Homologous and unique G protein alpha subunits in the nematode Caenorhabditis elegans

Michael A. Lochrie,*† Jane E. Mendel,*‡ Paul W. Sternberg,* † and Melvin I. Simon*§
*Howard Hughes Medical Institute
†Division of Biology
California Institute of Technology
Pasadena, California 91125

A cDNA corresponding to a known G protein alpha subunit, the alpha subunit of Gα (Gα,α), was isolated and sequenced. The predicted amino acid sequence of C. elegans Gα,α is 80–87% identical to other Gα,α sequences. An mRNA that hybridizes to the C. elegans Gα,α cDNA can be detected on Northern blots. A C. elegans protein that crossreacts with an antibovine Gα,α antibody can be detected on immunoblots. A cosmid clone containing the C. elegans Gα,α gene (goa-1) was isolated and mapped to chromosome I. The genomic fragments of three other C. elegans G protein alpha subunit genes (gpa-1, gpa-2, and gpa-3) have been isolated using the polymerase chain reaction. The corresponding cosmids were isolated and mapped to disperse locations on chromosome V. The sequences of two of the genes, gpa-1 and gpa-3, were determined. The predicted amino acid sequences of gpa-1 and gpa-3 are only 48% identical to each other. Therefore, they are likely to have distinct functions. In addition they are not homologous enough to G protein alpha subunits in other organisms to be classified. Thus C. elegans has G proteins that are identifiable homologues of mammalian G proteins as well as G proteins that appear to be unique to C. elegans. Study of identifiable G proteins in C. elegans may result in a further understanding of their function in other organisms, whereas study of the novel G proteins may provide an understanding of unique aspects of nematode physiology.

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, phospholipases, 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). 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 (Jones et al., 1990; Mumby et al., 1990). G protein α subunits are also phosphorylated (Gunderson and Devreotes, 1990; reviewed by Sagi-Eisenberg, 1989).

G proteins are members of a diverse family of proteins with a variety of functions and tissue distributions. Gα, 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. Gα 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), human (Lau et al., 1988), and Xenopus (Olate et al., 1989) but not in yeast or the slime mold Dictostelium. In vertebrates, Gα is concentrated in the neuropl (Gabrion et al., 1989; Worley et al., 1986) and growth cones (Strittmatter et al., 1990). In Drosophila, Gα has been found in brain and ovaries (deSousa et al., 1989; Schmidt et al., 1989; Thambi et al., 1989; Yoon et al., 1989). Gα may function in several signal transduction pathways (Serventi et al., 1990). In many cases

© 1991 by The American Society for Cell Biology

135
it appears to be able to interact with the same receptors as the closely related G proteins, although there are some exceptions (Senogles et al., 1990; Ueda et al., 1990). However, Gs may regulate a set of effectors distinct from those regulated by the Gi proteins. The Gi proteins regulate adenylate cyclase and atrial potassium channels, whereas there is evidence that Gs 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). Gs has no direct effect on adenylate cyclase activity.

To understand the role of G proteins in 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. C. elegans is a model organism for studying development and the function of the nervous system (Kenyon, 1988; Wood, 1988). It is the only multicellular organism for which a complete cell lineage is known ( Sulston, 1988; Sulston et al., 1988). The hermaphrodite contains 959 somatic nuclei; of these, 302 are neuronal. The entire nervous system and neuronal wiring diagram is known (White et al., 1986, 1988; Chalfie and White, 1988). C. elegans exhibits many behaviors (reviewed by Chalfie and White, 1988). It has many of the neurotransmitters and signaling systems found in coelomates (Chalfie and White, 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 facile (Fire, 1986; C. Mello, V. Ambros, J. Kramer, and D. Stinchcomb, personal communication) and an ordered set of cosmids clones covering most of the genome is available to facilitate gene cloning and mapping (Coulson et al., 1986, 1988).

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 et al., 1990) and an α subunit gene (gpa-2; Fino Silva and Plasterk, 1990) has been reported. This paper reports the molecular cloning and sequencing of a C. elegans Gαα protein, cDNA and the genes encoding the two apparently novel C. elegans G protein α subunits (gpa-1 and gpa-3). These cloned genes can provide tools to study their role in C. elegans biology.

