Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 4620–4624, July 1988
Biochemistry

Sequences of the bovine and yeast ADP-ribosylation factor and comparison to other GTP-binding proteins

(cholera toxin)

JENNY L. SEWELL AND RICHARD A. KAHN

Laboratory of Biological Chemistry, Division of Cancer Treatment, National Cancer Institute, Building 37, Room 5D-02, Bethesda, MD 20892

Communicated by Alfred G. Gilman, February 16, 1988 (received for review December 16, 1987)

ABSTRACT **The ADP-ribosylation factor (ARF) is a 21-kDa GTP-binding protein that serves as the cofactor in the cholera toxin-catalyzed activation of the stimulatory guanine nucleotide-binding protein of adenylate cyclase (G_s). An oligonucleotide probe based on the partial amino acid sequence was used to clone ARF from a bovine adrenal chromaffin cDNA library. The yeast (*Saccharomyces cerevisiae*) ARF gene was then cloned from a YCp50 genomic library by cross-species hybridization by using the coding region of the bovine gene. RNA gel blots of poly(A)$^+$ RNA indicate that only one ARF message size (900 and 2000 base pairs) is present in yeast and cows, respectively. Comparison of the cDNA-derived amino acid sequences of ARF to other GTP-binding proteins reveals a structural relationship between ARF and the ras family of proteins. A slightly better structural relationship is detected when ARF is compared to the α subunits of the trimeric GTP-binding proteins, including $G_{s\alpha}$. All of the biochemical characteristics of the purified ARF, including the lack of GTPase activity and the posttranslational myristoylation, are consistent with the derived sequences. Comparison of the ARF sequences to that of the chicken processed pseudogene (CPS-1), previously reported as a ras homologue, reveals that CPS-1 is actually an ARF-derived gene. These results demonstrate that ARF is a GTP-binding protein with structural features of both the ras and the trimeric GTP-binding protein families.**

Covalent modification (ADP-ribosylation) of the stimulatory guanine nucleotide-binding protein of adenylate cyclase (G_s) by cholera toxin results in the irreversible activation of adenylate cyclase (1). Efficient intoxication of purified G_s has been shown to require another GTP-binding protein found in both membrane (2) and cytosolic (3) fractions. The activity of the ADP-ribosylation factor (ARF) is that of a cofactor, which is in turn regulated by GTP. The working model for the toxin reaction includes the rate-limiting binding of GTP to ARF, which allows its interaction with G_s to form the GTP–ARF–$G_{s\alpha}$ complex (where $G_{s\alpha}$ is the α subunit of G_s), the presumed substrate for cholera toxin (4). Thus, ARF appears to interact with at least one component of the adenylate cyclase complex. ARF is a protein found in every eukaryotic cell examined so far, including man, yeast, and slime mold. It is present in high abundance ($\approx$1% of cell protein) in brain tissue. The occurrence of ARF in such a diverse array of organisms suggests another, more fundamental, role in cellular physiology. Such a role has not yet been identified.

Studies on the regulation of adenylate cyclase activity and visual transduction in the retina led to the identification of a family of trimeric GTP-binding regulatory proteins (G proteins) (5, 6). Members of the G protein family include G_s and G_i, the stimulatory and inhibitory proteins of adenylate cyclase, respectively; G_o, of unknown function; and G_t, also called transducin, the activator of the light-sensitive phosphodiesterase in the retina. The α subunit of each G protein is distinct and contains a single high-affinity binding site for guanine nucleotides. It is likely that each of these membrane proteins acts as a transducer of information between the cell and the external milieu. These functions have been reviewed elsewhere (6, 7).

The *ras* genes all code for 21-kDa GTP-binding proteins, termed p21 (8). The *ras* gene family is composed of three members, termed Ha-*ras*, Ki-*ras*, and N-*ras*, which are expressed to different extents in different tissues and cell types. Viral or activated forms of p21 can cause transformation in cultured cells and have been detected in an estimated 20% of human tumors (9). Work on these oncogenes has led to the identification of normal cellular homologues of the *ras* genes, which are present in all eukaryotic cells examined. Ras proteins are thought to function as signal transducers although the activators and effectors of ras activity have yet to be identified. In the yeast *Saccharomyces cerevisiae*, the protein products of the *RAS* genes appear to regulate adenylate cyclase (10).

