MOLECULAR AND CELLULAR BIOLOGY, Mar. 1990, p. 910–917
0270-7306/90/030910-08$02.00/0

Vol. 10, No. 3

Two Forms of *Drosophila melanogaster* Gsα Are Produced by Alternate Splicing Involving an Unusual Splice Site

FRANKLIN QUAN* AND MICHAEL A. FORTE

Vollum Institute for Advanced Biomedical Research, Oregon Health Sciences University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97201-3098

Received 26 September 1989/Accepted 5 December 1989

G proteins are responsible for modulating the activity of intracellular effector systems in response to receptor activation. The stimulatory G protein Gs is responsible for activation of adenylate cyclase in response to a variety of hormonal signals. In this report, we describe the structure of the gene for the α subunit of *Drosophila melanogaster* Gs. The gene is approximately 4.5 kilobases long and is divided into nine exons. The exon-intron structure of the *Drosophila* gene shows substantial similarity to that of the human gene for Gsα. Alternate splicing of intron 7, involving either use of an unusual TG or consensus AG 3′ splice site, results in transcripts which code for either a long (DGsαL) or short (DGsαS) form of Gsα. These subunits differ by inclusion or deletion of three amino acids and substitution of a Ser for a Gly. The two forms of *Drosophila* Gsα differ in a region where no variation in the primary sequence of vertebrate Gsα subunits has been observed. In vitro translation experiments demonstrated that the *Drosophila* subunits migrate anomalously on sodium dodecyl sulfate-polyacrylamide gels with apparent molecular weights of 51,000 and 48,000. Additional Gsα transcript heterogeneity reflects the use of multiple polyadenylation sites.

G proteins are a family of guanine nucleotide-binding proteins involved in the regulation of intracellular effector systems in response to extracellular signals (6, 30). G proteins are heterotrimers consisting of α, β, and γ subunits. The α subunit binds guanine nucleotides and also possesses intrinsic GTPase activity. When the nucleotide-binding site is occupied by GDP, the G protein is an inactive heterotrimer. Upon interaction with the appropriate agonist-receptor complex, the GDP is exchanged for GTP and the α subunit is released from both the receptor and the β-γ subunit complex. The freed α subunit then modulates the activity of the appropriate effector enzymes or ion channels. After GTP hydrolysis, the α subunit reassociates with the β-γ subunit complex to form the inactive heterotrimer. The α subunit is, therefore, responsible for the specific interaction with both the receptor and effector molecules. The β-γ subunit complex has been thought to be responsible for inactivation of the α subunit and anchoring of the G protein to the membrane. Recent studies demonstrating that the β-γ subunit complex is able to open cardiac K^+ channels, possibly by activating membrane-bound phospholipase A2, suggest additional functions for these subunits (12, 15).

The stimulatory G protein Gs was first described as participating in the activation of adenylate cyclase in response to a variety of hormonal signals (β-adrenergic agonists, glucagon, serotonin, etc.). Recently, purified Gs and Gsα subunits have also been shown to activate cardiac and skeletal muscle T-tubule voltage-gated Ca^{2+} channels (20, 33). Thus, Gs, like other classes of G protein (34), is capable of interacting with more than one effector protein.

Vertebrate Gsα proteins are heterogeneous, consisting of two species with apparent molecular weights of 45,000 and 52,000 on sodium dodecyl sulfate (SDS)-polyacrylamide gels (21). This heterogeneity is the result of variations in the primary sequence near the amino termini of the subunits. For example, human tissues contain four species of Gsα mRNA which code for long and short subunits that differ after amino acid residue 70 by inclusion or deletion of 15 amino acids. Additional heterogeneity results from inclusion or deletion of a single serine residue in this region (3). On the basis of the organization of the human Gsα gene, the transcripts that encode the long and short forms of Gsα are thought to arise from the variable use of exon 3. Inclusion or deletion of the codon for the additional Ser residue results from use of either an unusual TG or consensus AG splice site at the 5′ end of exon 4 (14).

The physiological significance of the multiple forms of Gsα is unknown. While only three of the four forms of Gsα have been tested, each appears to be capable of interacting with both adenylate cyclase and voltage-gated Ca^{2+} channels (20). Some quantitative differences among the various forms have been observed in vitro in their interaction with adenylate cyclase, but the significance of this is unknown (7).

The sophisticated genetics and ease of manipulation of *Drosophila melanogaster* make it a particularly attractive system in which to study the role of G proteins in complex biological processes, such as development and behavior (27). As a part of the study of G proteins in this organism, we have recently reported the isolation and sequencing of a cDNA that codes for a Gsα protein preferentially expressed in the *Drosophila* nervous system (24). Southern blot analysis indicated the presence of a single Gsα gene in the *Drosophila* genome. In this report, we present the organization of the gene for *Drosophila* Gsα and demonstrate that multiple species of *Drosophila* Gsα are produced by alternate splicing involving an unusual splice junction. This observation is consistent with the results of Western blotting (immunoblotting) of *Drosophila* neuronal membranes probed with Gsα-specific peptide antisera, indicating the existence of multiple molecular weight forms of Gsα (W. J. Wolfgang, F. Quan, P. Goldsmith, C. Unson, A. Spiegel, and M. Forte, J. Neurosci., in press). Interestingly, *Drosophila* Gsα protein exhibits variability in a region different from that of vertebrate Gsα. Additional variability in the expression of the *Drosophila* Gsα gene results from the use of multiple polyadenylation sites.

* Corresponding author.

MATERIALS AND METHODS

Screening of cDNA and genomic libraries. A genomic library constructed in λ Charon4A from sheared Canton S embryonic DNA (17) was provided by T. Maniatis. An adult *Drosophila* head cDNA library constructed in λ gt11 was provided by P. Salvaterra (City of Hope Research Institute, Duarte, Calif.). Gsα recombinants were isolated by screening libraries with the nick-translated insert of pDGs12. pDGs12 contains a previously described 1.7-kilobase (kb) *Drosophila* Gsα cDNA (24) subcloned into the *Eco*RI site of pBS(M13−)KS (Stratagene). Libraries were plated at a density of 50,000 plaques per 150-mm-diameter dish. Bacteriophage DNA was transferred to nitrocellulose filters (Schleicher & Schuell Inc.) as previously described (16). Hybridizations were done in 50% (vol/vol) formamide–3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–5× Denhardt solution (1× Denhardt solution is 0.02% Ficoll, 0.02% bovin serum albumin, and 0.02% polyvinylpyrrolidone)–50 mM sodium phosphate (pH 6.8)–10% (wt/vol) dextran sulfate at 37°C. Filters were washed in 0.1× SSC–0.5% SDS at 65°C. After plaque purification, phage DNA was prepared from liquid lysates as previously described (16). Restriction fragments were subcloned into pBS(M13−)KS for further analysis.

