

# Molecular cloning and sequence determination of four different cDNA species coding for $\alpha$ -subunits of G proteins from *Xenopus laevis* oocytes

Juan Olate<sup>1</sup>, Sixta Martinez<sup>2</sup>, Patricia Purcell<sup>1</sup>, Hugo Jorquera<sup>1</sup>, Juan Codina<sup>3</sup>, Lutz Birnbaumer<sup>3</sup> and Jorge E. Allende<sup>1</sup>

<sup>1</sup>Departamento de Bioquímica, Facultad de Medicina, Universidad de Chile, Casilla 70086, Santiago 7, Chile, <sup>2</sup>Departamento de Química, Universidad Nacional, Bogotá, Colombia and <sup>3</sup>Department of Cell Biology, Baylor College of Medicine, Texas Medical Center, Houston, TX 77030, USA

Received 9 May 1990

A cDNA library prepared from *Xenopus laevis* oocytes in  $\lambda$ gt10 was screened with a mixture of three oligonucleotide probes designed to detect sequences found in different mammalian genes coding for  $\alpha$ -subunits of G-proteins. In addition to a clone coding for a  $G_{\alpha o}$ -type subunit previously reported [(1989) FEBS Lett. 244, 188–192] four additional clones have been found coding for different  $G_{\alpha}$  protein subunits. By comparison with mammalian  $\alpha$ -subunits, these oocyte cDNAs correspond to two closely related  $G_{\alpha s}$ -1a, to a  $G_{\alpha i}$ -1 and to a  $G_{\alpha i}$ -3 species. The derived amino acid sequences showed that both  $G_{\alpha s}$  species contain 379 residues, corresponding to the short species without the serine residue and with a calculated  $M_r$  of 42720. The  $G_{\alpha i}$ -1 gene encodes a 354 amino acid protein with an  $M_r$  of 39000 and the  $G_{\alpha i}$ -3 encodes an incomplete open reading frame of 345 residues, lacking the first 9 amino acid residues at the  $NH_2$  terminus. All these  $G_{\alpha}$ -subunits showed high identity with their respective mammalian counterparts (75–80%), indicating a great degree of conservation through the evolution and the important cellular regulatory function that they play.

G-protein; cDNA cloning; Nucleotide sequence; *Xenopus laevis* oocyte

## 1. INTRODUCTION

The transduction of many external signals towards the interior of the cells involves trimeric proteins that bind guanine nucleotides and that are known as G-proteins [1]. There is a large family of these proteins since more than 16 different G-proteins have been isolated from different species and tissues [2–5]. Although the function of some of these G-proteins has been elucidated in particular signal transduction pathways, there are still many questions open as to the role that each one of these may play in different systems.

The *Xenopus laevis* oocyte has become a popular system for researchers studying the function of receptors and who have isolated mRNAs coding for these receptor proteins. The reason for this popularity is the fact that the oocyte microinjected with these mRNAs has shown itself to be capable of both translating these receptors and also of coupling the newly synthesized receptors to transducing systems. The microinjected oocyte thus acquires the capacity to respond physiologically to the agonist that binds to that particular receptor [6–9].

It has become important, therefore, to study the endogenous transducing systems of the oocyte in order to be able to determine the entities that participate in the mechanism of action of various signals and that couple to their respective effector systems. These considerations have induced us to clone the genes coding for different G-proteins that are expressed in this amphibian oocyte. In a previous communication [10], we reported the cloning of the cDNA coding for the oocyte  $G_{\alpha o}$ -type subunit which showed a high degree of identity to the mammalian  $G_{\alpha o}$ .

In this report, we present the cloning and sequencing of four other different cDNAs from *Xenopus laevis* oocytes coding for  $\alpha$ -subunits highly analogous to  $G_{\alpha s}$ ,  $G_{\alpha i}$ -1 and  $G_{\alpha i}$ -3 of mammalian systems. These results indicate that this single cell type has at least 5 different types of G-proteins.

## 2. MATERIALS AND METHODS

### 2.1. cDNA library

A *Xenopus laevis* oocyte cDNA library constructed in the vector gt10 (kindly donated by Dr D.A. Melton of Harvard University) was utilized [11].

### 2.2. Screening of the cDNA library

Close to  $2 \times 10^5$  recombinant plaques were screened by plaque hybridization [12] with three synthetic probes labeled at the 5'-end with <sup>32</sup>P. The probes used for this purpose and their respective se-

Correspondence address: J. Olate, Departamento de Bioquímica, Facultad de Medicina, Universidad de Chile, Casilla 70086, Santiago 7, Chile

quences were the same utilized in the previous work [10]. Phages from five positive lytic plaques from the first screening round were plaque-purified through secondary and tertiary screening. Four of these clones were fully sequenced. Hybridizations were done overnight at 40°C in a solution containing 6 × SSC, pH 6.8, 100 μg of heat-denatured salmon sperm DNA per ml and 0.1% SDS. Filters were washed three times at 40°C in 6 × SSC and then three times at 45°C for 15 min. Films were exposed overnight at room temperature.

### 2.3. cDNA sequence analysis

Nucleotide sequences were determined by using the M13mp19 and the dideoxynucleotide chain-termination method [13] as described in the 'Sequenase' booklet provided by the US Biochemical Corp. (USB).

## 3. RESULTS

### 3.1. Screening of the $\lambda$ gt10 *Xenopus laevis* oocyte cDNA library

Using the same strategy described previously by Olate et al. we screened about 2 × 10<sup>5</sup> lysis plaques and five positive clones were obtained. All these clones were subjected to a secondary and tertiary screening, their DNAs purified and finally analyzed by nucleotide sequencing of the cDNA inserts following the strategy shown in Fig. 1.

