rT_r\alpha$
BLUE SEPHAROSE	50%	1	0-0.5 M KCl	$rT_r\alpha$ peaks at 0.1 M, 0.3 M
AFFIGEL BLUE	<5%	N. D.	0-0.5 M KCl	low recovery of $rT_r\alpha$
BLUE A	40%	1.4	0.5 M NaCl	batch elution
BLUE B	50%	1	0.5 M NaCl	batch elution
GREEN	40%	1.5	0.5 M NaCl	batch elution
RED	20%	1.2	0.5 M NaCl	batch elution
ORANGE	60%	1	0.5 M NaCl	batch elution
WHITE	70%	1	0.5 M NaCl	batch elution
A.S. precipitation	80%	4	-	
HYDROXYLAPATITE	5-10%	1	0.0- 0.5 M KPO_4	low recovery
HYDROXYLAPATITE	<5%	N.D.	0.0-2.0 M NaCl	low recovery
RHODOPSIN	<1%	N. D.	0.1 mM GTP	$rT_r\alpha$ did not bind

Figure 1. Diagram of pML 1. The bovine cone transducin alpha subunit cDNA ($rT_C\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{tac}, trp/lac promoter; TET^S, truncated tetracycline resistance gene.

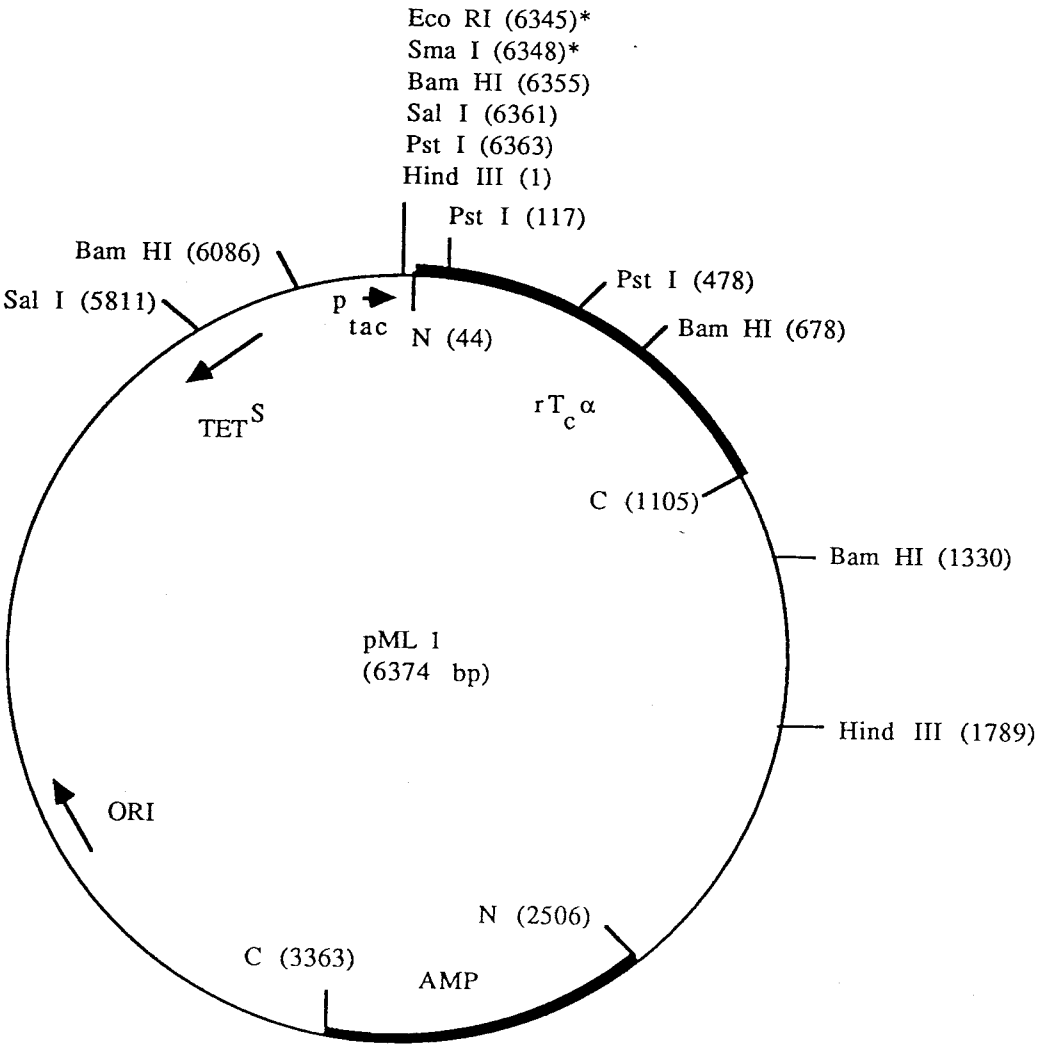

Eco RI (6345)*
Sma I (6348)*
Bam HI (6355)
Sal I (6361)
Pst I (6363)
Hind III (1)
Pst I (117)
Pst I (478)
Bam HI (678)
Bam HI (6086)
Sal I (5811)
p
tac
N (44)
TET^S
rT_c α
C (1105)
Bam HI (1330)
pML 1
(6374 bp)
Hind III (1789)
ORI
N (2506)
C (3363)
AMP

Figure 2. Sequence of the promoter region of pML 1. The sequence was derived from Brosius and Holy (1984) and Lochrie *et al.* (1985). The slash indicates the fusion point between the trp and lac promoters which creates the tac promoter (p_{tac}). The -35 region, -10 region, transcription initiation site, and the ribosome binding site (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram. The ATG preceeding the $rT_c\alpha$ coding region is in bold-faced lettering.

Hinc II

TGTTGACAATTAATCATC/GGCTCGTATAATGTGTGGAATTGTGAGCGGAT (6314)

-35 region (Ptac) **-10 region (Ptac)** **Transcription initiation site (Ptac)**

Eco RI Sma I Bam HI Sal I Pst I

AACAATTTCACACAGGAAACAGAATTCCCGGGGATCCGTCGACCTGCAGC (6364)

RBS

Hind III

CAAGCTTAAGGGCTGGAGAAAGCTGCCGAGGAGGAGACGG**ATG**AAAGATG (48)

MET

$rT_C\alpha$

Figure 3. Diagram of pML 4. The bovine cone transducin alpha subunit cDNA ($rT_C\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{tac}, trp/lac promoter; TET^S, truncated tetracycline resistance gene.

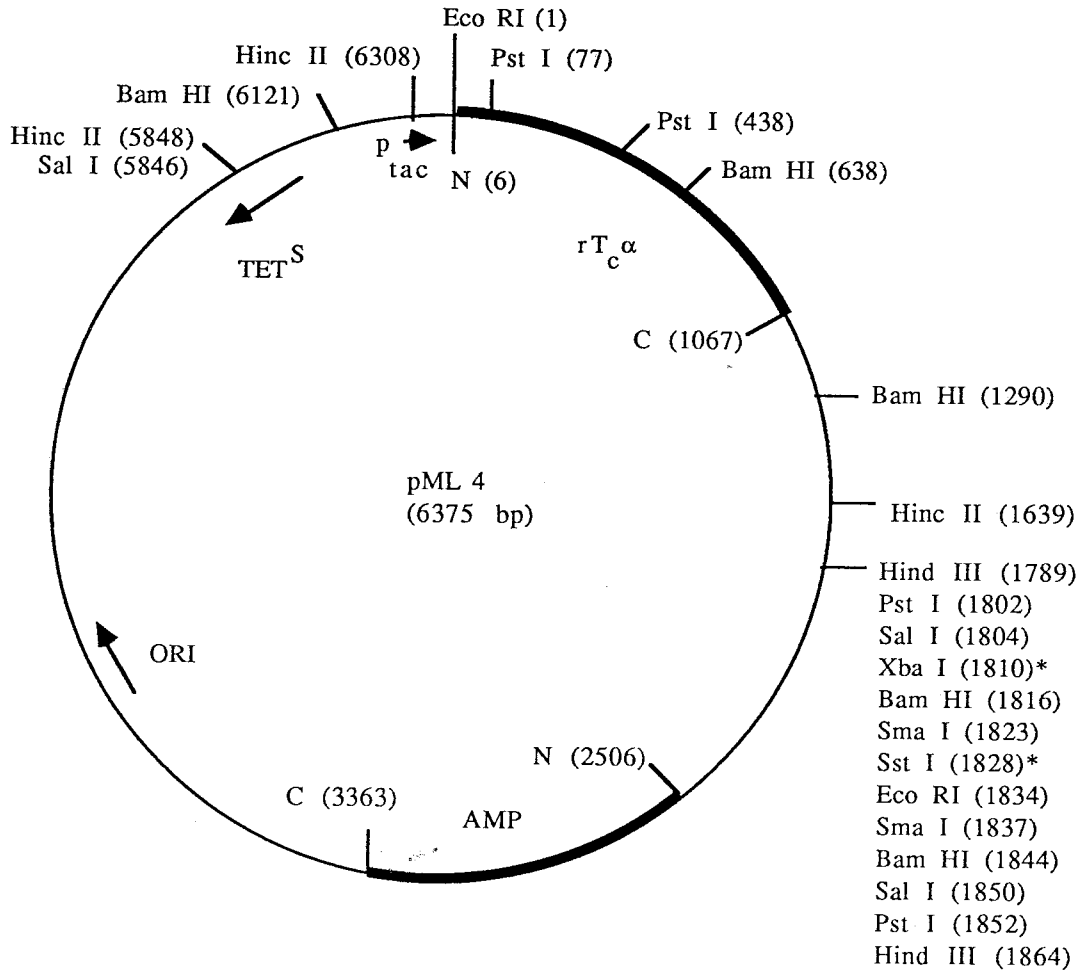

Eco RI (1)
Hinc II (6308)
Pst I (77)
Bam HI (6121)
Pst I (438)
Hinc II (5848)
Sal I (5846)
p
tac
N (6)
Bam HI (638)
TET^S
rT_cα
C (1067)
Bam HI (1290)
pML 4
(6375 bp)
Hinc II (1639)
Hind III (1789)
Pst I (1802)
Sal I (1804)
Xba I (1810)*
Bam HI (1816)
Sma I (1823)
Sst I (1828)*
Eco RI (1834)
Sma I (1837)
Bam HI (1844)
Sal I (1850)
Pst I (1852)
Hind III (1864)
ORI
N (2506)
C (3363)
AMP

Figure 4. Sequence of the promoter region of pML 4. The sequence was derived from Amman *et al.* (1983) and Lochrie *et al.* (1985). The slash indicates the fusion point between the trp and lac promoters which creates the tac promoter (p_{tac}). The -35 region, -10 region, transcription initiation site, and the ribosome binding site (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

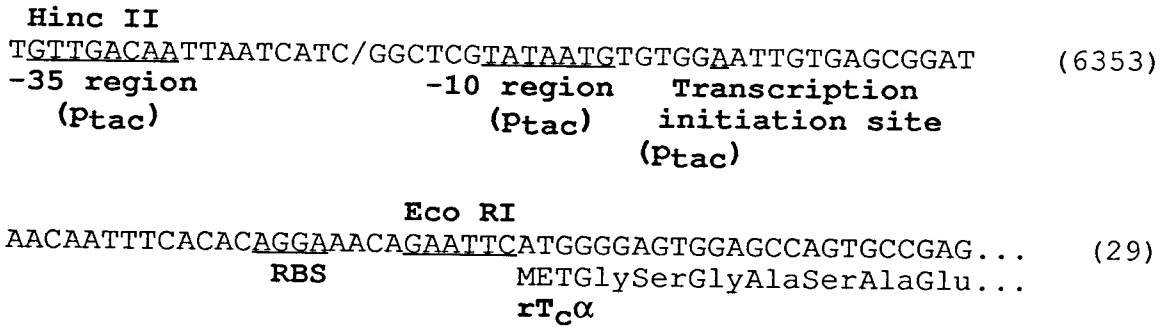

Hinc II
TGTTGACAATTAATCATC/GGCTCGTATAATGTGTGGAATTGTGAGCGGAT (6353)
-35 region (Ptac)
-10 region (Ptac)
Transcription initiation site (Ptac)
Eco RI
AACAATTTCACACAGGAAACAGAATTCATGGGGAGTGGAGCCAGTGCCGAG... (29)
RBS
METGlySerGlyAlaSerAlaGlu...
rTcα

Figure 5. Diagram of pML 7. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; TET^S, truncated tetracycline resistance gene.

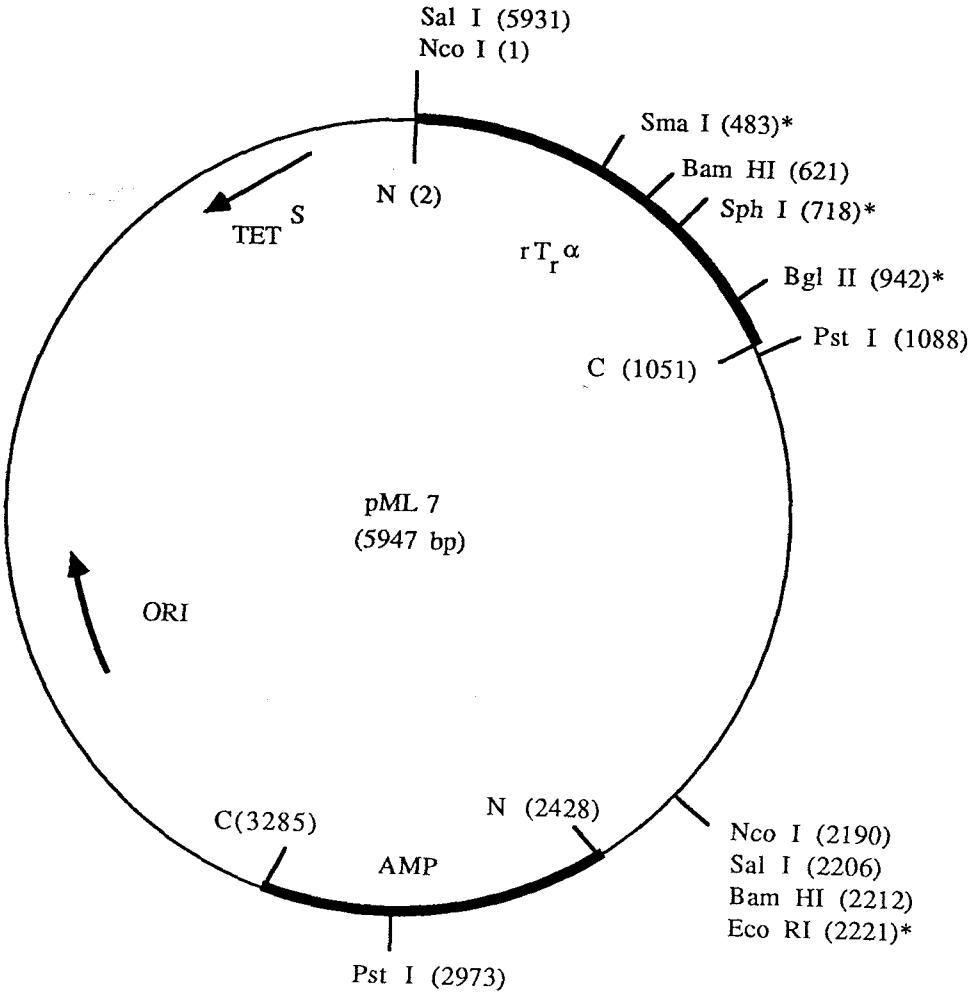

Sal I (5931)
Nco I (1)
Sma I (483)*
Bam HI (621)
Sph I (718)*
Bgl II (942)*
Pst I (1088)
C (1051)
N (2)
TET S
rT r α
pML 7
(5947 bp)
ORI
C(3285)
N (2428)
AMP
Nco I (2190)
Sal I (2206)
Bam HI (2212)
Eco RI (2221)*
Pst I (2973)

Figure 6. Diagram of pML 8. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{SP6}, bacteriophage SP6 RNA polymerase promoter.

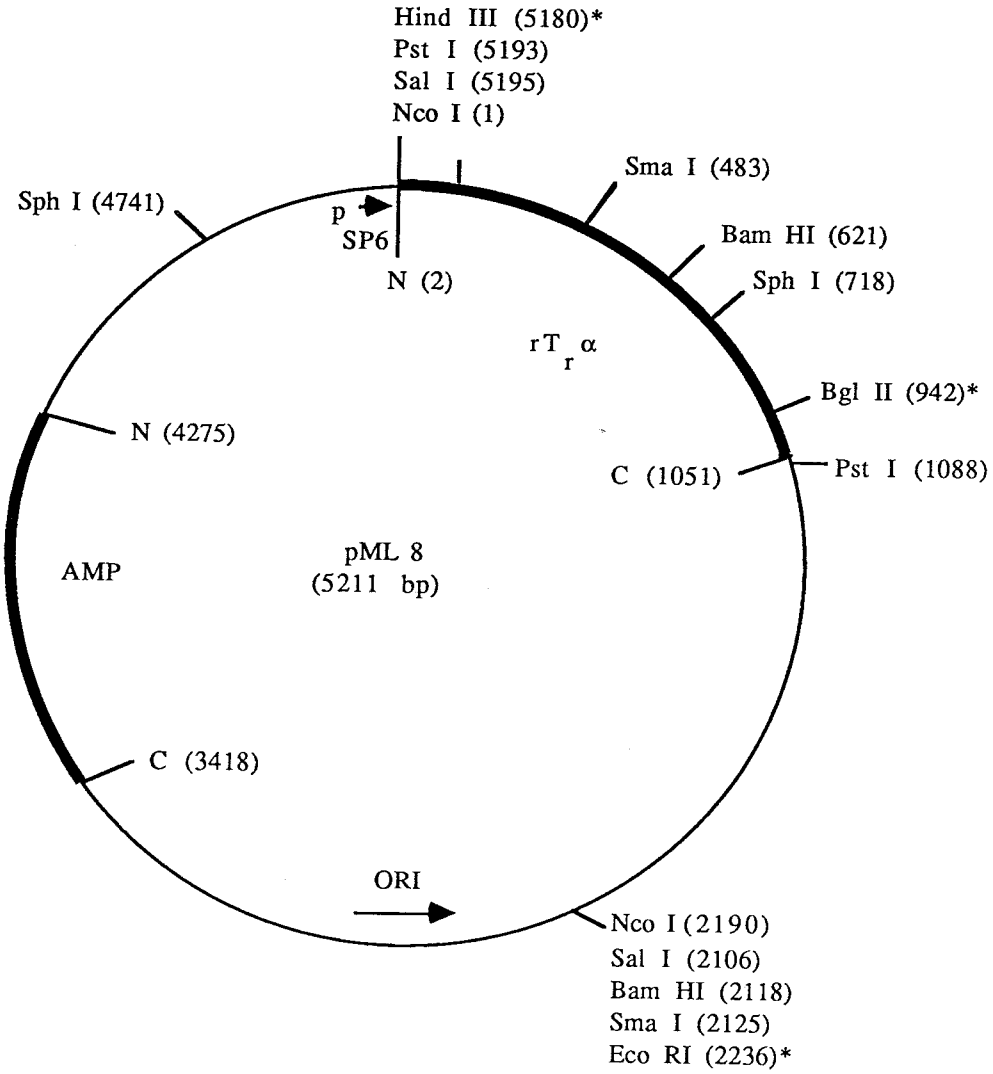

Hind III (5180)*
Pst I (5193)
Sal I (5195)
Nco I (1)
Sma I (483)
Bam HI (621)
Sph I (718)
Bgl II (942)*
Pst I (1088)
C (1051)
N (2)
p
SP6
Sph I (4741)
N (4275)
AMP
C (3418)
ORI
Nco I (2190)
Sal I (2106)
Bam HI (2118)
Sma I (2125)
Eco RI (2236)*
pML 8
(5211 bp)
rT r α

Figure 7. Diagram of pML 10. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{SP6}, bacteriophage SP6 RNA polymerase promoter.

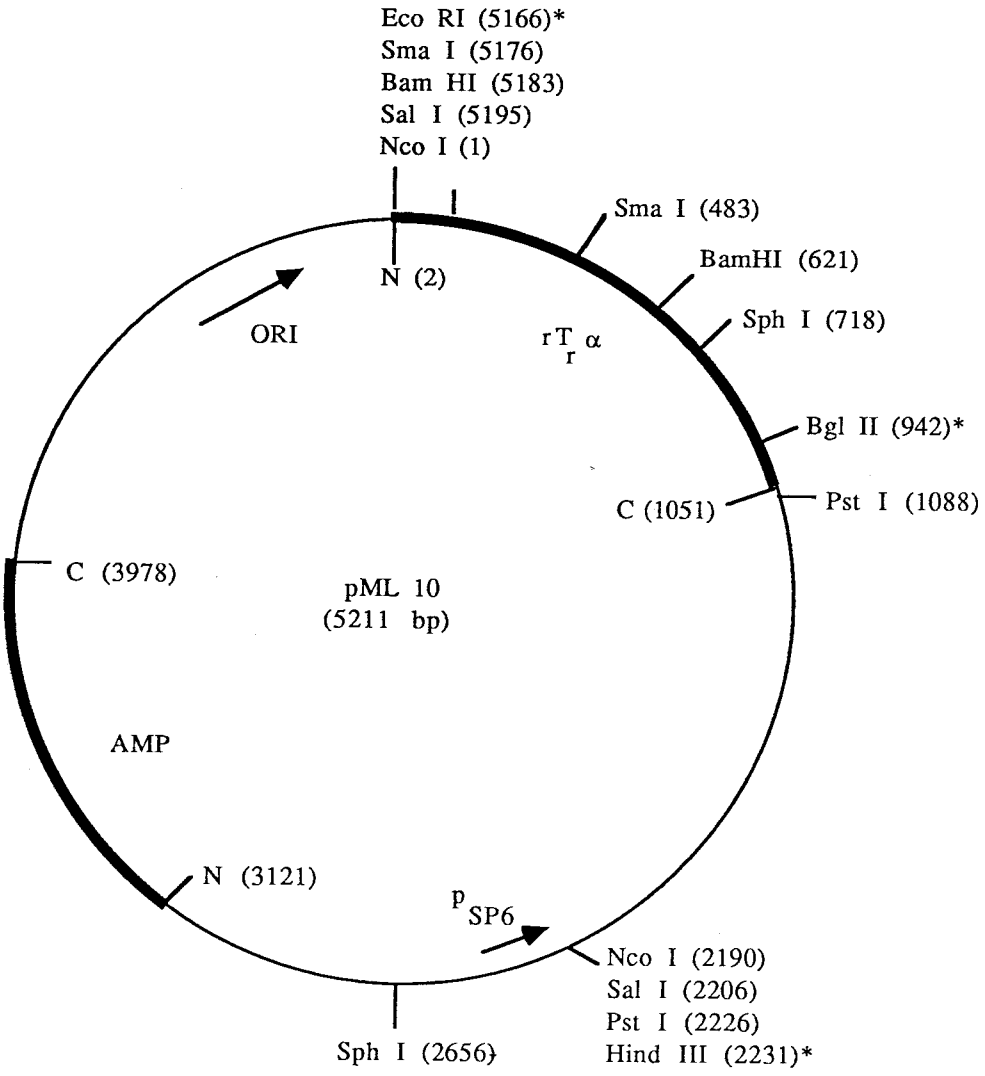

Eco RI (5166)*
Sma I (5176)
Bam HI (5183)
Sal I (5195)
Nco I (1)
Sma I (483)
BamHI (621)
Sph I (718)
Bgl II (942)*
Pst I (1088)
C (1051)
N (2)
ORI
rTr α
C (3978)
pML 10
(5211 bp)
AMP
N (3121)
pSP6
Nco I (2190)
Sal I (2206)
Pst I (2226)
Hind III (2231)*
Sph I (2656)

Figure 8. Diagram of pML 11. The bovine cone transducin alpha subunit cDNA ($rT_c\alpha$) and the yeast LEU 2 gene are represented as solid thick lines. DNA derived from the yeast 2μm plasmid are stippled and DNA related to the *E. coli* plasmid pBR322 are striped. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985). Their approximate positions are indicated in brackets and their exact positions are in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{ADH}, the yeast alcohol dehydrogenase promoter; AMP, ampicillin resistance gene; TET^S, truncated tetracycline resistance gene.

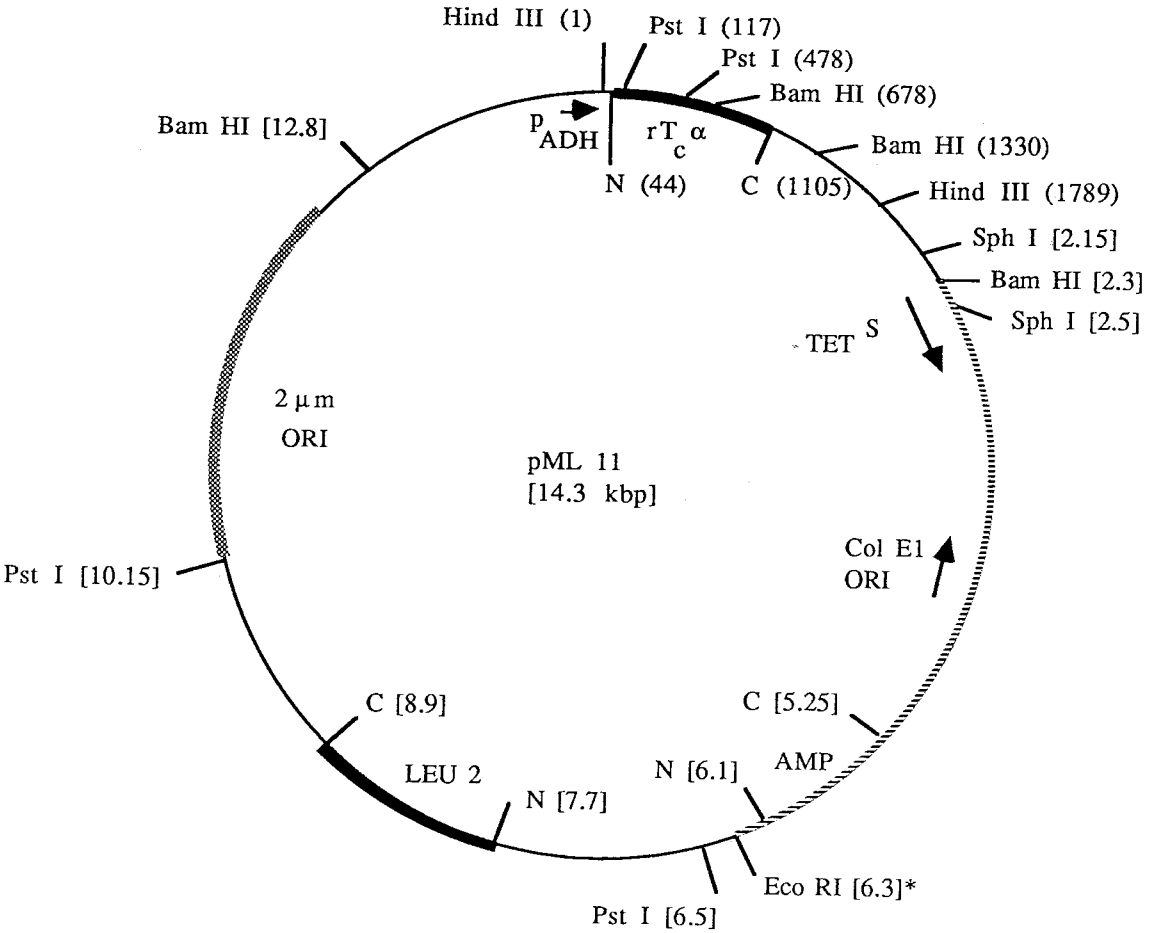

Hind III (1)
Pst I (117)
Pst I (478)
Bam HI (678)
Bam HI [12.8]
P ADH
rT c α
Bam HI (1330)
N (44)
C (1105)
Hind III (1789)
Sph I [2.15]
Bam HI [2.3]
Sph I [2.5]
TET S
2 μm
ORI
pML 11
[14.3 kbp]
Col E1
ORI
Pst I [10.15]
C [8.9]
C [5.25]
LEU 2
N [6.1]
AMP
N [7.7]
Eco RI [6.3]*
Pst I [6.5]

Figure 9. Sequence of the promoter region of pML 11. The sequence was derived from Bennetzen and Hall (1982), Ammerer (1983), and Lochrie *et al.* (1985). The TATA-box and transcription initiation sites are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

AGTTTGCCGCTTTGCTATCAAGTATAAATAGACCTGCAATTATTAATCTT [14200]
TATA-box

TTGTTTCCTCGTCATTGTTCTCGTTCCCTTTCTTCCTTCTTTCTTTTTCT [14250]

Hind III
GCACAATATTTCAAGCTATACCAAGCATACAATCAACTCCAAGCTTAAGG (9)
Transcriptional initiation sites (p_{ADH})

GCTGGAGAAAGCTGCCGAGGAGGAGACGGATGAAAGATGGGGAGTGGAGCC... (60)
METGlySerGlyAla...
$rT_C\alpha$

Figure 10. Diagram of pML 12. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{tac}, trp/lac promoter; TET^S, truncated tetracycline resistance gene.