Nucleotide sequencing. Nucleotide sequencing was done by the dideoxynucleotide method (28) using Sequenase (United States Biochemicals) and [^{35}S]dATP on double-stranded templates (13). Nucleotide sequencing was done by using commercially available primers or *Drosophila* Gsα-specific oligonucleotides synthesized with an Applied Biosystems nucleic acid synthesizer. Oligonucleotide 652 (5′-TCCGGCAAATCAACCAT-3′) corresponds to nucleotides +160 to +176 of the previously reported sequence of pDGs12 (24) and allows synthesis of the sense strand. Oligonucleotide 918 (5′-CGCATGTGAAATGTGGATAG-3′) is complementary to nucleotides +1050 to +1069 and allows synthesis of the antisense strand. To locate exons, genomic restriction fragments were sequenced after generation of nested deletions with exonuclease III (Bethesda Research Laboratories, Inc.) and S1 nuclease (Boehringer Mannheim Biochemicals) as previously described (8).

PCRs. Polymerase chain reactions (PCRs) were done by using oligonucleotides 652 and 918 described above. First-strand cDNA synthesis was performed in 50-μl reactions containing 10 pmol of oligonucleotide 918, 50 mM Tris (pH 8.3), 5 mM $MgCl_2$, 5 mM dithiothreitol, 50 mM KCl, 0.4 M dATP, 0.4 M dCTP, 0.4 M dGTP, 0.4 M dTTP, 20 U of RNasin (Promega Inc.), 5 μg of poly(A)$^+$ RNA from either whole flies or bodies, and 32 U of avian myeloblastosis virus reverse transcriptase (Life Sciences Inc.). Reactions were performed at 42°C for 90 min. PCRs were done in 25-μl volumes and contained 2 μl of the reverse transcription reaction, 20 pmol each of oligonucleotides 652 and 918, 10 mM Tris hydrochloride (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 0.5 M dATP, 0.5 M dCTP, 0.5 M dGTP, 0.5M dTTP, 0.2% W1 detergent (Bethesda Research Laboratories), and 5 U of *Taq* polymerase (Bethesda Research Laboratories). Thirty cycles of denaturation at 95°C for 0.5 min, annealing at 60°C for 1 min, and extension at 72°C for 1 min were followed by a final extension step at 72°C for 3 min. The products of the reaction were analyzed by agarose gel electrophoresis and Southern blot hybridization with pDGs12 as described above.

In vitro transcription and translations. cDNAs that code for the long and short forms of *Drosophila* Gsα were subcloned into the *Eco*RI site of pBS(M13−)KS to yield pDGs4 and pDGs9, respectively. pDGs4 and pDGs9 begin at residues −213 and −216, respectively, relative to the full-length cDNAs (pDGs12) and thus do not include the two upstream ATG codons found at positions −242 and −252 (24). Sense transcripts were synthesized by using T3 RNA polymerase (Bethesda Research Laboratories) in the absence of the labeled nucleotide and cap. After removal of the template with RQI DNaseI (Promega), the RNA was extracted twice with phenol-chloroform, ethanol precipitated, and suspended in DEPC-treated H_2O. Translations were performed in rabbit reticulocyte lysates (Promega Biotech) in [^{35}S]methionine for 1 h at 30°C. Labeled proteins were visualized by autoradiography after separation on 10% acrylamide–0.13% bisacrylamide–SDS gels.

RNase protection mapping. Total RNA was prepared from whole fly heads and bodies, and poly(A)$^+$ RNA was selected by oligo(dT) cellulose chromatography (Boehringer Mannheim) (16). Antisense RNA transcripts were synthesized by using a kit purchased from Promega and [α-^{32}P]UTP. To examine alternate splicing at the 3′ splice site of intron 7, a 166-base-pair (bp) *Bgl*II-*Ssp*I fragment spanning this splice site (see Fig. 3) was subcloned into the *Bam*HI-*Hin*cII sites of pBS(M13−)KS and antisense transcripts were synthesized by using T3 RNA polymerase. Approximately 10^5 cpm of labeled RNA was precipitated with 10 μg of poly(A)$^+$ RNA and suspended in 30 μl of 80% (vol/vol) formamide–0.4 M NaCl–40 mM PIPES piperazine-*N*,*N*′-bis(2-ethanesulfonic acid; pH 6.4)–1 mM EDTA. Hybridizations were done at 37°C overnight after denaturation at 85°C for 5 min. RNase A and T_1 digestions were done as previously described at 30°C (1). The products were analyzed on an 8% polyacrylamide gel containing 7 M urea.

RESULTS

Characterization of *Drosophila* Gsα genomic clones. A *Drosophila* genomic library in λ Charon4A (17) was screened with the previously isolated 1.7-kb *Drosophila* Gsα cDNA insert of pDGs12 (24). Approximately 28 kb of genomic DNA was isolated in six overlapping recombinants (Fig. 1). After preliminary restriction analysis, overlapping fragments were subcloned into pBS(M13−)KS for further analysis. Genomic restriction fragments containing Gsα sequences were identified by Southern analysis, and exons were located by nucleotide sequencing and by comparison to previously sequenced cDNAs (24). The *Drosophila* Gsα gene is approximately 4.5 kb long and is split into nine exons by eight introns (Fig. 1). Introns range in size from 56 bp to approximately 1.4 kb, the largest intron separating exons 1 and 2.

The nucleotide sequence of the *Drosophila* Gsα gene, excluding that of introns 1 and 4, is shown in Fig. 2. With two exceptions, the sequence of the exons is identical to that of the cDNA, pDGs12. Nucleotides −296 and 1263 of the previously reported cDNA sequence (24) are C and T in exons 1 and 8 instead of G and C, respectively. These differences occur in the 5′-untranslated region and position 3 of the codon for Asp-321, respectively. Thus, these changes do not alter the deduced amino acid sequence of the protein. Nucleotide sequencing of a number of cDNA isolates has demonstrated the presence of both C and T at position 1263. This difference, therefore, most likely represents an allelic difference. The G at position −296 may be a cloning artifact as it is only four nucleotides from the 5′ end of the cDNA.