### 3.2. Nucleotide sequences of the *Gai-1* cDNA clone

Fig. 2 shows the nucleotide sequence of the cDNA encoding the *Gai-1* type  $\alpha$ -subunit. The sequence is 2759 nucleotides long and predicts an open reading frame of a 354 amino acid protein ( $M_r$  40200), with a 5'-untranslated region of 184 nucleotides and a 3'-untranslated region of 1510 nucleotides. The sequence contains two poly(A) addition sequences, AATAA, at positions 1979 and 2539 and it ends with a 14 residue poly(A) tail. The deduced amino acid sequence showed a high degree of identity (85%) with the human *Gai-1* [14].

### 3.3. Nucleotide sequence of the *Gai-3* cDNA clone

Fig. 3 shows the nucleotide sequence analysis of an insert with great similarity to human *Gai-3*. The sequence is 2037 nucleotides long and it predicts a single open reading frame of 1035 nucleotides coding for a continuous sequence of a 345 amino acid protein. Since no ATG initiation codon was found, we assume that this sequence contains the partial sequence for a  $G\alpha$ -type protein missing a short stretch of the NH<sub>2</sub>-terminus. At the protein level, comparison of corresponding residues between the oocyte and human *Gai*

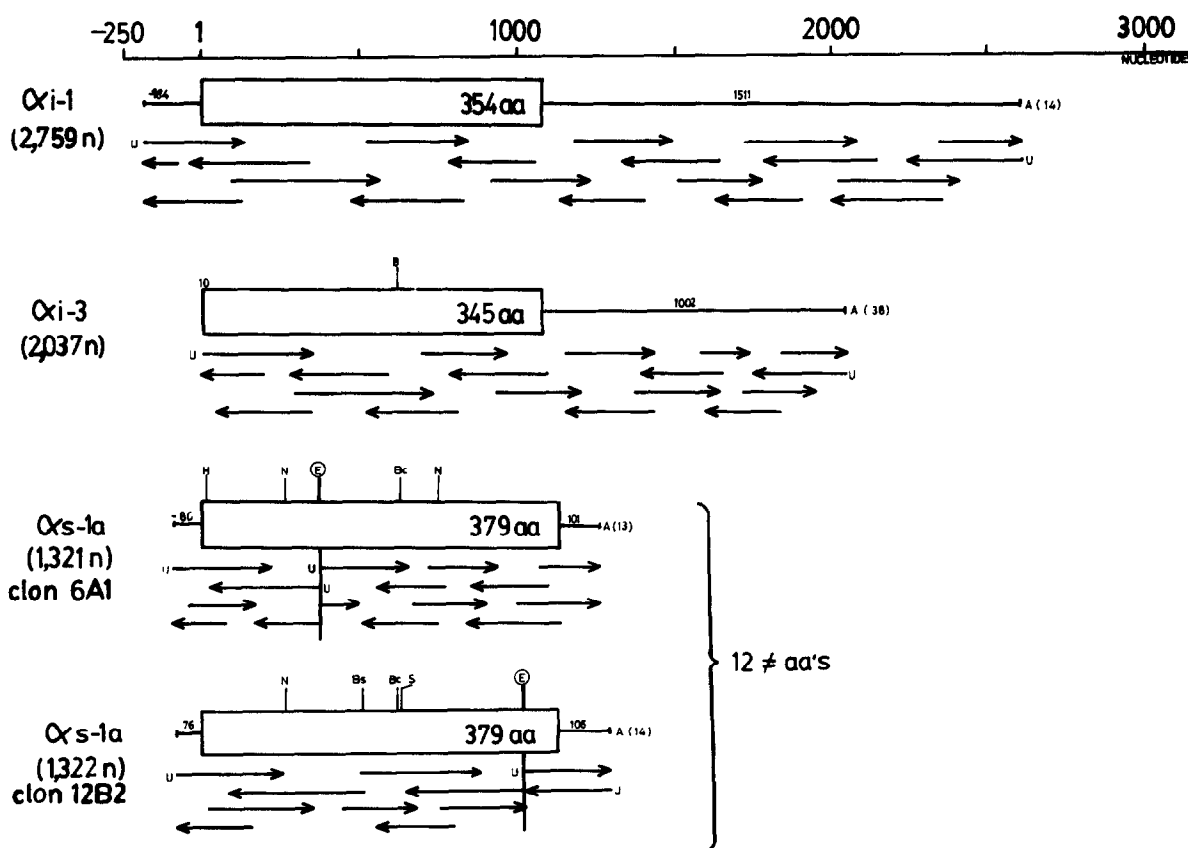


Fig. 1. Sequencing strategy and partial restriction endonuclease map for the *Xenopus laevis*  $G\alpha$  cDNAs. The top scale indicates cDNA length in nucleotides. Open boxes show the open reading frames (ORF) for the different proteins. The thin black lines show the 5'- and 3'-untranslated regions of the mRNA. The arrows indicate extent and direction of sequencing obtained with the oligonucleotide primers. All the cDNA inserts were sequenced in their complete length after subcloning them into the M13mp19 vector. The numbers in parentheses correspond to the length of each cDNA. The restriction endonuclease sites are denoted by one letter. B, *Bam*HI; H, *Hind*III; N, *Nci*I; E, *Eco*RI; Bc, *Bcl*I; S, *Sph*I; Bs, *Bst*I.