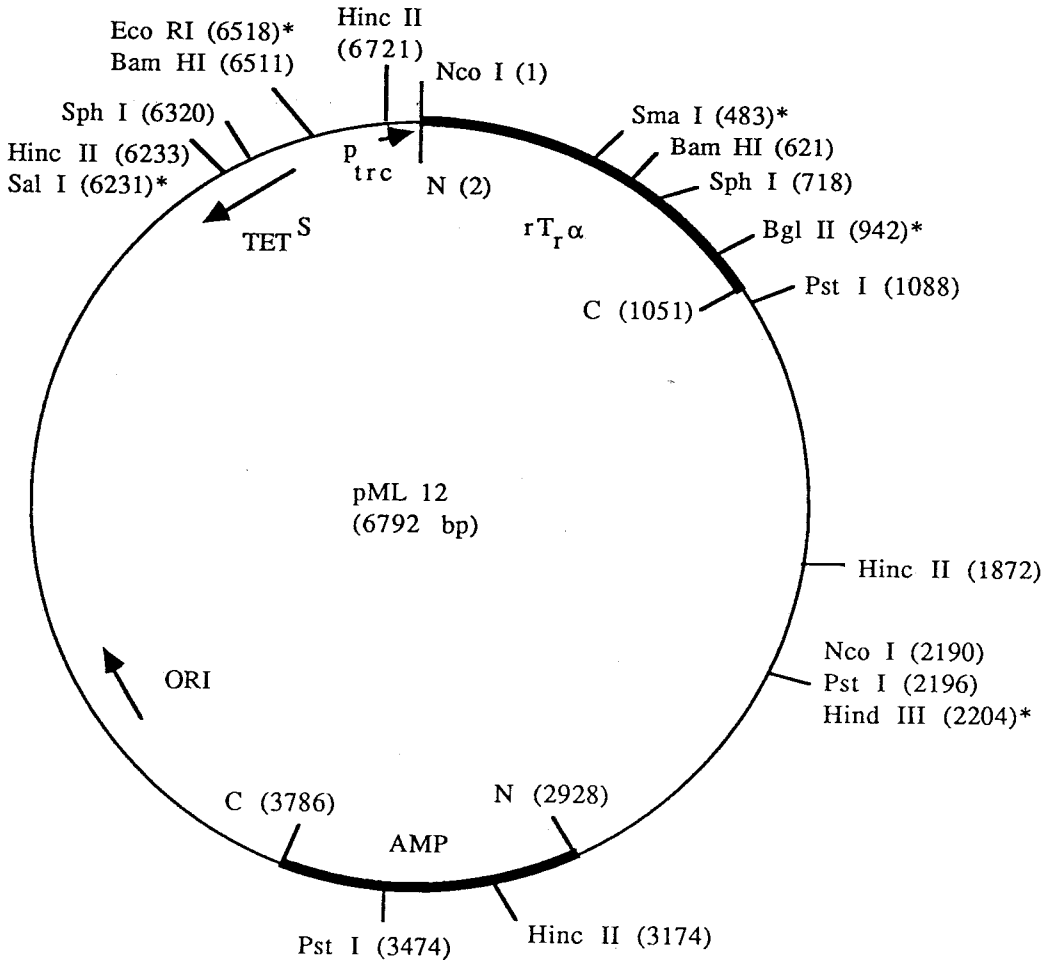

Eco RI (6518)*
Bam HI (6511)
Hinc II
(6721)
Nco I (1)
Sph I (6320)
Hinc II (6233)
Sal I (6231)*
p trc
N (2)
Sma I (483)*
Bam HI (621)
Sph I (718)
TET S
rT r α
Bgl II (942)*
Pst I (1088)
C (1051)
pML 12
(6792 bp)
Hinc II (1872)
Nco I (2190)
Pst I (2196)
Hind III (2204)*
ORI
C (3786)
N (2928)
AMP
Pst I (3474)
Hinc II (3174)

Figure 11. Sequence of the promoter region of pML 12. The sequence was derived from Amman *et al.*, 1983, Amman and Brosius, 1985, and Tanabe *et al.*, 1985. A cytosine between the -10 and -35 regions which makes the length between them one base longer than the same region of the p_{tac} promoter is underlined. The -35 region, -10 region, transcription initiation site, and the ribosome binding site (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

Bam HI
GGATCCG (6516)

Eco RI
GAATTCTCATGTTTGACAGCTTATCATCGACTGCACGGTGCACCAATGCT (6566)

TCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCGTA (6616)

AATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAAT (6666)

GTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGC (6716)

Hinc II
TGTTGACAATTAATCATCCGGCTCGTATAATGTGTGGAATTGTGAGCGGA (6766)
-35 Region (Ptrc) **-10 Region (Ptrc)** **Transcription initiation site (Ptrc)**

Nco I
TAACAATTTCACACAGGAAACAGACCATGGGGGCTGGGGCCAGCGCTGAG... (25)
RBS METGlyAlaGlyAlaSerAlaGlu...
$rT_r\alpha$

Figure 12. Diagram of pML 15. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. The bacteriophage λcI_{857} gene is represented as a striped line and Δbio256 as a stippled line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985). Asterisks denote unique restriction sites. The positions of the restriction sites are indicated in brackets or parentheses. Sites in parentheses are derived from the vector pKK233-2. Sites in brackets are derived from bacteriophage λ and are numbered according to the bacteriophage λ DNA sequence (Sanger *et al.*, 1982). The Nco I and Pvu II sites in the Δbio256 region were determined by restriction mapping and their precise location is not known. Because of space limitations the following bacteriophage λ genes have been omitted from the figure: cro [38041-38239], cII [38360-38651], and O [38686-39168]/(-285 to -192). Consult Figure 18 for the location of these genes. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{trc}, trp/lac hybrid promoter; p_L, the leftward bacteriophage λ promoter; p_R, the rightward bacteriophage λ promoter; λcI_{857}, the temperature sensitive bacteriophage λ repressor; TET^S, truncated tetracycline resistance gene; λ, bacteriophage λ N gene; Δbio256, a region of bacteriophage λ that was deleted and substituted with DNA from the *E. coli* bio operon during the formation of λbio256. The precise extent of Δbio256 is unknown. I have estimated the border may be near λ coordinate 34500. Hedgpeth (1978) suggested it may extend into the carboxyl terminus of the N gene (λ coordinate 35039) since the N gene activity of λbio256 is measurable but altered.

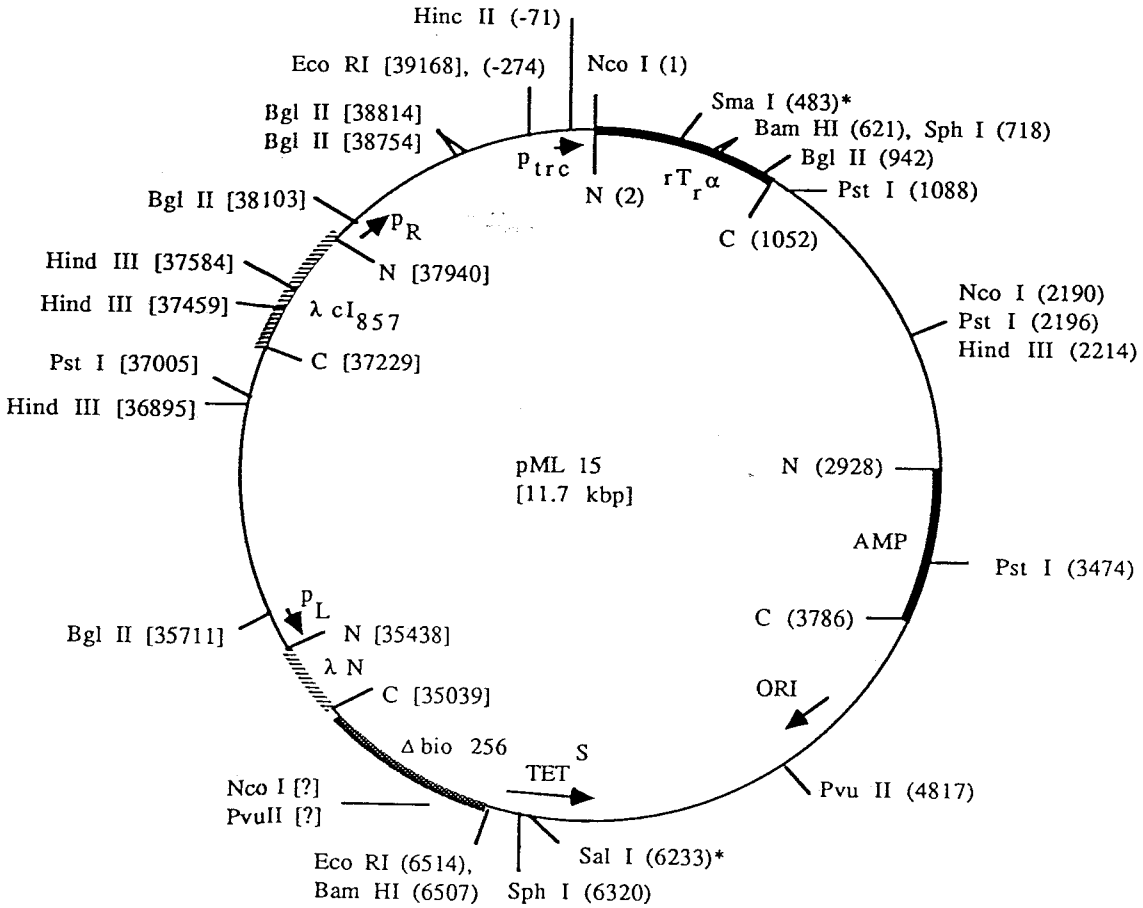

Hinc II (-71)
Eco RI [39168], (-274)
Nco I (1)
Sma I (483)*
Bam HI (621), Sph I (718)
Bgl II [38814]
Bgl II [38754]
Bgl II (942)
Pst I (1088)
Bgl II [38103]
Hind III [37584]
Hind III [37459]
Pst I [37005]
Hind III [36895]
Nco I (2190)
Pst I (2196)
Hind III (2214)
pML 15
[11.7 kbp]
N (2928)
AMP
Pst I (3474)
C (3786)
ORI
Pvu II (4817)
Bgl II [35711]
N [35438]
C [35039]
Δ bio 256
TET
S
Nco I [?]
PvuII [?]
Eco RI (6514),
Bam HI (6507)
Sal I (6233)*
Sph I (6320)
N [37940]
C [37229]
C (1052)
N (2)

Figure 13. Sequence of the region upstream of the $rT_r\alpha$ coding region in pML 15 and pML 25. The sequence was derived from Sanger *et al.* (1982), Amman *et al.* (1983), Amman and Brosius (1985), and Tanabe *et al.* (1985). A cytosine between the -10 and -35 regions, which makes the length between them one base longer than the same region of the p_{tac} promoter, is underlined. The -35 region, -10 region, transcription initiation site, and the ribosome binding site (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the pML 25 plasmid diagram.

Eco RI
GAATTCTCATGTTTGACAGCTTATCATCGACTGCACGGTGCACCAATGCT (5996)
AsnSerHisVal---
ΔO

TCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCGTA (6046)

AATCACTGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAAT (6096)

GTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGC (6146)

Hinc II
TGTTGACAATTAATCATCCGGCTCGTATAATGTGTGGAATTGTGAGCGGA (6196)
-35 Region (Ptrc) **-10 Region (Ptrc)** **Transcription initiation site (Ptrc)**

Nco I
TAACAATTTCACACAGGAAACAGACCATGGGGGCTGGGGCCAGCGCTGAG... (25)
RBS METGlyAlaGlyAlaSerAlaGlu...
$rT_r\alpha$

Figure 14. Diagram of pML 17. The bovine rod transducin alpha subunit cDNA ($rT_{r}\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{trc}, trp/lac hybrid promoter; p_{L}, the leftward bacteriophage λ promoter; TET^{s}, truncated tetracycline resistance gene.

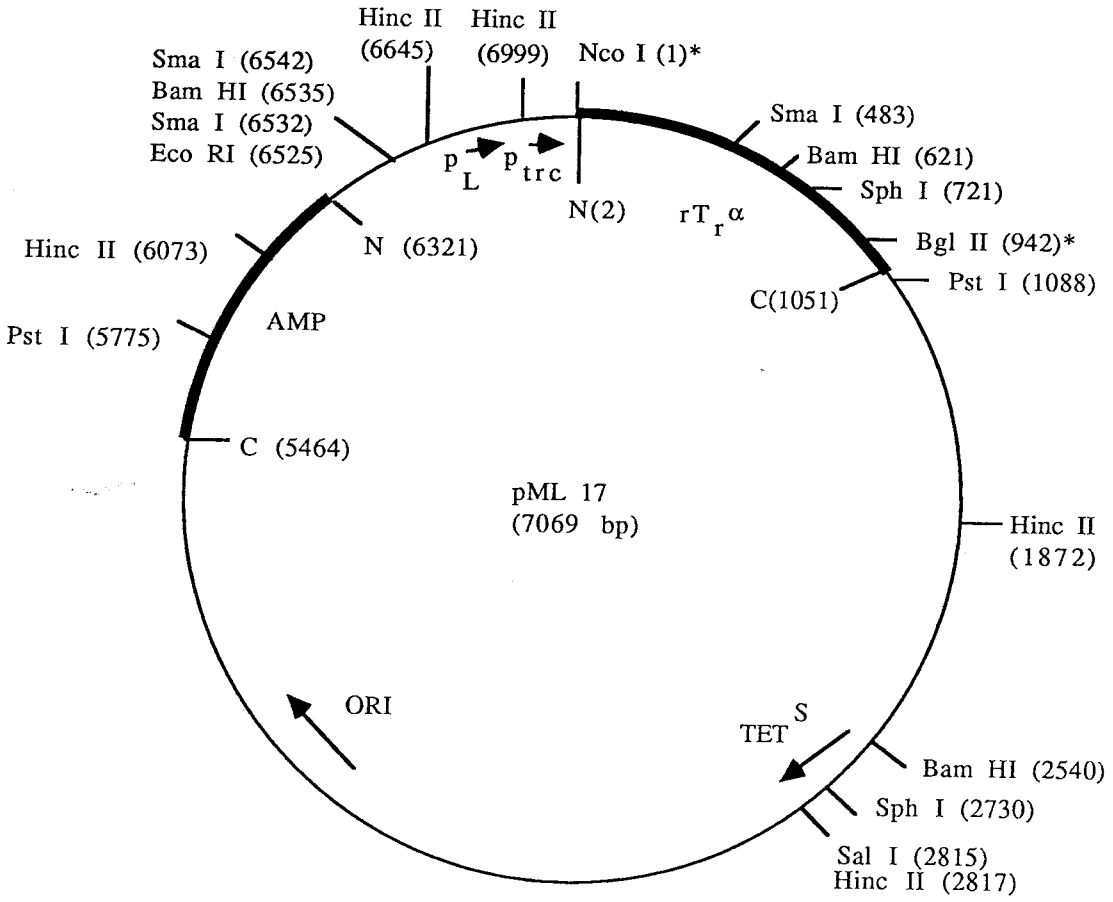

Hinc II (6645)
Hinc II (6999)
Nco I (1)*
Sma I (6542)
Bam HI (6535)
Sma I (6532)
Eco RI (6525)
Sma I (483)
Bam HI (621)
Sph I (721)
Bgl II (942)*
Pst I (1088)
C(1051)
N(2)
rT r α
N (6321)
Hinc II (6073)
Pst I (5775)
AMP
C (5464)
pML 17
(7069 bp)
Hinc II (1872)
ORI
TET S
Bam HI (2540)
Sph I (2730)
Sal I (2815)
Hinc II (2817)

Figure 15. Sequence of the promoter region of pML 17. The sequence was derived from pPL (Pharmacia) and Tanabe *et al.* (1985). The leftward bacteriophage λ operators (OL1, OL2, and OL3), which are binding sites for the λ repressor, are underlined. The -35 region, -10 region, transcription initiation site, and the ribosome binding site (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

Eco RI Sma I Bam HI Sma I
GAATTCCCGGGGATCCCCGGGGATCTCTCACCTACCAAACAATGCCCCCC (6573)

TGCAAAAAATAAATTCATATAAAAAACATACAGATAACCATCTGCGGTGA (6623)
OL3

TAAATTATCTCTGGCGGTGTTGACATAAATACCACTGGCGGTGATACTGA (6673)
OL2 **OL1**

GCACATCAGCAGGACGCACTGACCACCATGAAGGTGACGCTCTTAAAAAT (6723)
Transcription initiation site (p_L)

TAAGCCCTGAAGAAGGGCAGCATTCAAAGCAGAAGGCTTTGGGGTGTGTG (6773)

ATACGAAACGAAGCATTGGCCGTAAGTGCGATTCCGGATTAGCTGCCAAT (6823)
ME
λN protein

GTGCCAATCGCGGGGGGTTTTCGTTCAGGACTACAACTGCCACACACCAC (6873)
TCysGlnSerArgGlyValPheValGlnAspTyrAsnCysHisThrProP

CAAAGCTAACTGACAGGAGAATCCAGATGGATGCACAAACACGCCGCCGC (6923)
roLysLeuThrAspArgArgIleGlnMetAspAlaGlnThrArgArgArg

GAACGTCGCGCAGAGAAACAGGCTCAATGGAAAGCAGCAAATCCCCTGTT (6973)
GluArgArgAlaGluLysGlnAlaGlnTrpLysAlaAlaAsnProLeuLe

Hinc II
GGTTGGGGTAAGCGCAAAACCAGTTGACAATTAATCATCCGGCTCGTATA (7023)
uValGlyValSerAlaLysProValAspIle---
-35 Region **-10 Region**

Nco I
ATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGACCATG (4)
Transcription initiation site (p_{trc}) **RBS** MET **$rT_r\alpha$**

GGGGCTGGGGCCAGCGCTGAGGAGAAGCACTCAAGGGAGCTGGAAAAGAAG... (55)
GlyAlaGlyAlaSerAlaGluGluLysHisSerArgGluLeuGluLysLys...

Figure 16. Diagram of pML 18. The bovine cone transducin alpha subunit cDNA ($rT_C\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{trc}, trp/lac hybrid promoter; p_L, the leftward bacteriophage λ promoter; TET^S, truncated tetracycline resistance gene.

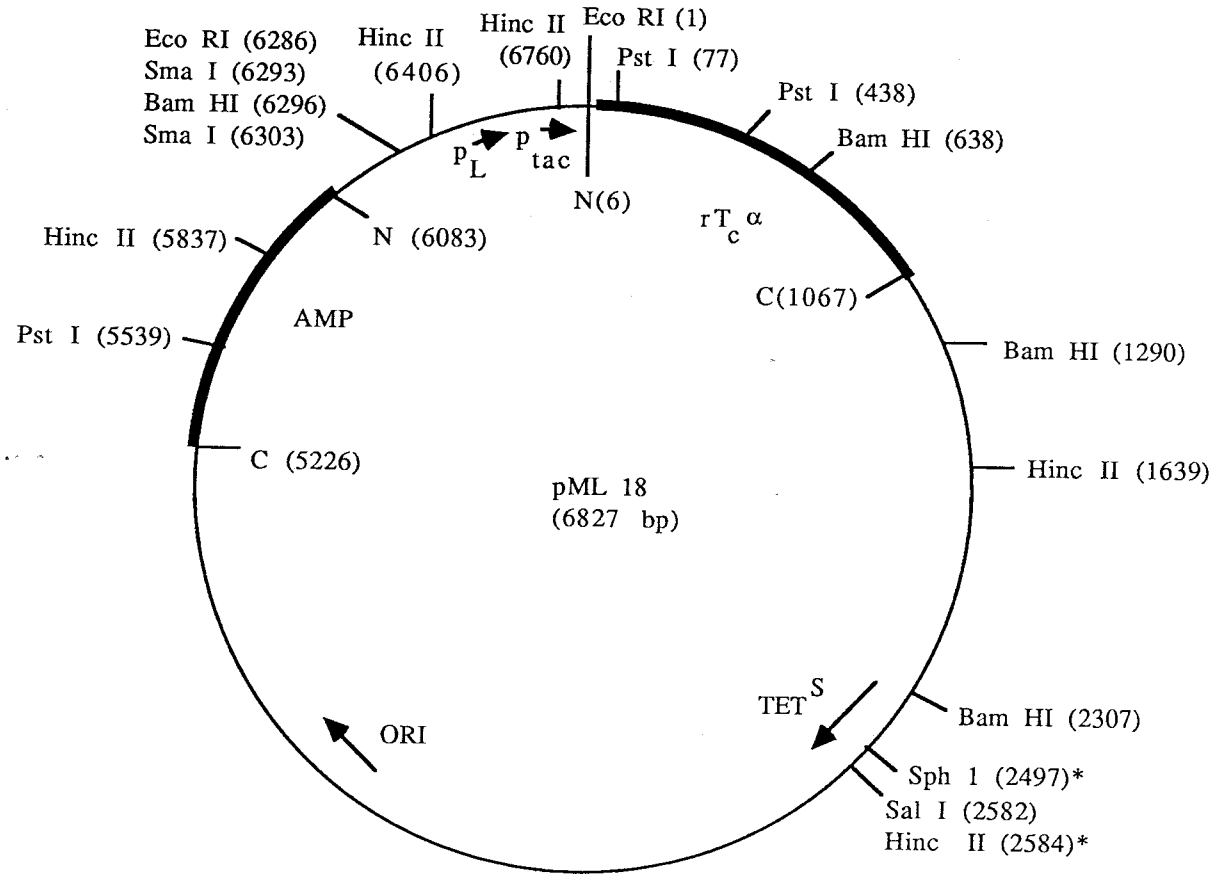

Eco RI (6286)
Sma I (6293)
Bam HI (6296)
Sma I (6303)
Hinc II (6406)
Hinc II (6760)
Eco RI (1)
Pst I (77)
Pst I (438)
Bam HI (638)
p_L
p tac
N(6)
$rT_c \alpha$
C(1067)
Bam HI (1290)
Hinc II (1639)
Bam HI (2307)
Sph 1 (2497)*
Sal I (2582)
Hinc II (2584)*
TET S
ORI
C (5226)
Pst I (5539)
AMP
Hinc II (5837)
N (6083)
pML 18
(6827 bp)

Figure 17. Sequence of the promoter region of pML 18. The sequence was derived from pPL (Pharmacia) and Lochrie *et al.* (1985). The leftward bacteriophage λ operators (OL1, OL2, and OL3), which are binding sites for the λ repressor, are underlined. The -35 region, -10 region, transcription initiation site, and the ribosome binding site (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

Eco RI Sma I Bam HI Sma I
GAATTCCCGGGGATCCCCGGGGATCTCTCACCTACCAAACAATGCCCCCC (6334)

TGCAAAAAATAAATTCATATAAAAAACATACAGATAACCATCTGCGGTGA (6384)
OL3

TAAATTATCTCTGGCGGTGTTGACATAAATACCACTGGCGGTGATACTGA (6434)
OL2 **OL1**

GCACATCAGCAGGACGCACTGACCACCATGAAGGTGACGCTCTTAAAAAT (6484)
Transcription initiation site (p_L)

TAAGCCCTGAAGAAGGGCAGCATTCAAAGCAGAAGGCTTTGGGGTGTGTG (6534)

ATACGAAACGAAGCATTGGCCGTAAGTGCGATTCCGGATTAGCTGCCAAT (6584)
ME
λN protein

GTGCCAATCGCGGGGGGTTTTCGTTCAGGACTACAACTGCCACACACCAC (6634)
TCysGlnSerArgGlyValPheValGlnAspTyrAsnCysHisThrProP

CAAAGCTAACTGACAGGAGAATCCAGATGGATGCACAAACACGCCGCCGC (6684)
roLysLeuThrAspArgArgIleGlnMetAspAlaGlnThrArgArgArg

GAACGTCGCGCAGAGAAACAGGCTCAATGGAAAGCAGCAAATCCCCTGTT (6734)
GluArgArgAlaGluLysGlnAlaGlnTrpLysAlaAlaAsnProLeuLe

Hinc II
GGTTGGGGTAAGCGCAAAACCAGTTGACAATTAATCATCGGCTCGTATAA (6784)
uValGlyValSerAlaLysProValAspIle---
-35 region **-10 region**

Eco RI
TGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGAATTC (5)
Transcription initiation site (p_{tac}) **RBS**

ATGGGGAGTGGAGCCAGTGCCGAGGACAAAGAACTGGCCAAGAGGTCC... (53)
METGlySerGlyAlaSerAlaGluAspLysGluLeuAlaLysArgSer...
$rT_C\alpha$

Figure 18. Diagram of pML 25. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Bacteriophage λ genes (cI_{857}, cro, cII, and λ O) are indicated by the stippled lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{trc}, trp/lac hybrid promoter; p_R, the rightward bacteriophage λ promoter; TET^S, truncated tetracycline resistance gene; ΔO, truncated bacteriophage λ O gene. The ΔO protein is 163 amino acids long. Of these 161 are derived from bacteriophage λ and 2 are derived from the parent vector pKK 233-2. The sequence of the region 5' to the initiation codon is the same as that of pML 15.