The sequences of the intron-exon junctions, with the

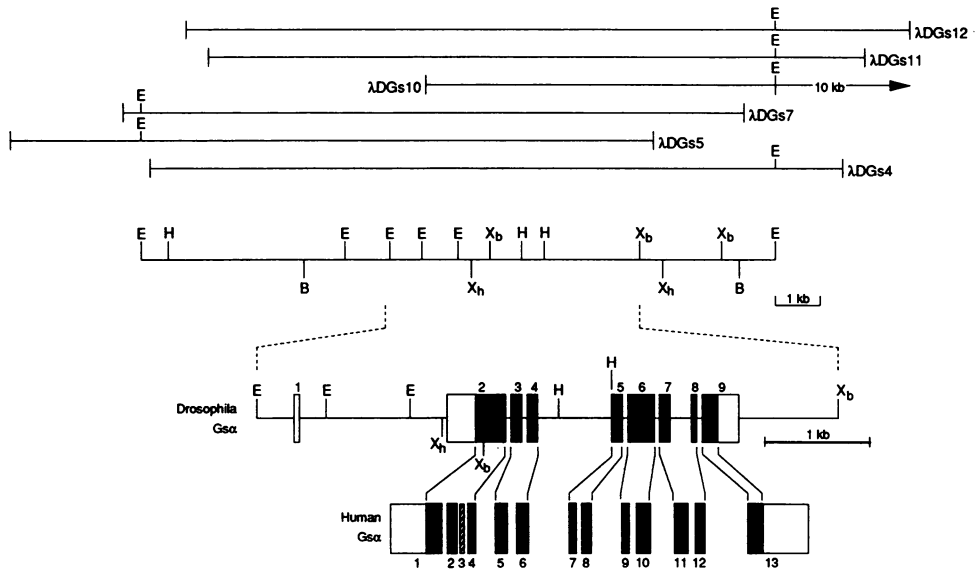


FIG. 1. Structure of the *Drosophila* gene for Gsα and comparison of the exon-intron structure with that of the human gene for Gsα. Exons are shown as boxes. Solid regions indicate amino acid-coding regions. The individual λ recombinants are shown above the restriction map. The restriction endonuclease sites shown are as follows: B, *Bam*HI; E, *Eco*RI, H, *Hin*dIII; X_b, *Xba*I; X_h, *Xho*I. The human gene for Gsα is shown below that of the *Drosophila* gene for Gsα and is not drawn to scale. Solid lines join intron-exon junctions that have been conserved between the two genes.

exception of the 3′ splice site of intron 7 (Fig. 2), are in close agreement with the GT and AG consensus sequences demarcating the 5′ and 3′ boundaries of introns (22). On the basis of the sequence of pDGs12, intron 7 appears to use an unusual 3′ splice site (Fig. 2). The sequence of this site, TG, deviates from the consensus sequence at a position which is thought to be invariant. As discussed below, intron 7 is the site of alternate splicing involving the use of both the unusual splice site and another splice site nine nucleotides downstream which conforms more closely to the splice consensus sequence.

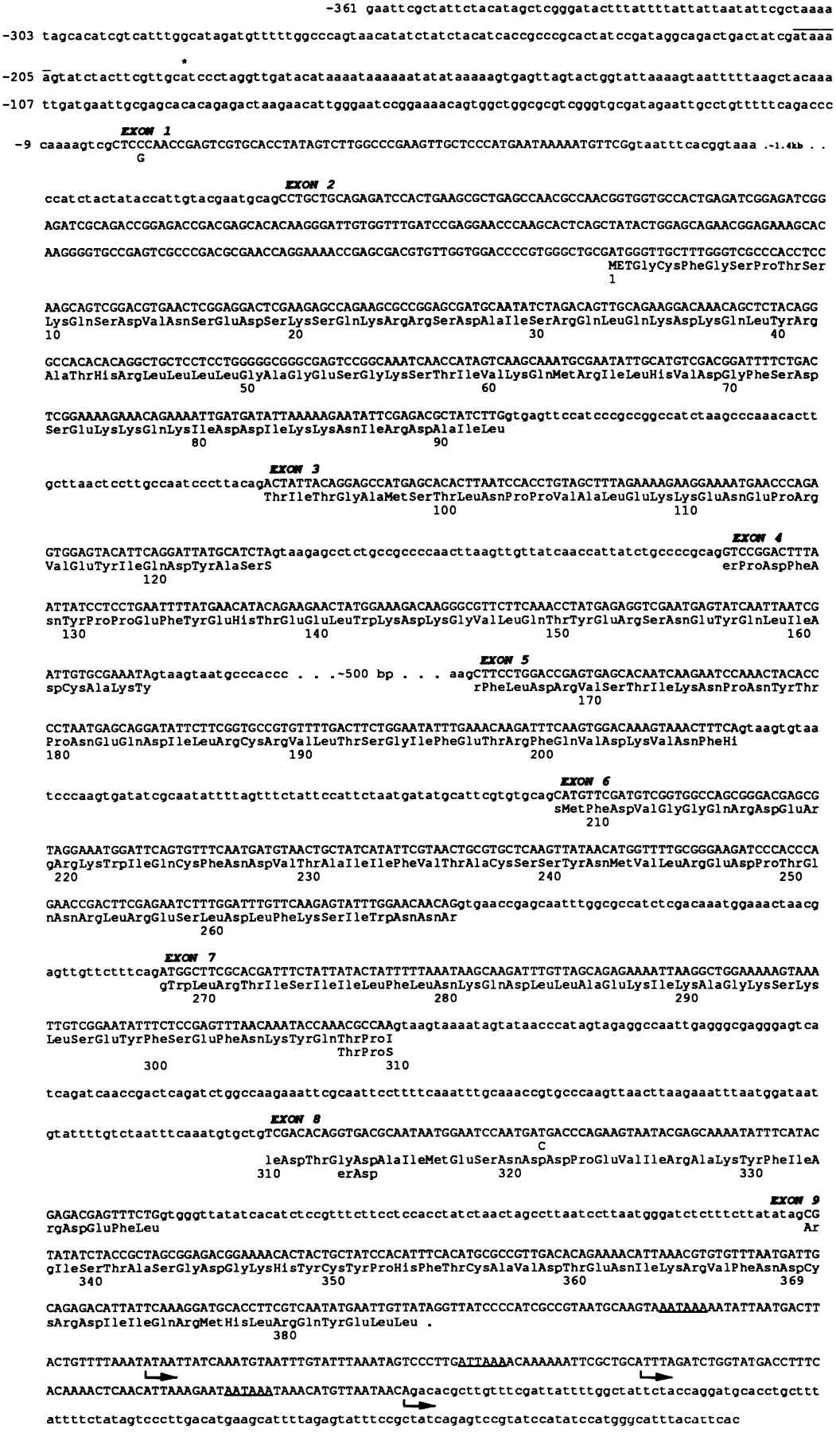
The transcription start site has not been precisely determined. On the basis of the sequence of pDGs12, the longest cDNA isolated, exon 1 is at least 64 bp long. The genomic sequence extending for an additional 361 bp upstream of exon 1 is shown in Fig. 2. Preliminary primer extension analysis (data not shown) suggests that the transcription start site is at an A residue at position −186, 23 bp downstream from the sequence ATAAAA (Fig. 2). The size of exon 1 would, therefore, be 250 bp. The ATAAAA sequence is similar to the TATA box homology and has been found in the promoters of other genes (4). However, we cannot exclude the possibility of an additional upstream intron.

Multiple DGsα transcripts. In vertebrates, heterogeneity of Gsα subunits is the result of differences in the primary sequence after amino acid residue 70 (3). Previous nucleotide sequence analysis has failed to demonstrate variation of *Drosophila* Gsα subunits in the corresponding region (after amino acid residue 74) (24). To confirm this finding, 12 additional Gsα cDNAs were isolated from a *Drosophila* head cDNA library. Nucleotide sequencing of these cDNAs by using Gsα-specific oligonucleotide 652 failed to reveal the existence of transcripts that code for Gsα subunits that differ after residue 74. After extensive restriction mapping, the only detectable difference between the Gsα cDNAs was the presence or absence of a *Bgl*II site in the 3′-untranslated region, 61 bp upstream of the polyadenylation site (data not shown).