**Results**

**Polymerase chain reactions**

To isolate C. elegans G protein α subunit genes, we used short, degenerate oligonucleotide primers in polymerase chain reactions on genomic DNA. The primers correspond to three closely spaced amino acid sequences that are highly conserved in G protein α subunits. They represent three of the six regions that are thought to be involved in guanine nucleotide binding and, therefore, would be expected to be present in most G protein alpha subunit sequences. The distances between the oMP19, oMP20/oMP21 and the oMP19, Ta29 primer pairs in a cDNA would be ~200 and 375 base pairs, respectively. Because genomic DNA was amplified in this study, the PCR products could be larger if they contain introns. PCR amplification of C. elegans genomic DNA with the oMP19 and oMP20 primer pair produced PCR products of 270, 525, and 1000 base pairs. Amplification using oMP19 and oMP21 produced PCR products of 260, 300, and 550 base pairs. Amplification using oMP19 and Ta29 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 Ta29 reaction, were subcloned. The sequences of several subclones of each PCR product were analyzed. The sequences fell into three classes designated as gpa-1, gpa-2, and gpa-3. These sequences contained 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 the oMP19 and Ta29 primers is an artifact of the PCR. Most of the subclones of this PCR product had truncated gpa-1 or gpa-3 sequences fused to unknown sequences. Two were gpa-3/gpa-1 or gpa-2/gpa-1 hybrid sequences. Out of a total 85 subclones that were sequenced, 86% were gpa-1, gpa-2, or gpa-3 sequences. All three sequences contained introns. The positions of these introns were identical to those of the sixth and seventh introns found in the mammalian Gαα, Gβγ, and Tαα genes (Kaziro et al., 1988; Raport et al., 1989).

**Gαα cDNA characterization**

A mixture of the three PCR-derived gene fragments was used as a probe to screen a C. ele-
Gans cDNA library at low stringency. Fifty-two of 45,000 plaques hybridized to the PCR-derived probes. Seven of these clones tested positive by PCR analysis for the presence of G protein α subunit related sequences using the primer pairs oMP19, oMP20/21 or oMP19, Ta29. 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. The open reading frame found in this sequence encodes a protein that is clearly homologous to Goα sequences identified in other organisms. The predicted amino acid sequence of C. elegans Goα is 87% identical to Drosophila Goα, 80% identical to Xenopus Goα, and 81–82% identical to human, rat, mouse, hamster, or bovine Goα (Figure 3). Two different forms of Goα, called Goα-A and Goα-B, have been described from hamster (Hsu et al., 1990) and mouse (Strathmann et al., 1990). They differ in their sixth and seventh exons as a result of alternative splicing patterns. However, the C. elegans Goα corresponds to neither form of Goα because it is as different from Goα-A as it is from Goα-B. All of the motifs involved in GTP binding and hydrolysis (Lochrie and Simon, 1988) are highly conserved in the C. elegans Goα sequence. A site for myristoylation on rat Goα, Glycine-2 (Mumby et al., 1990), is present. Amino acids that are substrates for cholera toxin (Arg-179) and pertussis toxin (Cys-351) are also found in C. elegans Goα, as they are in other Goα proteins. The region of C. elegans Goα that is most different from other Goα sequences is in the region of amino acids 90–140 and 290–320. This divergence would be expected because this region is also the most variable between any other pair of G protein α subunits.

No sequences resembling the SL1 or SL2 spliced leader sequences, which are spliced posttranscriptionally onto certain C. elegans mRNAs (Krause and Hirsh, 1987; Huang and Hirsh, 1989), were found at the 5' end of the Goα cDNA. However, the Goα cDNA sequence in Figure 2 is probably not full length, because the size of the mRNA detected on a Northern blot is ~2.2 kilobase pairs (Figure 4). Thus, excluding the poly-A" tail, the sequence may be missing ~400 base pairs. The results of Southern blot analysis indicate that the C. elegans Goα gene is present in a single copy in the C. elegans genome (data not shown).

To identify a candidate protein encoded by the C. elegans Goα cDNA, we performed an immunoblot analysis with a heterologous antibody. A protein of 40 kDa can be detected on immuno- blots of total C. elegans extracts with an affinity-purified antibody to bovine Goα (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 that produces ~30% males. Thus Goα is probably not male-specific. 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 (data not shown). This sequence is found as the last 10 amino acids of all Goα proteins except Xenopus Goα and rodent Goα-B.