-184

CTCTGGGAGG CTTCTGACGC CATTGTGATC CGCGCCACC CCCTGCTGCT CCGCTTCTCT CGCCTTCCGG CTGACCGAG AACCGTGGCC

1 ATG GGA TGT ACE CTG AGC GCC GAA GAC AAG GCA GCC GTG GAG AGG AGC AAA ATG ATC GAT 20  
 Met Gly Cys Thr Leu Ser Ala Glu Asp Lys Ala Ala Val Glu Arg Ser Lys Met Ile Asp 30  
 61 AGG AAC CTT AGC GAG GAC GGA GAG AAG GCT CGC GCG GAG GTG AAG CTG CTT CTG CTC GGC 40  
 Arg Asn Leu Arg Glu Asp Gly Phe Val Leu Ala Arg Glu Val Lys Leu Leu Leu Gly 50  
 121 GCT GGG GAA TCT GGC AAA AGC ACA ATT GTA AAA CAA ATG AAA ATC ATC CAT GAA GCC GGA 60  
 Ala Gly Glu Ser Gly Lys Ser Thr Ile Val Lys Gln Met Lys Ile Ile His Glu Ala Gly 70  
 181 TAC TCA GAA GAA GAA TGC AAA CAG TAC AAG GCA GTT CTT TAC ACT AAC ACA ATT CAA TCC 80  
 Tyr Ser Glu Glu Glu Cys Lys Gln Tyr Lys Ala Val Val Tyr Ser Asn Thr Ile Gln Ser 90  
 241 ATT ATT GCC ATT ATT CGG GCA ATG GGC AGA CTG AAG ATA GAT TTT GGT GAT CCC TCA AGA 100  
 Ile Ile Ala Ile Ile Arg Ala Met Gly Arg Leu Lys Ile Asp Phe Gly Asp Pro Ser Arg 110  
 301 GCG GAT GAC GCA CGC CAG CTT TTT GTA TTG GCT GGA GCA GCA GAA GAA GGT TTT ATG ACT 120  
 Ala Ser Glu Ala Arg Glu Lys Gln Tyr Lys Ala Val Val Tyr Ser Asn Thr Ile Gln Ser 130  
 361 GCA GAA CTA GCT GGA GTT ATA AAA AGA TTA TGG AAG GAT GGT GGT GTA CAG CCG TGT TTC 140  
 Ala Glu Leu Ala Gly Val Ile Lys Arg Leu Trp Lys Asp Gly Glv Val Gln Ala Cys Phe 150  
 421 AAC AGG TCA AGA GAA TAT CAC CTC AAT GAC TCT GCA GCA TAT TAT CTT AAC GAT TTG GAC 160  
 Asn Arg Ser Arg Glu Tyr Lys Gln Leu Asn Asp Ser Ala Ala Tyr Tyr Leu Asn Asp Leu Asp 170  
 481 AGG ATA GCA CAG AAC AGT TAC ATA CCA ACT CAA CAG GAT GCT CTC AGC ACT AGA GTG AAA 180  
 Arg Ile Ala Gln Asn Ser Tyr Ile Pro Thr Gln Gln Asp Val Leu Arg Tyr Arg Val Lys 190  
 541 ACT ACG GGC ATA GTA GAA ACT ATT ACT TTC AAG GAC CTT CAT TTC AAA ATG TTT GAT 200  
 Thr Thr Gly Ile Val Glu Thr His Phe Thr Phe Lys Asp Leu His Phe Lys Met Phe Asp 210  
 601 GTG GGA GGC CAA AGA TCT GAA AGA AAA AAA TGC ATT CAT TGC TTT GAG GCG CTC ACA GCA 220  
 Val Gly Glu Lys Arg Ser Glu Arg Lys Lys Trp Ile His Cys Phe Glu Gly Thr Ala 230  
 661 ATA ATT TTC TGT GTA GCA CTG AGT GAT TAT GAT CTA GTT TTA GCT GAA GAT GAG GAA ATG 240  
 Ile Ile Phe Cys Val Ala Leu Ser Asp Tyr Asp Leu Val Leu Ala Glu Asp Glu Glu Met 250  
 721 AAC CCG ATG CAT GAA AGC ATG AAA CTA TTC GAT AGT ATC TGC AAT AAC AAG TGG TTT ACA 260  
 Asn Arg Met His Glu Ser Met Lys Leu Phe Asp Ser Ile Cys Asn Asn Lys Phe Thr 270  
 781 GAC ACT TCC ATT ATT CTC TTT CTA AAT AAA AAA GAT CTT TTT GAG GAG AAA ATC AAG AGA 280  
 Asp Thr Ser Ile Ile Leu Phe Leu Asn Lys Lys Asp Leu Phe Glu Glu Lys Ile Lys Arg 290  
 841 AGT CCT TTA ACA ATT TGT TAC CCA GAA TAT CCA GGT TCA AAC ACA TAT GAA GAG GCG CCT 300  
 Ser Pro Leu Thr Ile Cys Tyr Pro Glu Tyr Pro Gly Ser Asn Thr Tyr Glu Glu Ala Ala 310  
 901 GCA TAC ATT CAG TGT CAG TTT GAA CAT CCT AAT AAA AGA AAG GAT ACA AAA GAA ATA TAC 320  
 Ala Tyr Ile Gln Cys Gln Phe Glu Asp Leu Asn Lys Arg Lys Asp Thr Lys Glu Ile Tyr 330  
 961 ACG CAT TTT ACA TGT GCT ACG GAT ACC AAG AAT GTS CAG TTT GTT GAT GCA GTA ACT 340  
 Thr His Phe Thr Cys Ala Thr Asp Thr Lys Asn Val Gln Phe Val Phe Asp Ala Val Thr 350  
 1021 GAT GTC ATC ATA AAA AAT AAT CTG AAA GAC TGT GGC CTT TTC TAA TACATGATTA 360  
 Asp Val Ile Ile Lys Asn Asn Leu Lys Asp Cys Gly Leu Phe STOP