In addition to these two well-defined families of GTP-binding proteins, a number of related genes have more recently been described that share structural features common to the ras proteins. These genes include *rho* (11), *ypt* (12), *ral* (13), and R-*ras* (14). These genes have all been cloned, and primary sequences of the proteins they encode have been determined. The structural relationships between members of the ras family and the G protein family have been described (15–17). This relationship is limited to short stretches of sequence that are known to be involved in the binding of guanine nucleotides.

As both ARF and *ras*-encoded p21 have tenuous functional relationships with the adenylate cyclase system and are 21-kDa GTP-binding proteins, it has been suggested that a structural or functional relationship may exist. However, the purified proteins are immunologically distinct, and ras proteins have no activity in the ARF assay (31). To further determine the relationship between ARF and other members of the superfamily of GTP-binding regulatory proteins, we have cloned and sequenced ARF from two different species.* Comparison of the predicted amino acid sequence of ARF to those of the other GTP-binding proteins reveals that ARF is

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ARF, ADP-ribosylation factor; G protein, guanine nucleotide-binding regulatory protein; G_s and G_i, G proteins that mediate stimulation and inhibition, respectively, of adenylate cyclase; G_o, a G protein of unknown function; G_t, a G protein (transducin) that activates the light-sensitive phosphodiesterase in the retina; $G_{i\alpha}$ and $G_{s\alpha}$, the α subunits of G_i and G_s, respectively.

*The sequences reported in this paper are being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession nos. J03275, bovine; J03276, yeast).

the only protein with a significant relationship to both the ras and G protein families.

MATERIALS AND METHODS

Reagents. Synthetic oligonucleotides, used both for screening the cDNA library and as primers in the sequencing reactions, were the generous gift of Michael Brownstein (Laboratory of Cell Biology, National Institutes of Health, Bethesda, MD). They were synthesized on an Applied Biosystems (Foster City, CA) model 380A oligonucleotide synthesizer. The bovine adrenal chromaffin cell cDNA library was constructed by Lee Eiden and Anna Iacangelo (Laboratory of Cell Biology, National Institutes of Health) in the Okayama–Berg vector. The yeast genomic library, which utilized the YCp50 vector, was constructed (18) and supplied by Mark Rose (Princeton University, Princeton, NJ). Sequencing was performed by using the dideoxy chain termination method of Sanger *et al.* (19) with deoxyadenosine 5′-[γ-[^{35}S]thio]triphosphate. Protein sequences were analyzed by using the ALIGN program (20). A gap penalty of 25 was used after empirically determining the optimal alignments for members of each family and minimizing random, single residue identities. No qualitative differences result from assigning a lower gap penalty.

Screening of Libraries. All hybridization protocols were performed in hybridization buffer containing 6× SSC (1× = 150 mM NaCl/15 mM sodium citrate, pH 7), 5× Denhardt's solution (5× = 0.1% Ficoll/0.1% polyvinylpyrrolidone/0.1% bovine serum albumin), 50 mM sodium phosphate (pH 6.5), 0.1% sodium dodecylsulfate, 100 μg of tRNA per ml, 100 μg of herring sperm DNA per ml, 1 mM sodium pyrophosphate, and 0–50% (vol/vol) formamide. Approximately 4 × 10^5 colonies were screened with a 47-base-pair (bp) oligonucleotide. The original screening was performed in buffer containing 20% formamide at 37°C. Filters were washed at room temperature in 0.2× SSC/0.1% sodium dodecylsulfate/0.05% sodium pyrophosphate (wash buffer). The final screening utilized more stringent washing conditions (1× SSC at 65°C).

The yeast genomic library was screened by cross-species hybridization with the 700-bp *Pst* I–*Pst* I fragment from the bovine cDNA, which includes all of the coding region of bovine ARF. Hybridizations were done in buffer containing 50% formamide at 25°C, and filters were washed in wash buffer at 37°C.