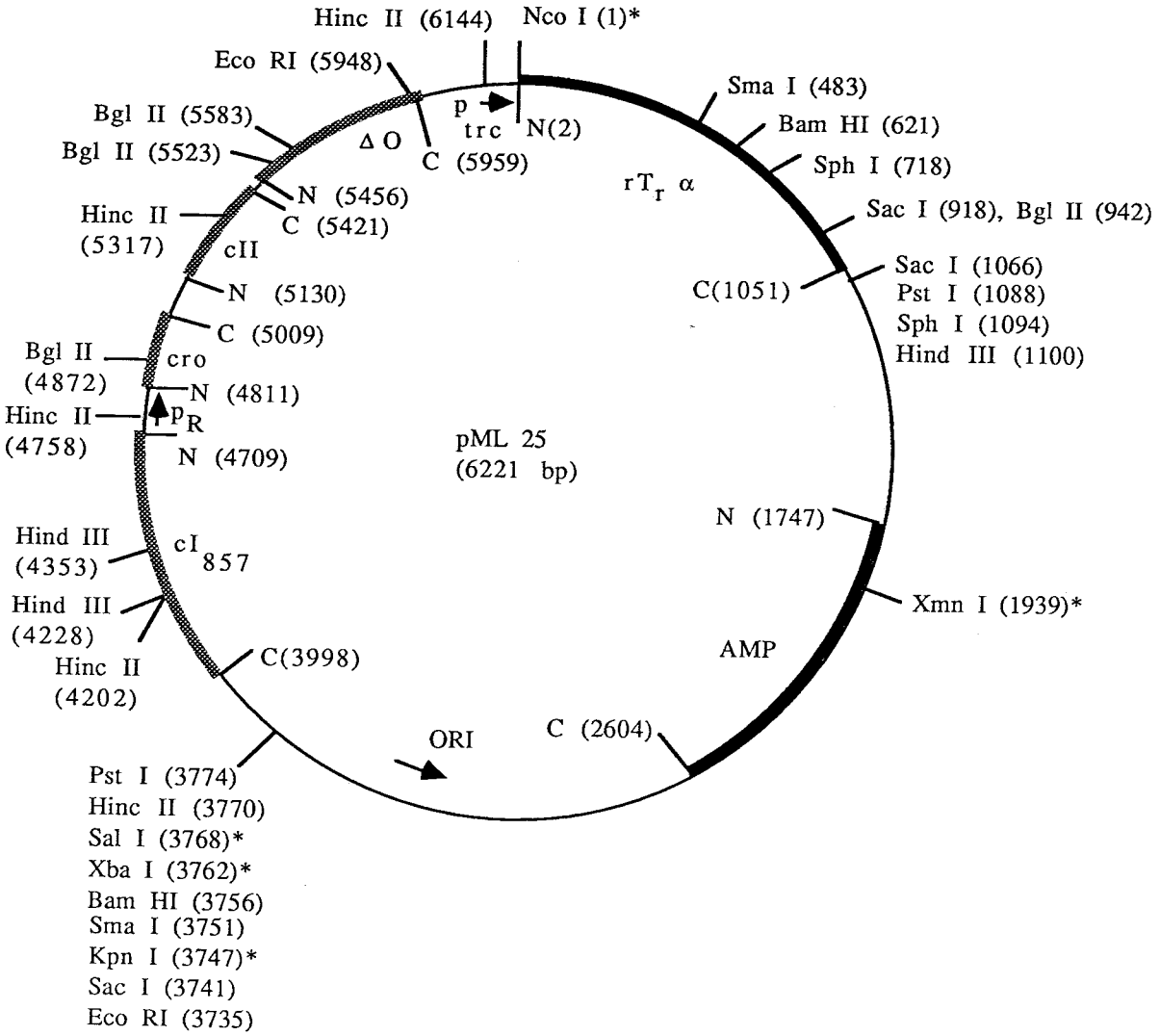
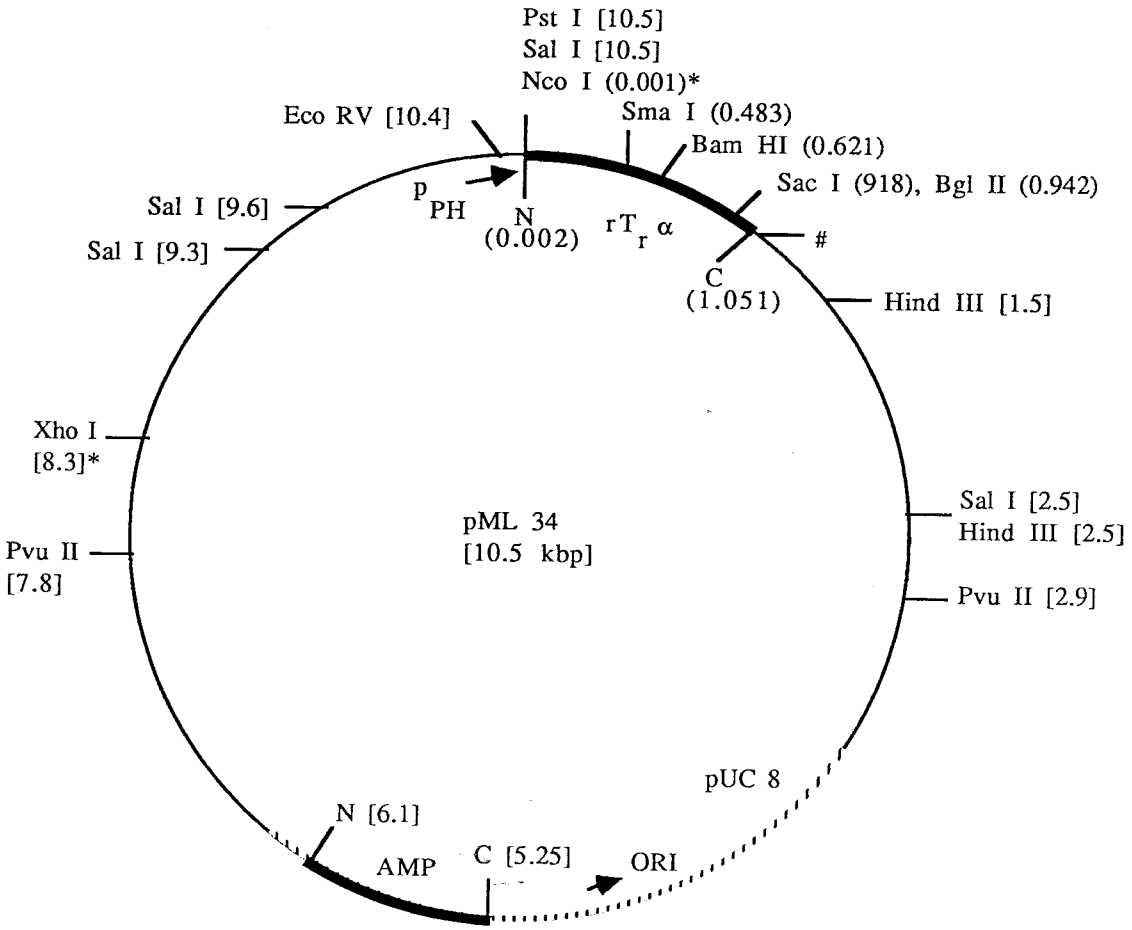

Hinc II (6144)
Nco I (1)*
Eco RI (5948)
Bgl II (5583)
Bgl II (5523)
Hinc II (5317)
p trc
N(2)
C (5959)
Δ O
N (5456)
C (5421)
cII
N (5130)
C (5009)
Bgl II (4872)
cro
N (4811)
Hinc II (4758)
p_R
N (4709)
Hind III (4353)
cI 857
Hind III (4228)
Hinc II (4202)
C(3998)
Pst I (3774)
Hinc II (3770)
Sal I (3768)*
Xba I (3762)*
Bam HI (3756)
Sma I (3751)
Kpn I (3747)*
Sac I (3741)
Eco RI (3735)
ORI
C (2604)
AMP
N (1747)
Xmn I (1939)*
C(1051)
Sac I (1066)
Pst I (1088)
Sph I (1094)
Hind III (1100)
Sac I (918), Bgl II (942)
Sph I (718)
Bam HI (621)
Sma I (483)
rT_r α
pML 25
(6221 bp)

Figure 19. Diagram of pML 34. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. DNA derived from the *E. coli* plasmid pUC 8 is shown as a striped line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985). Exact positions of the restriction sites are indicated in parentheses and approximate positions are indicated in brackets. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{PH}, baculovirus polyhedrin promoter.



#= Sac I (1.066), Pst I (1.088), Sal I (1.090), Hinc II (1.092), Xba I (1.096)*, Bam HI (1.102), Sma I (1.109), Sac I (1.114), Eco RI (1.120)

Figure 20. Sequence of the promoter region of pML 34. The sequence was derived from Possee and Howard (1987), Luckow and Summers (1988), and Tanabe *et al.* (1985). The CAAT-box, TATA-box, and the transcription initiation site are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

Eco RV

CTATCAATATATAGTTGCTGATATCATGGAGATAATTAAAATGATAACCA [10400]

CAAT-box **TATA-box**

TCTCGCAAATAAATAAGTATTTTACTGTTTTCGTAACAGTTTTGTAATAA [10450]

Transcription initiation site

Nco I

AAAAACCTGCCCCCCTGCAGGTCGACTTCTTGAAAACCATGGGGGCTGGG... (13)

METGlyAlaGly...

$rT_r\alpha$

Figure 21. Diagram of pML 35. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) is represented as a solid thick line. DNA sequences derived from yeast are shown as striped lines. DNA sequences derived from the *E. coli* plasmid pBR 322 are indicated as stippled lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985). Exact positions of the restriction sites are indicated in parentheses and approximate positions are indicated in brackets. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; 2μm ORI, yeast 2μm plasmid origin of DNA replication; bp, base pairs; p_{AF}, yeast alpha factor promoter; PPAF, prepro α-factor leader peptide.

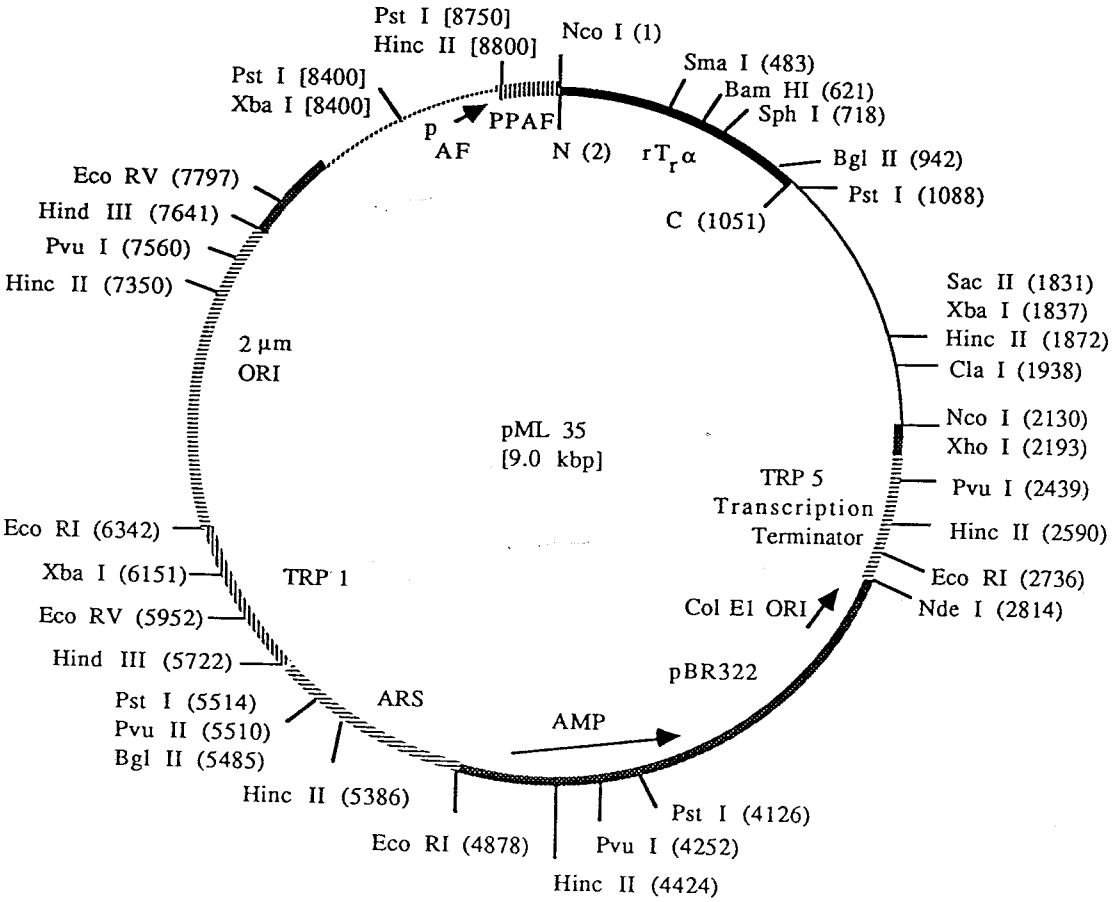

Pst I [8750]
Hinc II [8800]
Nco I (1)
Sma I (483)
Bam HI (621)
Sph I (718)
Pst I [8400]
Xba I [8400]
PPAF
p AF
N (2)
rT$_r$α
Bgl II (942)
Pst I (1088)
C (1051)
Eco RV (7797)
Hind III (7641)
Pvu I (7560)
Hinc II (7350)
Sac II (1831)
Xba I (1837)
Hinc II (1872)
Cla I (1938)
2 µm
ORI
pML 35
[9.0 kbp]
Nco I (2130)
Xho I (2193)
Pvu I (2439)
TRP 5
Transcription
Terminator
Hinc II (2590)
Eco RI (6342)
Xba I (6151)
TRP 1
Eco RI (2736)
Nde I (2814)
Col E1 ORI
Eco RV (5952)
Hind III (5722)
pBR322
Pst I (5514)
Pvu II (5510)
Bgl II (5485)
ARS
AMP
Hinc II (5386)
Pst I (4126)
Eco RI (4878)
Pvu I (4252)
Hinc II (4424)

Figure 22. Sequence of the promoter region of pML 35. The sequence was derived from Kurjan and Herskowitz (1982), Miyajima *et al.* (1985), and Tanabe *et al.* (1985). The sequence from [8921] to (25) was confirmed by DNA sequence analysis. The TATA-box and the glycosylation sites are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

CAAGAAAACCAAAAAGCAACAACAGGTTTTGGATAAGTACATATATAAGA [8600]
TATA box

GGGCCTTTTGTTCCCATCAAAAATGTTACTGTTCTTACGATTCATTTACG [8650]

ATTCAAGAATAGTTCAAACAAGAAGATTACAAACTATCAATTTCATACAC [8700]

Pst I
AATATAAACGACCAAAAGAATGAGATTTCCTTCAATTTTTACTGCAGTTT [8750]
METArgPheProSerIlePheThrAlaValL
prepro α-factor

Hinc II
TATTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAA [8800]
euPheAlaAlaSerSerAlaLeuAlaAlaProValAsnThrThrThrGlu
Glycosylation site

GATGAAACGGCACAAATTCCGGCTGAAGCTGTCATCGGTTACTCAGATTT [8850]
AspGluThrAlaGlnIleProAlaGluAlaValIleGlyTyrSerAspLe

AGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAATTA [8900]
uGluGlyAspPheAspValAlaValLeuProPheSerAsnSerThrAsnA
Glycosylation site

ACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAA [8950]
snGlyLeuLeuPheIleAsnThrThrIleAlaSerIleAlaAlaLysGlu
Glycosylation site

Nco I
GAAGGGGTATCTTTGGATAAAAGGCCCATGGGGGCTGGGGCCAGCGCTGAG... (25)
GluGlyValSerLeuAspLysArgProMETGlyAlaGlyAlaSerAlaGlu...
$rT_r\alpha$
Kex 2 cleavage site

Figure 23. Diagram of pML 44. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. DNA derived from the *E. coli* plasmid pUC 8 is shown as a striped line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985). Exact positions of the restriction sites are indicated in parentheses and approximate positions are indicated in brackets. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{PH}, baculovirus polyhedrin promoter. The sequence of the promoter region is the same as that of pML 34.

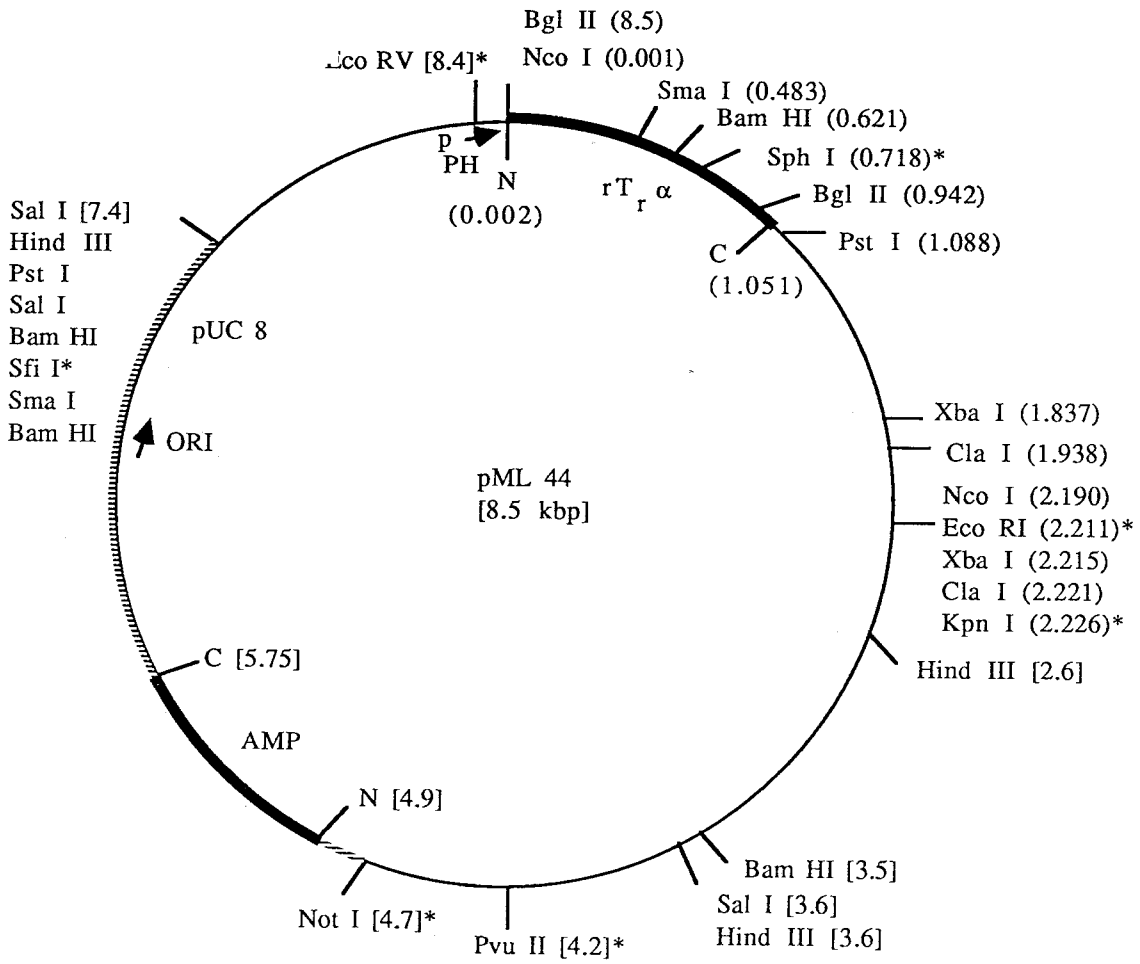

Bgl II (8.5)
Nco I (0.001)
Eco RV [8.4]*
Sma I (0.483)
Bam HI (0.621)
Sph I (0.718)*
Bgl II (0.942)
Pst I (1.088)
C
(1.051)
rT r α
p
PH
N
(0.002)
Sal I [7.4]
Hind III
Pst I
Sal I
Bam HI
Sfi I*
Sma I
Bam HI
pUC 8
ORI
pML 44
[8.5 kbp]
Xba I (1.837)
Cla I (1.938)
Nco I (2.190)
Eco RI (2.211)*
Xba I (2.215)
Cla I (2.221)
Kpn I (2.226)*
Hind III [2.6]
C [5.75]
AMP
N [4.9]
Bam HI [3.5]
Sal I [3.6]
Hind III [3.6]
Not I [4.7]*
Pvu II [4.2]*

Figure 24. Diagram of pML 45. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. DNA derived from the *E. coli* plasmid pUC 8 is shown as a striped line. The sequence derived from the baculovirus polyhedrin protein is stippled. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985). Exact positions of the restriction sites are indicated in parentheses and approximate positions are indicated in brackets. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{PH}, baculovirus polyhedrin promoter.

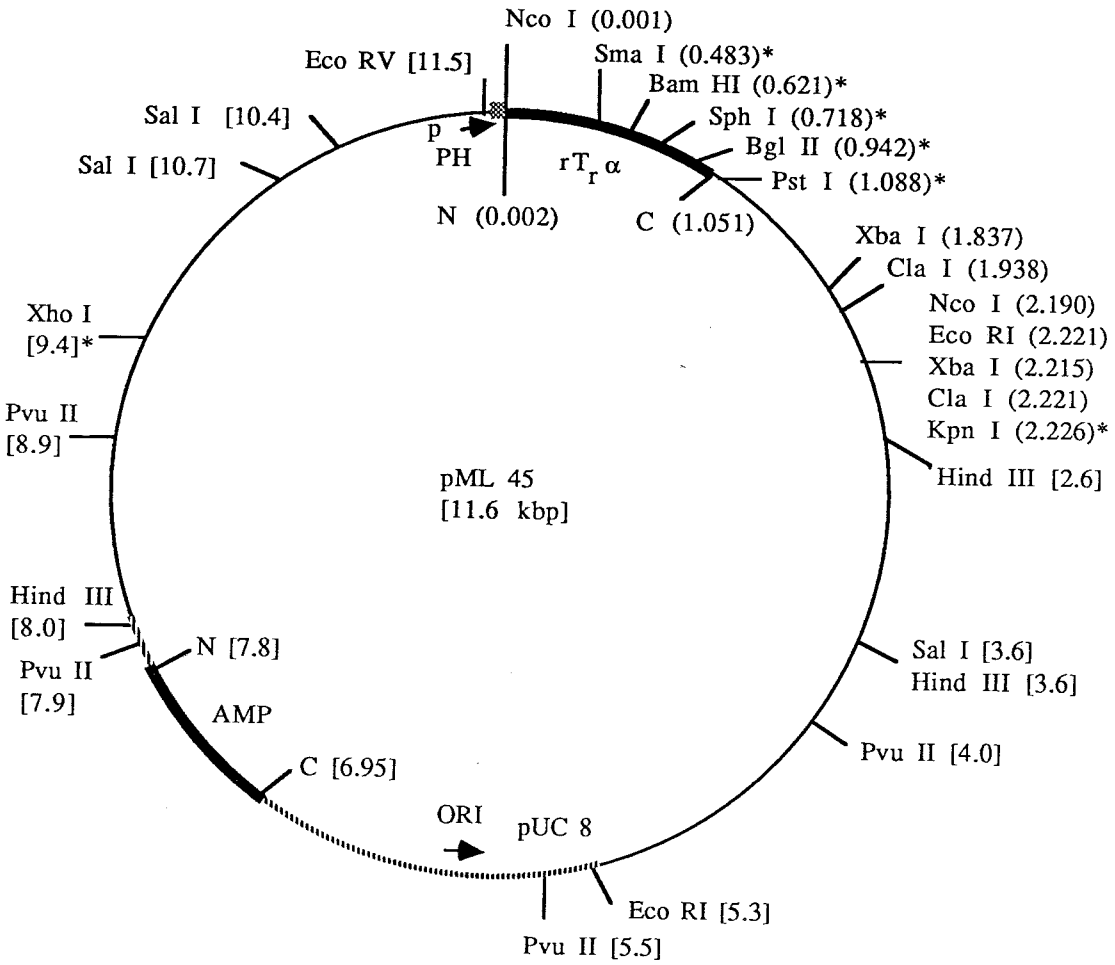

Nco I (0.001)
Sma I (0.483)*
Bam HI (0.621)*
Sph I (0.718)*
Bgl II (0.942)*
Pst I (1.088)*
C (1.051)
Xba I (1.837)
Cla I (1.938)
Nco I (2.190)
Eco RI (2.221)
Xba I (2.215)
Cla I (2.221)
Kpn I (2.226)*
Hind III [2.6]
Sal I [3.6]
Hind III [3.6]
Pvu II [4.0]
Eco RI [5.3]
Pvu II [5.5]
ORI
pUC 8
C [6.95]
AMP
N [7.8]
Hind III
[8.0]
Pvu II
[7.9]
Pvu II
[8.9]
Xho I
[9.4]*
Sal I [10.7]
Sal I [10.4]
Eco RV [11.5]
p
PH
rT$_r$ α
N (0.002)
pML 45
[11.6 kbp]

Figure 25. Sequence of the promoter region of pML 45. The sequence was derived from Possee and Howard (1987), Summers and Smith (1987), and Tanabe *et al.* (1985). The sequence is numbered according to the corresponding plasmid diagram.

Eco RV

CTATCAATATATAGTTGCTGATATCATGGAGATAATTAAAATGATAACCA [8400]

CAAT-box **TATA-box**

TCTCGCAAATAAATAAGTATTTTACTGTTTTCGTAACAGTTTTGTAATAA [8450]

Transcription initiation site

AAAAACCTATAAATATGCCGGATTATTCATACCGTCCCACCATCGGGCCG [8500]

METProAspTyrSerTyrArgProThrIleGlyPro

Polyhedrin

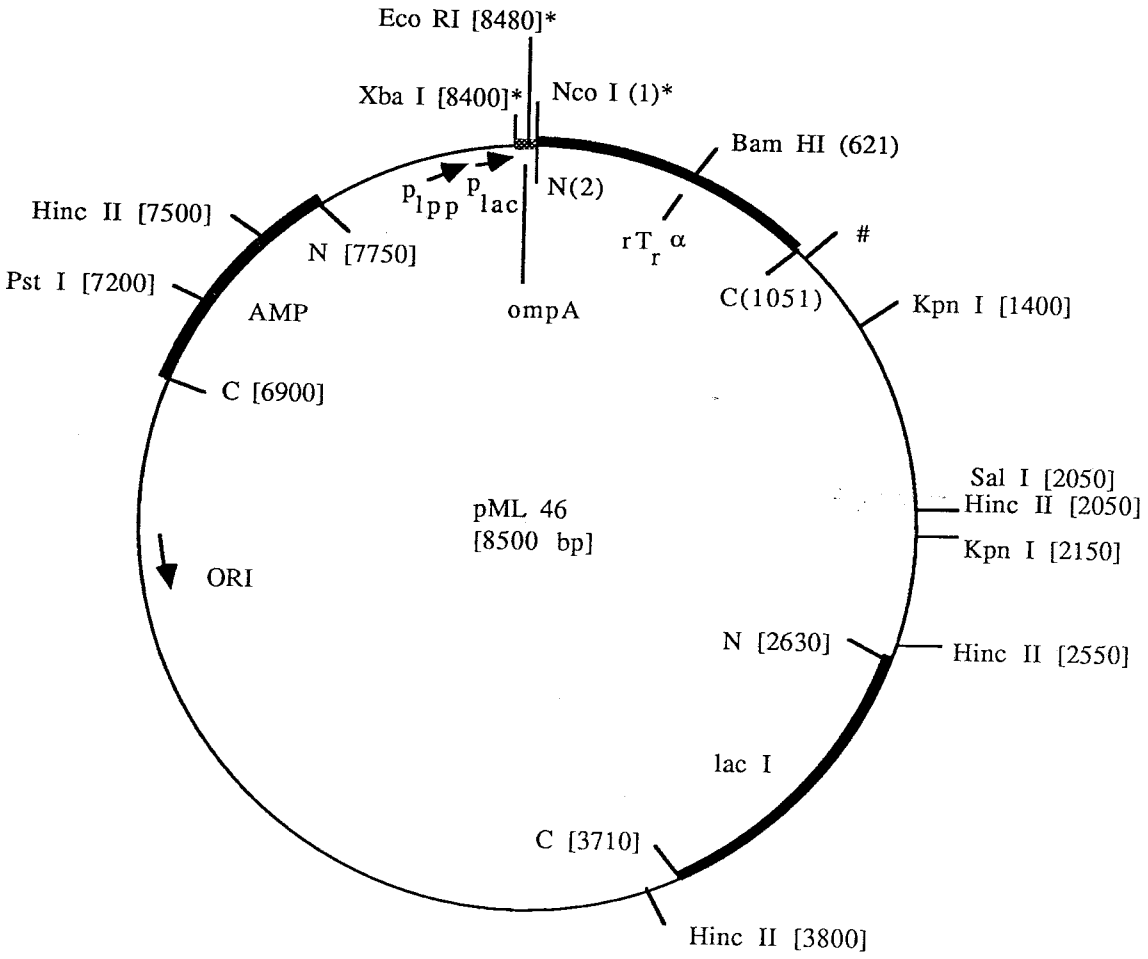
Nco I

GATCTCGACTTCTTGAAAACCATGGGGGCTGGGGCCAGCGCTGAGGAGAAG... (31)

AspLeuAspPheLeuLysThrMETGlyAlaGlyAlaSerAlaGluGluLys...

Polylinker **$rT_r\alpha$**

Figure 26. Diagram of pML 46. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$), the lac I repressor gene, and the ampicillin gene (AMP) are represented as solid thick lines. The ompA peptide is represented as a stippled line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985). Exact positions of restriction sites are indicated in parentheses and approximate positions are indicated in brackets. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{lac}, *E. coli* lactose operon promoter; p_{lpp}, *E. coli* lipoprotein promoter; ompA, *E. coli* outer membrane protein leader peptide.



#= Pst I (1088), Sal I (1090),
Hinc II (1092), Sna BI* (1103),
Hind III* (1106), Bam HI (1112)

Figure 27. Sequence of the promoter region of pML 46 and 50. The sequence was derived from Nakamura and Inouye (1979), Nakamura and Inouye (1982), Masui *et al.* (1983), Ghrayheb *et al.* (1984), Tanabe *et al.* (1985), and Gloria Dalbadie-McFarland, (personal communication). The -35 regions, -10 regions, transcription initiation sites, and the ribosome binding sites (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

CCATCAAAAAAATATTCTCAACATAAAAAACTTTGTGTAATACTTGTAAC [8250]
-35 Region (P_{lpp}) -10 Region (P_{lpp})

GCTACATGGAGATTAACTCAATCTAGCACTCATTAGGCACCCCAGGCTTA [8300]
Transcription initiation site (P_{lpp})

CACTTTATGCTTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAAC [8350]
-35 Region (P_{lac}) -10 Region (P_{lac}) Transcription initiation site (P_{lac})

Xba I
AATTTCACACAGGAAACAGCTCTAGATAACGAGGGCAAAAAATGAAAAAG [8400]
RBS METLysLys
ompA

ACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTAGCGCA [8450]
ThrAlaIleAlaIleAlaValAlaLeuAlaGlyPheAlaThrValAlaGl

Eco RI Nco I
GGCCGCGAATTCAGATTTAAATAAGGAGGAATAATCCATGGGGGCTGGGG (14)
nAlaAlaAsnSerAspLeuAsnLysGluGlu--- MetGlyAlaGlyA
RBS $rT_r\alpha$

CCAGCGCTGAGGAGAAGCACTCAAGGGAGCTGGAAAAGAAGCTGAAAGAA... (64)
laSerAlaGluGluLysHisSerArgGluLeuGluLysLysLeuLysGlu...