1076

TATATGATGG CATTGACCTT CACCCGTGTA TAOCCTGATG GCTTTTGGCC TAACTAAGA TTCTGTATGA  
 ACAGGGGACC AGTACTGTAC TTGCCAGTCT TATTAGCTTT ATTTATGTTC ATGCTTGTGA AATTTTTAAA  
 ACTAACTGCT TGTAGGCCAC AAAAAGAAAG GGTATTTTGA TTGTATGTAT ACTGTAATTC TAGGAATGTT  
 ATTTGTACAG ACATTGAACA GAATATTTTA ATAGATATGAA TTGTCAAAGG GATCACTTTC TTTTCTAAAA  
 TCGTGTAGA GATTTTTAAT TTGCCCTTTT CAGTATTTTA AAGAAACCAT GTACATTATC CTTTTTGTTA  
 CCCTTTATG CATGCATGCT GCTTATTTCT GTTTTAGGCC TTTTATGCT AAGCTCGAAG ATAAATGACT  
 TGTTTAAAT TGTGACATGT CTTACAGTTC CACAACACC CACTCAGCA TGTAAACAAA AATCACTACT  
 TGACCTCAGC AGTCAGTGTG CATTTTTTAT CTCTACCTAT TACTGTGCA TTGCTTAAA GGTGGGSCA  
 AAATAGCTT TTTTATACA TTTTAAAGCT AGTGGTGA TTATGTAAA TAGTPTTAA GGTGGGSCA  
 GTTGTAGCA GGTGTTCAG ATATAGGCC AGTGAATGG TGTGTGAA ATGCAATGCT TTGAAATGTT  
 GTTWTGCGC TCAATTTCTT AGATGACCC TCTCATTTCT TAAAAATATG ACACAACT GCAACATTTA  
 AGTCACGCTT CCCCATTCAA AGTTCATG ACCTTTGTTT GTGGCTGTGT AATGCTGTG AATGCTTCA  
 GCTCTACAA GCTTTTATCA TTCTTTCTCT TTAGACTT TATGTGTAAA CTTTAGGCC ASA TATAAA  
 GGTCTGTGAT ATGTAAAGTA TTCAATTAAC GAATACCATT GAAAGTATAG CAAATATAAA CCTCATCTCT  
 GACATTTAT TTAAGCTCT CTAAGAAATC TGTTTAGAAA CAGTTCATTT ATTTATATA CTTATTTTAA  
 TCAATTTAT TAAATTCAC TTGATGATC AGTTTGGTTC ACTATTAATA ACAGTGTCA TTTCACATTT  
 TTAACATGTC ATAAGTAACT GAGACATGT ATAAATTAAC TGTACTCTT GTAACGTCAAT  
 TAGGCTTTT TGTACTCAC CCTAAGGAAT ACATTATAG TGAATACGAA TGTACTAAGT CACCACTGTC  
 TAAATCTTT GTGCCCTCAA TTATATTTT TATTTAATTT AAATTAACIT GSAAGATTAC GCTGCTATA  
 CTGTATATGA TATACTGAT TGAATGAAA TCAATTTGTC AATTTACCA TATAAAAAAG CACACATGT  
 TGTAAATTT TACTGAAGA TCTCTTGACT TTATCCTGTA AAACGTGTT GCTAAATAAT CCTTATAAC  
 CTTCAATTT TTTGAGAAA AAAAAAAA

2575

Fig. 2. Nucleotide and predicted amino acid sequence of the cDNA insert for *Gai-1*. Numbers indicate the position of nucleotides or amino acid residues starting at the initiation codon. Sequences enclosed in open boxes correspond to the poly(A) addition sequences. Nucleotide sequence recognized by the common 27-base probe is underlined [10].

proteins indicated a high degree of identity (87%) with the human *Gai-3* [15], so this protein was classified as a *Gai-3* oocyte protein. The sequence has a 3'-untranslated region of 1002 nucleotides, it contains a poly(A) addition sequence at position 2010 followed by a stretch of 38 adenosines.

3.4. Nucleotide sequence of a *Gas*-type cDNA

Fig. 4 shows the nucleotide and predicted amino acid

10

CRG GAG GCT GCC GAG CGA AGT AAA ATG ATC GAT 20  
 Arg Glu Ala Ala Glu Arg Ser Lys Met Ile Asp 30  
 61 CGG AAC CTT CGG GAA GAT GGG GAA AAG GCA TCC AAG GAG GTG AAA CTG CTG CTA CTC GGT 40  
 Arg Asn Leu Arg GCT Glu Asp Gly Glu Lys Ala Ser Lys Glu Val Lys Leu Leu Leu Gly 50  
 121 GCT GGT GAG TCT GGG AAA AGC AGC ATT CTG AAG CAA ATG AAA ATT ATC CAT GAG GAT GEA 60  
 Ala Gly Glu Ser Gly Lys Ser Thr Ile Val Lys Gln Met Lys Ile Ile His Glu Asp Gly 70  
 181 TAC TCC GAG GAA GAA TGC CGG CAG TAC AAA GTG GTC GTG TAC AGT AAC ACT ATT CAG TCA 80  
 Tyr Ser Glu Glu Glu Cys Arg Gln Tyr Lys Val Val Val Tyr Ser Asn Thr Ile Gln Ser 90  
 241 ATC ATC GCT ATA ATC CGA GCC ATG GGA AGG CTA AGE ATT GAT TTT GGA GAT GTC GCT AGA 100  
 Ile Ile Ala Ile Ile Arg Ala Met Gly Arg Leu Arg Ile Asp Phe Gly Asp Val Ala Arg 110  
 301 GCT GAT GAT GCT CGA CAG CTC TTT GTG TGG GCC AGT AGT GCT GAG GAG GGA GTT ATG TCT 120  
 Ala Asp Asp Ala Arg Gln Leu Phe Val Leu Ala Ser Ser Ala Glu Glu Gly Val Met Ser 130  
 361 GCA GAA CTT GCA GGT GTA ATT CAG AGC CTG TRG GAA GAT TCT GGA GTT CAG GGC TGT TTC 140  
 Pro Glu Leu Ala Gly Val Ile Glu Arg Lys Trp Glu Met Lys Trp Glu Gln Ala Cys Phe 150  
 421 AGC CGT TCC CGT GAA TAC CAA CTT AAT GAC TCT GCT TCA TAC TAT CTG ACT GAC ATT GAA 160  
 Ser Arg Ser Arg Glu Tyr Gln Leu Asn Asp Ser Ala Ser Tyr Tyr Leu Ser Asp Ile 170  
 481 ARG ATT GCC CAA GGC AGC TAC ATT CCT ACT CAG CAG GAT GTA CTT CGA ACG GCG TTC AAG 180  
 Arg Ile Ala Gln Gly Ser Tyr Ile Pro Thr Gln Gln Asp Val Leu Tyr Thr Arg Val Lys 190  
 541 ACT ACA GGG ATT GTG GAG ACC CAT TTC ACT TTC AAG GAC CTG TAT TTT AAA ATG TTT GAT 200  
 Thr Thr Gly Ile Val Glu Thr His Phe Thr Phe Lys Asp Leu His Phe Lys Met Phe Asp 210  
 601 GTG GGT GGT CAG AGG TCA GAA AGG AAG AAA TGC ATC CAT TGT TTT GAG CGA CTT ACT GCA 220  
 Val Gly Gly Gln Arg Ser Glu Arg Lys Lys Trp Ile His Cys Phe Glu Gly Thr Ala 230  
 661 ATC ATT TTT TGT GTG GCA CTC AGT GAT TAC GAC TTA CTG CTG GCT GAA GAT GAG GAG ATG 240  
 Ile Ile Phe Cys Val Ala Leu Ser Asp Tyr Asp Leu Leu Leu Ala Glu Asp Glu Glu Met 250  
 721 AAT CCG ATG CAT GAA AGC ATG AAA TTG TTT GAT AGC ATC TGC AAC AAC AAG TGG TTC ATA 260  
 Asn Arg Met His Glu Ser Met Lys Leu Phe Asp Ser Ile Cys Asn Asn Lys Thr Phe Ile 270  
 781 GAC ACC TCA ATC ATC CTC TTC CTT AAC AAA AAG GAC CTG TTT GAA GAA AAG ATC TCC AGG 280  
 Asp Thr Ser Ile Ile Leu Phe Leu Asn Lys Lys Asp Leu Phe Glu Glu Lys Ile Ser Arg 290  
 841 ACC CCC CTT ACT ATT TGC TAC CCA GAA TAC TCA GGG TCT AAC ACC TAT GAA GAA GGT GCA 300  
 Ser Pro Leu Thr Ile Cys Tyr Pro Glu Tyr Pro Gly Ser Asn Thr Tyr Glu Glu Ala Ala 310  
 901 GCC TAC ATT CAG TGC CAG TTT GAG GAC TTG AAC CGC AGG AAA GAC ACA AAG GAA ATA TAC 320  
 Ala Tyr Ile Gln Cys Gln Phe Glu Asp Leu Asn Arg Arg Lys Asp Thr Lys Glu Ile Tyr 330  
 961 ACA CAT TTC ACA TGT GCC ACG GAT ACC AAG AAC GTT CAG TTT GTA TTT GAT GCA GTC ACA 340  
 Thr His Phe Thr Cys Ala Thr Asp Thr Lys Asn Val Gln Phe Val Phe Asp Ala Val Thr 350  
 1021 GAT GTC ATA ATT AAG AGC AAT CTA ATG GAG TGT GGC TTG TAT TAG TCCGTGATGAT 360  
 Asp Val Ile Ile Lys Ser Asn Leu Met Glu Cys Gly Leu Tyr STOP