Hybridizations for DNA and RNA gel blotting were performed in buffer containing 50% formamide at 42°C and washed in wash buffer at 56°C, unless otherwise indicated.

RESULTS AND DISCUSSION

Cloning the Bovine ARF cDNA. Purified peptides, generated by cyanogen bromide digestion of purified bovine brain ARF, were sequenced as previously described (31). Three peptides representing approximately one-third of the protein (59 residues) were sequenced. Based on the sequence of one of these peptides (Met-Val-Gly-Leu-Asp-Ala-Ala-Gly-Lys-Thr-Trp-Ile-Leu-Tyr-Lys-Leu), a synthetic oligonucleotide of 47 bp with an 8-fold degeneracy (AGCTTGTA$^{C}_{G}$AG-GATCCAG$^{G}_{A}$TCTTGCCAGCAGCATC$^{C}_{G}$AGGCCCAC-CAT) was used to screen a bovine adrenal chromaffin cell cDNA library. Four independent clones each containing inserts of ≈2000 bp were obtained in this manner. Three of the clones had identical restriction maps, and the fourth was similar but lacked ≈150 bp at the 5′ end. One of the longer clones (pARF1) was selected for DNA sequencing. The restriction map and sequencing scheme are depicted in Fig. 1*A*. The cDNA sequence and translated open reading frame are shown in Fig. 1*B*. The cDNA sequence begins immediately after the guanosine tail used in the construction of the Okayama–Berg library and ends a short distance before the poly(A) tail. DNA sequencing was not possible 3′ to the *Acc* I site presumably due to the presence of the poly(A) tail. The long open reading frame in pARF1 codes for a protein with 181 amino acids having a calculated molecular mass of 20,699 Da. The proposed initiating methionine conforms well to the consensus sequence for eukaryotic initiation sites (21). The presence of all three peptide sequences, obtained by protein sequencing, in the cDNA-derived amino acid sequence confirms the identity of the cDNA as that of ARF. One would predict only the three peptides previously described to result from cyanogen bromide digestion of the protein sequence shown in Fig. 1*B*.

The 700-bp *Pst* I–*Pst* I fragment of pARF1 includes the entire coding region of bovine ARF (see Fig. 1*A*). Oligo(dT)-selected RNA from bovine adrenal chromaffin cells was hybridized with the labeled *Pst* I–*Pst* I fragment in RNA gel blots. A single band of about 2000 bp was observed, which was not seen in the poly(A)$^-$ RNA (not shown). Thus, it is likely that the inserts are full length.

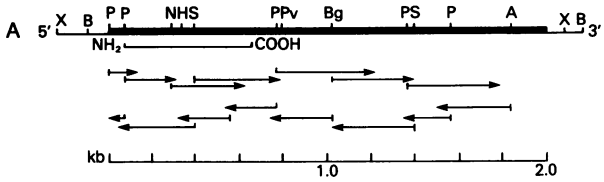
Cloning the Yeast ARF Gene. The yeast ARF gene was obtained by cross-species hybridization using the 700-bp *Pst* I–*Pst* I fragment of the bovine ARF gene to screen a genomic yeast library constructed in the YCp50 vector. Each of the positive clones obtained in this screening included the restriction map shown in Fig. 1*C* as part of the insert. The sequencing strategy and location of the coding region of clone yARF1 are also shown in Fig. 1*C*, and the DNA and translated protein sequences are shown in Fig. 1*D*. The open reading frame codes for a protein with 181 amino acids and a predicted molecular mass of 20,532 Da. The proposed translation initiation start site is similar to other eukaryotic start regions, including an adenosine at position −3, a guanosine at +4, and a ribosylthymine at +6 (21). The 5′ untranslated region contains four TATAA regions, at positions −34, −100, −223, and −418.

The 660-bp *Sst* I–*Xba* I fragment from yARF1 was used as a probe of RNA gel blots. Gel blots of oligo(dT)-selected yeast RNA revealed a single band at ≈900 bp.