Figure 28. Diagram of pML 49. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; f1(+) ORI, bacteriophage f1 origin of single stranded DNA replication; bp, base pairs; p_{T3}, bacteriophage T3 RNA polymerase promoter; p_{T7}, bacteriophage T7 RNA polymerase promoter. The antisense strand of the $rT_r\alpha$ cDNA is produced when pML 49 is rescued with a M13 helper phage.

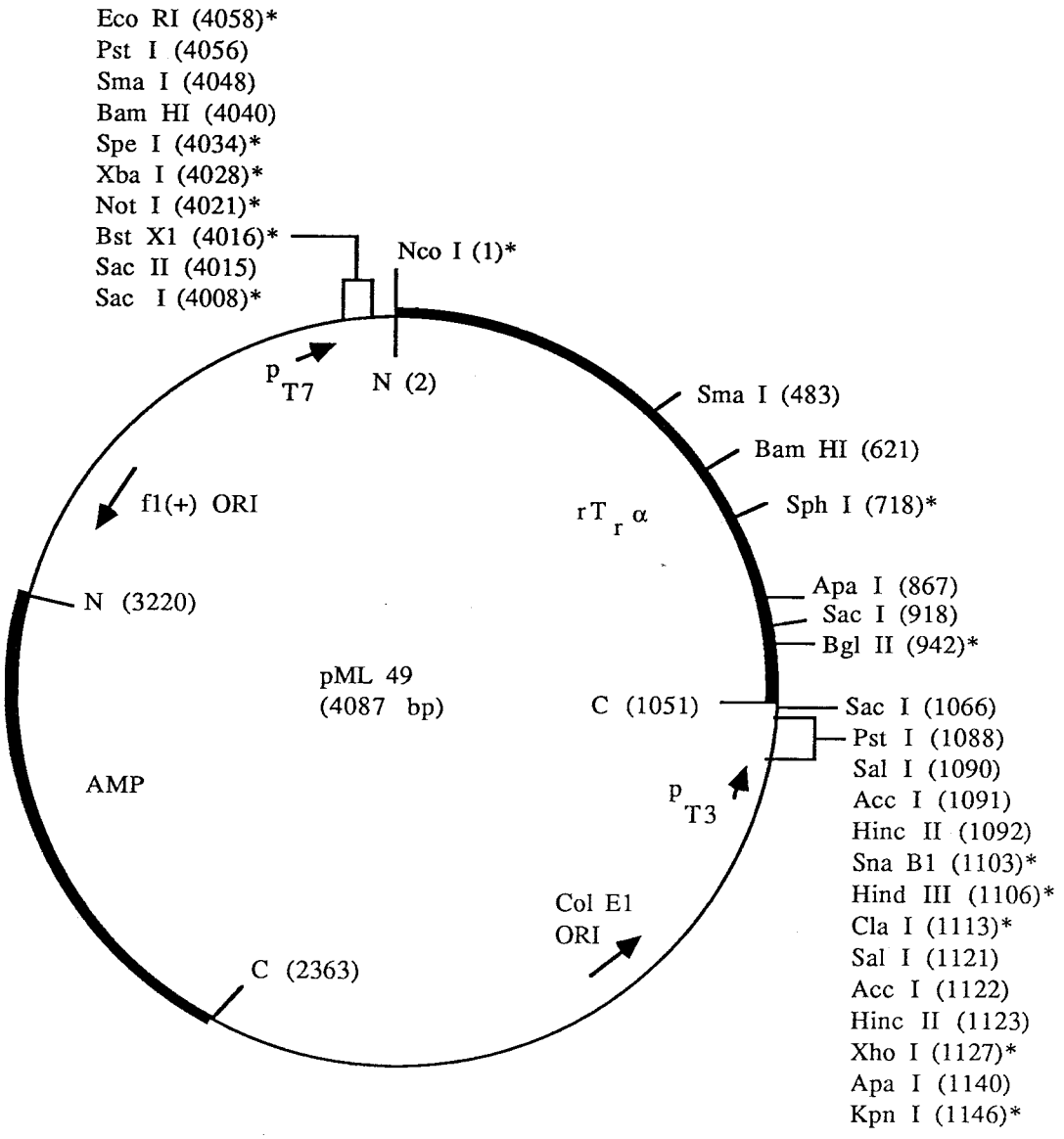

Eco RI (4058)*
Pst I (4056)
Sma I (4048)
Bam HI (4040)
Spe I (4034)*
Xba I (4028)*
Not I (4021)*
Bst X1 (4016)*
Sac II (4015)
Sac I (4008)*
Nco I (1)*
p T7
N (2)
Sma I (483)
Bam HI (621)
Sph I (718)*
rT r α
Apa I (867)
Sac I (918)
Bgl II (942)*
f1(+) ORI
N (3220)
pML 49
(4087 bp)
C (1051)
Sac I (1066)
Pst I (1088)
Sal I (1090)
Acc I (1091)
Hinc II (1092)
Sna B1 (1103)*
Hind III (1106)*
Cla I (1113)*
Sal I (1121)
Acc I (1122)
Hinc II (1123)
Xho I (1127)*
Apa I (1140)
Kpn I (1146)*
AMP
p T3
Col E1
ORI
C (2363)

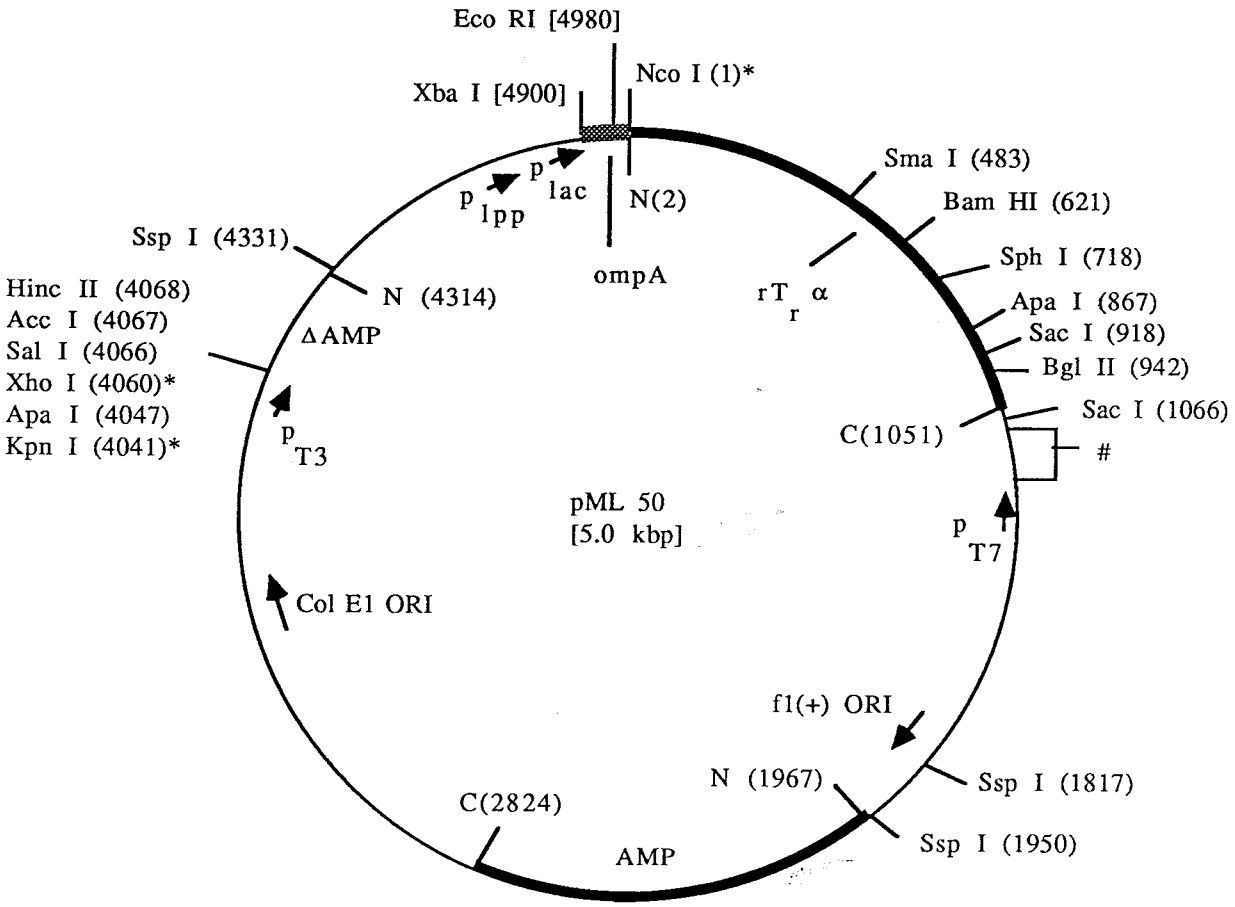
Figure 29. Sequence of the promoter region of pML 49. The sequence was derived from pBluescript/KS+ (Stratagene) and pML 46. The T7 promoter is in bold-faced lettering. The transcription initiation site and the ribosome binding site (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

GTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCC (4024)
T7 promoter Transcription initiation site
(P_{T7})

Xba I Eco RI
GCTCTAGAACTAGTGGATCCCCCGGGCTGCAGAATTCAGATTTAAATAAG (4074)

Nco I
GAGGAATAATCCATGGGGGCTGGGGCCAGCGCTGAGGAGAAGCACTCAAGG... (40)
MetGlyAlaGlyAlaSerAlaGluGluLysHisSerArg...
RBS $rT_r\alpha$

Figure 30. Diagram of pML 50. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. The ompA peptide is represented as a stippled line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; Col E1 ORI, Col E1 origin of DNA replication; f1(+) ORI, bacteriophage f1 origin of single stranded DNA replication; bp, base pairs; p_{lac}, *E. coli* lactose operon promoter; p_{lpp}, *E. coli* lipoprotein promoter; p_{T3}, bacteriophage T3 RNA polymerase promoter; p_{T7}, bacteriophage T7 RNA polymerase promoter; ompA, *E. coli* outer membrane protein leader peptide; ΔAMP, truncated ampicillin resistance gene. The sense strand of the $rT_r\alpha$ cDNA is produced when pML 50 is rescued with a M13 helper phage. The sequence of the promoter region of pML 50 is the same as that of pML 46.



#=Pst I (1088), Sal I (1090), Acc I (1091), Hinc II (1092), Sna BI (1103)*, Hind III (1106)*, Eco RV (1110)*, Eco RI (1118), Pst I (1128), Sma I (1132), Bam HI (1140), Spe I (1146)*, Xba I (1152)*, Not I (1159)*, Bst XI (1164)*, Sac II (1172)*, Sac I (1179)

Figure 31. Diagram of pML 52. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$) and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; f1(+) ORI, bacteriophage f1 origin of single-stranded DNA replication; bp, base pairs; p_{T3}, bacteriophage T3 RNA polymerase promoter; p_{T7}, bacteriophage T7 RNA polymerase promoter; ompA, *E. coli* outer membrane proteinleader peptide. The antisense strand of the $rT_r\alpha$ cDNA is produced when pML 52 is rescued with a M13 helper phage.

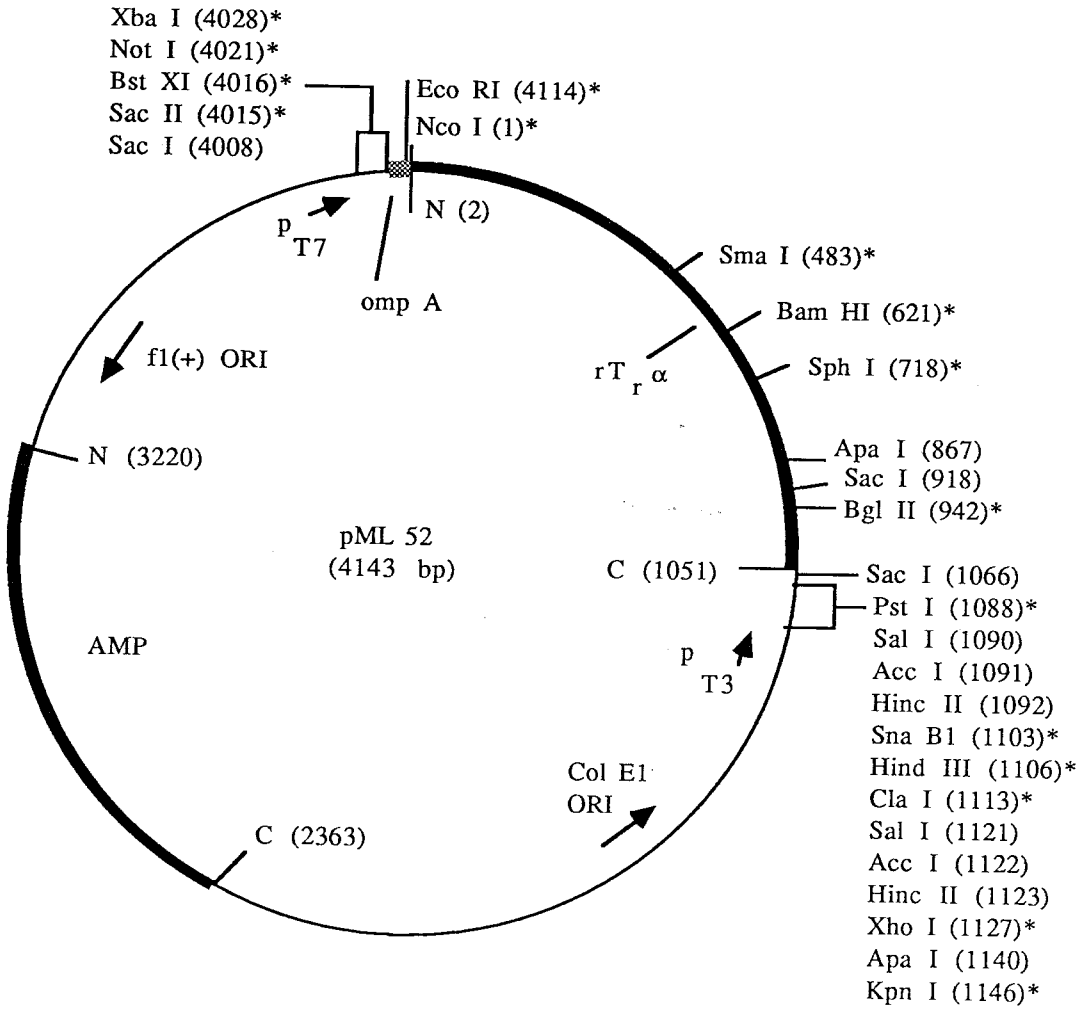

Xba I (4028)*
Not I (4021)*
Bst XI (4016)*
Sac II (4015)*
Sac I (4008)
Eco RI (4114)*
Nco I (1)*
N (2)
p T7
omp A
Sma I (483)*
Bam HI (621)*
Sph I (718)*
rT r α
f1(+) ORI
N (3220)
Apa I (867)
Sac I (918)
Bgl II (942)*
pML 52
(4143 bp)
C (1051)
Sac I (1066)
Pst I (1088)*
Sal I (1090)
Acc I (1091)
Hinc II (1092)
Sna B1 (1103)*
Hind III (1106)*
Cla I (1113)*
Sal I (1121)
Acc I (1122)
Hinc II (1123)
Xho I (1127)*
Apa I (1140)
Kpn I (1146)*
AMP
p T3
Col E1
ORI
C (2363)

Figure 32. Sequence of the promoter region of pML 52. The sequence was derived from pBluescript/KS+ (Stratagene) and pML 46. The region from positions (4036) to (31) was confirmed by DNA sequence analysis. The T7 promoter is in bold-faced lettering. The transcription initiation site and the ribosome binding sites (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

TAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCG (4025)
T7 promoter **Transcription initiation site**
(p_{T7})

Xba I
CTCTAGATAACGAGGGCAAAAAATGAAAAAGACAGCTATCGCGATTGCAG (4075)
RBS METLysLysThrAlaIleAlaIleAlaV
ompA

Eco RI
TGGCACTGGCTGGTTTCGCTACCGTAGCGCAGGCCGCGAATTCAGATTTA (4125)
alAlaLeuAlaGlyPheAlaThrValAlaGlnAlaAlaAsnSerAspLeu

Nco I
AATAAGGAGGAATAATCCATGGGGGCTGGGGCCAGCGCTGAGGAGAAG... (31)
AsnLysGluGlu--- MetGlyAlaGlyAlaSerAlaGluGluLys...
RBS **$rT_r\alpha$**

Figure 33. Diagram of pML 53. The bovine cone transducin alpha subunit cDNA ($rT_C\alpha$), the lac I repressor gene, and the ampicillin gene (AMP) are represented as solid thick lines. The ompA peptide is represented as a stippled line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985). Exact positions of restriction sites are indicated in parentheses and approximate positions are indicated in brackets. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{lac}, *E. coli* lactose operon promoter; p_{lpp}, *E. coli* lipoprotein promoter; ompA, *E. coli* outer membrane protein leader peptide.

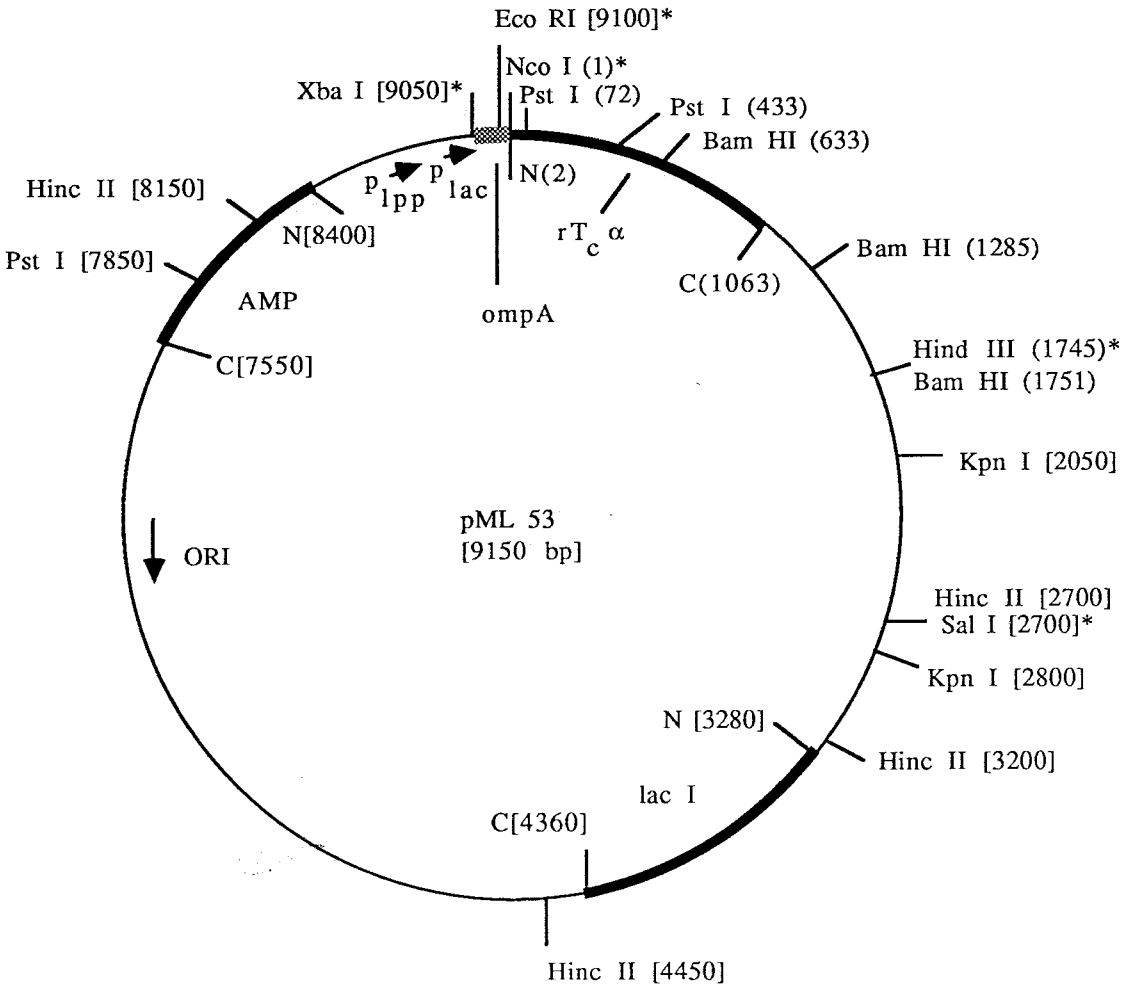

Eco RI [9100]*
Nco I (1)*
Xba I [9050]*
Pst I (72)
Pst I (433)
Bam HI (633)
Hinc II [8150]
p_{lpp} p_{lac}
N(2)
N[8400]
rT_c α
Pst I [7850]
Bam HI (1285)
C(1063)
AMP
ompA
C[7550]
Hind III (1745)*
Bam HI (1751)
Kpn I [2050]
pML 53
[9150 bp]
ORI
Hinc II [2700]
Sal I [2700]*
Kpn I [2800]
N [3280]
Hinc II [3200]
lac I
C[4360]
Hinc II [4450]

Figure 34. Sequence of the promoter region of pML 53. The sequence was derived from Nakamura and Inouye (1979), Nakamura and Inouye (1982), Masui *et al.* (1983), Ghrayheb *et al.* (1984), Lochrie *et al.* (1985), and Gloria Dalbadie-McFarland, (personal communication). The -35 regions, -10 regions, transcription initiation sites, and the ribosome binding sites (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

CCATCAAAAAAATATTCTCAACATAAAAAACTTTGTGTAATACTTGTAAC [8925]
-35 Region (P_{lpp}) -10 Region (P_{lpp})

GCTACATGGAGATTAACTCAATCTAGCACTCATTAGGCACCCCAGGCTTA [8975]
Transcription initiation site (P_{lpp})

CACTTTATGCTTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAAC [9025]
-35 Region (P_{lac}) -10 Region (P_{lac}) Transcription initiation site (P_{lac})

Xba I
AATTTCACACAGGAAACAGCTCTAGATAACGAGGGCAAAAAAATGAAAAAG [9075]
RBS METLysLys
ompA

ACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTAGCGCA [9125]
ThrAlaIleAlaIleAlaValAlaLeuAlaGlyPheAlaThrValAlaGl

Eco RI Nco I
GGCCGCGAATTCAGATTTAAATAAGGAGGAATAATCCATGGGGAGTGGAG (14)
nAlaAlaAsnSerAspLeuAsnLysGluGlu--- MetGlySerGlyA
RBS $rT_C\alpha$

Figure 35. Diagram of pML 55. The *S. japonicum* glutathione S-transferase (GST) gene, the lac I^Q repressor gene, and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{tac}, *E. coli* trp/lac hybrid promoter; ΔlacZ, truncated *E. coli* lacZ gene. The ΔlacZ protein consists of 174 amino acids. Of these 148 are derived from lacZ and 26 are derived from the p_{tac11} promoter region. The length of GST is 237 amino acids.

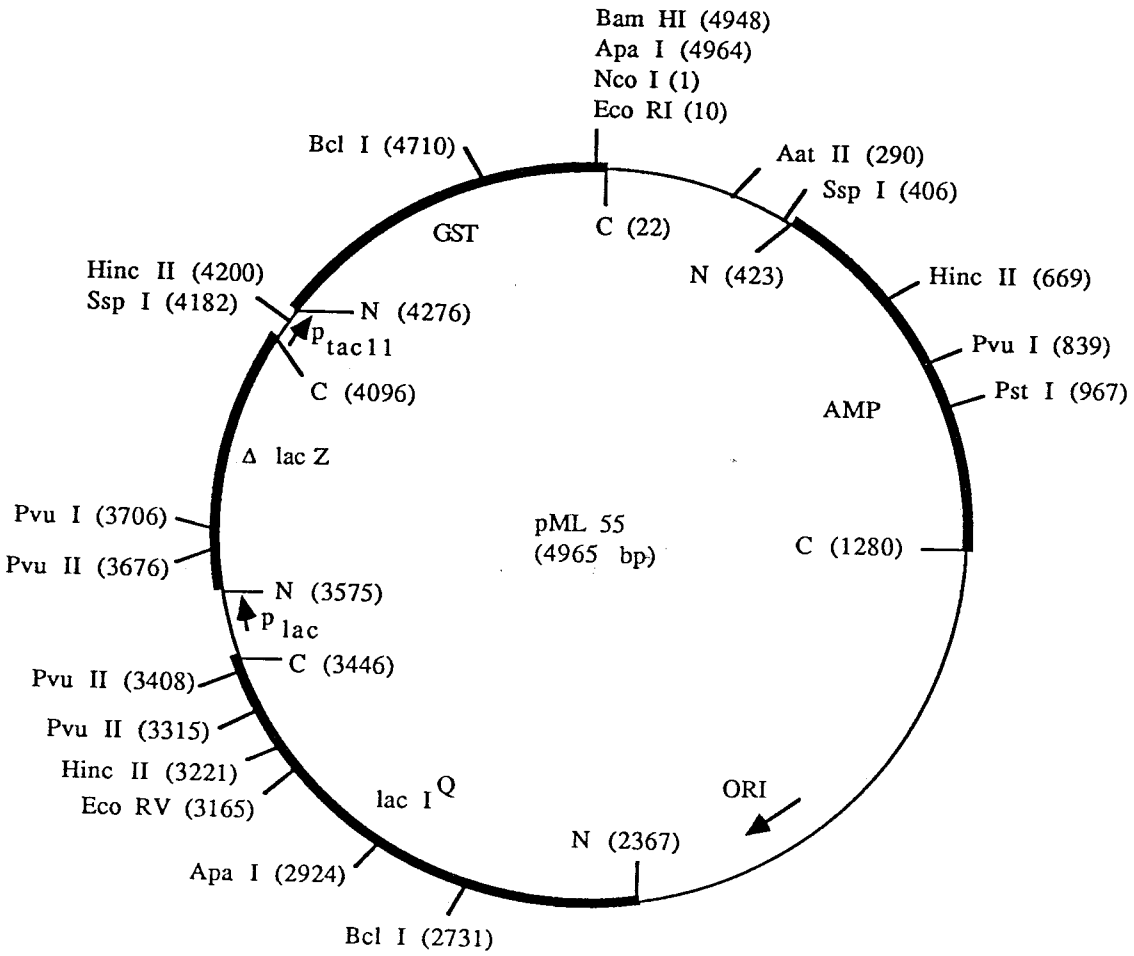

Bam HI (4948)
Apa I (4964)
Nco I (1)
Eco RI (10)
Bcl I (4710)
Aat II (290)
Ssp I (406)
GST
C (22)
N (423)
Hinc II (4200)
Ssp I (4182)
Hinc II (669)
N (4276)
Pvu I (839)
p_{tac11}
C (4096)
Pst I (967)
AMP
Δ lac Z
Pvu I (3706)
pML 55
(4965 bp)
C (1280)
Pvu II (3676)
N (3575)
p_{lac}
C (3446)
Pvu II (3408)
Pvu II (3315)
Hinc II (3221)
Eco RV (3165)
lac I^{Q}
ORI
N (2367)
Apa I (2924)
Bcl I (2731)

Figure 36. Sequence of the polylinker region of pML 55. The sequence was determined by DNA sequence analysis. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