**Isolation and mapping of cosmid clones**

The PCR-derived fragments of gpa-1, gpa-2, and gpa-3 and the Goα cDNA insert were used as

---

**Figure 1. Restriction map and DNA sequencing strategy of C. elegans Goα cDNA.** The C. elegans Goα cDNA in λCe6-2 is represented as a horizontal line. The Goα 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, BamHI; Bc, Bcl I; Cl, Cla I; E, EcoRI; H2, HincII; Nc, Nco I; Ns, Nsi I; S, Sph I; T, Taq I. Other abbreviations used are: N, amino terminus; C, carboxyl terminus.
Figure 2. Sequence of the C. elegans Gα, cDNA. Amino acid abbreviations used are Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Gin, 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.

probes to isolate the corresponding cosmid clones. The number of cosmids that hybridized to the gpa-1, gpa-2, gpa-3, and Gα, probes was three, one, three, and one, respectively. The cosmids that contain the gpa-1, gpa-2, gpa-3, and Gα (goa-1) genes mapped to unique po-
Gene sequences of gpa-1 and gpa-3

The restriction map, intron/exon organization, and DNA sequencing strategy of the gpa-1 and gpa-3 genes are 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 seven introns and the gpa-3 coding region is divided by five introns. Intron boundaries were 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 C. elegans consensus splice junction sequences (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’ TTTTCAG 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 3’ junctions in the gpa-3 gene. The sequence TTTCAG 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 during mRNA splicing—is also found in most of the putative introns in the gpa-1 and gpa-3 genes, with the exception of 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 >65% AT except for introns 1, 5, and 7 in the gpa-1 gene. 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).

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 in the gpa-1, gpa-2, and gpa-3 genes are conserved relative to the mammalian Gα, Gα, and Tα genes (Itoh et al., 1988; Kaziro et al., 1988; Weinstein et al., 1988; Rapport et al., 1989). Those 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 near amino acid sequences that are well conserved (i.e., those involved in guanine
nucleotide binding). The introns in the \textit{gpa-1}, \textit{gpa-2}, and \textit{gpa-3} genes that are extra or missing relative to the mammalian G\textsubscript{a}, G\textsubscript{G\alpha}, and T\textsubscript{a} genes occur within the region that is most variable between G protein alpha subunits.

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 (1447 base pairs) and the gpa-2 gene (1020 base pairs) are also large (Fino Silva and Plasterk, 1990; Van der Voorn et al., 1990). This is unusual because in C. elegans most of the introns are ~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 ~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 2012–2350 within the fifth intron. The elements of the repeat are ~95% identical. The repeats are ~330 base pairs long and are ~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 consensus CAAT-box, TATA-box, and transcription initiation site sequences (Breathnach and Chambon, 1981). The gpa-1 gene has sequences at positions 50–55 (GCTAAT), 86–92 (TATATA), and 113 that conform well to the spacing constraints and consensus sequences for the CAAT-box, TATA-box, and transcription initiation site. In addition, a consensus 3' splice junction acceptor site (TTTCAG) is found immediately preceding the initiation codon. This site may be where a trans-splicing reaction 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 (Fino Silva and Plasterk, 1990).

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 to 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 38–43 (TTTATG) 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 (TTTTCAG), 253–257 (TTTAG), 320–326 (TTTTTAC), 694–700 (TTTTTAT), and 730–736 (TTACAG). However, no sequences resembling the consensus 5' splice junction sequence are found within this region.

The sequences 5' to the initiation codon were compared with 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

---

**Figure 5. Immunoblot analysis of C. elegans Gα protein.**

An affinity-purified rabbit anti-bovine Gαα antisera was used at a dilution of 1:1000 to detect Gαα proteins as described in Methods. Lane 1, 1 μg purified bovine Gαα; Lane 2, 50 μg total protein from C. elegans N2; Lane 3, 50 μg total protein from C. elegans him-5 (e1490).
of the gpa-1, gpa-2, gpa-3, and gpb-1 genes did not reveal any striking homologies. Each 5' region has several repeated sequences, but none of these are common to any pair of sequences. 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 with 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 (Proudfoot and Brownlee, 1979; Wickens and Stephenson, 1984). This sequence was found in the 3' region of the gpa-1 gene at positions 3340–3345. It was not found in the gpa-3 3' region, although minor variants that might be functional are found at positions 2825–2830 (AATACA) and 2891–2896 (AATATA).