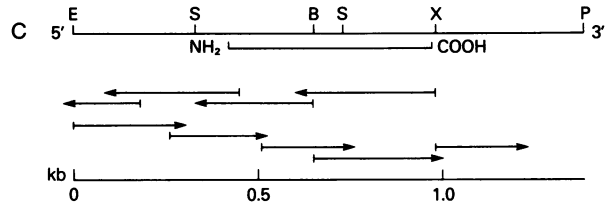
As the bovine and yeast ARF proteins are each 181 amino acids in length, alignment of proteins and nucleic acids in the coding region was straightforward. Nucleic acid sequences of the coding regions are 65% identical (352 out of 543). At the amino acid level, the two ARF proteins share 76.8% (139 out of 181) identity, with many conservative substitutions (see Fig. 3). The extent of the identities between yeast and bovine ARF is greater than that for the other GTP-binding proteins having homologues in both yeast and higher eukaryotes (ras, rho, or $G_{i\alpha}$/GPA1).

Structural Features of ARF. ARF is the only known GTP-binding protein with a significant structural relationship to both the G protein α subunit and ras families. A more extensive relationship is seen between ARF and the α subunits of the G proteins than with the other 21-kDa GTP-binding proteins. However, the extent of the relationship is not sufficient to classify ARF as a member of either family of proteins. The significance of this level of structural relationship, intermediate between two defined families, is of undetermined significance as far as function or evolutionary origin of ARF.