As discussed above, intron 7 of the *Drosophila* gene for Gsα uses an unusual TG 3′ splice site. Since intron 7 interrupts the amino acid-coding region, it was possible that heterogeneity in Gsα proteins could be produced by use of another, more conventional 3′ splice site at this intron-exon junction. To investigate this possibility, the exon 7-8 regions of a number of cDNAs were sequenced. Nucleotide sequencing was done by using Gsα-specific oligonucleotide 918 complementary to sequences downstream of the exon 7-8 junction. This analysis resulted in the identification of cDNAs that differ from pDGs12 by deletion of nine nucleotides at the exon 7-8 junction (Fig. 2). These cDNAs code for a shorter Gsα subunit form that replaces a Gly with a Ser and has three amino acids deleted (Fig. 3). This observation suggested the existence of long (DGsαL) and short (DGsαS) forms of Gsα in *D. melanogaster*. These two forms appear to be equally abundant, as determined by their representation in this cDNA library. As expected, transcripts that code for the shorter protein appear to result from use of a consensus 3′ AG splice site (Fig. 2).

Alternate splicing of the *Drosophila* gene for Gsα. The sequences of several cDNA clones suggested that intron 7 of the *Drosophila* gene for Gsα is alternately spliced to both an unusual TG and a consensus AG 3′ splice site (Fig. 2). To

FIG. 2. Nucleotide sequence and deduced amino acid sequence of the *Drosophila* gene for Gsα. Exon sequences are shown in uppercase, and the predicted amino acid sequence is shown below. The amino acid sequence of DGsαS is indicated below that of DGsαL at the site of alternate splicing. The translation termination codon is indicated by a dot. Polyadenylation signals are underlined, and polyadenylation sites are indicated by arrows below the nucleotide sequence. A possible promoter sequence (overlined) and transcription initiation site (*) are indicated.


-361 gaattcgctattctacatagctcgggatactttatttattattaatattcgctaaaa
-303 tagcacatcgtcatttggcatagatgtttttggcccagtaacatatctatctacatcaccgcccgcactatccgataggcagactgactatcgataaa
-205 agtatctacttcgttgcatccctaggttgatacataaaataaaaaatatataaaaagtgagttagtactggtattaaaagtaatttttaagctacaaa
-107 ttgatgaattgcgagcacacagagactaagaacattgggaatccggaaaacagtggctggcgcgtcgggtgcgatagaattgcctgtttttcagaccc
EXON 1
-9 caaaagtcgCTCCCAACCGAGTCGTGCACCTATAGTCTTGGCCCGAAGTTGCTCCCATGAATAAAAATGTTCGgtaatttcacggtaaa .~1.4kb . .
G
EXON 2
ccatctactataccattgtacgaatgcagCCTGCTGCAGAGATCCACTGAAGCGCTGAGCCAACGCCAACGGTGGTGCCACTGAGATCGGAGATCGG
AGATCGCAGACCGGAGACCGACGAGCACACAAGGGATTGTGGTTTGATCCGAGGAACCCAAGCACTCAGCTATACTGGAGCAGAACGGAGAAAGCAC
AAGGGGTGCCGAGTCGCCCGACGCGAACCAGGAAAACCGAGCGACGTGTTGGTGGACCCCGTGGGCTGCGATGGGTTGCTTTGGGTCGCCCACCTCC
METGlyCysPheGlySerProThrSer
1
AAGCAGTCGGACGTGAACTCGGAGGACTCGAAGAGCCAGAAGCGCCGGAGCGATGCAATATCTAGACAGTTGCAGAAGGACAAACAGCTCTACAGG
LysGlnSerAspValAsnSerGluAspSerLysSerGlnLysArgArgSerAspAlaIleSerArgGlnLeuGlnLysAspLysGlnLeuTyrArg
10 20 30 40
GCCACACACAGGCTGCTCCTCCTGGGGGCGGGCGAGTCCGGCAAATCAACCATAGTCAAGCAAATGCGAATATTGCATGTCGACGGATTTTCTGAC
AlaThrHisArgLeuLeuLeuLeuGlyAlaGlyGluSerGlyLysSerThrIleValLysGlnMetArgIleLeuHisValAspGlyPheSerAsp
50 60 70
TCGGAAAAGAAACAGAAAATTGATGATATTAAAAAGAATATTCGAGACGCTATCTTGgtgagttccatcccgccggccatctaagcccaaacactt
SerGluLysLysGlnLysIleAspAspIleLysLysAsnIleArgAspAlaIleLeu
80 90
EXON 3
gcttaactccttgccaatcccttacagACTATTACAGGAGCCATGAGCACACTTAATCCACCTGTAGCTTTAGAAAAGAAGGAAAATGAACCCAGA
ThrIleThrGlyAlaMetSerThrLeuAsnProProValAlaLeuGluLysLysGluAsnGluProArg
100 110
EXON 4
GTGGAGTACATTCAGGATTATGCATCTAgtaagagcctctgccgccccaacttaagttgttatcaaccattatctgccccgcagGTCCGGACTTTA
ValGluTyrIleGlnAspTyrAlaSerS erProAspPheA
120
ATTATCCTCCTGAATTTTATGAACATACAGAAGAACTATGGAAAGACAAGGGCGTTCTTCAAACCTATGAGAGGTCGAATGAGTATCAATTAATCG
snTyrProProGluPheTyrGluHisThrGluGluLeuTrpLysAspLysGlyValLeuGlnThrTyrGluArgSerAsnGluTyrGlnLeuIleA
130 140 150 160
EXON 5
ATTGTGCGAAATAgtaagtaatgcccaccc . . .~500 bp . . . aagCTTCCTGGACCGAGTGAGCACAATCAAGAATCCAAACTACACC
spCysAlaLysTy rPheLeuAspArgValSerThrIleLysAsnProAsnTyrThr
170
CCTAATGAGCAGGATATTCTTCGGTGCCGTGTTTTGACTTCTGGAATATTTGAAACAAGATTTCAAGTGGACAAAGTAAACTTTCAgtaagtgtaa
ProAsnGluGlnAspIleLeuArgCysArgValLeuThrSerGlyIlePheGluThrArgPheGlnValAspLysValAsnPheHi
180 190 200
EXON 6
tcccaagtgatatcgcaatattttagtttctattccattctaatgatatgcattcgtgtgcagCATGTTCGATGTCGGTGGCCAGCGGGACGAGCG
sMetPheAspValGlyGlyGlnArgAspGluAr
210
TAGGAAATGGATTCAGTGTTTCAATGATGTAACTGCTATCATATTCGTAACTGCGTGCTCAAGTTATAACATGGTTTTGCGGGAAGATCCCACCCA
gArgLysTrpIleGlnCysPheAsnAspValThrAlaIleIlePheValThrAlaCysSerSerTyrAsnMetValLeuArgGluAspProThrGl
220 230 240 250
GAACCGACTTCGAGAATCTTTGGATTTGTTCAAGAGTATTTGGAACAACAGgtgaaccgagcaatttggcgccatctcgacaaatggaaactaacg
nAsnArgLeuArgGluSerLeuAspLeuPheLysSerIleTrpAsnAsnAr
260
EXON 7
agttgttctttcagATGGCTTCGCACGATTTCTATTATACTATTTTTAAATAAGCAAGATTTGTTAGCAGAGAAAATTAAGGCTGGAAAAAGTAAA
gTrpLeuArgThrIleSerIleIleLeuPheLeuAsnLysGlnAspLeuLeuAlaGluLysIleLysAlaGlyLysSerLys
270 280 290
TTGTCGGAATATTTCTCCGAGTTTAACAAATACCAAACGCCAAgtaagtaaaatagtataacccatagtagaggccaattgagggcgagggagtca
LeuSerGluTyrPheSerGluPheAsnLysTyrGlnThrProI
ThrProS
300 310
tcagatcaaccgactcagatctggccaagaaattcgcaattccttttcaaatttgcaaaccgtgcccaagttaacttaagaaatttaatggataat
EXON 8
gtattttgtctaatttcaaatgtgctgTCGACACAGGTGACGCAATAATGGAATCCAATGATGACCCAGAAGTAATACGAGCAAAATATTTCATAC
C
leAspThrGlyAspAlaIleMetGluSerAsnAspAspProGluValIleArgAlaLysTyrPheIleA
310 erAsp 320 330
EXON 9
GAGACGAGTTTCTGgtgggttatatcacatctccgtttcttcctccacctatctaactagccttaatccttaatgggatctctttcttatatagCG
rgAspGluPheLeu Ar
TATATCTACCGCTAGCGGAGACGGAAAACACTACTGCTATCCACATTTCACATGCGCCGTTGACACAGAAAACATTAAACGTGTGTTTAATGATTG
gIleSerThrAlaSerGlyAspGlyLysHisTyrCysTyrProHisPheThrCysAlaValAspThrGluAsnIleLysArgValPheAsnAspCy
340 350 360 369
CAGAGACATTATTCAAAGGATGCACCTTCGTCAATATGAATTGTTATAGGTTATCCCCATCGCCGTAATGCAAGTAAATAAAAATATTAATGACTT
sArgAspIleIleGlnArgMetHisLeuArgGlnTyrGluLeuLeu .
380
ACTGTTTTAAATATAATTATCAAATGTAATTTGTATTTAAATAGTCCCTTGATTAAAACAAAAAATTCGCTGCATTTAGATCTGGTATGACCTTTC
ACAAAACTCAACATTAAAGAATAATAAATAAACATGTTAATAACAgacacgcttgtttcgattattttggctattctaccaggatgcacctgcttt
attttctatagtcccttgacatgaagcattttagagtatttccgctatcagagtccgtatccatatccatgggcatttacattcac