1076

GTTCGACGGG AGACTGTATT GTCTCAGTTT GATTCAGGA CCATTTCCTT GAGTAAAGG GAATCCCTGC  
 AGTGAAGAT GTCCCTGTTT TAGTTTTCAT TGGGAATCCT GTCTTGACA CTGAAGGTG CTGACCCCTGA  
 GCTGTAGACC CCAATTTTTC TATGACATTA CCATCACTAG CACACCAAGC ATTTCTCCCA CCTATGAAA  
 TGGACTGTTT CTGAACCTTT GTAATAGCAG CATCACTACT TTCCACTGTT ATTAGAATAT CAGAAGAAAT  
 CSTATGATCA CATCTTTTGT TAGGACCCCT CTGTTTTTAA AGTAGGCCGC AGTCAGACTG TTTCTCTTC  
 TTTTCTCGTG GCCAGGATTT GTCCCTGAGC TATGCACTTG TCAGAGATCA CAGGACCCAT ATGAATACTG  
 GACTTTATAA CTGTTTCTTG CACTCACCAT GATTACTTAA CCTGTCTTGT GCATTTGTT TATACATAAA  
 CTTCTCCTTA AAACGGGAGG CAACCTTTCA TFCACATTA GCGAGATAT GCTGTACTT TCCTCTGTA GTTACATGAA  
 ATGATTCCTC TGTCTTACTA TTGTGCCTCA CGAGATATCT GCTGTACTT TCCTCTGTA AATTTGTTTC  
 AGTCCTTCTC TCTCTCTTTC TCTGTGATTT ATTTCTTAC ATTTCTTAC TTGAGCCAT TTTGAGCCAT CACTGTAAC  
 AATAAATA ATAAATGCT TCTATTTCTT GCCCGTTTGT TGCAGTGGCC ACTTCCCTTC TCGCTGATC  
 GCTTTCTGCT CTTGTGCCAA ACTCCCATGT GATGGTTTT TTTTITTAG TTTTAAATG TTTTATCAT  
 TCTTAAATG GTCCACTGCG ACACATACAT TAATCAATA TATTTTTCC CACTGGCAT TCATTAATCA  
 TTCTTGGCC ATCCATTCAA CAGTATAATA ATGTGTGAGC ATGCAAAAA AAAAAACAAA AAAAAAAA  
 AAAAAAAAA AA

2067

Fig. 3. Nucleotide and predicted amino acid sequence of the cDNA insert for *Gai-3*. Numbers indicate the position of nucleotides or amino acid residues starting at the hypothetical initiation codon. Sequence enclosed in an open box correspond to the poly(A) addition sequence. Nucleotide sequence recognized by the common 27-base probe is underlined [10].

sequence for one of the *Gas* cDNA (clone 6A1). The sequence is 1321 nucleotides long and predicts an open reading frame of 379 amino acid residues ( $M_r$  42720) and the protein presents more similarity (90%) to the short species of human *Gas-1a* [16]. At the nucleotide level, the 5'-non-coding region contains 80 nucleotides and a small 3'-non-coding region of 101 residues containing a poly(A) addition sequence at position 1226 and a poly(A) tail of 13 adenosines. Clone *Gas* 12B2 is very similar to clone 6A1 at the nucleotide level, showing only 42 differences which generate different restriction sites (see *EcoRI* site in Fig. 1) and 12 different amino acid residues (Fig. 5).