These relationships can be detected by examination of the aligned sequences, as seen in Fig. 2, or by more complex algorithms, such as the ALIGN program used to generate the data in Table 1. ARF contains regions homologous to all four of the conserved regions identified by Halliday (23) for




```
B -96  TCTTGTGGGAGCGAAACCAACGCCTGCCTCCGAGCAGCATCTGCGCCGCGGCCGCCGAGGGTTTCTCTGGCCAGTGTTGCCCCCACCGGTCTGCAGCATG   3
                                                                                                        Met

  4    GGGAATATCTTTGCAAACCTCTTCAAGGGCCTTTTTGGCAAAAAAGAAATGCGCATTCTCATGGTGGGCCTAGATGCTGCGGGAAAGACCACCATCCTGT  103
       GlyAsnIlePheAlaAsnLeuPheLysGlyLeuPheGlyLysLysGluMetArgIleLeuMetValGlyLeuAspAlaAlaGlyLysThrIleLeuT

104    ACAAACTGAAGCTGGGTGAAATTGTGACCACCATTCCCACTATTGGCTTCAACGTGGAGACCGTGGAGTACAAGAACATCAGCTTCACTGTGTGGGACGT  203
       yrLysLeuLysLeuGlyGluIleValThrThrIleProThrIleGlyPheAsnValGluThrValGluTyrLysAsnIleSerPheThrValTrpAspVa

204    GGGTGGCCAGGACAAGATCCGGCCTCTGTGGCGCCACTACTTTCAAAACACACAAGGTCTCATCTTTGTGGTTGACAGCAATGACAGAGAGCGTGTGAAT  303
       lGlyGlyGlnAspLysIleArgProLeuTrpArgHisTyrPheGlnAsnThrGlnGlyLeuIlePheValValAspSerAsnAspArgGluArgValAsn

304    GAGGCCCGTGAGGAGCTCATGAGGATGTTGGCAGAAGACGAGCTCAGGGACGCCGTTCTGCTCGTGTTTGCTAACAAACAGGACCTCCCTAATGCCATGA  403
       GluAlaArgGluGluLeuMetArgMetLeuAlaGluAspGluLeuArgAspAlaValLeuLeuValPheAlaAsnLysGlnAspLeuProAsnAlaMetA

404    ACGCGGCCGAGATCACGGACAAGCTGGGCCTGCACTCTCTGCGCCACAGGAACTGGTACATTCAGGCCACCTGTGCCACCAGTGGGGACGGGCTTTACGA  503
       snAlaAlaGluIleThrAspLysLeuGlyLeuHisSerLeuArgHisArgAsnTrpTyrIleGlnAlaThrCysAlaThrSerGlyAspGlyLeuTyrGl

504    GGGACTGGACTGGCTGTCCAATCAGCTCCGGAACCAGAAATGAGCTGTACTCCTCTCTCCTCCCCTCTCACTCCCTCCTTCTACCCTCGCTTTACTCTCA  603
       uGlyLeuAspTrpLeuSerAsnGlnLeuArgAsnGlnLysEnd

604    TGTGGCAAACGTGCCCCTGTGGTGTGAGTGCCAGAAGCTGTCTCCACGGTCGGTCACAGTGTGCTTCACCATGCTGTAAATGTGCAGACGCAGCCTGCAG  703

704    CTGGGTTTTTATTTAATGTAAATAGTTTTTGTTCCCAATGAGGCAGTTTCTGGTACTCCTATGCAATATTACTCAGCTTTTTTTATTGTAAAAAGAAAAT  803

804    CAACCCACTGTTTAGTACTGAGAAGGGATTTTAGGCACACGGGTCATCCAGGAGTTGTGTGTCAGACGAGCCCAATACTCCTCCTCTGGGCTTTAGATCT  903

904    GTGTTGAGATCCATTTTGGTGGTTGGTTTTTAACCCAAACTCAGTGCATTTTTTTAAAATAGTTACAAATACAAAATAAGGAGAACACTTGAACACACAG 1003

1004   AAGGGAGAAATGTGCCTAATGTAGGTTCTGCGGTAACGGCCTGAGTCCAGTCCACCGGTGGTCTGTTTCCCGTTGGGAACATGAAACGGGGCAGAACCAG 1103

1104   CCATGATATCCATTCTGCATGGTCACAATAGGGTCCCTGTGATTTGCTCTGTCTCGGGTCCTCCCACACCCCATAGCGTTTGTGCTTGTATTTGTGCATA 1203

1204   CATGGCCGCCAGGAGATGGGGTGGAGCTGCAGCCACCGTTCTCAGAGCCCTTGGAAGAGCCCCGGGCTTTTCTCGGGTCAGCATGGGCACTTTGGTTGGG 1303