DGsαL	305-333	KYQTPIDTGDAIMESNDDPEVIRAKYFIR
DGsαS	305-330	KYQTP SDAIMESNDDPEVIRAKYFIR
Bovine Gsα	317-342	RYTTP EDATPEPGEDDPRVTRAKYFIR

FIG. 3. Amino acid sequences of the variable regions of DGsαL and DGsαS and comparison with that of vertebrate Gsα. The amino acid sequence of the variable region of the long (DGsαL) and short (DGsαS) *Drosophila* Gsα subunit forms are shown in the single-letter amino acid code. Residues identical to those of the corresponding region of bovine Gsα are boxed.

verify that both splice junctions are used in vivo, RNase protection studies were done by using a 166-bp *Bgl*II-*Ssp*I fragment spanning the 3′ splice junction of intron 7 (Fig. 4). The *Ssp*I site is located in exon 8, 60 and 51 bp downstream of the expected splice sites. Uniformly labeled antisense RNA transcripts were prepared and hybridized to both head and body poly(A)$^+$ RNAs overnight at 37°C. The resulting hybrids were digested with RNase A and T_1 and size fractionated on a denaturing polyacrylamide gel. Both head and body poly(A)$^+$ RNAs protected bands of 60 and 49 to 51 bp from RNase digestion (Fig. 4). These results confirm the use of both splice sites in the generation of two mRNAs that

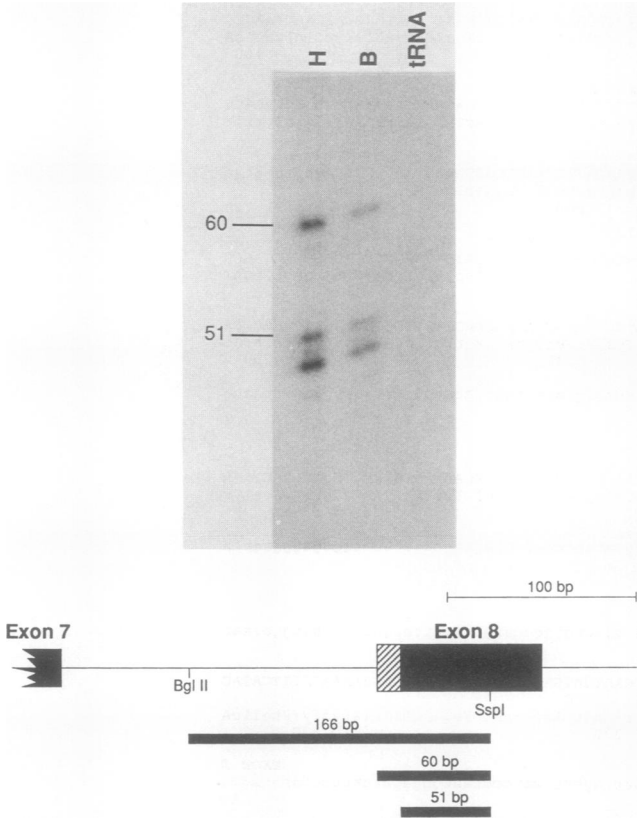


FIG. 4. Demonstration of alternate splicing at the 3′ end of intron 7 by RNase protection mapping. Exons are indicated by solid boxes. The portion of exon 8 which is alternately spliced is hatched. RNase protection studies were done with the uniformly labeled 166-bp RNA probe indicated, as described in Materials and Methods. The source of the protected fragments is also indicated. RNase protection studies were done with poly(A)$^+$ RNAs isolated from adult fly heads (H) or bodies (B) only. *Saccharomyces cerevisiae* tRNA was used as a control. Numbers to the left indicate apparent sizes in nucleotides based on the mobility of ^{32}P-labeled standards produced by transcription of pBS plasmids linearized in the polylinker region.