The predicted amino acid sequences of the gpa-1, gpa-2, and gpa-3 proteins are compared in Figure 11. The percent sequence identity observed between them is gpa-1 versus gpa-2: 41%, gpa-1 versus gpa-3: 48%, and gpa-2 versus gpa-3: 58%. The gpa-1, gpa-2, and gpa-3 proteins are 49%, 46%, and 53% identical to the C. elegans G_α sequence, respectively. All three gpa proteins have the six regions implicated in binding guanine nucleotides (Lochrie and Simon, 1988). Also, all three have the consensus myristoylation signal Met-Gly-X-X-Ser at the amino terminus (Towler et al., 1988). Moreover, 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 (Fino 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.

Discussion

We have determined the sequence of a cDNA encoding a C. elegans G_α protein that is very...
similar to that of Goalpha proteins found in other species. Therefore, it is likely that C. elegans Goalpha is similar in function to other Goalpha proteins. In vertebrates and Drosophila, Goalpha is found pre-dominantly in neurons. However, its role in signal transduction pathways is not well delineated. Thus, the presence of a Goalpha homologue in C. elegans may provide new approaches for understanding its function in other organisms.

The results of Southern blot analysis indicate 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 Goalpha cDNA as a probe. It is, therefore, likely that the protein detected on the immunoblot is the same as that encoded by the Goalpha cDNA described here and not some other closely related G protein alpha subunit. However, the possibility that there may be multiple Goalpha proteins in C. elegans cannot be excluded until the genomic sequence of goa-1 has been determined. Multiple species of Goalpha, which arise as a result of alternative mRNA splicing patterns, have been found in Drosophila (deSouza et al., 1989; Thambi et al., 1989; Yoon et al., 1989) and rodents (Hsu et al., 1990; Strathmann et al., 1990).

In contrast to C. elegans Goalpha, the gpa-1, gpa-2, and gpa-3 proteins cannot be classified according to their relatedness to other G protein alpha subunits. They are 40–50% identical to any other non-C. elegans G protein alpha subunit sequence. G protein alpha subunits that are >80% identical are generally considered to be in the same class (Lochrie and Simon, 1988) and can activate the same effector proteins (Gillespie and Beavo, 1988; Yatani et al., 1988). G protein alpha subunits that are as little as 60–70% identical (e.g., Goalpha, Goalpha, and Talpha) 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 alpha subunits that are 40–60% identical have not been observed to interact with the same receptor or effector and are generally considered to be distinct in function. Therefore, gpa-1, gpa-2, and gpa-3 are likely to have separate functions. This divergence does not exclude the possibility that they may operate in signaling systems similar to those in mammals, because the receptors and effectors that gpa-1, gpa-2, and gpa-3 interact with may also be different in sequence from their mammalian counterparts.

Although the sequences of the gpa-1, gpa-2, and gpa-3 proteins provide no clues about their function, several approaches are available in C. elegans for investigating their function. One approach initiated in this study is the mapping of genes near known mutations with visible phenotypes, followed by a determination of which of these mutations, if any, correspond to mutations in G protein alpha subunit genes. This determination 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 existing mutations will be in a G protein gene, because the genetic map is not saturated with mutations. If so, there are a number of strategies to identify mutant phenotypes for these genes. For example, recessive mutations might be isolated by a PCR-based screen for transposon insertion mutations, using methods similar to those developed in Drosophila (Ballinger and Benzer, 1989; Kaiser and Goodwin, 1990). Dominant mutations can be generated by transformation with mutants of the gpa genes constructed in vitro. Based on studies of mutant G protein alpha subunits found in human tumor cells (Landis et al., 1989) and the biochemical properties of G protein alpha subunits synthesized in E. coli (Freissmuth and Gilman, 1989; Graziano and Gilman, 1989) and mammalian cells (Masters et al., 1989), certain amino acid changes in the C. elegans G protein alpha subunit sequences might be expected to result in a dominant phenotype. A determination of gpa gene expression patterns could also provide information about the function of the gpa proteins. However, our initial attempts to detect gpa gene expression by Northern blotting were unsuccessful. In addition, we were unable to isolate cDNA clones for gpa-1, gpa-2, or gpa-3.
Figure 7. DNA sequence of *C. elegans* gp-1 gene. The predicted amino acid sequence of the gp-1 protein is shown below the DNA sequence. The standard three-letter amino acid abbreviations are given in the legend to Figure 2. Coding sequences are shown in boldface lettering. The primers used in PCR reactions are underlined. oMP19 corresponds to nucleotide positions 2564-2581, oMP20/oMP21 to positions 2791-2808, and Ta29 to positions 3140-3157.
differences could have a profound effect on their biological properties. There may be G protein genes in *C. elegans* other than those reported to date. A *C. elegans* homologue of the mammalian Gsa gene (gsa-1) has been isolated (S. Ohshima, S. Tashiro, J.-H. Park, T. Tani, and Y. Ohshima, personal communication). This gene could not have been iso-
In conclusion, C. elegans has at least five G proteins. Two of these (Gα and Gβ) are identifiable homologues 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 Gα and Gβ. However, unicellular organisms, such as yeast and Dictyostelium, do not have Gα and Gβ.