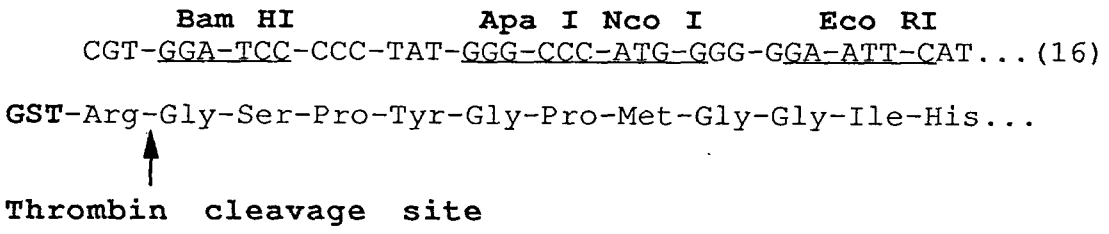
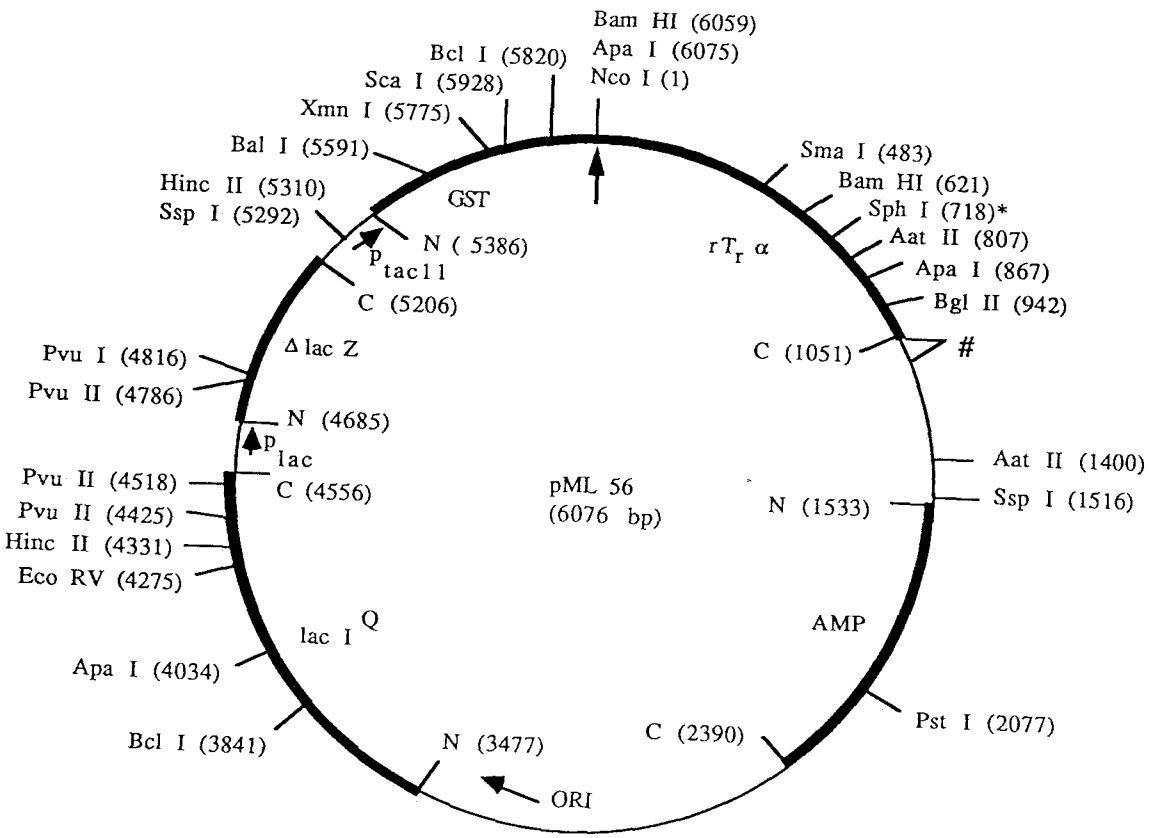

Bam HI
Apa I Nco I
Eco RI
CGT-GGA-TCC-CCC-TAT-GGG-CCC-ATG-GGG-GGA-ATT-CAT...(16)
GST-Arg-Gly-Ser-Pro-Tyr-Gly-Pro-Met-Gly-Gly-Ile-His...
Thrombin cleavage site

Figure 37. Diagram of pML 56. The bovine rod transducin alpha subunit cDNA ($rT_r\alpha$), the *S. japonicum* glutathione S-transferase (GST) gene, the lac I^Q repressor gene, and the ampicillin gene (AMP) are represented as solid thick lines. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. The fusion point between GST and $rT_r\alpha$ is indicated by an arrow. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{tac}, *E. coli* trp/lac hybrid promoter; p_{lac}, *E. coli* lactose operon promoter; ΔlacZ, truncated *E. coli* lacZ gene. The ΔlacZ protein consists of 174 amino acids. Of these 148 are derived from lacZ and 26 are derived from the p_{tac11} promoter region. The length of GST is 230 amino acids and the length of $rT_r\alpha$ is 350 amino acids.



= Sac I (1066), Pst I (1088), Sal I (1192), Xba I (1096), Bam HI (1102), Sma I (1109), Sac I (1114), Eco RI (1120).

Figure 38. Sequence of the protein fusion junction region of pML 56. The sequence was determined by DNA sequence analysis. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

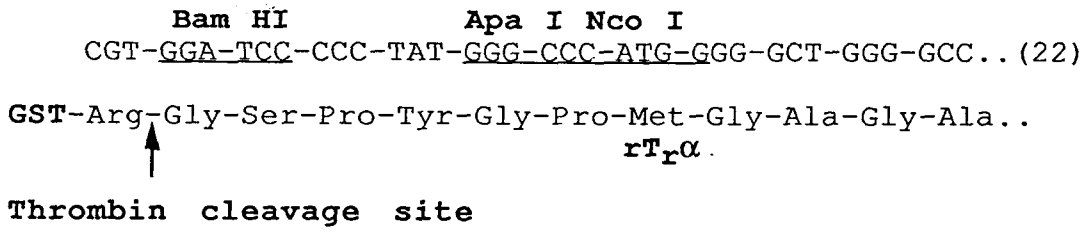

Bam HI
Apa I Nco I
CGT-GGA-TCC-CCC-TAT-GGG-CCC-ATG-GGG-GCT-GGG-GCC..(22)
GST-Arg-Gly-Ser-Pro-Tyr-Gly-Pro-Met-Gly-Ala-Gly-Ala..
rTrα.
Thrombin cleavage site

Figure 39. Diagram of pGD 108. The lac I repressor gene and the ampicillin gene (AMP) are represented as solid thick lines. The ompA peptide is represented as a stippled line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985). Exact positions of restriction sites are indicated in parentheses and approximate positions are indicated in brackets. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; ORI, Col E1 origin of DNA replication; bp, base pairs; p_{lac}, *E. coli* lactose operon promoter; p_{lpp}, *E. coli* lipoprotein promoter; ompA, *E. coli* outer membrane protein leader peptide.

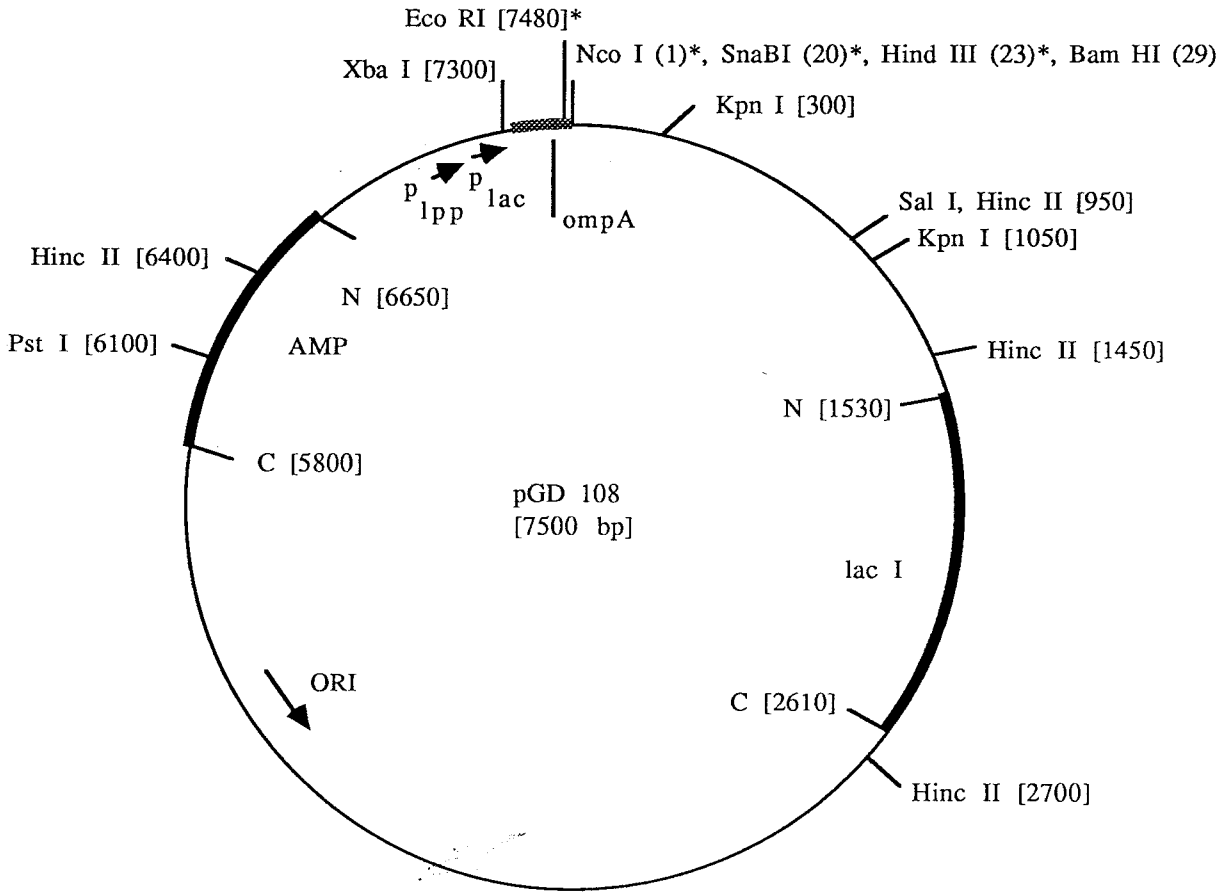

Eco RI [7480]*
Nco I (1)*, SnaBI (20)*, Hind III (23)*, Bam HI (29)
Xba I [7300]
Kpn I [300]
p lpp
p lac
ompA
Sal I, Hinc II [950]
Kpn I [1050]
Hinc II [6400]
N [6650]
Pst I [6100]
AMP
Hinc II [1450]
N [1530]
C [5800]
pGD 108
[7500 bp]
lac I
ORI
C [2610]
Hinc II [2700]

Figure 40. Sequence of the promoter region of pGD 108. The sequence was derived from Nakamura and Inouye (1979), Nakamura and Inouye (1982), Masui *et al.* (1983), Ghrayheb *et al.* (1984), and Gloria Dalbadie-McFarland, (personal communication). The -35 region, -10 region, transcription initiation site, and the ribosome binding sites (RBS) are underlined and indicated below the sequence. Relevant restriction sites are indicated above the sequence. The sequence is numbered according to the corresponding plasmid diagram.

CCATCAAAAAAATATTCTCAACATAAAAAACTTTGTGTAATACTTGTAAC [7250]
-35 Region (Plpp) -10 Region (Plpp)

GCTACATGGAGATTAACTCAATCTAG/CACTCATTAGGCACCCCAGGCTTA [7300]
Transcription initiation site (Plpp)

CACTTTATGCTTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAAC [7350]
-35 Region (Plac) -10 Region (Plac) Transcription initiation site (Plac)

Xba I
AATTTCACACAGGAAACAGCTCTAGATAACGAGGGCAAAAAATGAAAAAG [7400]
RBS METLysLys
ompA

ACAGCTATCGCGATTGCAGTGGCACTGGCTGGTTTCGCTACCGTAGCGCA [7450]
ThrAlaIleAlaIleAlaValAlaLeuAlaGlyPheAlaThrValAlaGl

Eco RI Nco I Spe I
GGCCGCGAATTCAGATTTAAATAAGGAGGAATAATCCATGGTTCATCACT (14)
nAlaAlaAsnSerAspLeuAsnLysGluGlu---
RBS

Sna BI Hind III Bam HI
AGTACGTAAGCTTGGATCC... (33)

Figure 41. Diagram of M13 ml 1. The bovine cone transducin alpha subunit cDNA ($rT_C\alpha$) is represented as a solid thick line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; bp, base pairs.

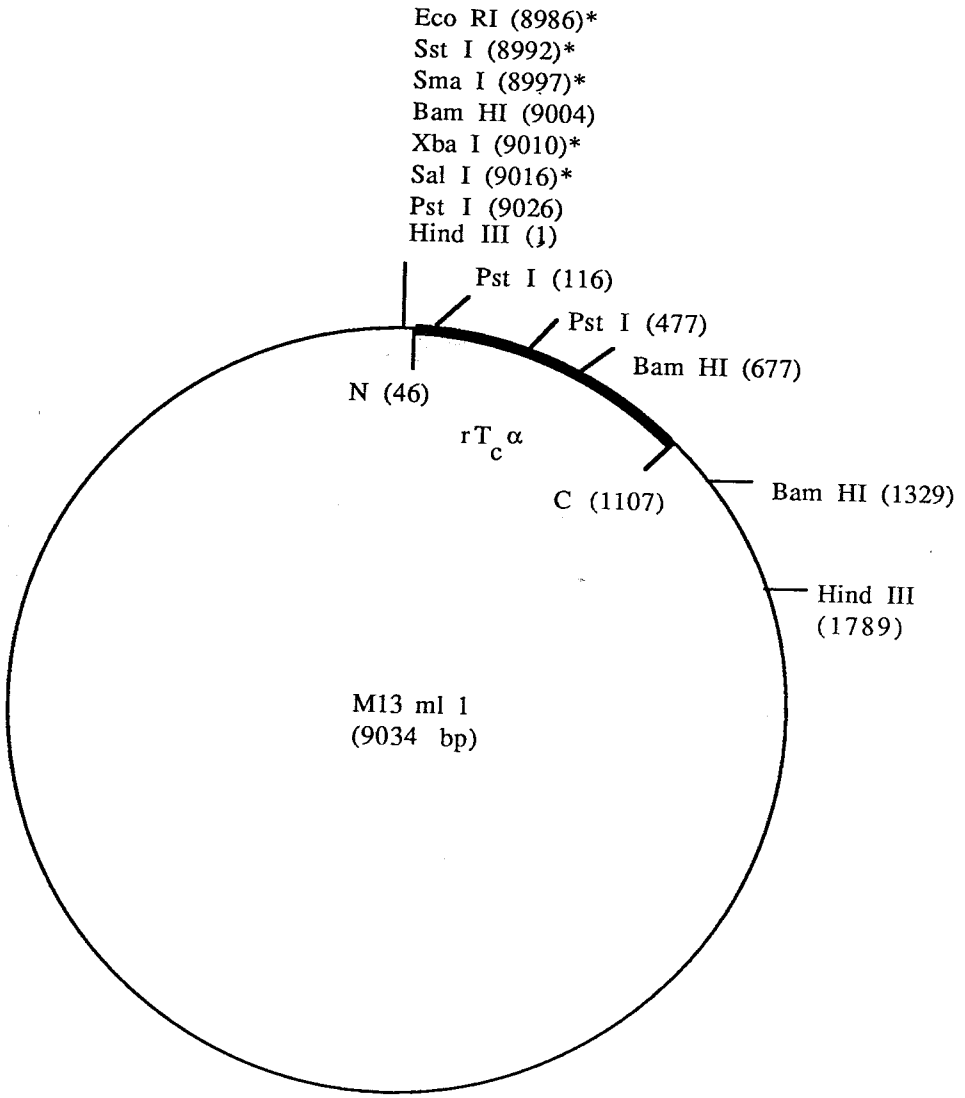

Eco RI (8986)*
Sst I (8992)*
Sma I (8997)*
Bam HI (9004)
Xba I (9010)*
Sal I (9016)*
Pst I (9026)
Hind III (1)
Pst I (116)
Pst I (477)
Bam HI (677)
N (46)
rTcα
C (1107)
Bam HI (1329)
Hind III
(1789)
M13 ml 1
(9034 bp)

Figure 42. Diagram of M13 ml 4. The bovine cone transducin alpha subunit cDNA ($rT_C\alpha$) is represented as a solid thick line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; bp, base pairs.

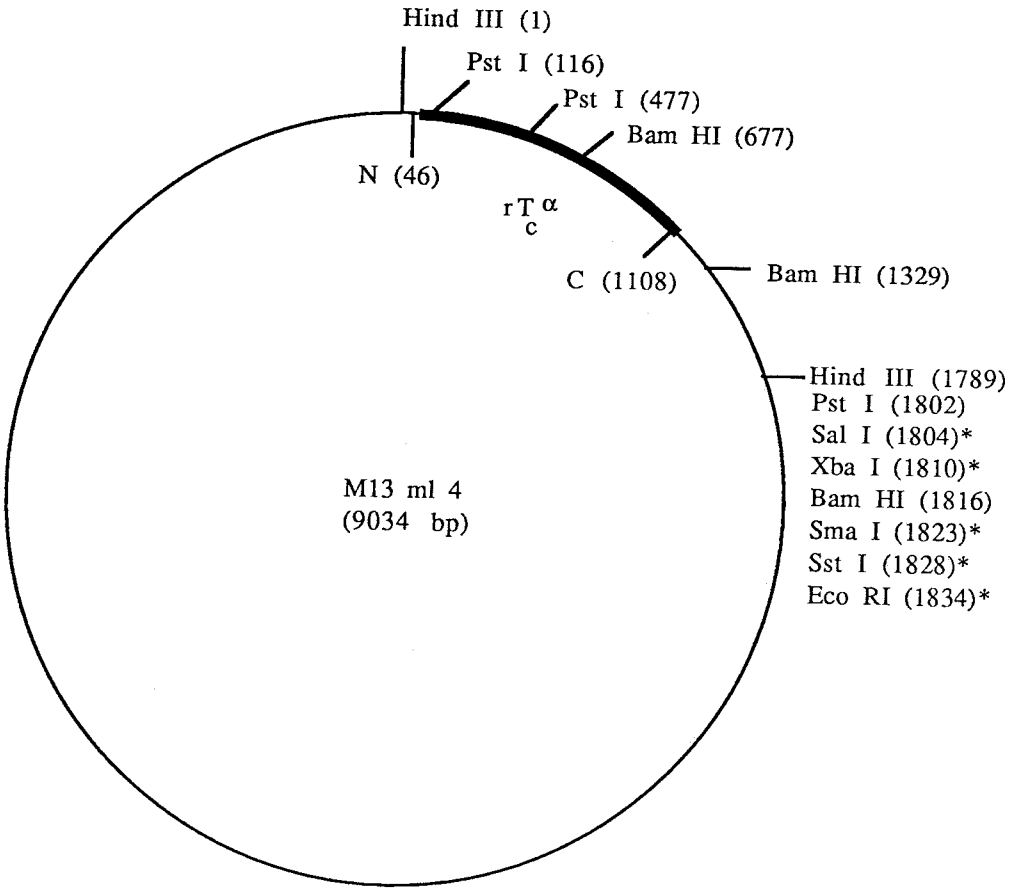

Hind III (1)
Pst I (116)
Pst I (477)
Bam HI (677)
N (46)
rTcα
C (1108)
Bam HI (1329)
Hind III (1789)
Pst I (1802)
Sal I (1804)*
Xba I (1810)*
Bam HI (1816)
Sma I (1823)*
Sst I (1828)*
Eco RI (1834)*
M13 ml 4
(9034 bp)

Figure 43. Diagram of M13 ml 1-Nco I. The bovine cone transducin alpha subunit cDNA ($rT_C\alpha$) is represented as a solid thick line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; bp, base pairs.

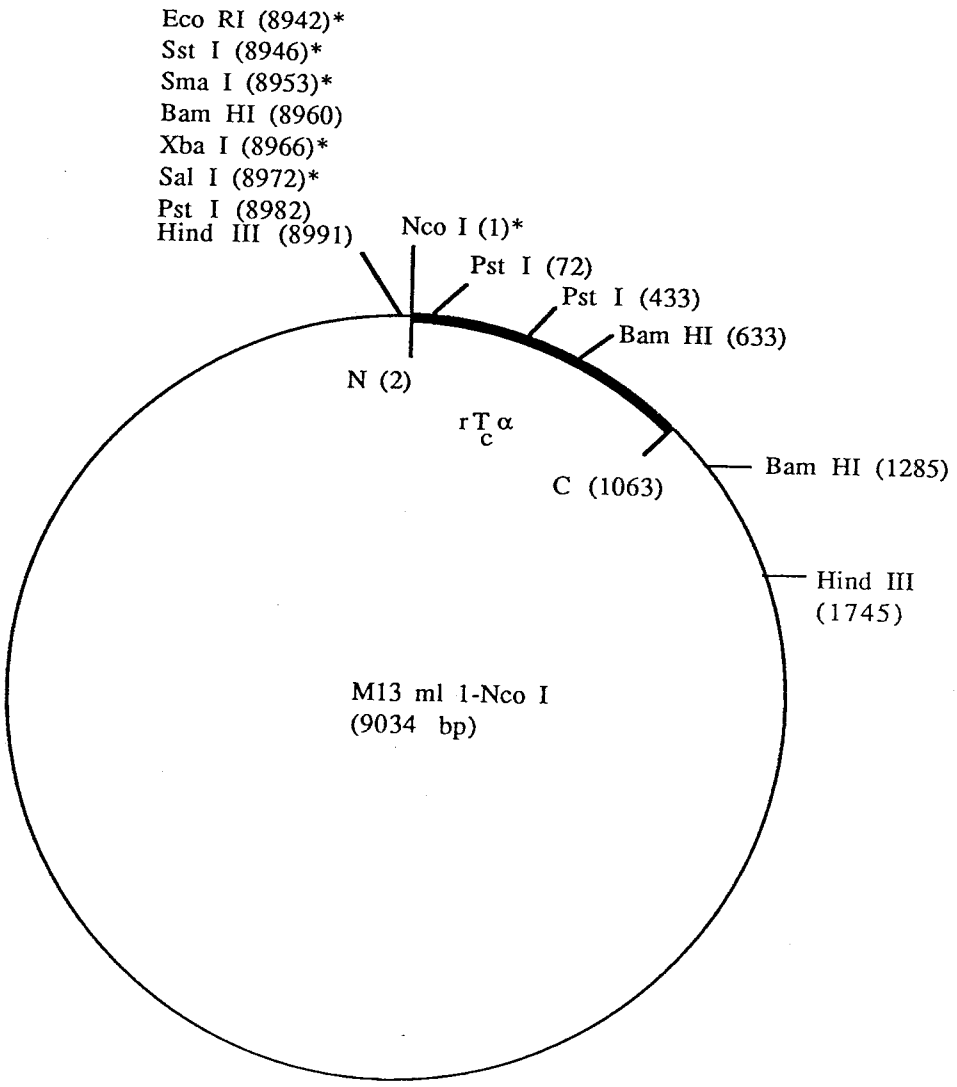

Eco RI (8942)*
Sst I (8946)*
Sma I (8953)*
Bam HI (8960)
Xba I (8966)*
Sal I (8972)*
Pst I (8982)
Hind III (8991)
Nco I (1)*
Pst I (72)
Pst I (433)
Bam HI (633)
N (2)
rTcα
C (1063)
Bam HI (1285)
Hind III
(1745)
M13 ml 1-Nco I
(9034 bp)

Figure 44. Diagram of M13 ml 91. The bovine cone transducin alpha subunit cDNA ($rT_C\alpha$) is represented as a solid thick line. Restriction enzyme cleavage sites are abbreviated according to standard nomenclature (Roberts, 1985) and their positions are indicated in parentheses. Asterisks denote unique restriction sites. Abbreviations used are: N, amino terminus; C, carboxyl terminus; bp, base pairs.

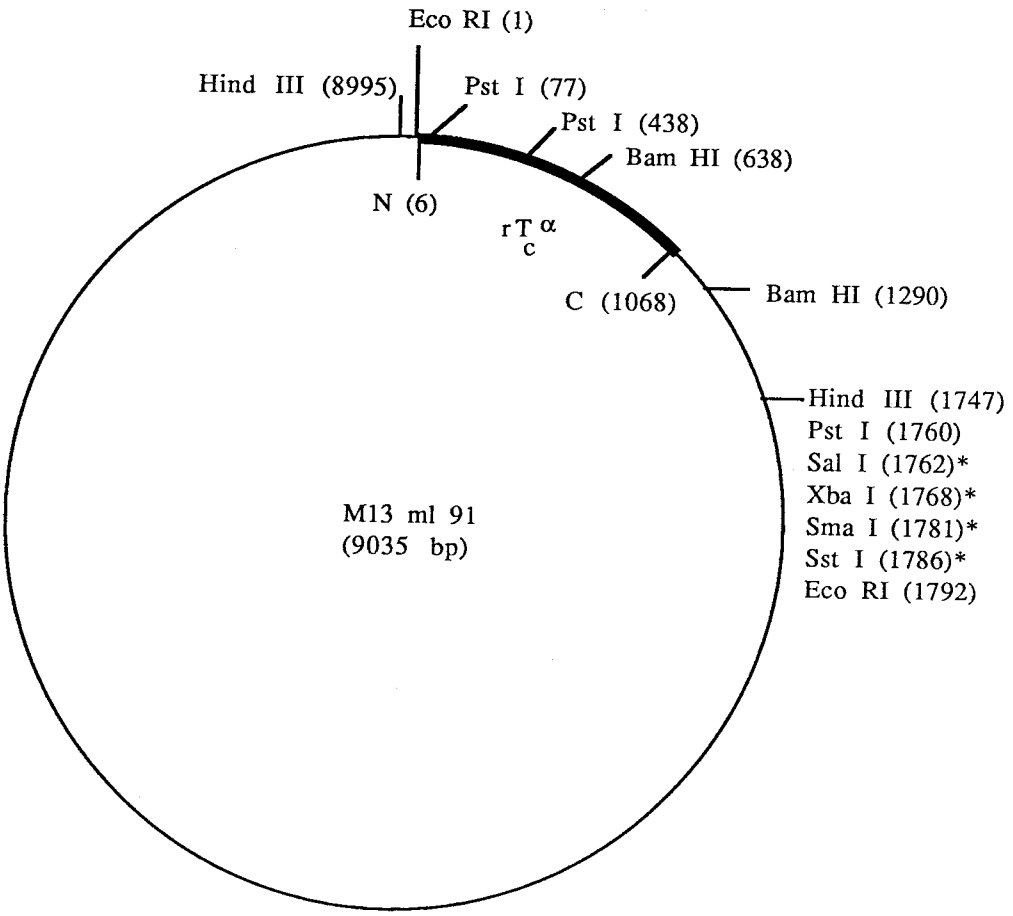

Eco RI (1)
Hind III (8995)
Pst I (77)
Pst I (438)
Bam HI (638)
N (6)
rTcα
C (1068)
Bam HI (1290)
Hind III (1747)
Pst I (1760)
Sal I (1762)*
Xba I (1768)*
Sma I (1781)*
Sst I (1786)*
Eco RI (1792)
M13 ml 91
(9035 bp)

Figure 45. Derivation history of plasmids containing the $rT_1\alpha$ cDNA.