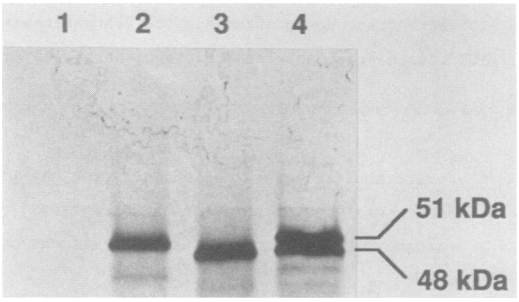


FIG. 5. Translation products synthesized in vitro from transcripts of cDNAs that code for the long and short forms of *Drosophila* Gsα. ^{35}S-labeled proteins were synthesized from transcripts of cDNAs that code for the long (lane 2) and short (lane 3) forms of *Drosophila* Gsα as described in Materials and Methods. Lane 4, equal amounts of the two forms of Gsα; lane 1, control translation reaction with no RNA. kDa, Kilodaltons.

code for alternate forms of Gsα. The smaller protected bands are likely due to intramolecular "breathing" of the hybrids. Protected bands indicating the use of other splice sites at this intron-exon junction were not seen.

In vitro translation of DGsαL and DGsαS. On the basis of their deduced amino acid sequences, DGsαS and DGsαL are predicted to have molecular weights of 44,704 and 45,003, respectively. By Western blot analysis (using Gsα peptide-specific antibodies), *Drosophila* Gsα proteins have been observed with relative molecular weights of 51,000, 48,000, and 44,000 (Wolfgang et al., in press). To determine the relative molecular weights of the protein products of the two forms of *Drosophila* Gsα mRNA, each Gsα cDNA was subcloned into the *Eco*RI site of pBS(M13−)KS and transcribed by using T3 RNA polymerase. These transcripts were then translated in rabbit reticulocyte lysates in [^{35}S]methionine, and the labeled translation products were analyzed by SDS-polyacrylamide gel electrophoresis and autoradiography. The relative molecular weights of DGsαS and DGsαL following in vitro translation was 48,000 and 51,000, respectively, identical to the higher-molecular-weight forms of *Drosophila* Gsα seen by Western blotting (Fig. 5).

Additional forms of *Drosophila* Gsα. As described above, alternate splicing produces two Gsα mRNAs whose products correspond to the two largest forms of Gsα protein observed on Western blots of *Drosophila* neuronal membranes. To look for additional transcripts, PCRs were performed by using poly(A)$^+$ RNAs isolated from whole flies and bodies. Oligonucleotides 652 and 918 were used as opposing primers to amplify the region of variability in both the vertebrate and *Drosophila* proteins. Oligonucleotide 652 is situated upstream of the vertebrate variable region, and oligonucleotide 918 is situated downstream of the *Drosophila* variable region. Amplification products of 900 and 909 bp were expected from transcripts that code for DGsαS and DGsαL, respectively, covering approximately 80% of the coding sequence for each form. Transcripts that code for additional species of Gsα were expected to yield amplification products of significantly different sizes. PCR products were analyzed by agarose gel electrophoresis and Southern hybridization by using pDGs12. In both PCRs, a single band of approximately 900 bp was seen against a background smear of low-molecular-weight material (Fig. 6). Southern hybridization confirmed that this band was the amplification product expected from transcripts that code for DGsαL and

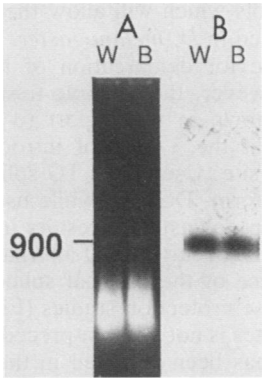


FIG. 6. Analysis of heterogeneity in the coding region of *Drosophila* Gsα transcripts by the PCR. PCRs were done by using poly(A)$^+$ RNAs isolated from whole adult flies (W) or bodies only (B) as described in Materials and Methods. (A) Analysis of PCR products by ethidium bromide staining after agarose gel electrophoresis. The approximately 900-bp product is indicated. (B) Southern blot analysis of PCR products with a ^{32}P-labeled insert of pDGs12.

DGsαS (Fig. 6). No additional amplification products were detected, even after long exposure times (data not shown).

Alternate polyadenylation sites. In addition to the transcript heterogeneity described above, nucleotide sequencing revealed the use of multiple polyadenylation sites. pDGs12 was previously reported to be polyadenylated at a CA dinucleotide 190 bp downstream of the translation termination codon (24). cDNAs polyadenylated at TA or CA dinucleotides 61 and 123 bp downstream of the TAG translation termination codon, respectively, were also found. Polyadenylation at these upstream sites results in cDNAs that lack the *Bgl*II site of the 3′-untranslated region. Each of the polyadenylation sites is downstream of a consensus (AATAAA) or variant (ATTAAA) polyadenylation signal (2) (Fig. 2). Each polyadenylation site is used in cDNAs that code for DGsαL and DGsαS.

DISCUSSION

The gene that codes for the *Drosophila* stimulatory G protein α subunit was isolated and characterized. This gene extends for approximately 4.5 kb and consists of nine exons separated by eight introns (Fig. 1). The organization of the gene was compared to a proposed model of the functional domains of G proteins (18). The four regions of homology among the G proteins, bacterial elongation factor EF-Tu, and the *ras* oncogene proteins (designated A, C, E, and G) (9) are located in separate exons in the *Drosophila* Gsα gene. Region A, which interacts with phosphoryl groups of the guanine nucleotide and may be responsible for the GTPase activity of the subunit, is found in exon 2. Region C, which may be involved in the conformational switch which occurs upon GDP-GTP exchange, and region E, which is thought to form a hydrophobic partial pocket for the guanine ring, are both found in exon 6. Region G, which is also thought to interact with the guanine nucleotide, is found in exon 7. Regions of the Gα subunits responsible for the effector and receptor interactions have not been precisely defined. By analogy to the *ras*-encoded proteins (29), the effector-binding region has been thought to be specified by the least conserved portion of the various Gα proteins. This region in the *Drosophila* gene for Gsα is distributed over part of exon 2 and exons 3 to 5. The Arg residue potentially ADP-ribosylated by cholera toxin (32) is found in exon 5. The C-terminal portion of the protein, thought to form part of the receptor-binding domain (19, 31), is coded for by exons 8 and 9. Recent studies have shown that at least part, if not all, of the effector interaction may also be specified by this region (19).

The intron-exon structure of the *Drosophila* gene for Gsα is more similar to the human gene for Gsα than it is to the *Drosophila* gene for Giα. The *Drosophila* gene for Giα consists of five exons separated by four introns (23). Only one intron location is conserved between the *Drosophila* genes. This intron is adjacent to region C in both genes and separates exons five and six of the *Drosophila* gene for Gsα. The human gene for Gsα consists of 13 exons separated by 12 introns (14). The locations of five introns are identical in the human and *Drosophila* genes for Gsα. These conserved introns were likely acquired early in evolution before the divergence of these genes.