Instead they have G proteins with novel sequences. Therefore, it may be that G proteins such as Gα and Gβ arose during the development of multicellular organisms and have been highly conserved across phylogeny because they have fundamental roles in intercellular signaling. On the other hand, the G proteins such as gpa-1, gpa-2, and gpa-3 that are not well conserved may have functions specific to a particular organism.

Methods

Nematodes

The wild-type hermaphrodite N2 (Bristol) strain is from Brenner (1974) and him-5 (e1490) is from Hodgkin et al. (1979).

Polymerase chain reactions (PCRs)

Twenty-five nanograms of C. elegans genomic DNA from the hermaphrodite wild-type N2 strain were amplified in a 10-μl reaction containing 50 mM KCl, 1.5 mM MgCl₂, 10 mM tris(hydroxymethyl)aminomethane (Tris)-HCl, pH 8.3, 0.01% gelatin, 200 μM deoxyribonucleotides (dATP, dCTP, dGTP, and dTTP), 2.5 U Taq polymerase (Perkin Elmer-Cetus, Norwalk, CT), 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

Figure 8. Restriction map, DNA sequencing strategy, and gene organization of C. elegans gpa-3 gene. (A) Restriction map of 3 kbp HindIII 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 NsiI 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, NsiI; H2, HindIII; H3, HindIII; 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, carboxy terminus.
strand that encodes the amino acid sequence Phe-Leu-Asn-Lys-Lys-Asp. The sequence of Ta29 is GAATTC(GATC)GT(GATC)GC(AG)CA(GATC)GT. This sequence contains an EcoRI restriction site at one end to facilitate subcloning and corresponds to the antisense strand that encodes the amino acid sequence Thr-Cys-Ala-Thr. Thirty amplification cycles consisting of 1 min at 92°C, 30 s at 37°C, and 1 min at 72°C were performed. At the end of the amplification cycles, the reactions were incubated another 10 min at 72°C to complete partially extended chains. The products of the reaction were analyzed on 3% NuSieve (FMC, Rockland, ME) agarose gels. The PCR products were excised from the gel, reamplified in a 100-μl reaction using the same primers, made blunt with Klenow, phosphorylated with T4 polynucleotide kinase, ligated to linkers, and subcloned (Maniatis et al., 1982) into pBluescript KS−/− (Stratagene, Burlingame, CA). The DNA sequences of the subclones were determined as described below.

**Isolation of Gαx cdNA**

A *C. elegans* cDNA library containing ~250 000 clones was constructed in λgt10 and kindly supplied by Stuart Kim (Stanford Univ.). The mRNA used to construct this library represented all developmental stages of the hermaphroditic wild-type N2 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× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0, 0.1% sodium dodecyl sulfate (SDS), 30% formamide, 100 μg/ml sonicated salmon sperm DNA, 5× Denhardt’s solution (1× Denhardt’s solution is 0.02% Ficoll 400, 0.02% bovine serum albumin (BSA), 0.02% polyvinylpyrrolidone) at 37°C for 2 h. Probes were added to the prehybridization solution and hybridized under the same conditions for 24 h. 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, Arlington Heights, IL) using α-32P-dCTP and hybridized to the nitrocellulose filter replicas in 1% BSA, 7% SDS, 1 mM EDTA, 0.5% formamide, 0.2 M NaPO4, pH 7.2 at 65°C for 12 h using 4× 105 cpm/ml of probe. The filters were washed in 2× SSC at 50°C for 30 min, 0.2× SSC at 50°C for 30 min, 0.2× SSC at 65°C for 30 min and then exposed to X-ray film. Positive 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 pattern of HindIII Sau3A1 restriction fragments of a cosmid to that of an overlapping set of cosmids (contigs) that now represents >95% of the *C. elegans* genome (Coulson et al., 1986; A. Coulson, J. Sulston, and R. Waterston, personal communication).