-80  
CGGCCGTCGG

CCTACTTGGT TGGTTGGCCT CCGCCTCCCG CGCTCTTTCC TGAGCGGCC GCCCAGCC CACACGGACC  
 1 10 20  
 ATG GGG TGT CTG GGC AAC AGC AAG ACC GAC GAT CAG CGC AAC CAG CAG AAG CTG CAG CCG  
 Met Gly Cys Leu Gly Asn Ser Lys Thr Glu Asp Gln Arg Asn Gln Gln Lys Val Gln Arg  
 61 30 40  
 GAG ACC AAC AAG AAG ATC GAG AAG CAG CTG CAG AAG GAC AAG CAG GTG TAT AGG GGC ACG  
 Glu Thr Asn Lys Lys Ile Glu Lys Gln Leu Gln Lys Asp Lys Gln Val Tyr Arg Ala Thr  
 121 50 60  
 CAC AGG CTT CTG CTG CTC GGT GCT GGA GAG TCT GST AAA AGC TCA ATT GTG AAG CAG ATG  
 His Arg Leu Leu Leu Leu Gly Ala Gly Glu Ser Gly Lys Ser Ser Ile Val Lys Gln Met  
 181 70 80  
 CGG ATC CTG CAC GTG AAT GGA TTC AAT GCA GAG GAG AAG AAA ACC AAA GTG CAA GAT ATA  
 Arg Ile Leu His Val Asn Gly Phe Asn Ala Glu Glu Lys Lys Thr Lys Val Gln Asp Ile  
 241 90 100  
 AAG AAT AAT ATT AAG GAA GCT ATA GAG ACA ATA GTT ACG GCA ATG GGC AAC CTC TCT CCC  
 Lys Asn Asn Ile Lys Glu Ala Ile Glu Thr Ile Val Thr Ala Met Gly Asn Leu Ser Pro  
 301 110 120  
 CCG GTG GAG CTA GTG AAT CCA GAA AAC CAG TTC CGA ATT GAC TAC ATC CTC AAC CTA CCC  
 Pro Val Glu Leu Val Asn Pro Glu Asn Gln Phe Arg Ile Asp Tyr Ile Leu Asn Leu Pro  
 361 130 140  
 AAT TAC AAA GAC TTT GAA TTC TCA CCA GAG TTC TAT GAA CAC ACA AAA ACG CTC TGG CAG  
 Asn Tyr Lys Asp Phe Glu Phe Ser Pro Glu Phe Tyr Glu His Thr Lys Thr Ser Lys Trp Gln  
 421 150 160  
 GAC GAG GGG GTG AGG GCG TGC TAC GAG CGA TCC AAC GAG TAC CAG CTG ATT GAC TGC GCA  
 Asp Glu Gly Val Arg Ala Cys Tyr Glu Arg Ser Asn Glu Tyr Gln Leu Ile Asp Cys Ala  
 481 170 180  
 CAG TAT TTT CTA GAC AAA ATT GAC ATT GTG AAA CAG AAC GAC TAC ACG CCT AGT GAT CAG  
 Gln Tyr Phe Leu Asp Lys Ile Phe Asp Val Glu Ser Gln Asn Asp Tyr Thr Pro Ser Asp Gln  
 541 190 200  
 GAC TTG CTG CGA TGC AGA GTT CTC ACG TCG GGA ATA TTC GAA ACC AAG TTT CAG GTG GAC  
 Asp Leu Leu Arg Cys Arg Val Leu Thr Ser Gly Ile Phe Glu Thr Lys Phe Gln Val Asp  
 601 210 220  
 AAA GTC AAT TTC CAC ATG TTC GAT GTT GGA GGC CAG CGA GAT GAG CGC AGG AAG TGG ATC  
 Lys Val Asn Phe His Met Phe Asp Val Gly Glu Lys Arg Asp Glu Arg Lys Thr Lys Trp Ile  
 661 230 240  
 CAG TGC TTT AAC GAT GTC ACA GCT ATA ATC TTC GTT GTA GCC AGC AGC AGC TAC AAC ATG  
 Gln Cys Phe Asn Asp Val Thr Ala Ile Ile Phe Val Val Ala Ser Ser Ser Tyr Asn Met  
 721 250 260  
 GTG ATC CGG GAG GAC AAT CAC ACC AAC AGA CTC CAG GAA GCA TTA AAC CTC TTC AAG AGT  
 Val Ile Arg Glu Asp Asn His Thr Asn Arg Leu Gln Glu Ala Leu Asn Leu Phe Lys Ser  
 781 270 280  
 ATC TGG AAC AAT AGG TGG CTA CGG ACC ATT TCA GTC ATT CTC TTC CTC AAT AAA CAA GAT  
 Ile Trp Asn Asn Arg Trp Leu Arg Thr Ile Ser Val Ile Leu Phe Leu Asn Lys Gln Asp  
 841 290 300  
 CTG CTC GCT GAA AAG GTT AAT GCT GGG AAA TCT AAA ATA GAG GAC TAC TTC CCT GAG TTT  
 Leu Leu Ala Glu Lys Val Asn Ala Gln Lys Ser Lys Ile Glu Asp Tyr Phe Leu Glu Phe  
 901 310 320  
 GCC CGA TAC ACC ACC CCA GAT GAT GCA ACT CCA GAA GTC GGC GAG GAT CCT CGA GTC ACA  
 Ala Arg Tyr Thr Thr Pro Asp Asp Ala Thr Pro Glu Val Gly Glu Asp Pro Arg Val Thr  
 961 330 340  
 AGG GCC AAG TAT TTT ATT AGA GAT GAG TTT CTT AGA ATC AGC ACA GCC AGT GGG GAC GGC  
 Arg Ala Lys Tyr Phe Ile Arg Asp Glu Phe Leu Arg Ile Ser Thr Ala Ser Gly Asp Gly  
 1021 350 360  
 CGC CAT TAT TGC TAC CCT CAT TTC ACA TGT GCA GTG GAT ACA GAA AAC ATC CGA AGG GTT  
 Arg His Tyr Cys Tyr Pro His Phe Thr Cys Ala Val Asp Thr Glu Asn Ile Arg Arg Val  
 1081 370 380  
 TTT AAT GAT TGT CGG GAC ATT ATC CAA AGG ATG CAC CTG CGC CAA TAT GAG CTG TTG TGA  
 Phe Asn Asp Cys Arg Asp Ile Gln Arg Met His Leu Arg Gln Tyr Glu Glu CTG TTG TGA  
 1141  
 TGGAGAGCAC TTTTGGTTTT TGGTCCCGCC CCCCCCCTC CTGTGTTC GGACTACAAA TTCTCATTT  
 AAAGAAAAAA ACCTAATAA AAAAAAAA A