Additional similarities exist between the *Drosophila* and human genes for Gsα. Three of the *Drosophila* exons encode amino acid sequences which are found in multiple exons in the human gene (Fig. 1). An example of this is seen in the splitting of the amino acid-coding region of exon 2 of the *Drosophila* gene into three exons of the human gene. These introns were apparently acquired by the human gene after divergence of the two genes. In addition, pairs of exons in both genes code for the same amino acid region but are interrupted by introns at different positions (Fig. 1). For example, this is true of *Drosophila* exons 3 and 4 and human exons 5 and 6. It appears likely that these introns were inserted into a conserved exon after divergence of the two. Exon 3 of the human gene for Gsα, which codes for the variable portion of the vertebrate Gsα subunits (14), has no counterpart in the *Drosophila* gene for Gsα (Fig. 1).

Comparisons of the organization of the genes for human Gi2α and Gi3α (11), *Drosophila* Giα (23), mouse Gtα (transducin) (25), and human Gsα (14) have revealed that all of these genes contain two conserved intron positions. These introns are located in or near regions A and C described above. The intron in region A is not found in the *Drosophila* gene for Gsα. The intron adjacent to region C is found in the *Drosophila* gene for Gsα and, as discussed above, corresponds to the only intron position conserved between the *Drosophila* genes for Gsα and Giα. The absence of substantial similarity in the exon-intron structures of the *Drosophila* genes for Gsα and Giα is consistent with the early divergence of these proteins. This implies either that intron-interrupting region A has been precisely lost from the *Drosophila* gene for Gsα so as not to impair function or that this intron was acquired independently by the human gene for Gsα.

Western blot analysis of *Drosophila* membrane preparations using Gsα-specific peptide antisera has previously suggested the existence of three molecular weight forms of Gsα with relative molecular weights of 44,000 to 51,000 on SDS-polyacrylamide gels (Wolfgang et al., in press). This report demonstrates that this heterogeneity is due, at least in part, to the existence of *Drosophila* mRNAs for Gsα which differ over an internal region of nine nucleotides in the amino acid-coding region. These transcripts code for two Gsα subunit forms, DGsαL and DGsαS, which differ by inclusion or deletion of three amino acids and substitution of a Gly for a Ser after residue 309. DGsαS is more similar to the vertebrate Gsα subunits than to DGsαL (Fig. 3). The extra amino acid residues of DGsαL are not found in the vertebrate subunits. In vitro translations of RNAs transcribed from plasmids that code for the two Gsα subunit forms have

demonstrated that these forms of Gsα can be separated on SDS-polyacrylamide gels and have relative molecular weights of 48,000 and 51,000, respectively (Fig. 5), identical to the higher-molecular-weight forms of Gsα identified by Western blotting. On the basis of the deduced amino acid sequences of DGsαS and DGsαL, the predicted molecular weights of these proteins are 44,704 and 45,003, respectively. The reason for the discrepancy between the predicted and observed molecular weights is unclear. It is interesting that the larger form of bovine Gsα also migrates anomalously on SDS-polyacrylamide gels (26).

Extensive restriction mapping and nucleotide sequence analysis have failed to identify transcripts that code for additional forms of Gsα protein. This result was confirmed by PCRs with poly(A)$^+$ RNA prepared from whole flies and bodies only with primers designed to amplify the region of variability in both vertebrate and *Drosophila* proteins. Thus, the origin of the 44-kilodalton band seen on Western blots of *Drosophila* membrane preparations is unclear. Although the possibility of additional transcripts exists, the additional protein band may also be the result of posttranslational modifications or proteolysis. The nature of the posttranslational modifications to which Gsα subunits are subject is unknown. Studies of recombinant vertebrate Gsα subunits synthesized in *Escherichia coli* suggest that eucaryotic posttranslational modifications are required for full activity (7). Specific proteolysis of Gα proteins in *Drosophila* membranes has been observed in relation to the 40-kilodalton protein modified by pertussis toxin (10). This protein, recently shown to be the *Drosophila* Goα homolog (31a), can be proteolyzed in neuronal membranes to a discrete 37-kilodalton protein.

Additional variability in the expression of the *Drosophila* gene for Gsα arises from the use of multiple polyadenylation sites. Multiple polyadenylation consensus sequences have been found in the 3′-untranslated region of human Gsα transcripts, but use of multiple polyadenylation sites has not been reported (14). Each polyadenylation site has been found in association with each *Drosophila* Gsα splice variant. The significance of the use of these multiple polyadenylation sites is unknown.

This report demonstrates that the vertebrate and *Drosophila* Gsα subunits are subject to primary-structure variations in different regions. This result is confirmed by the structure of the *Drosophila* gene for Gsα. The N-terminal portion of the human Gsα subunit is encoded by multiple exons, and variability in this region is the result of alternate splicing involving variable use of exon 3 (14). Alternate splicing cannot occur in this region of the *Drosophila* gene, since these sequences are found in a single exon. Since the *Drosophila* and human Gsα proteins are subject to variation in widely separated regions, it is possible that different aspects of G protein function are affected. Vertebrate Gsα subunits are variable in or close to the proposed effector region (3). In vertebrates, the functional significance of this heterogeneity is unknown, although some quantitative differences in their interactions with adenylate cyclase has been found (7). It is tempting to speculate that in *D. melanogaster*, the variation in the primary structure affects receptor interaction. The different *Drosophila* Gsα proteins may interact with different receptors or have different affinities for receptors. This may result in a different tissue distribution for the two spliced forms. The validity of this speculation awaits further study of the extent of the region of Gsα which is involved in receptor interactions and the development of tools which will allow the two forms to be cleanly discriminated in *D. melanogaster*.

The GT-AG rule for demarcation of introns is almost invariant (22). However, the multiple forms of *Drosophila* Gsα have been shown in this report to be the result of alternate splicing at the 3′ end of intron 7 involving an unusual TG splice site. Use of the TG splice site results in the larger subunit form, DGsαL, while use of a consensus AG splice site 9 bp downstream results in transcripts that code for DGsαS. The occurrence of alternate splicing at this junction and the use of the unusual splice site have been confirmed by RNase protection studies (Fig. 4). The use of unusual 3′ splice sites is not without precedent. The unusual 3′ splice site CG has been reported in the *Drosophila per* gene (5). Interestingly, the human gene for Gsα also appears to use a TG 3′ splice site as part of the mechanism by which multiple forms of human Gsα are generated (14). Since alternate splicing takes place at different positions in the human and invertebrate genes, this mechanism of generating diversity in Gsα proteins appears to have arisen independently in the two genes during evolution.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grants from the National Institutes of Health to M.F. F.Q. is the recipient of a Canadian Medical Research Council postdoctoral fellowship.