1304   ACGCCCTGGCTGGAGCGGGACCCTGCAGCCCCCTTGCTTGCCTGCTCCGGGCTCTCTGGGGTCTCACCAGCAGGAGATGCGGGTGGTCATCCTGGAGCCC 1403

1404   CAGCCCCTCAGAGCCCCTGTCCACATGTGCTTTCACCCCCTCAGCCTGCAAGGGCCAGATTTGCCATTGAAGATGATAACCTCTACTTCTTTCTTTTGTA 1503

1504   TTTTGATAAACACTGAAGAAGCTGGAGCTGTTAAACTTTATCTTGGGGAAAACCTCAGAGCTGGTTTATTTGGTGTCGTGGAACCTCTTACTGCTTTCAG 1603

1604   TACACAATTAGTAATCAACTGTTTT 1628
```




```
D -545  ATGTAAAGTGAATGTTGGTAAATGATACTTAAGAGCTAAACGAAAAGATAAGTATCTCGCTCAACCTTTTAATTTTGCACAGTTGAATTCTCGTCTGATA  -446

-445   CTTAAGAAAGTCATTGCCCACATATATAACAGTAGTAATAATGATAATGATAGCAATAGTTTTAAGCTAGCTACTTAGTTGCATTTTTCAATAGTTTAGT  -346

-345   AAAAAAAGTCACGCAATAAGCTCTCAAGAAGCCACTAATACCGTAATGATAGCAGTTTATTGTAGAAAAACCATGTTATTACCCTTCCCTTTTTATTTCT  -246

-245   TTTCGCGTTGCAAATCACATATAACGAGGTGGCTTGTATTTGTCAAACCAAAAAAAAAAATGAAAATCGAAAAATGGAAAAACAGAGAGAGAAACGGAAT  -146

-145   CTTTGACACGTTCTCGATCCACTTGTTTATCGAGGTGGTTTTTATAAGTCTTACTAGATAGAAAGTTCATTTTGTTTTGAAACTTTTTGGAACCTCTGGC  -46

-45    ATTGAAGGTATAAGAAAGAACTCAAACAGGTTTAATAGAATTAAAAATGGGTTTGTTTGCCTCTAAGTTGTTCAGTAACCTTTTTGGTAACAAAGAAATG  54
                                                                                                  MetGlyLeuPheAlaSerLysLeuPheSerAsnLeuPheGlyAsnLysGluMet

55     CGTATTCTTATGGTTGGTCTTGATGGTGCTGGTAAGACCACCGTTTTGTACAAGTTGAAATTGGGTGAAGTTATCACTACCATTCCAACAATTGGTTTCA  154
       ArgIleLeuMetValGlyLeuAspGlyAlaGlyLysThrThrValLeuTyrLysLeuLysLeuGlyGluValIleThrThrIleProThrIleGlyPheA

155    ACGTTGAAACTGTCCAATATAAGAACATTTCATTCACTGTCTGGGATGTCGGTGGACAAGACAGAATTAGATCTCTATGGAGACACTACTACAGAAACAC  254
       snValGluThrValGlnTyrLysAsnIleSerPheThrValTrpAspValGlyGlyGlnAspArgIleArgSerLeuTrpArgHisTyrTyrArgAsnTh

255    TGAAGGTGTTATCTTTGTTGTCGATTCTAACGATAGATCGCGTATTGGTGAAGCTAGAGAAGTTATGCAAAGAATGTTGAACGAAGATGAATTGAGAAAC  354
       rGluGlyValIlePheValValAspSerAsnAspArgSerArgIleGlyGluAlaArgGluValMetGlnArgMetLeuAsnGluAspGluLeuArgAsn

355    GCCGCTTGGTTGGTGTTCGCTAACAAGCAAGATTTGCCAGAAGCCATGTCTGCTGCTGAAATCACTGAAAAACTAGGTTTACATTCTATTAGAAACCGTC  454
       AlaAlaTrpLeuValPheAlaAsnLysGlnAspLeuProGluAlaMetSerAlaAlaGluIleThrGluLysLeuGlyLeuHisSerIleArgAsnArgP

455    CATGGTTTATCCAAGCCACGTGTGCTACCTCCGGTGAAGGTTTGTATGAAGGTTTGGAATGGTTAAGTAACAGTTTGAAAAACTCAACTTAAAAATTCTA  554
       roTrpPheIleGlnAlaThrCysAlaThrSerGlyGluGlyLeuTyrGluGlyLeuGluTrpLeuSerAsnSerLeuLysAsnSerThrEnd

555    GAATATGGATCAAATACGCTTGTATAAACTAAATGAACATAAAGATTAAGAACTTAAGAGGCCAACGTCGATGGATTTATTGACGATCACCAGCCAACAC  654

655    ATATAGATTTTAGTGTAAAAGCAATAAAAACCAAGATAATAAAATAAAAAAATACTGAAGAAGCCTAACTAGTATAAACTACTTTAACTAATAATGGCAA  754

755    TTTGATATAGAAACAAAGAAACATGATATATTTAGGATATTATACAACGCATTTTCATTTGTTTTACAGCACCCCTGCGTGAATCATATATTGA  848
```