1241

Fig. 4. Nucleotide and predicted amino acid sequence of the cDNA for *Gαs* (clone 6A1). Numbers indicate the position of nucleotides or amino acid residues starting at the initiation codon. Sequence enclosed by the open box corresponds to the poly(A) addition sequence. Nucleotide sequence recognized by the common 27-base probe is underlined [10] and the *EcoRI* site is indicated by a line over the sequence GAATTC.

3.5. Comparison of the deduced amino acid sequences of the *Xenopus laevis* oocyte *Gα* subunits

Fig. 5 shows the alignment of the predicted protein sequences of each of the cDNA clones reported here. The sequences share great similarity among themselves and with their mammalian counterparts. The regions of identity are enclosed in boxes. Also the figure shows the sequences labelled A, C, E and G which are supposedly involved in the binding of GTP.

4. DISCUSSION

The results presented in this communication and in the previous publication [10] demonstrate that one single cell type, the amphibian oocyte, contains at the mRNA level at least 5 different types of  $\alpha$  G-proteins.

Since our search has not been exhaustive, it is quite possible that there are more different  $\alpha$ -subunits not yet detected.

The analysis of the nucleotide and deduced amino acid sequences allows us to classify the oocyte subunits according to their analogy to the structure of mammalian G-protein  $\alpha$ -subunits. Thus it has been established that two of the isolated clones correspond to  $\alpha s$ -type subunits, and the other two clones correspond to  $\alpha i$ -1 and  $\alpha i$ -3 subunits.

The two  $\alpha s$  clones correspond to the 'short' version of the mammalian  $\alpha s$  which arises through alternative splicing and elimination of exon 3, present in the 'long'  $\alpha s$  [17]. Another variant found in mammals is that the short version can be found with or without a serine in position 72. The two oocyte clones are short versions without this serine. This is to our knowledge the first time that two different  $\alpha s$  coding genes have been found in the same cell type and differ in other characteristics than those mentioned above. Since the 5'- and 3'-non-coding regions of the two clones are quite similar, it seems probable that they arose through a rather recent gene duplication. The small number of differences that account for 12 amino acid changes are all scattered through the polypeptide and therefore cannot be generated through alternative splicing. One of the changes alters an *EcoRI* site, a finding which may be useful for future analysis of the function of these proteins. One of the most interesting differences between the two  $\alpha s$  clones is the conservative change of serine-178 present in the 6A1 clone for threonine in the clone 12B2. On the basis of the results obtained with mutations done in rats, the region encompassing residues 192-196 in long species of the  $\alpha s$  subunits has been proposed to interact with the catalytic subunit of adenylyl cyclase and any mutation in this region could cause a change in its activity [18]. It is interesting to note that in this highly conserved region, all the  $\alpha$ -subunits that are not stimulatory (non- $\alpha s$ -type subunits) contain a threonine in position 178 (short  $\alpha s$  species) while all previously sequenced  $\alpha s$  subunits have a serine in this position. The only other exception to this latter generality is the short  $\alpha s$ -type found in rat olfactory neurons which contains a threonine in position 180 [19]. Interestingly, in the same tissue, olfactory neuroepithelium, a different long  $\alpha s$  species from non-neuronal origin was found by the same group and this contained a serine in position 193 [20]. The presence of these two types of  $\alpha s$  subunits in the oocyte suggests that they may have different functions in this cell.

Several G-protein regions share considerable similarity with the guanine nucleotide binding regions of elongation factor Tu and p21<sup>ras</sup> [21]. Consistent with this role, these regions (Fig. 5, labeled A, C, E and G) are all highly conserved within the four oocyte G-proteins. Other regions, as the one implicated in receptor,  $G\beta\gamma$ , and effector protein interaction [22] are not