**Northern blot analysis**

Total RNA was isolated by the guanidine isothiocyanate procedure described by Maniatis et al. (1982) from the N2 strain. The RNA was denatured, fractionated on a 1% formaldehyde gel, and blotted to a nitrocellulose filter. The insert from the Gαx cdNA was isotopically labeled using α-32P-dCTP by the method of Feinberg and Vogelstein (1983). The filter was hybridized to the probe in 50% formamide, 5× SSC, 5× Denhardt’s solution, 25 mM NaPO4, 0.1% SDS, and 0.25 mg/ml salmon sperm DNA at 42°C. The filter was washed in 0.2× SSC, 0.1% SDS at 60°C and exposed to X-ray film.

**Immunoblot analysis**

Mixed populations of N2 nematodes at various developmental stages were grown on 10-cm NGM plates (Sulston and Hodgkin, 1988) and washed off with M9 buffer. The nematodes were pelleted and 5× concentrated gel sample buffer (Laemmli, 1970) was added to a final concentration of 1×. The nematodes were boiled for 5 min and the dissolved proteins were fractionated by SDS electrophoresis 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 h. The filter was blocked in TBS (20 mM Tris-HCl, pH 7.5, 500 mM NaCl/3% BSA for 30 min and incubated in the primary antibody for 1 h at room temperature in TBS/1% BSA. The filter was washed with TBS for 30 min and then incubated in the secondary antibody for 1 h at room temperature. Goat anti-rabbit IgG (Fc) conjugated to alkaline phosphatase (Promega, Madison, WI) was used as the secondary antibody at a dilution of 1:7500 in TBS/1% BSA. After washing in TBS for 10 min, we developed the blot with nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) according to the manufacturer’s instructions.

**Isolation and mapping of cosmid clones**

A cosmid library was constructed from N2 (Bristol) genomic DNA that was partially digested with XhoI II. The DNA was size-fractionated by field inversion gel electrophoresis and subcloned into the BarnHI site of plWKeAn (J. Mendel, unpublished). The resulting cosmid library contains about three or four 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 with a Multiprime kit (Amersham) using α-32P-dCTP and hybridized to the nitrocellulose filter replicas in 1% BSA, 7% SDS, 1 mM EDTA, 0.5% formamide, 0.2 M NaPO4, pH 7.2 at 65°C for 12 h using 4× 105 cpm/ml of probe. The filters were washed in 2× SSC at 50°C for 30 min, 0.2× SSC at 50°C for 30 min, 0.2× SSC at 65°C for 30 min and then exposed to X-ray film. Positive 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 pattern of HindIII Sau3A1 restriction fragments of a cosmid to that of an overlapping set of cosmids (contigs) that now represents >95% of the *C. elegans* genome (Coulson et al., 1986; A. Coulson, J. Sulston, and R. Waterston, personal communication).

**DNA sequence analysis**

DNA sequencing reactions were performed using the Sequenase, version 2.0 kit according to the instructions provided by the manufacturer (United States Biochemical, Cleveland, OH). 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 h 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) were added. After incubation at room temperature for 1 min, the mixture was boiled for 2 min and immediately microfuged at room temperature for 5 min. 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 min. The supernatant was aspirated and the pellet was washed in 1 ml 70% ethanol, 1 ml 100% ethanol, then dried in a Spin-Vac. The pellet was resuspended in 50 μl H2O. The plasmid DNA was denatured