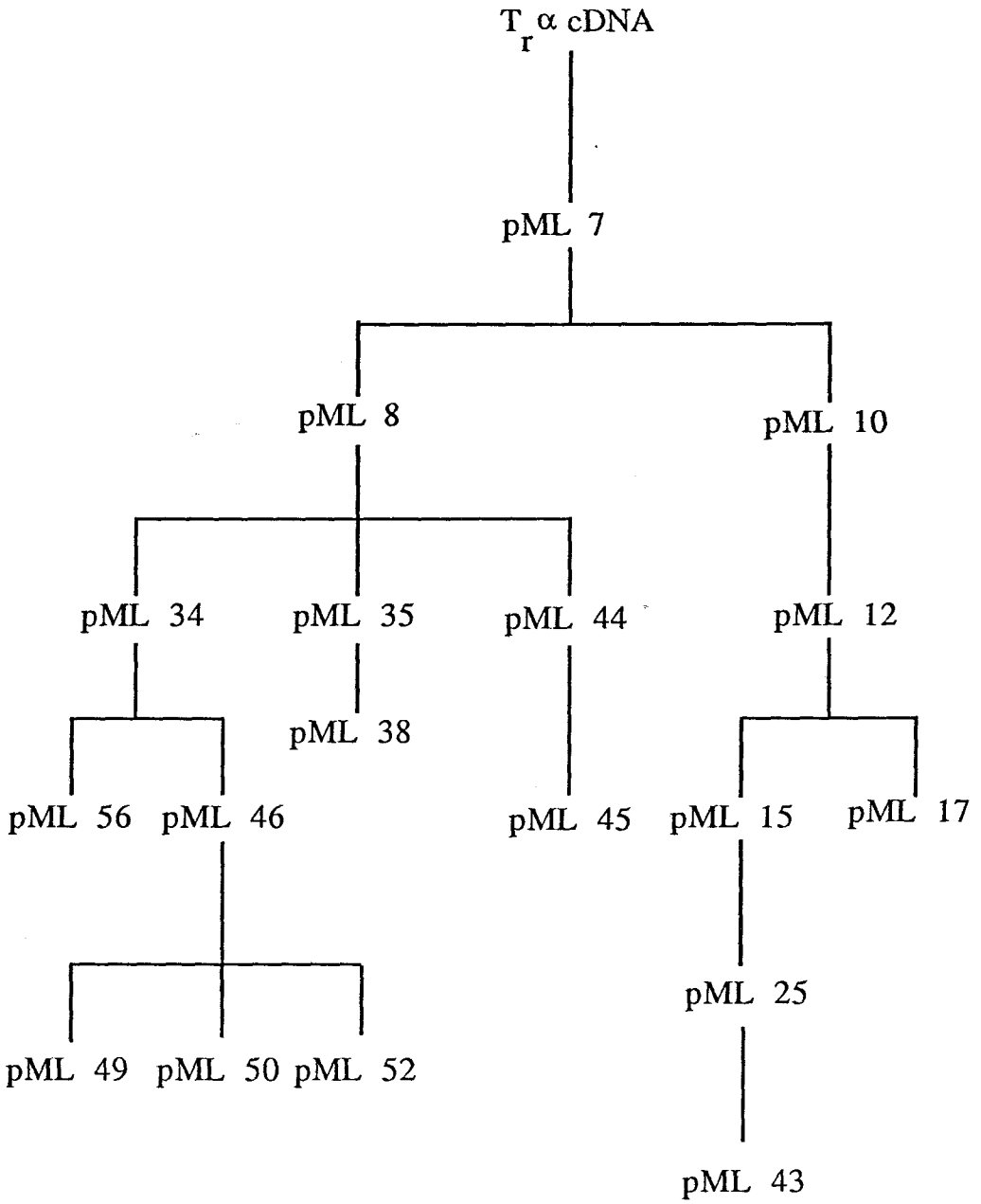

T r α cDNA
pML 7
pML 8
pML 10
pML 34
pML 35
pML 44
pML 12
pML 38
pML 56
pML 46
pML 45
pML 15
pML 17
pML 25
pML 49
pML 50
pML 52
pML 43

Figure 46. Derivation history of plasmids containing the $rT_C\alpha$ cDNA.

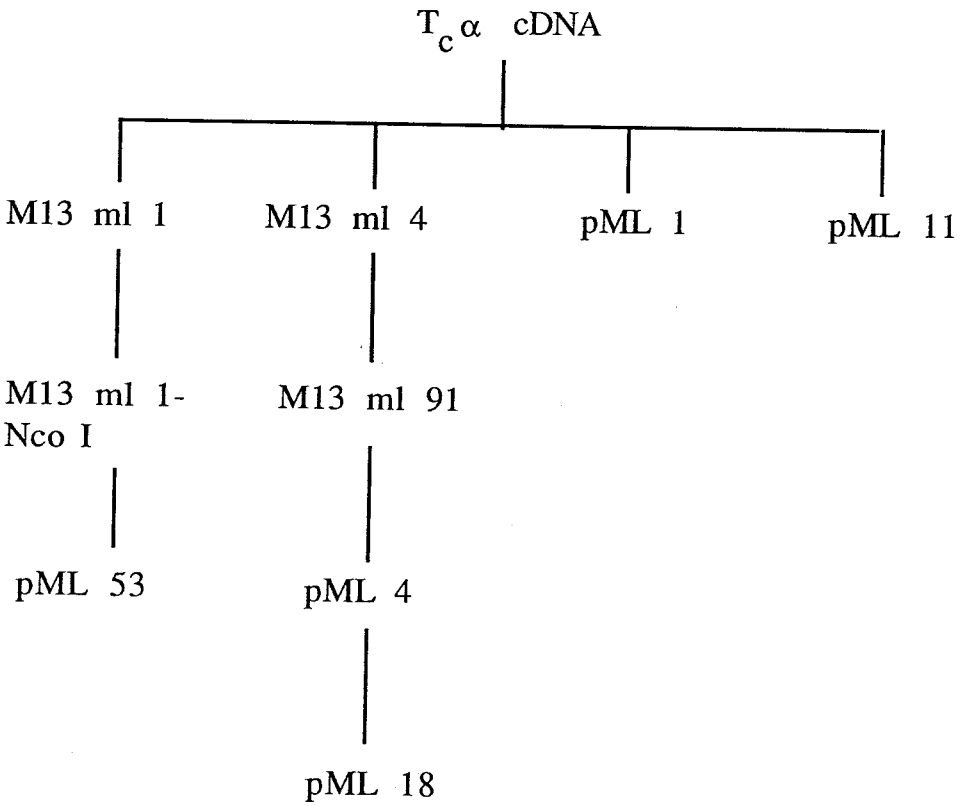
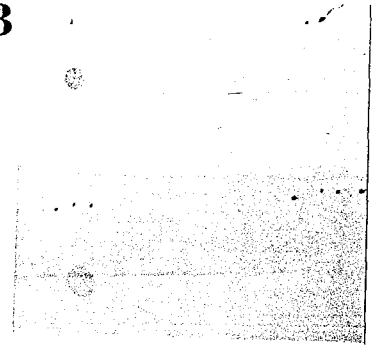

$T_c\alpha$ cDNA
M13 ml 1
M13 ml 4
pML 1
pML 11
M13 ml 1-Nco I
M13 ml 91
pML 53
pML 4
pML 18

Figure 47. Dot blots using $T_r\alpha$ and $T_c\alpha$ antipeptide antisera. Dots with 1 μg of retinal $T_r\alpha$ or 10 μg of each peptide were made on nitrocellulose squares in the arrangement indicated in (A). The blots were subjected to Western blotting as described in Materials and Methods using Tα1A sera (B), Tα2C sera (C), Tα2A sera (D), TαE sera (E), or TαF sera (F).

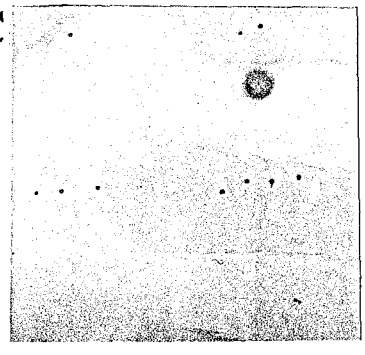
A

retinal $T_r\alpha$	Tα2CD peptide
Tα1AB peptide	TαF peptide

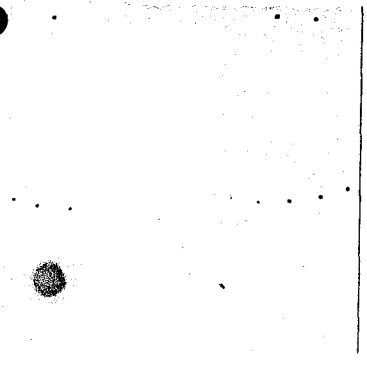
B



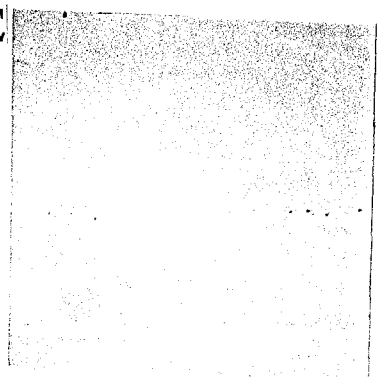
C



D



E



F

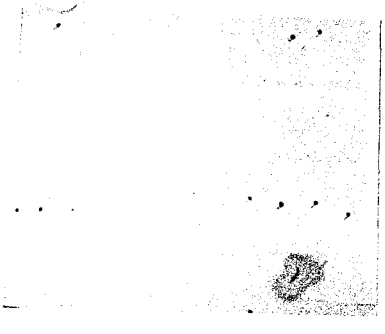


Figure 48. Specificity of antipeptide antisera. The preimmune Tα1AB sera (A), Tα1AB sera (B), preimmune Tα2CD sera (C), and Tα2CD sera (D) were used on Western blots of retinal $T_r\alpha$ (Lane 1), $rT_r\alpha$ (Lane 2), or $T_c\alpha$ (Lane 3) as described in Materials and Methods.

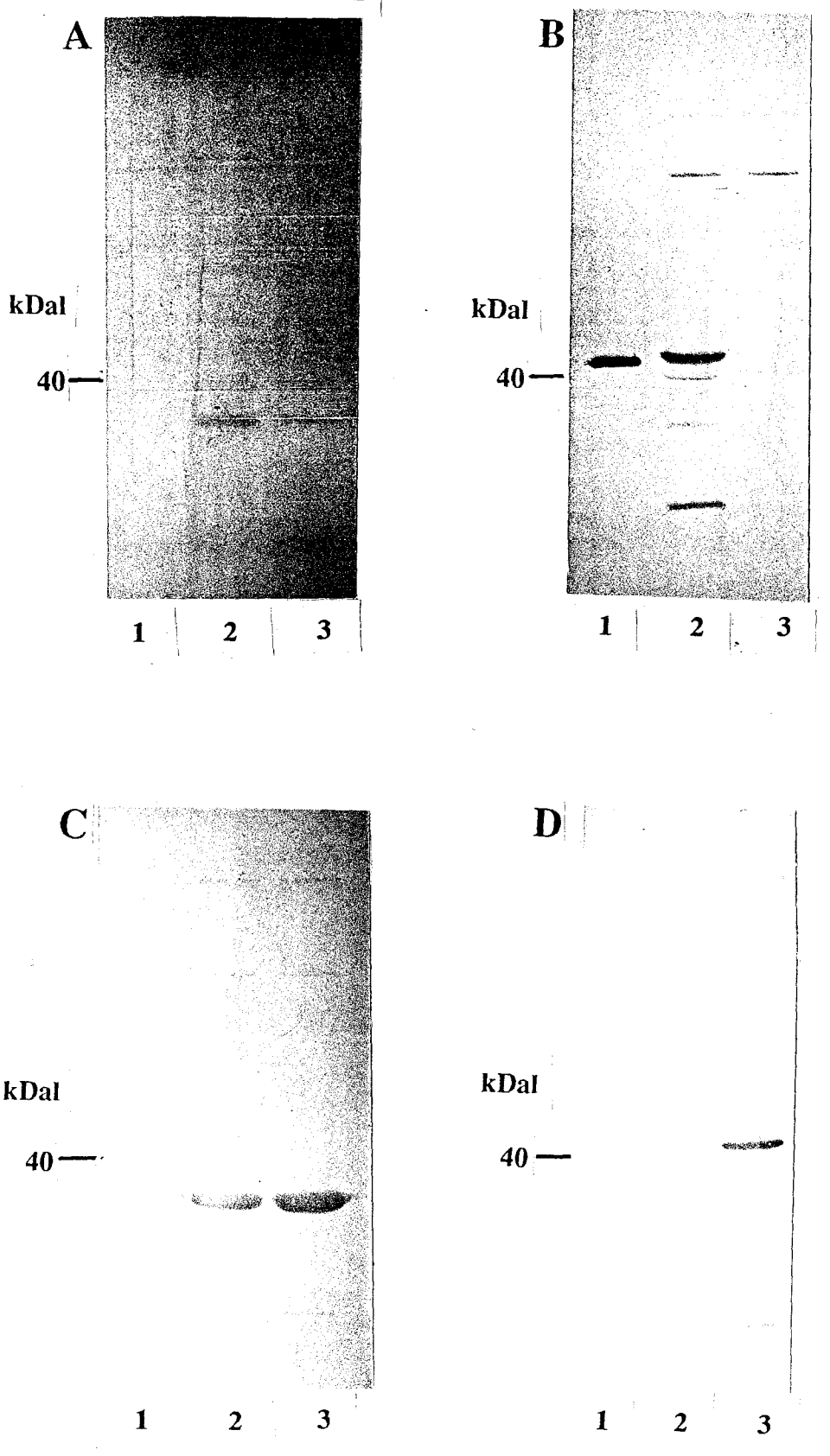
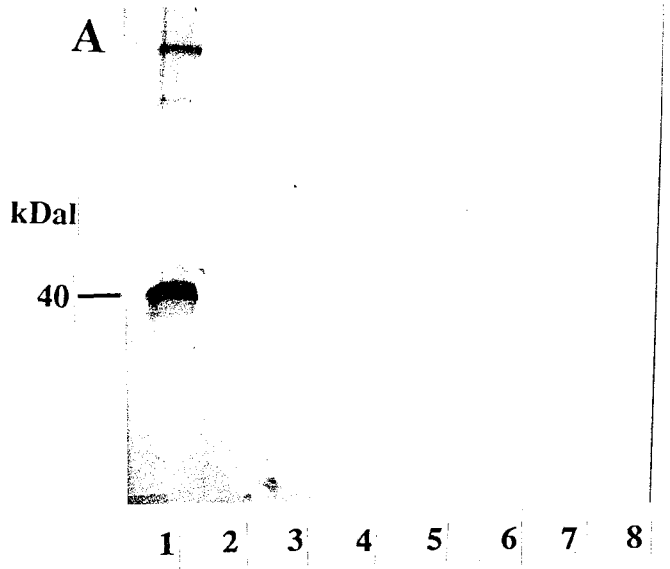

A
kDal
40
1
2
3
B
kDal
40
1
2
3
C
kDal
40
1
2
3
D
kDal
40
1
2
3

Figure 49. Western blots of yeast strains 20B-12 pML 35 and NNY105 pML 35. The cell extracts were analyzed in panel A and the media were analyzed in panel B. The Tα1B antisera was used. Lane 1, 500 ng retinal $T_r\alpha$; Lane 2, 20B-12; Lane 3, 20B-12 pMFα8; Lnae 4, 20B-12 pML 38; Lane 5, NNY105; Lane 6, NNY105 pMFα8; Lane 7, NNY105 pML 35; Lane 8, NNY105 pML 38.


A
kDal
40
1 2 3 4 5 6 7 8

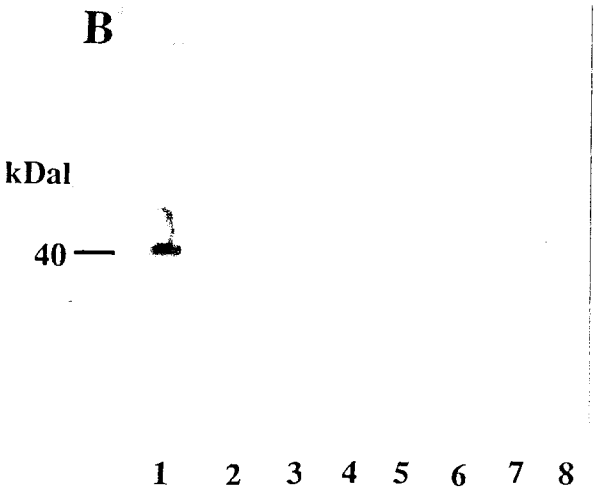
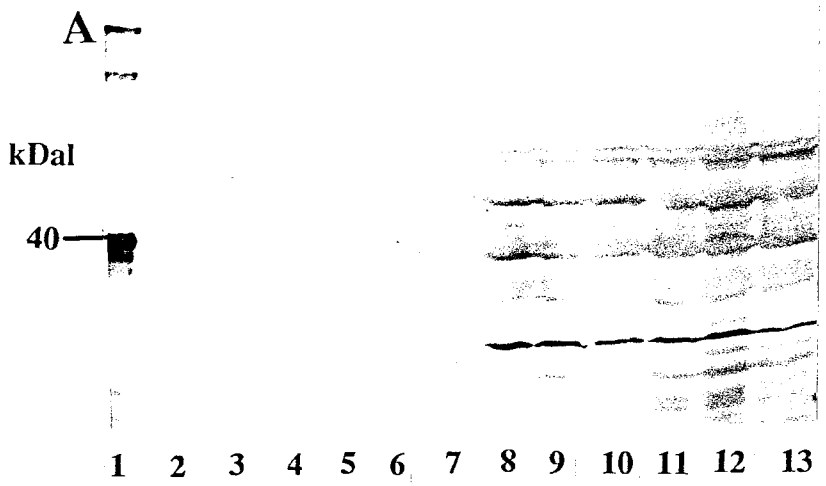

B
kDal
40
1 2 3 4 5 6 7 8

Figure 50. Western blots of yeast strains HR125 pML 35, EJ102 PML 35, and DSR865-11α pML 35. The Tα1B antisera was used for Western blotting and yeast extracts were prepared as described in Materials and Methods. A) Lane 1, 2 μg retinal $T_r\alpha$; Lanes 2 and 8, DSR865-11α; Lanes 3 and 9, DSR865-11α pML 35; Lanes 4 and 10, DSR865-11α pML 35; Lanes 5 and 11, EJ102; Lanes 6 and 12, EJ102 pML 35; Lanes 7 and 13, EJ102 pML 35. The media was analyzed in lanes 2-7 and the cell lysate was analyzed in lanes 8-13. B) Lane 1, 1 μg retinal $T_r\alpha$; Lanes 2 and 8, HR125a pMFα8; Lanes 3 and 9, HR125α pMFα8; Lanes 4 and 10, HR125a pML 35; Lanes 5 and 11, HR125α pML 35; Lanes 6 and 12, HR125a pML 38; Lanes 7 and 13, HR125α pML 38. The cell lysates were analyzed in lanes 2-7 and the media were analyzed in lanes 8-13.


A
kDal
40
1 2 3 4 5 6 7 8 9 10 11 12 13

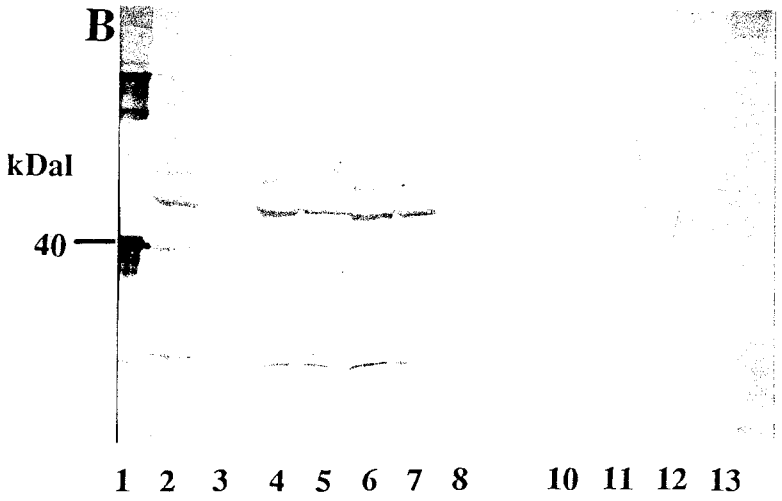

B
kDal
40
1 2 3 4 5 6 7 8 10 11 12 13

Figure 51. Western blots of yeast strains containing pML 11. Yeast strains were grown in nonselective YPD media (Lanes 2-7) or in selective SD media without leucine (Lanes 8-13) for two days. Yeast cell extracts were prepared and analyzed by Western blotting as described in Materials and Methods using the Tα2C antisera. Extracts analyzed were from: Lane 1, *E. coli* CAG1139 pML 4; Lanes 2 and 8, SF838-10α pML 11(-); Lanes 3 and 9 SF838-10α pML 11(+); Lanes 4 and 10, SR741-3B pML 11(-); Lanes 5 and 11, SR741-3B pML 11(+); Lanes 6 and 12, SR741-3B; Lanes 7 and 13, SF838-10α.

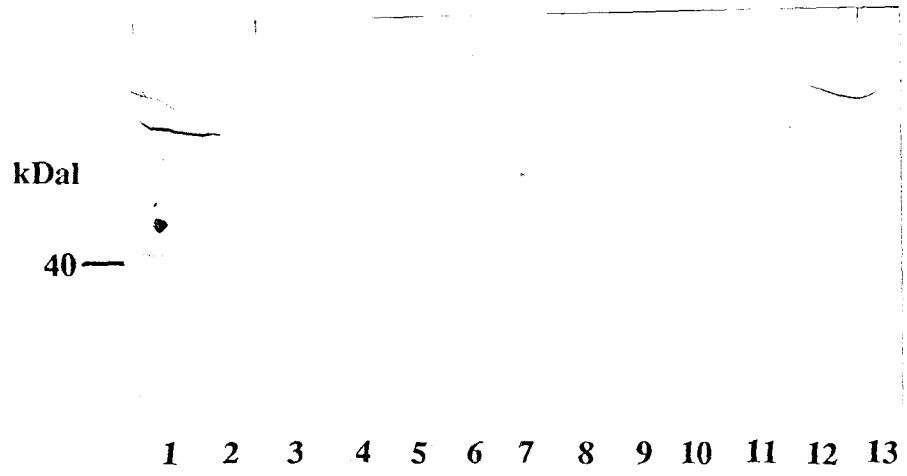

kDal
40
1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 52. SDS gel analysis of proteins in Sf9 insect cells infected with baculoviruses. A) Sf9 cells were infected with wild-type AcNPV (Lane 2) or AcNPV-Ac360βgal (Lane 3). The cells were harvested 48 hours post-infection, dissolved in gel sample buffer, and 50 μg of protein was subjected to SDS gel electrophoresis as described in Materials and Methods. Molecular weights standards are in Lane 1. B) Sf9 cells were infected with AcNPV-ML45 (Lane 2) and 10 μg of the insoluble pellet obtained after ultracentrifugation of a crude lysate at 100, 000 X *g* was analyzed. Molecular weight standards are in Lane 1.

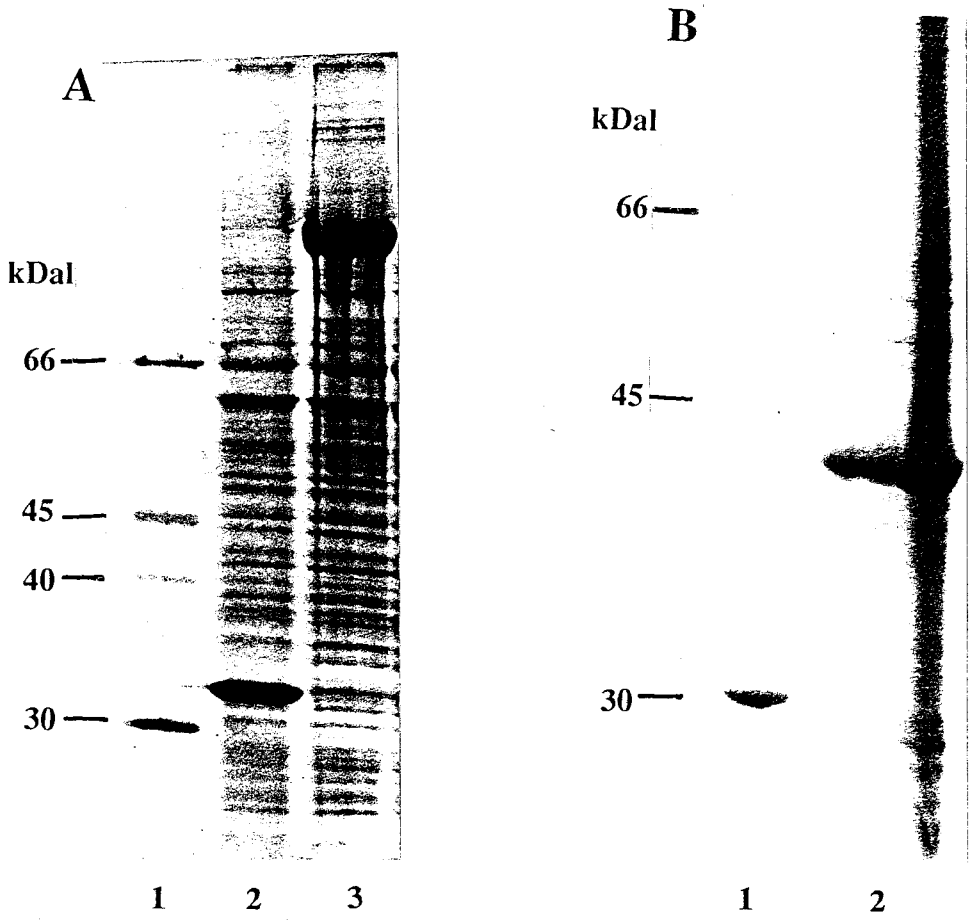

A
kDal
66
45
40
30
1
2
3
B
kDal
66
45
30
1
2

Figure 53. Plaques formed by AcNPV-Ac360βgal. About 6 X 10^6 Sf9 cells were seeded in a 10 cm petri dish and infected with AcNPV-Ac360βgal for one hour. The media was removed and the cells were overlayed with Grace's complete media containing 1% agarose and 0.15 mg/ml XGAL. The plate was incubated for 3 days at 28°C. A photograph of a plaque was taken at a magnification of 100 X.

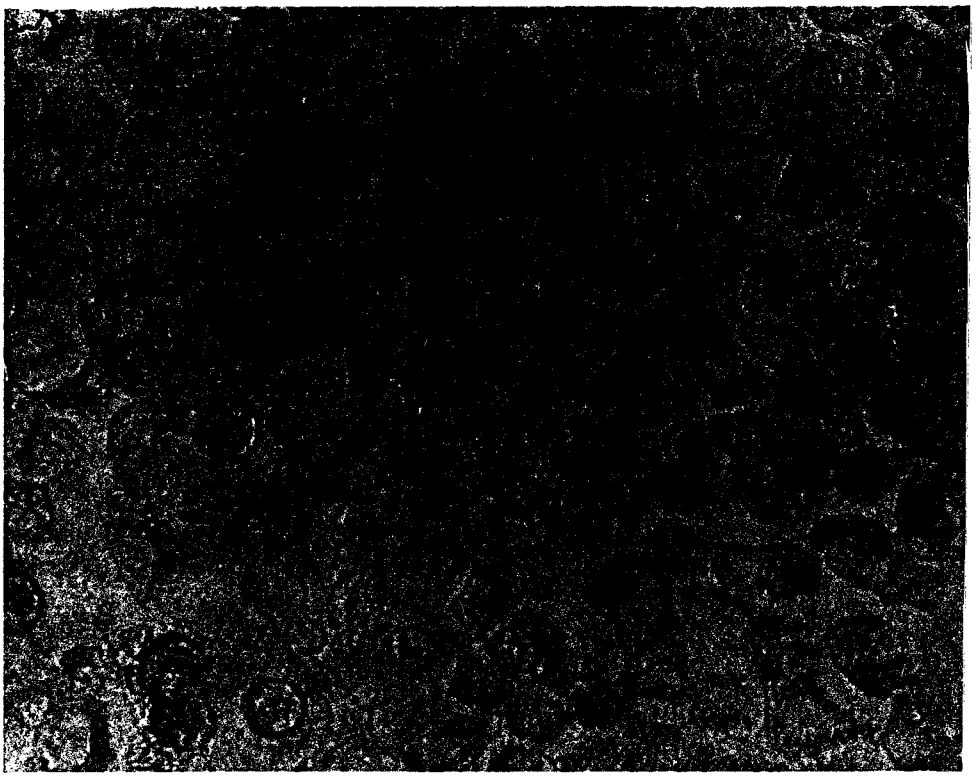


Figure 54. Time course of $rT_r\alpha$ expression in Sf9 insect cells infected with AcNPV-ML34. Sf9 cells were infected with AcNPV-ML34, harvested at 12, 24, 36, 48, 60, and 72 hour intervals post-infection, and cell lysates were prepared. The lysates were subjected to Western blotting using the Tα1B antisera to detect $rT_r\alpha$. After developing the blot it was scanned with a laser densitometer.