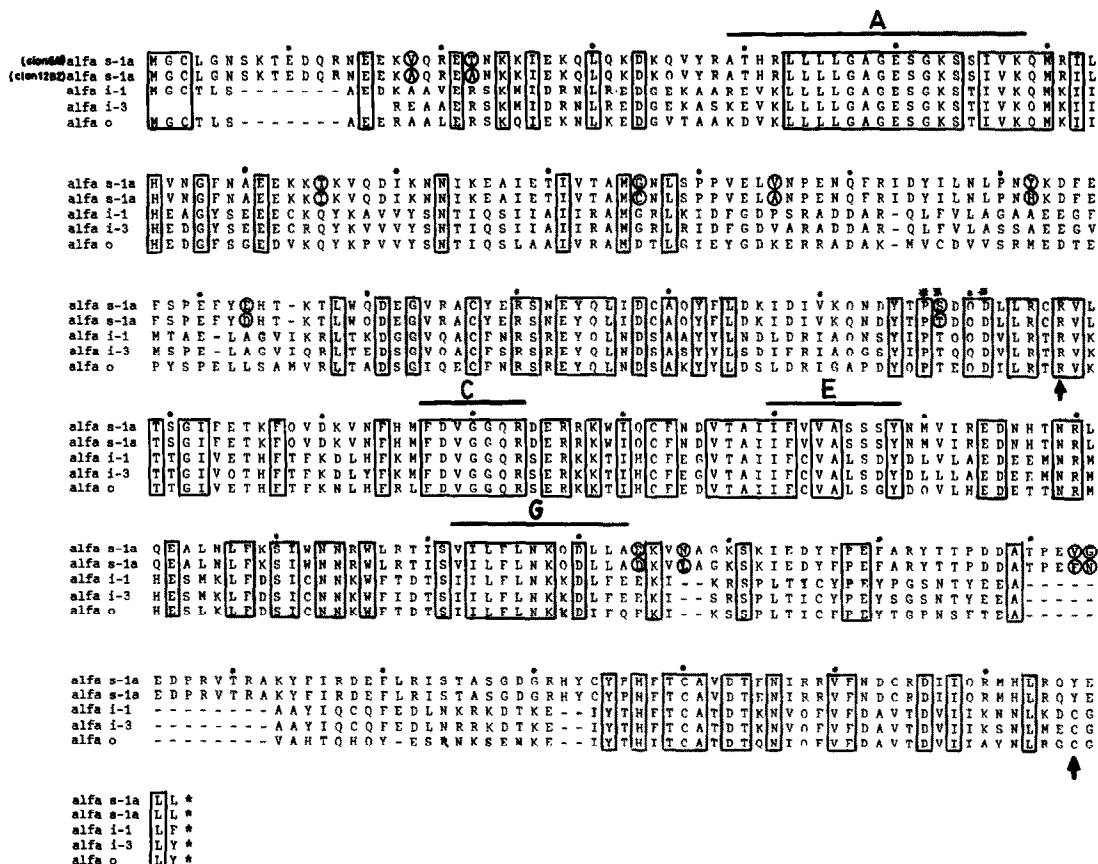


Fig. 5. Alignment of the amino acid sequences of the oocyte  $G\alpha$  subunits. Amino acid sequences are presented by the standard one-letter abbreviation code. The oocytes *Gas*, *Gai-1*, *Gai-3* and the already published *Gao* [10] are shown. Amino acid regions that are identical are enclosed by open boxes. The arrows indicate the arginine and cysteine residues that are ADP-ribosylated by *Cholera* and *Pertussis* toxins, respectively. The 12 amino acid differences between clones 12B2 and 6A1 of *Gas* are enclosed by circles. The amino acid regions that participate in the binding of GTP are overlined. The residues marked by asterisks correspond to the *Gas* region proposed to be important for the adenylyl cyclase activation [18].

as conserved as the guanine nucleotide binding regions, reflecting their role in the independent modulation of different signal transduction pathways.

The proteins encoded by *Gai-1* and *Gai-3* cDNAs are potential substrates for pertussis toxin (PTX)-catalyzed ADP-ribosylation because they have a cysteine residue at the appropriate site near the carboxyl-terminus. Also the two *Gas* clones contain the arginine that is modified by cholera toxin (Fig. 5, arrowheads).

Currently, we are expressing these proteins in *E. coli* and in an in vitro system in order to study some of their G-protein properties and functions.

REFERENCES

[1] Birnbaumer, L., Codina, J., Mattera, R., Yatani, A., Scherer, N., Jose-Toro, M. and Brown, A. (1987) *Kidney International* 32, S14-S37.  
 [2] Suki, W., Abramowitz, J., Mattera, R., Codina, J. and Birnbaumer, L. (1987) *FEBS Lett.* 220, 187-192.  
 [3] Matsuoka, M., Itoh, H., Kozara, T. and Kaziro, Y. (1988) *Proc. Natl. Acad. Sci. USA* 85, 5384-5388.  
 [4] Strathmann, M., Wilkie, T. and Simon, M. (1989) *Proc. Natl. Acad. Sci. USA* 86, 7407-7409.  
 [5] Gilman, A. (1989) *J. Am. Med. Assoc.* 262, 1819-1825.

[6] Kobilka, B., MacGregor, C., Daniel, C., Kobilka, T., Caron, M. and Lefkowitz, R. (1987) *J. Biol. Chem.* 262, 15796-15802.  
 [7] Snutch, T. (1988) *Trends Neurol. Sci.* 11, 250-256.  
 [8] Kline, D., Simoncini, L., Mandel, G., Maue, R., Kado, R. and Jaffe, L. (1988) *Science* 241, 464-467.  
 [9] Moriarty, T., Sealfon, S., Carty, D., Roberts, J., Iyengar, R. and Landau, E. (1989) *J. Biol. Chem.* 264, 13524-13530.  
 [10] Olate, J., Jorquera, H., Purcell, P., Codina, J., Birnbaumer, L. and Allende, J. (1989) *FEBS Lett.* 244, 188-192.  
 [11] Rabagliati, M., Weeks, D., Harvey, R. and Melton, D. (1985) *Cell* 42, 769-777.  
 [12] Abramowitz, J., Mattera, R., Liao, C., Olate, J., Perez-Ripoll, E., Birnbaumer, L. and Codina, J. (1988) *J. Rec. Res.* 8, 561-588.  
 [13] Sanger, F., Nicklen, S. and Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.  
 [14] Bray, P., Carter, A., Guo, V., Puckett, C., Karn Holz, J., Spiegel, A. and Nirenberg, M. (1987) *Proc. Natl. Acad. Sci. USA* 84, 5115-5119.  
 [15] Codina, J., Olate, J., Abramowitz, J., Mattera, R., Cook, R. and Birnbaumer, L. (1988) *J. Biol. Chem.* 263, 6746-6750.  
 [16] Mattera, R., Codina, J., Crozat, A., Kidd, V., Woo, S. and Birnbaumer, L. (1986) *FEBS Lett.* 206, 36-42.  
 [17] Kozara, T., Itoh, H., Tsukamoto, T. and Kaziro, Y. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2081-2085.  
 [18] McCormick, F. (1989) *Nature* 340, 678-679.  
 [19] Jones, D. and Reed, R. (1989) *Science* 244, 790-795.  
 [20] Jones, D. and Reed, R. (1987) *J. Biol. Chem.* 262, 14241-14249.