Vol. 2, February 1991
Figure 9
Ga

Gin C. elegans

ATTAATGTATTATTGAACATTTTCAGCTGAATCGAACATCTTTTATCACTCATTATTATG

ATATTTGTTCATTTTCAATGCGTCACAGCTTGGTGTCGTTTTGTTGAATTGGCAGCACCA

TTTCTTGTGAAAAATGCTTGCTTGCTTCTTATTATGGAGATCTCGTTAATCGCGCTTGGT

TTTCTTTCCTTTATTTTCTGTTTGAAACTTCTGCGATATACGCGGTCTGCT

TTCATTTTTCATTTTTGTTTTATTTTCAACATTCATTTTCAACAT

MetGlyLeuCysGlnSerAla

GAGGACAAAGAAGTACGAGTTGAAATCAAAGCAGATCGATAGGAATGGAACATCTTTTAT

GluAspLysGluLeuThrLeuLysSerLysAlaIleAspLysGluMetMetGlnAsnHis

ATGTCACAGCAGAGTTGAGGCTTCTGCTTTTTGGGTAAATTTAAAACG

MetSerGlnGlnLysValValLysGlnGlnGlnValValLeuLeuLeuLeuG

GGAGCCGAGCTCCGACGAGCTAGAATATGGGTAATCTGTCATATGACTCGAGAAGG

GluAspLysGluLeuThrLeuLysSerLysAlaIleAspLysGluMetMetGlnAsnHis

ATGTCACAGCAGAGTTGAGGCTTCTGCTTTTTGGGTAAATTTAAAACG

MetSerGlnGlnLysValValLysGlnGlnGlnValValLeuLeuLeuLeuG

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 legend to Figure 2. Coding sequences are shown in boldface lettering. The primers used in PCR reactions are underlined. oMP19 corresponds to nucleotide positions 2235-2252, oMP20/oMP21 to positions 2460-2477, and Tα29 to positions 2674-2691.
with alkali by adding 5 µl 2 M NaOH and 2 mM EDTA and incubating at room temperature for 5 min. 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 min, and microfuging for 10 min 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 H2O. For DNA sequencing reactions, 7 µl denatured template was annealed to 1 µl (10 ng) primer with 2 µl 5X Sequenase buffer at 37°C for 20 min. For DNA sequence analysis of the Gα cDNA, restriction fragments of λC6E2 were subcloned into pBluescript KS/−. Most of the sequence of these inserts was determined using sequencing primers that are complementary to pBluescript KS/−. The remaining sequence was determined using synthetic oligonucleotides as insert specific sequencing primers. A T7 DNA polymerase stop obscuring the sequence of positions 171–181 was resolved by the use of the TaqTrack DNA sequencing system (Promega) at a reaction temperature of 80°C. For DNA sequence analysis of the gpa-1 and gpa-3 genes, HindIII restriction fragments of the gpa-1 and gpa-3 cosmids that hybridize to the gpa-1 and gpa-3 PCR-derived probes were identified by Southern blotting and subcloned into pBluescript KS/− (see Figures 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 HindIII 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 (Strathmann et al., 1991). The location of the γδ transposons was determined by PCR, and DNA sequences were determined by the use of transposon-specific primers corresponding to each of the different ends of γδ. Because 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; T7 polymerase stops were resolved by the use of the TaqTrack system 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.

Acknowledgments

We thank Mike Strathmann for advice on PCR reactions and for the provision of materials for the γδ transposon-
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 3. Amino acids that are found in all three sequences at the same position are boxed.

Based sequencing method, the Caltech Microchemical Facility for oligonucleotide synthesis, Andrea Holboke for C. elegans genomic DNA, Alan Coulson and John Sulston for gene mapping data, Eva Neer and Graeme Milligan for Gα antibodies, Stuart Kim and Russell Hill for the C. elegans cDNA library, and Yasumi Ohshima and Ron Plasterk for generous communication of results before publication. P.W.S. is an investigator of the Howard Hughes Medical Institute. Supported by a Developmental Biology Grant from the Lucille P. Markey Foundation and by USPHS grants GM34236 to M.I.S. and HD23890 to P.W.S.

Received: October 5, 1990.
Revised and accepted: November 16, 1990.

References


Cerione, R.A., Regan, I.W., Nakata, H., Codina, J., Benovic, J.L., Gierschik, P., Somers, R., Spiegel, A.M., Birnbaumer,


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 protein, G\textsubscript{1} and