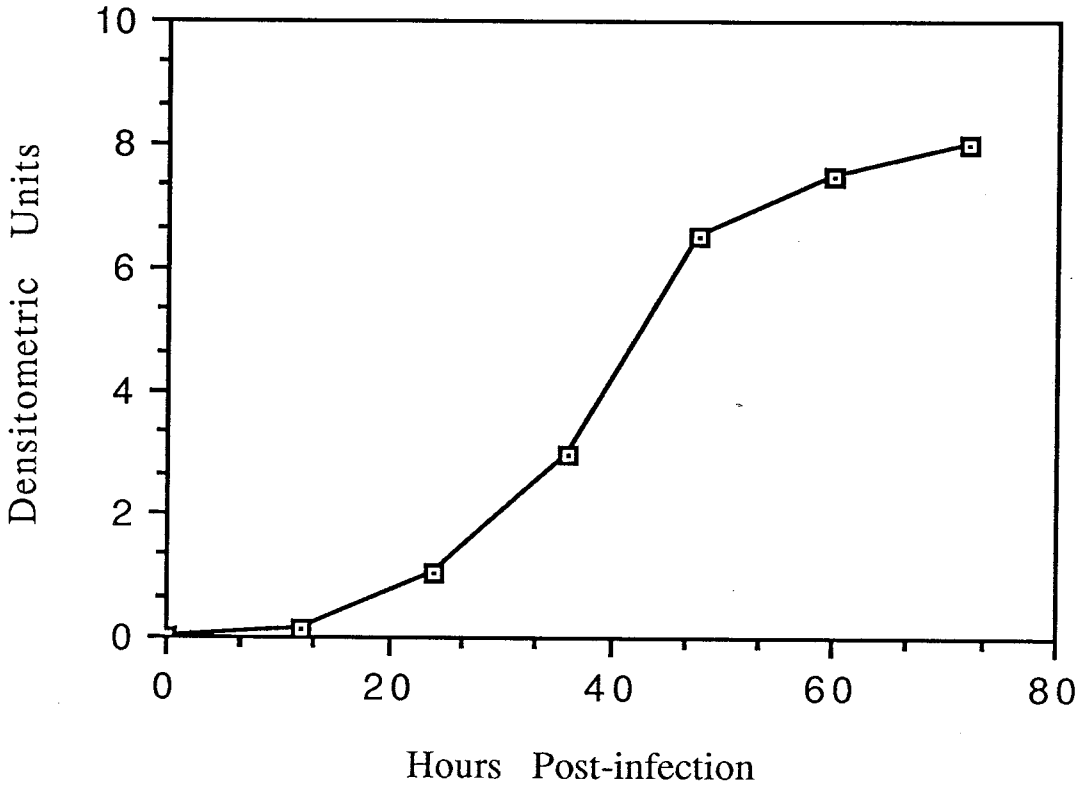

10
8
6
4
2
0
Densitometric Units
0
20
40
60
80
Hours Post-infection

Figure 55. Uninfected Sf9 cells. Sf9 cells were fixed as described in Materials and Methods and the photograph was taken using phase contrast optics at a magnification of 100 X.

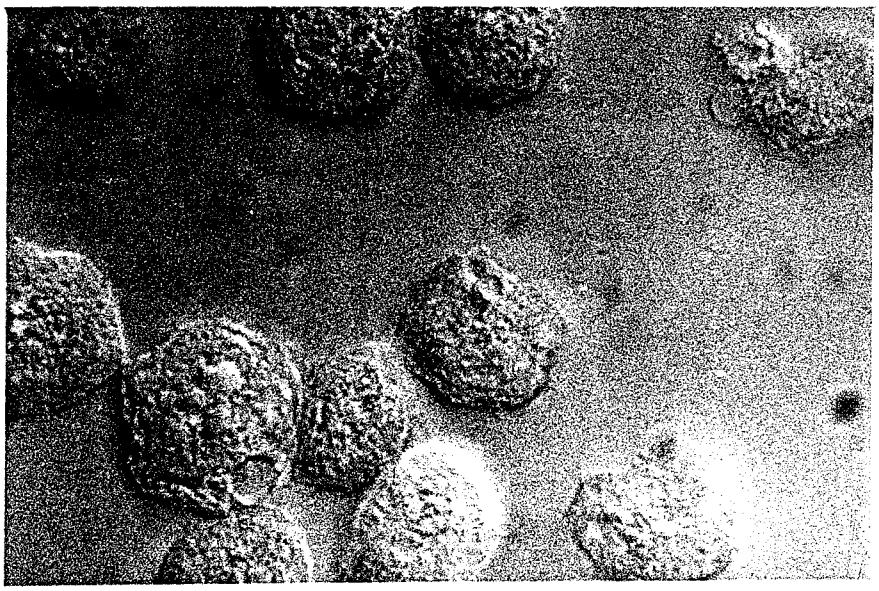
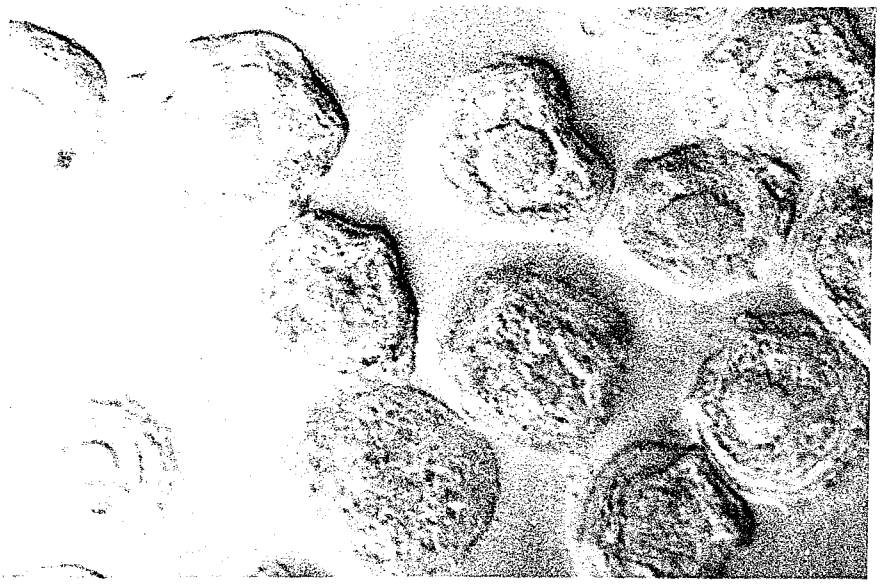


Figure 56. Sf9 cells infected with AcNPV-ML34(+). Sf9 cells were infected with AcNPV-ML34(+) and fixed 48 hours post-infection as described in Materials and Methods. The photographs were taken using phase contrast (A) or fluorescence (B) optics at a magnification of 100 X.

A



B

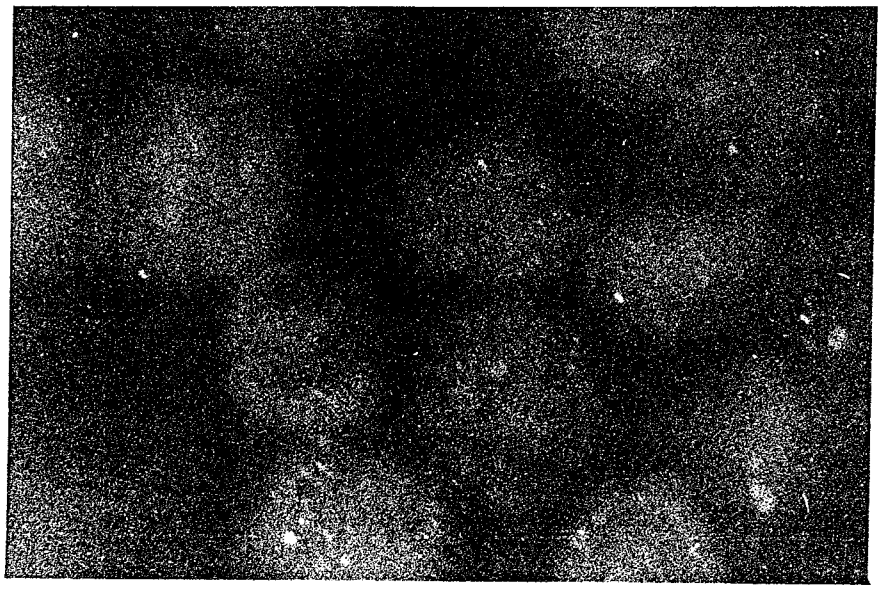


Figure 57. Sf9 cells infected with AcNPV-ML45 at 24 hours post-infection. Sf9 cells were infected with AcNPV-ML45 and fixed 24 hours post-infection as described in Materials and Methods. The photograph was taken using fluorescence optics at a magnification of 100 X.

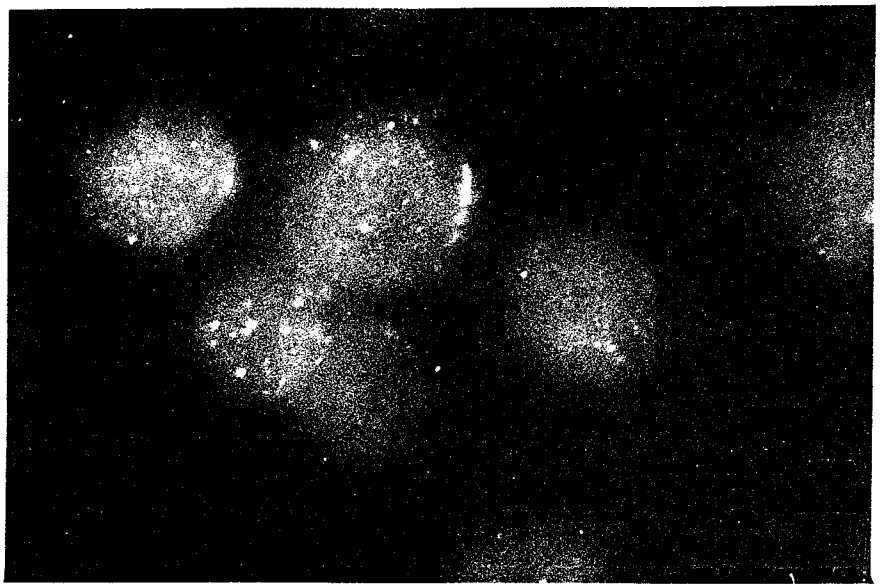
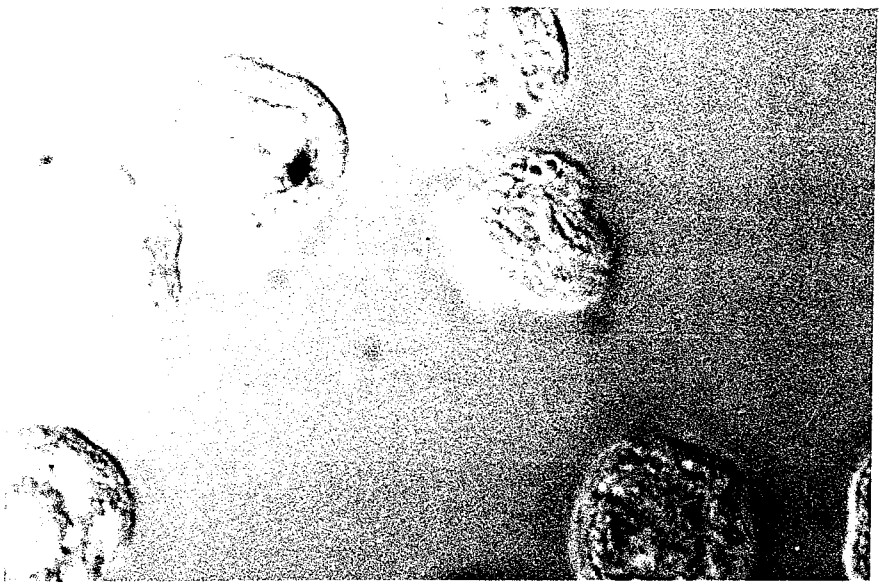


Figure 58. Sf9 cells infected with AcNPV-ML45 at 36 hours post-infection. Sf9 cells were infected with AcNPV-ML45 and fixed 36 hours post-infection as described in Materials and Methods. The photographs were taken using phase contrast (A) or fluorescence (B) optics at a magnification of 100 X.

A



B

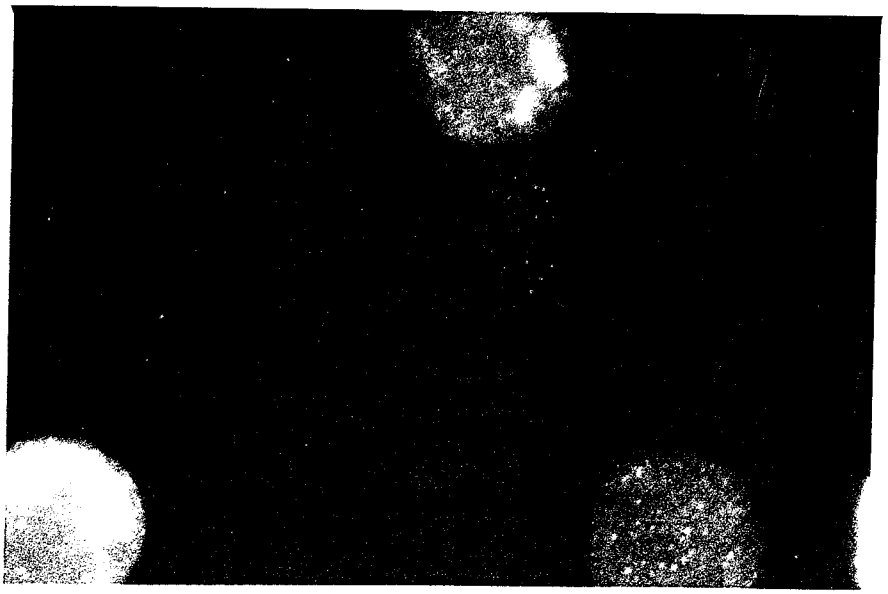


Figure 59. Sf9 cells infected with AcNPV-ML45 at 48 hours post-infection. Sf9 cells were infected with AcNPV-ML45 and fixed 48 hours post-infection as described in Materials and Methods. The photographs were taken using phase contrast (A) or fluorescence (B) optics at a magnification of 100 X.

A

B

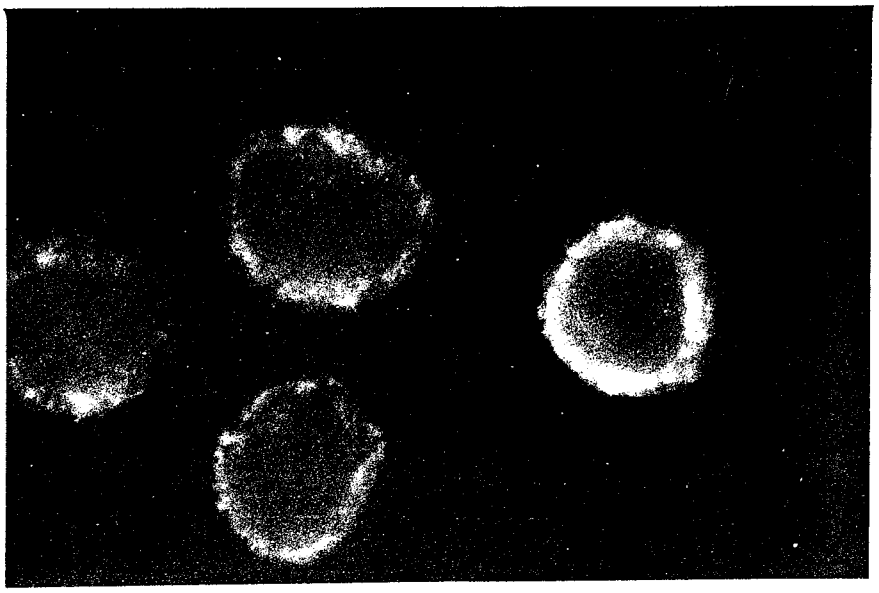
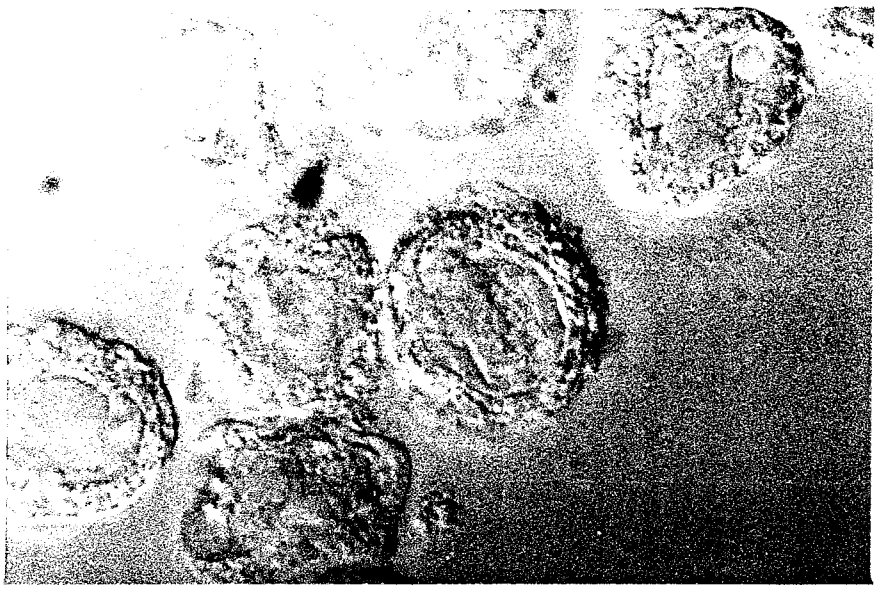


Figure 60. Sf9 cells infected with AcNPV-ML45 at 72 hours post-infection. Sf9 cells were infected with AcNPV-ML45 and fixed 72 hours post-infection as described in Materials and Methods. The photographs were taken using phase contrast (A) or fluorescence (B) optics at a magnification of 100 X.

A



B

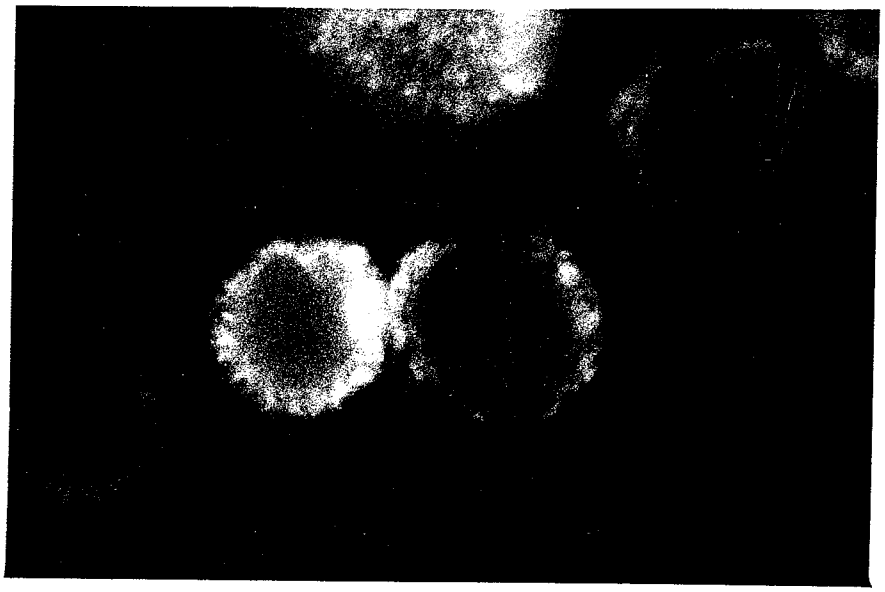
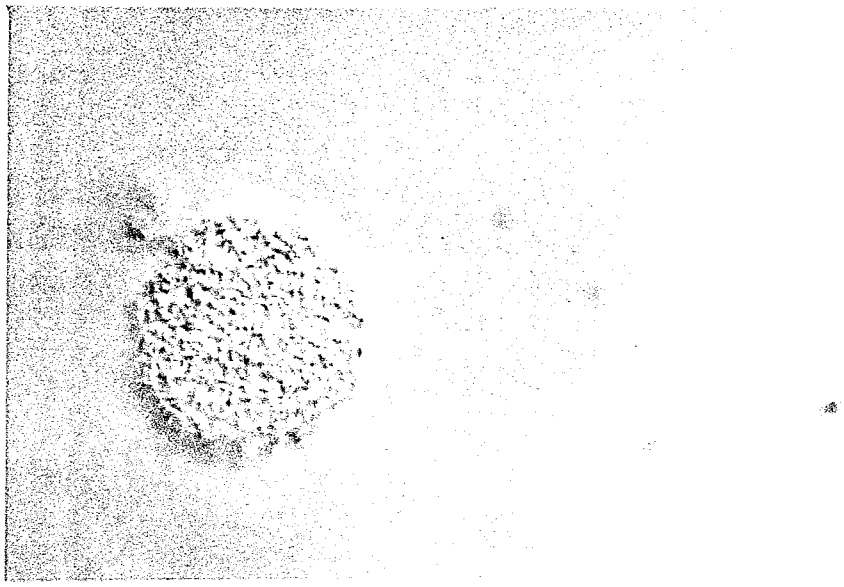


Figure 61. Sf9 cells infected with AcNPV-ML45 at 120 hours post-infection. Sf9 cells were infected with AcNPV-ML45 and fixed 120 hours post-infection as described in Materials and Methods. The photographs were taken using phase contrast (A) or fluorescence (B) optics at a magnification of 100 X.

A



B

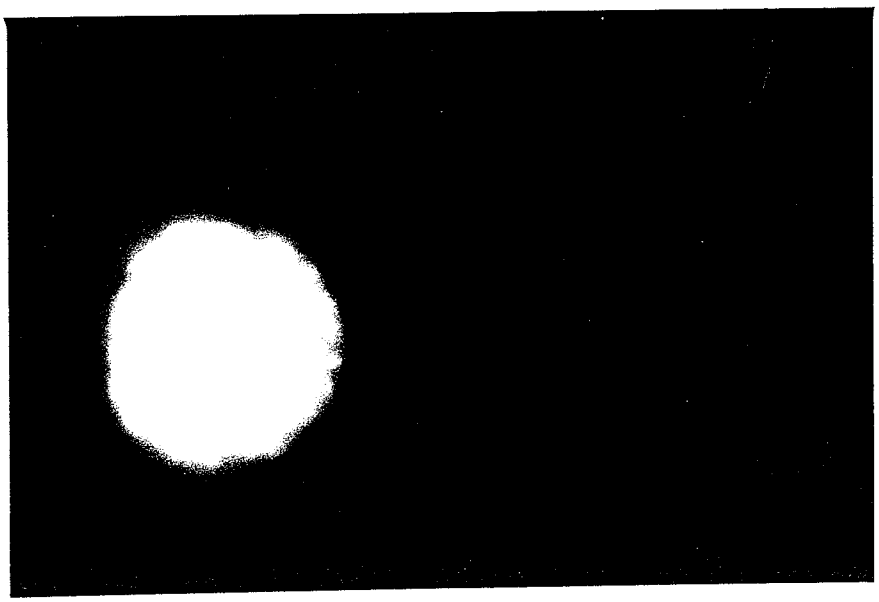
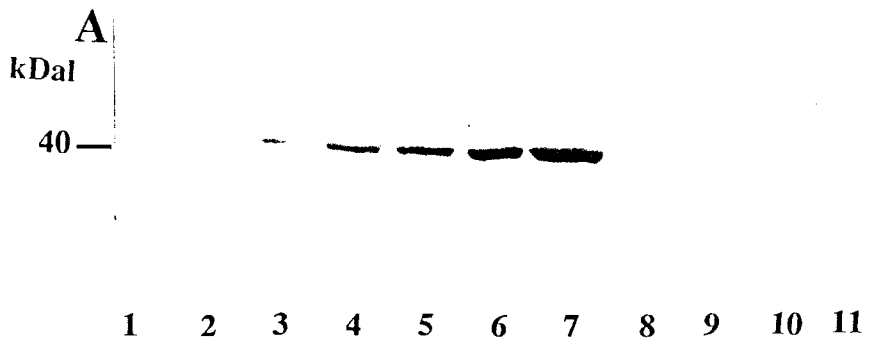


Figure 62. Distribution of polyhedrin-r$T_r\alpha$ in soluble and insoluble fractions of AcNPV-ML45-infected Sf9 cells. Sf9 cells were infected with AcNPV-ML45, harvested 48 hours post-infection, lysed in a French pressure cell, and centrifuged at 100,000 X *g* for one hour to generate a soluble supernatant and an insoluble pellet fraction. A) Lane 1, 16 μg supernatant; Lane 2, 160 μg supernatant; Lane 3, 2.4 μg pellet; Lane 4, 4.8 μg pellet; Lane 5, 9.6 μg pellet; Lane 6, 19.2 μg pellet; Lane 7, 38.4 μg pellet; Lane 8, 100 ng retinal $T_r\alpha$; Lane 9, 250 ng retinal $T_r\alpha$; Lane 10, 500 ng retinal $T_r\alpha$; Lane 11, 750 ng retinal $T_r\alpha$. B) Lane 1, 500 ng retinal $T_r\alpha$; Lane 2, 5% of supernatant by weight, Lane 3, 5% of pellet by weight.


A
kDal
40
1 2 3 4 5 6 7 8 9 10 11

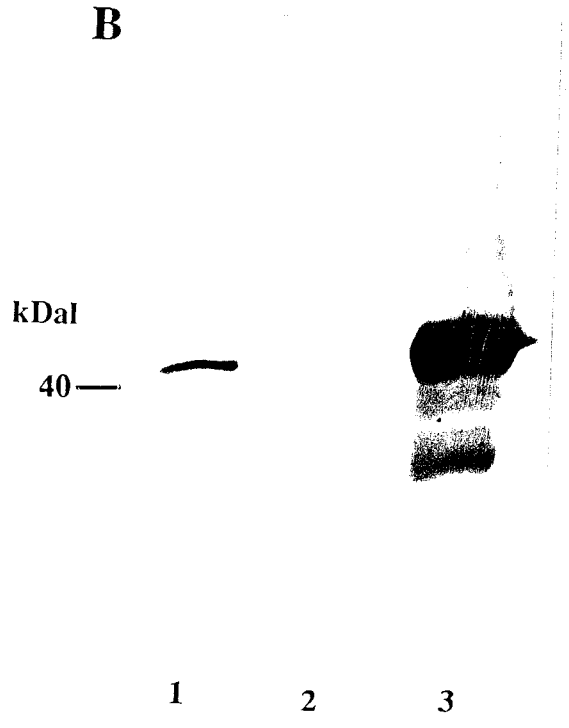
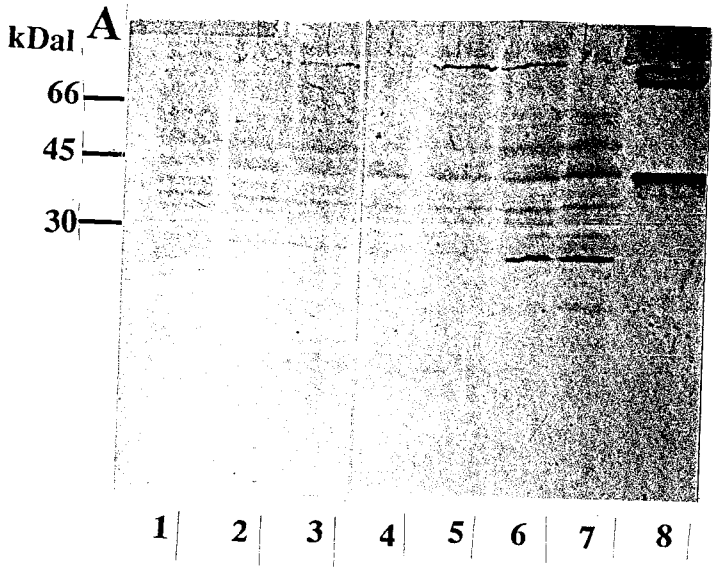

B
kDal
40
1 2 3

Figure 63. Expression of $rT_r\alpha$ in strains containing pML 12, pML 15, and pML 25. Total cell extracts were prepared from *E. coli* grown at 37°C and induced with 1 mM IPTG. Extracts were analyzed by Western blotting using the Tα1B antisera as described in Materials and Methods. A) Lane 1, D1210 pKK233-2 induced for 12 hours; Lane 2, D1210 pML 12(-) induced for 12 hours; Lane 3, D1210 pML 12(+) induced for 12 hours; Lane 4, MC1061 pML 15 uninduced; Lane 5, MC1061 pML 15 induced for 15 minutes; Lane 6, MC1061 pML 15 induced for 120 minutes; Lane 7, MC1061 pML 15 induced for 12 hours; Lane 8, 1 μg retinal $T_r\alpha$. B) Lane 1, 5 μg retinal $T_r\alpha$; Lane 2, JM101 pML 25 induced for 12 hours; Lane 3, JM101 pML 15 induced for 12 hours.


kDal
A
66
45
30
1
2
3
4
5
6
7
8

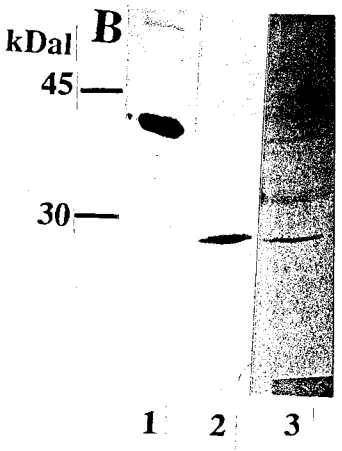
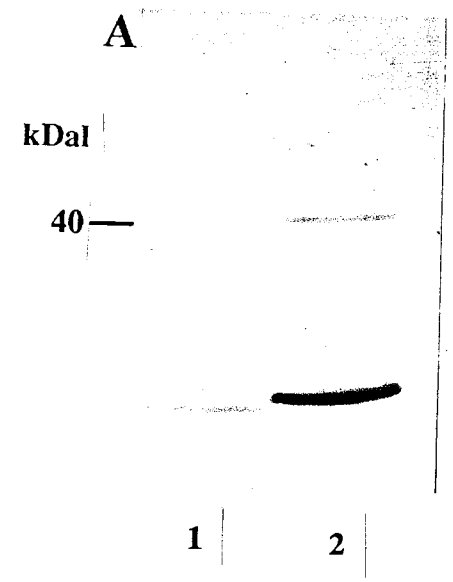

kDal
B
45
30
1
2
3

Figure 64. Analysis of $rT_r\alpha$ expression in strains containing pML 25. Cell extracts were prepared by French pressure cell lysis from cultures grown at 37°C and induced with 1 mM IPTG for 12 hours. A) Lane 1, MC1061 pML 25; Lane 2, CAG1139 pML 25. B) Lane 1, AR58 pML 25.