LITERATURE CITED

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. S. Struhl.** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
2. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site! Cell **41**:349–359.
3. **Bray, P., A. Carter, C. Simons, V. Guo, C. Puckett, J. Kamholz, A. Spiegel, and M. Nirenberg.** 1986. Human cDNA clones for four species of Gαs signal transduction protein. Proc. Natl. Acad. Sci. USA **83**:8893–8897.
4. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
5. **Citri, Y., H. V. Colot, A. C. Jacquier, Q. Yu, J. C. Hall, D. Baltimore, and M. Rosbash.** 1987. A family of unusually spliced biologically active transcripts encoded by a *Drosophila* clock gene. Nature (London) **326**:42–47.
6. **Gilman, A. G.** 1987. G proteins: transducers of receptor-generated signals. Annu. Rev. Biochem. **56**:615–649.
7. **Graziano, M. P., M. Freissmuth, and A. G. Gilman.** 1989. Expression of Gsα in *Escherichia coli* purification and properties of two forms of the protein. J. Biol. Chem. **264**:409–418.
8. **Guo, L., R. C. A. Yang, and R. Wu.** 1983. An improved strategy for rapid direct sequencing of both strands of long DNA molecules cloned in a plasmid. Nucleic Acids Res. **11**:5521–5540.
9. **Halliday, K.** 1984. Regional homology in GTP-binding proto-oncogene products and elongation factors. J. Cyclic Nucleotide Res. **9**:435–448.
10. **Hopkins, R. S., M. A. Stamnes, M. I. Simon, and J. B. Hurley.** 1988. Cholera toxin and pertussis toxin substrates and endogenous ADP-ribosyltransferase activity in *Drosophila melanogaster*. Biochim. Biophys. Acta **970**:355–362.
11. **Itoh, H., R. Toyama, T. Kozasa, T. Tsukamoto, M. Matsuoka, and Y. Kaziro.** 1988. Presence of three distinct molecular species of Gi protein: α subunit structure of rat cDNAs and human genomic DNAs. J. Biol. Chem. **263**:6656–6664.
12. **Kim, D., D. L. Lewis, L. Graziadei, E. J. Neer, D. Bar-Sagi, and D. E. Clapham.** 1989. G-protein βλ-subunits activate the cardiac muscarinic K^+-channel via phospholipase A_2. Nature (London) **337**:557–560.

13. **Korneluk, R. G., F. Quan, and R. A. Gravel.** 1985. Rapid and reliable dideoxy sequencing of double-stranded DNA. Gene **40**:317–323.
14. **Kozasa, T., H. Hoh, T. Tsukamoto, and Y. Kaziro.** 1988. Isolation and characterization of the human Gsα gene. Proc. Natl. Acad. Sci. USA **85**:2081–2085.
15. **Kurachi, Y., H. Ito, T. Sugimoto, T. Shimizu, I. Miki, and M. Ui.** 1989. Arachidonic acid metabolites as intracellular modulators of the G protein-gated cardiac K^+ channel. Nature (London) **337**:555–557.
16. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. **Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, and D. Quon.** 1978. The isolation of structural genes from libraries of eucaryotic DNA. Cell **15**:687–701.
18. **Masters, S. B., R. M. Stroud, and H. R. Bourne.** 1986. Family of G protein α chains: amphipathic analysis and predicted structure of functional domains. Protein Eng. **1**:47–54.
19. **Masters, S. B., K. A. Sullivan, R. T. Miller, B. Beiderman, N. G. Lopez, J. Ramachandran, and H. R. Bourne.** 1988. Carboxyl terminal domain of Gsα specifies coupling of receptors to stimulation of adenylyl cyclase. Science **241**:448–451.
20. **Mattera, R., M. P. Graziano, A. Yatani, Z. Zhou, R. Graf, J. Codina, L. Birnbaumer, A. G. Gilman, and A. M. Brown.** 1989. Splice variants of the α subunit of the G protein Gs activate both adenylyl cyclase and calcium channels. Science **243**: 804–807.
21. **Northup, J. K., P. C. Sternweis, M. D. Smigel, L. S. Schleifer, E. M. Ross, and A. G. Gilman.** 1980. Purification of the regulatory component of adenylate cyclase. Proc. Natl. Acad. Sci. USA **77**:6516–6520.
22. **Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55**:1119–1150.
23. **Provost, N. M., D. E. Somers, and J. R. Hurley.** 1988. A *Drosophila melanogaster* G protein α subunit gene is expressed primarily in embryos and pupae. J. Biol. Chem. **263**:12070–12076.
24. **Quan, F., W. J. Wolfgang, and M. Forte.** 1989. The *Drosophila* gene coding for the α subunit of a stimulatory G protein is preferentially expressed in the nervous system. Proc. Natl. Acad. Sci. USA **86**:4321–4325.
25. **Raport, C. J., B. Dere, and J. B. Hurley.** 1989. Characterization of the mouse rod transducin α subunit gene. J. Biol. Chem. **264**:7122–7128.
26. **Robishaw, J. D., M. D. Smigel, and A. G. Gilman.** 1986. Molecular basis for two forms of the G protein that stimulates adenylate cyclase. J. Biol. Chem. **261**:9587–9590.
27. **Rubin, G.** 1988. *Drosophila melanogaster* as an experimental organism. Science **240**:1453–1459.
28. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
29. **Sigal, I. S., J. B. Gibbs, J. S. D'Alonzo, and E. M. Scolnick.** 1986. Identification of effector residues and a neutralizing epitope of Ha-*ras*-encoded p21. Proc. Natl. Acad. Sci. USA **83**:4725–4729.
30. **Stryer, L., and H. Bourne.** 1986. G proteins: a family of signal transducers. Annu. Rev. Cell Biol. **2**:391–419.
31. **Sullivan, K. A., R. T. Miller, S. B. Masters, B. Beiderman, W. Heideman, and H. R. Bourne.** 1987. Identification of receptor contact site involved in receptor-G protein coupling. Nature (London) **330**:758–760.

31a. **Thambi, N. C., F. Quan, W. J. Wolfgang, A. Spiegel, and M. A. Forte.** 1989. Immunological and molecular characterization of Goα-like proteins in the *Drosophila* central nervous system. J. Biol. Chem. **264**:18552–18560.

32. **Van Dop, C., M. Tsubokawa, H. R. Bourne, and J. Ramachandran.** 1984. Amino acid sequence of retinal transducin at the site ADP-ribosylated by cholera toxin. J. Biol. Chem. **259**:696–698.
33. **Yatani, A., J. Codina, Y. Imoto, J. P. Reeves, L. Birnbaumer, and A. M. Brown.** 1987. A G protein directly regulates mammalian cardiac calcium channels. Science **238**:1288–1292.
34. **Yatani, A., R. Mattera, J. Codina, R. Graf, K. Okabe, E. Padrell, R. Iyengar, A. M. Brown, and L. Birnbaumer.** 1988. The G protein-gated atrial K^+ channel is stimulated by three distinct Giα-subunits. Nature (London) **336**:680–682.