A
kDal
40
1
2

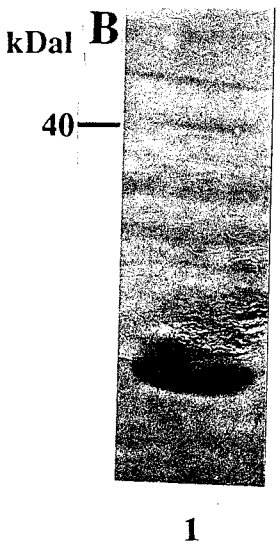

kDal
B
40
1

Figure 65. Analysis of $rT_C\alpha$ expression in CAG1139 pML 4 and CAG1139 pML 18. Cell extracts were prepared by French pressure cell lysis from cultures grown at 37°C and induced with 0.1 mM IPTG for 30 minutes. The crude extract was separated into a soluble and an insoluble fraction by centrifugation at 100,000 X g for one hour. Lane 1, $rT_C\alpha$ molecular weight standard; Lane 2, supernatant from CAG1139 pML 4; Lane 3, pellet from CAG1139 pML 4; Lane 4, supernatant from CAG1139 pML 18; Lane 5, pellet from supernatant from CAG1139 pML 18; Lane 6, $rT_T\alpha$ molecular weight standard.

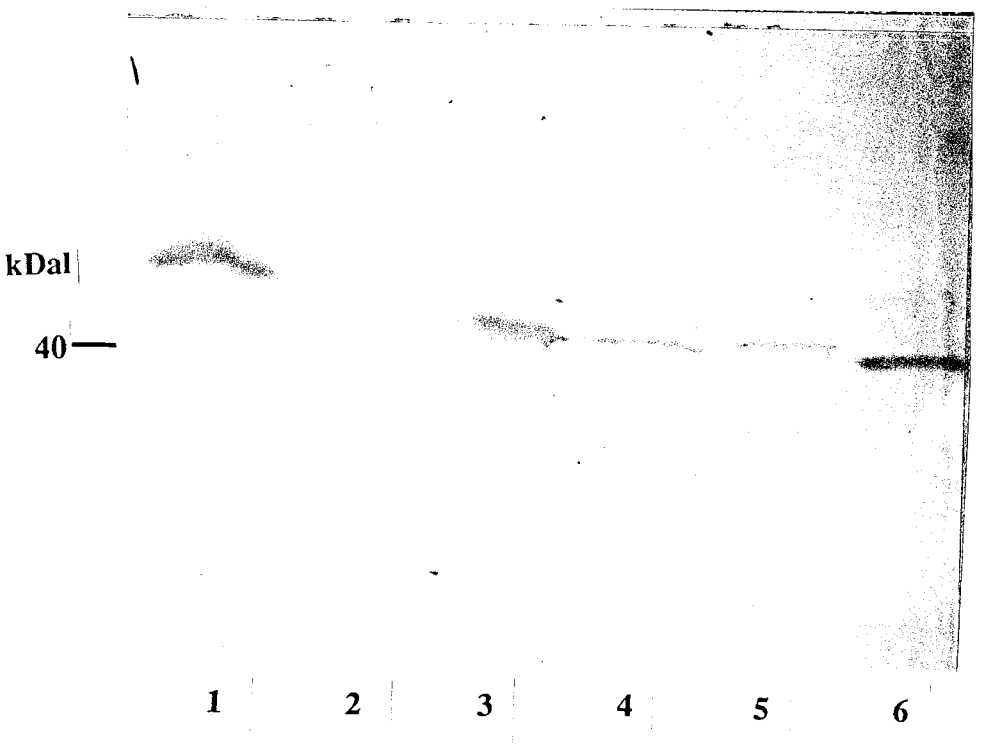

kDal
40
1
2
3
4
5
6

Figure 66. Diagram of Glutagene protein expression system. Cells containing a Glutagene vector are induced with IPTG and lysed. The glutathione S-transferase (GST)-transducin α (Tα) fusion protein that is present in the crude lysate is bound to glutathione-agarose beads. The beads are washed and the fusion is eluted with 5 mM glutathione in the buffer of choice. The purified fusion protein is then cleaved with thrombin to separate GST and Tα. GST is removed by binding to glutathione-agarose beads leaving purified Tα.

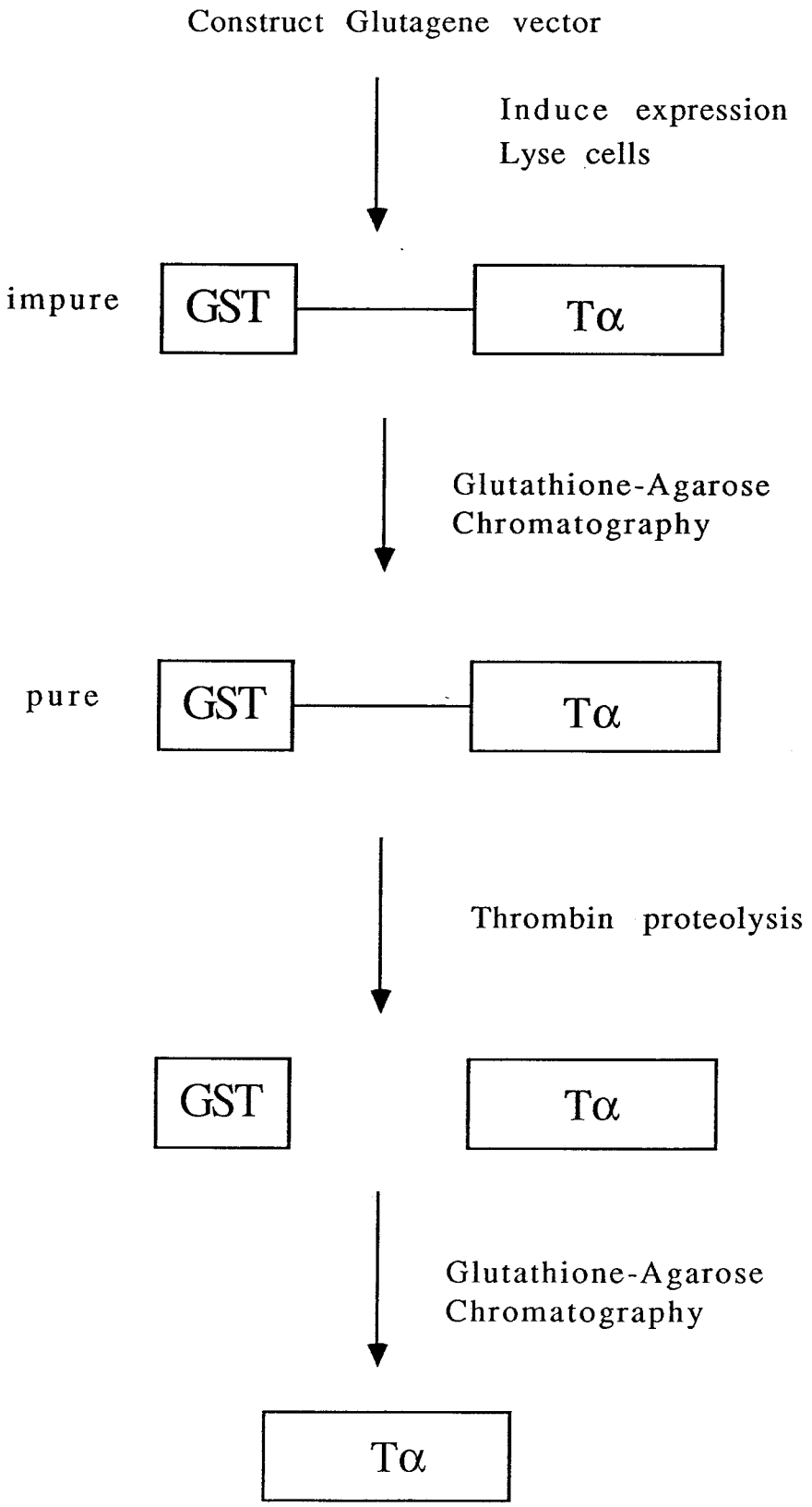
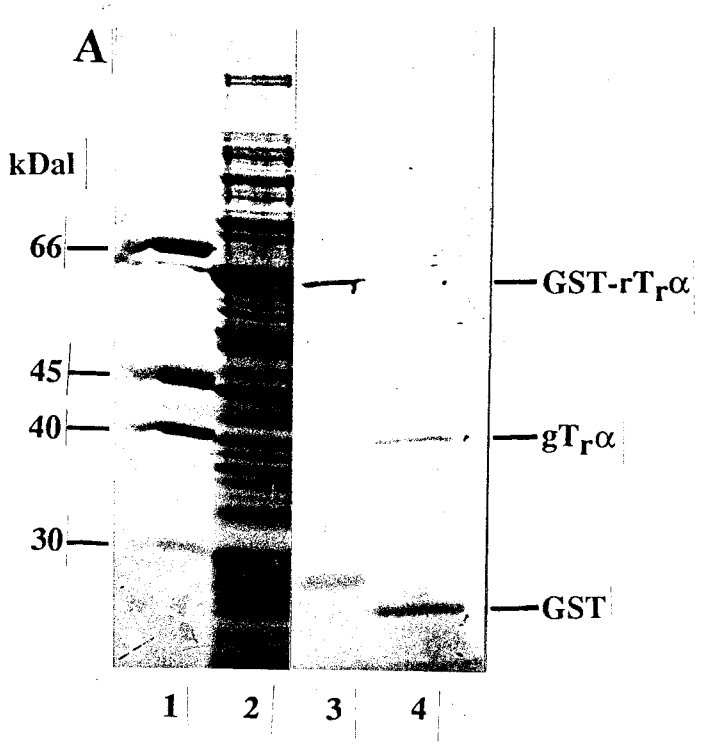

Construct Glutagene vector
Induce expression
Lyse cells
impure
GST
Tα
Glutathione-Agarose
Chromatography
pure
GST
Tα
Thrombin proteolysis
GST
Tα
Glutathione-Agarose
Chromatography
Tα

Figure 67. Analysis of $gT_r\alpha$ produced using the Glutagene system. A) Fractions derived from the preparation of $gT_r\alpha$ (see Materials and Methods for details). Lane 1, molecular weight standards; Lane 2, crude extract; Lane 3, GST-$rT_r\alpha$ eluted from glutathione-agarose beads; Lane 4, thrombin cleaved GST-$rT_r\alpha$. B) Trypsin cleavage of $gT_r\alpha$. Lane 1, 1 μg retinal $T_r\alpha$; Lane 2, 1 μg retinal $T_r\alpha$ treated with trypsin as described in Materials and Methods; Lane 3, 1 μg $gT_r\alpha$; Lane 4, 1 μg $gT_r\alpha$ treated with trypsin; Lane 5, 10 μg $gT_r\alpha$; Lane 6, 10 μg $gT_r\alpha$ treated with trypsin.


A
kDal
66
45
40
30
GST-rTrα
gTrα
GST
1
2
3
4

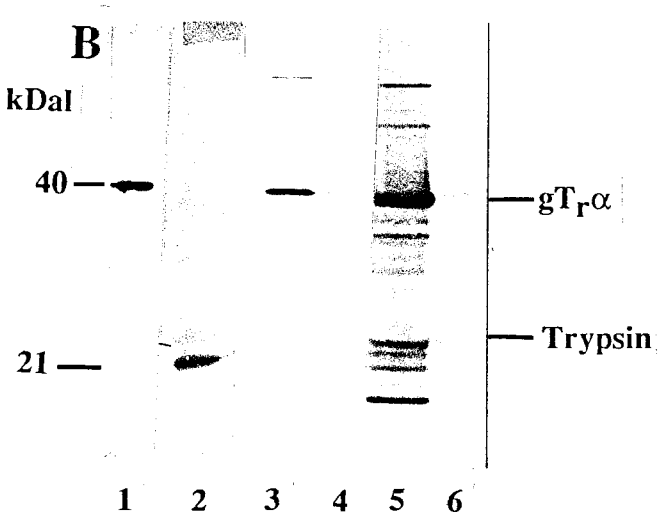
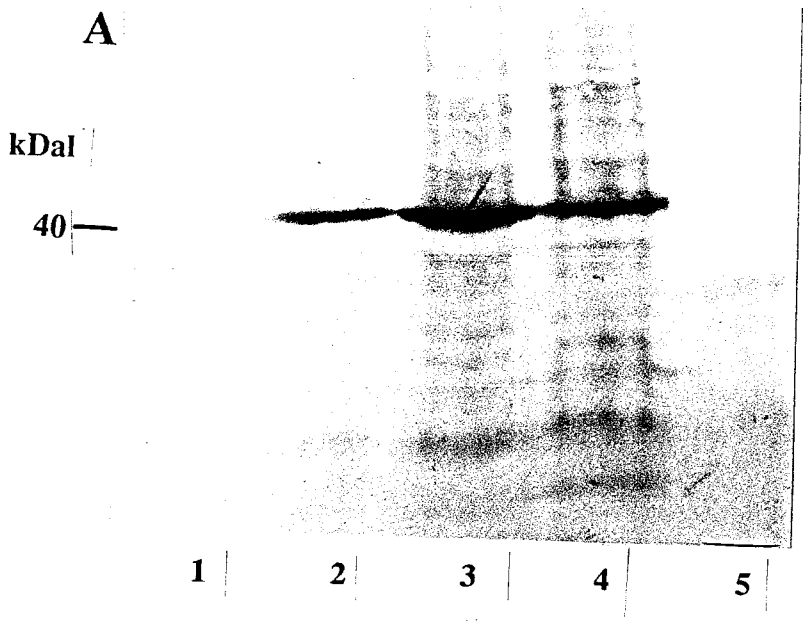

B
kDal
40
21
gTrα
Trypsin
1
2
3
4
5
6

Figure 68. Western blot analysis of $rT_r\alpha$ in extracts from CAG1139 pML 46 and CAG1139 pML 50 cells. Cells were grown in LB plus ampicillin and induced in 1 mM IPTG at 37°C. A) Time course of $rT_r\alpha$ expression. Lane 1, uninduced; Lane 2, 15-minute induction; Lane 3, 60-minute induction; Lane 4, 240-minute induction; Lane 5, 24-hour induction. B) Distribution of $rT_r\alpha$ in soluble and insoluble fractions. The fractions were prepared by centrifugation of a crude extract at 100,000 X *g* for one hour. Lane 1, 500 ng retinal $T_r\alpha$; Lane 2, 0.2% of the pellet from CAG1139 pML 46 cells; Lane 3, 0.8% of the supernatant from CAG1139 pML 46 cells; Lane 4, 0.2% of the pellet from CAG1139 pML 50 cells; Lane 5, 0.2% of the pellet from CAG1139 pML 50 cells.


A
kDal
40
1
2
3
4
5

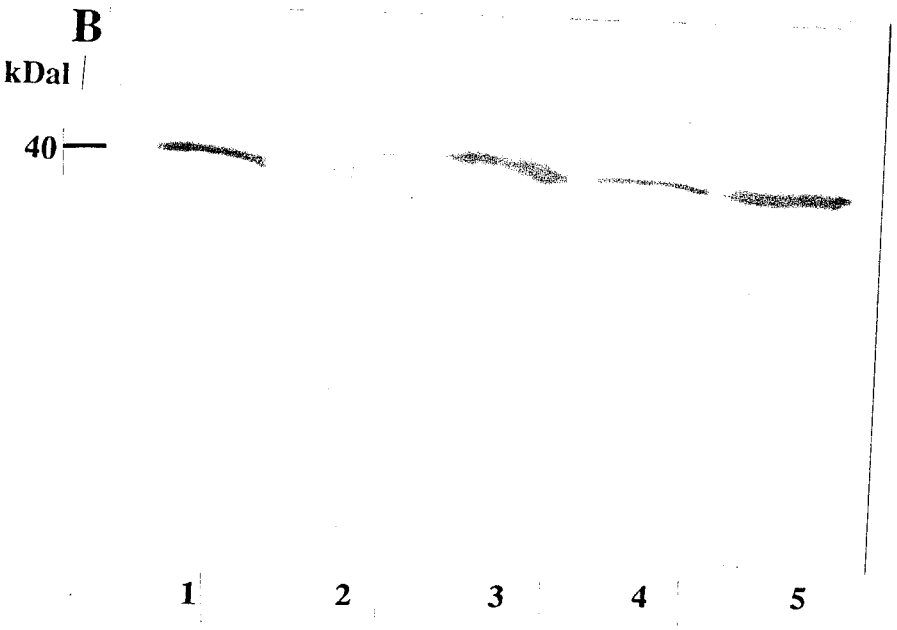

B
kDal
40
1
2
3
4
5

Figure 69. Mobility of $rT_r\alpha$ in an SDS gel. Total extracts were prepared from *E. coli* strains as described in Materials and Methods and subjected to Western blotting using a mixture of the Tα1B and Tα2C sera. Lane 1, 500 ng retinal $T_r\alpha$; Lane 2, $rT_c\alpha$ from CAG1139 pML 18; Lane 3, $rT_r\alpha$ from CAG1139 pML 46; Lane 4, $rT_r\alpha$ from CAG1139 pML 43; Lane 5, polyhedrin-$rT_r\alpha$ from AcNPV-ML45-infected Sf9 cells.

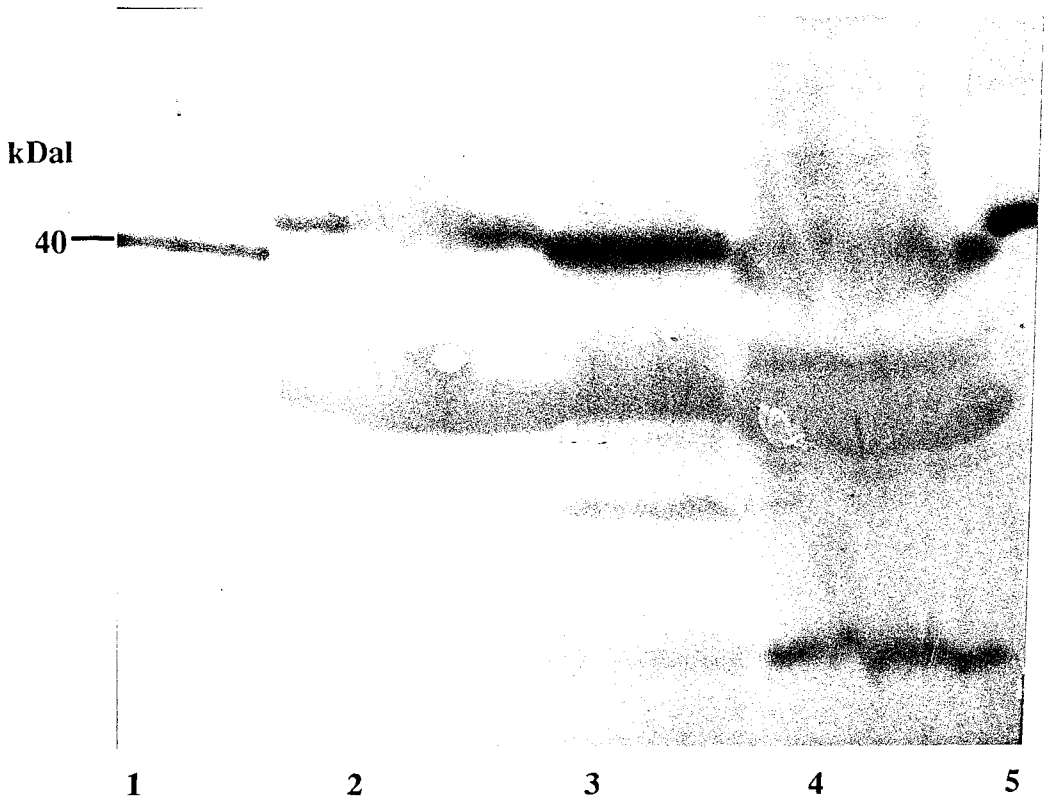
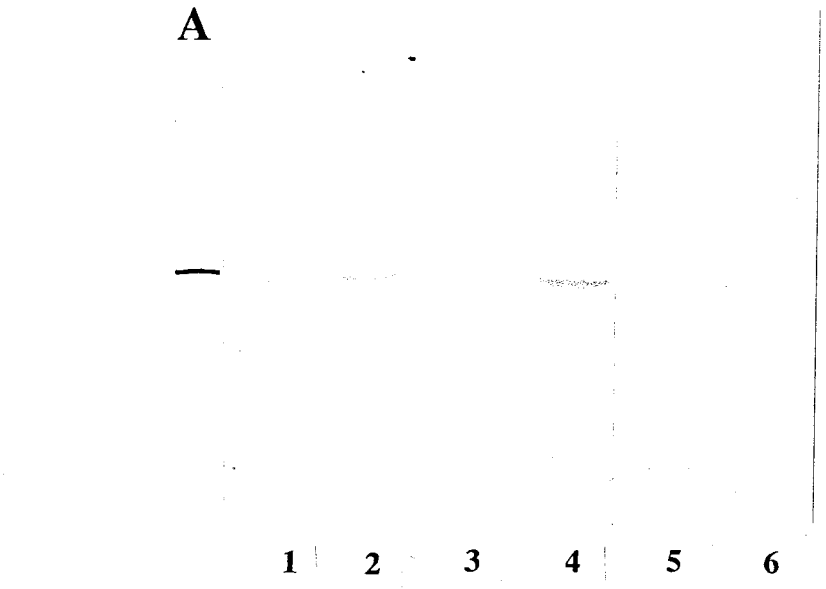
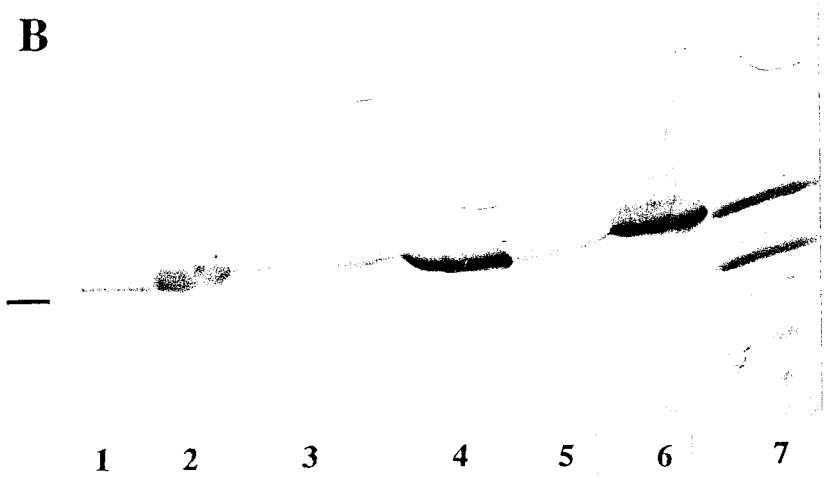

kDal
40
1
2
3
4
5

Figure 70. Partial purification of $rT_r\alpha$ and $rT_r\alpha$. $rT_r\alpha$ and $rT_r\alpha$ were partially purified from CAG1139 pML 46 or CAG1139 pML 53, respectively, as described in Materials and Methods. A) Partial purification of $rT_r\alpha$. Lane 1, soluble supernatant; Lane 2, insoluble pellet; Lane 3, 40% ammonium sulfate supernatant; Lane 4, 40% ammonium sulfate pellet; Lane 5, 0.0 M heptyl agarose buffer eluate; Lane 6, transducin buffer eluate. B) Partial purification of $rT_c\alpha$. Lane 1, soluble supernatant; Lane 2, insoluble pellet; Lane 3, 40% ammonium sulfate supernatant; Lane 4, 40% ammonium sulfate pellet; Lane 5, 0.5 M heptyl agarose buffer eluate; Lane 6, 0.0 M heptyl agarose buffer eluate; Lane 7, transducin buffer eluate.


A
1
2
3
4
5
6


B
1
2
3
4
5
6
7

CONCLUSION

In the past G proteins have been studied primarily by using classical biochemical techniques. This thesis demonstrates that various tools of molecular biology can also be used to study G proteins. Molecular cloning of the $T_c\alpha$ cDNA revealed the first complete primary sequence of a G protein α subunit. The cloning of the $T_c\alpha$ cDNA was significant in several respects. It provided a precise determination of the relationship of G proteins to the *ras* family of GTP-binding protein which at that time was unclear. It allowed identification of some of the GTP-binding regions of the protein. It led to the prediction that G protein α subunits would be myristoylated, which now has been proven true. The discovery that $T_c\alpha$ was expressed in cones and not rods led to the ideasthat the G protein family might be larger and more diverse than expected. This has also been found to be the case. Prior to the beginning of this thesis not even a single amino acid of any G protein subunit was known. Now there have been over 75 published reports of their sequences. Clearly, knowledge of the complex structure of biological molecules that are under study is a key element in a complete understanding of their function.

The existence of $T_c\alpha$ raised a number of questions. How is highly restricted cell-type specific expression of a G protein attained? What might the effect of mutation in a G protein be on the physiology of an organism? How does G protein structure correlate to function? Using the tools of molecular biology this thesis has developed systems for addressing each of these issues.

Molecular cloning revealed that G proteins exist in lower eukaryotes such as *Caenorhabditis elegans*. Currently it appears that G proteins will be confined to eukaryotes but will be found in all eukaryotes including plants. However G proteins genes were not identified in *Caenorhabditis elegans* solely to add another sequence to a comparison table.

In fact, I believe that the information gained from tabulation of G protein sequences is reaching the point of saturation. Rather, G protein genes were characterized in *Caenorhabditis elegans* so that powerful genetic tools could be used to understand G protein function. Since some of the G proteins found in *Caenorhabditis elegans* are the same as those found in mammals, this system may provide a better understanding of G proteins found in more complex organisms. In particular it may allow a detailed assessment of the role of G proteins in development. Such "surrogate organism" approaches have been successful in many other cases.

Using the Northern blotting technique,retinoblastoma cells were identified that express cone photoreceptor-specific genes. These cells may be useful for studying the promoter elements necessary to achieve expression specifically in cone cells. They could also be used to test expression of cone-specific promoters prior to making the committment of introduction fusion genes containing those promoters into transgenic mice.

The technique of heterologous expression was used to develop a system for studying in depth the relationship of G protein structure to function through a mutagenesis approach. Information obtained from *in vitro* reconstitution of signal transduction systems will complement information obtained from *C. elegans* or transgenic mice.

Together genetics, biochemistry, and molecular biology should ultimately provide a comprehensive picture of the role of G proteins in the life of an organism.