FIG. 1. (*Legend appears at the bottom of the opposite page.*)

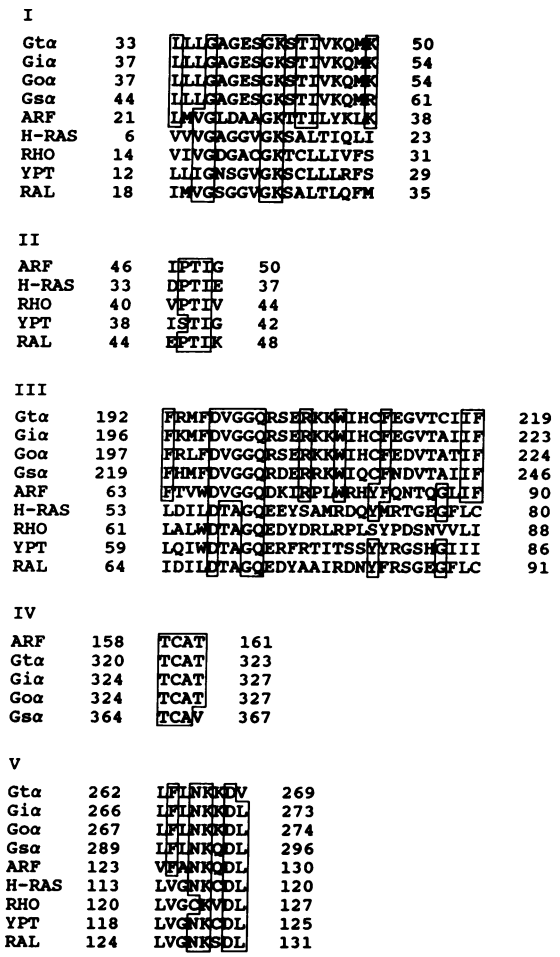


FIG. 2. Alignment of regions of bovine ARF and eight other GTP-binding proteins. Amino acid identities between ARF and at least three other proteins are boxed. Alignments were made with the aid of the ALIGN program. $G_{t\alpha}$ and $G_{o\alpha}$, the α subunits of G_t and G_o, respectively. Amino acids are identified by the single-letter code.

GTP-binding proteins. Group I in Fig. 2 shows the alignment of the different proteins around the first GTP-binding domain, which is known to be involved in controlling the rate of GTP hydrolysis. Asp-26 of ARF corresponds to Gly-12 of ras proteins, which is an important determinant of GTPase activity and transformation potency (24). The lack of a glycine at this position is found only in ypt and ARF and may explain the lack of GTPase activity of purified ARF (4). Regions II and IV show the alignment of the ARF sequence with those in p21s or the G protein α subunits, respectively. In these cases, there is no region in the G protein α subunits that corresponds to region II and no region in the p21s that corresponds to region IV. These two regions are of interest because region II has been implicated as a potential effector binding region in ras proteins, and region IV contains the only cysteine in the ARF sequence. Region III shows the longest stretch of conservation without any gaps. Included in this region is the sequence Asp-Val-Gly-Gly-Gln, which is absolutely conserved in all of the G proteins and ARF. This region

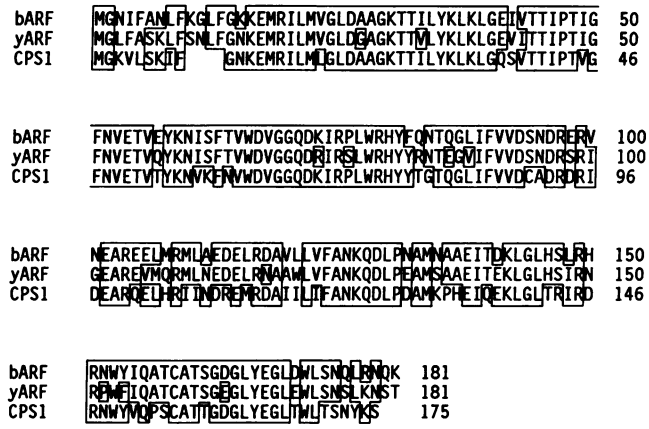


FIG. 3. Alignment of predicted primary sequences of bovine and yeast ARF and the chicken processed pseudogene CPS-1. Amino acid identities are boxed. Amino acids are identified by the single-letter code.

aligns with the sequence Asp-Thr-Ala-Gly-Gln, which is absolutely conserved in all forms of the ras proteins and all of the other 21-kDa GTP-binding proteins described to date. ARF lacks the cysteine near the carboxyl terminus, present in most of the other GTP-binding proteins, which is the site of both palmitoylation of p21s and ADP-ribosylation by pertussis toxin in all of the α subunits of the G proteins except G_s. Each ARF sequence contains a single potential glycosylation site (Asn-Xaa-Ser/Thr) at residues 60–62.

ARF also contains the consensus sequence for amino-terminal myristoylation (Met_{init}-Gly-Xaa-Xaa-Xaa-Ser/Ala/Thr) (25). Of the GTP-binding proteins mentioned above, only G_o, G_i (26, 27), and ARF are known to contain myristic acid as an amino-terminal blocking group. These are also the only GTP-binding proteins listed that contain the consensus sequence. This type of acylation has been shown to be functionally important (28, 29), although the specific role of the fatty acid in cellular localization or enzyme activity has not been demonstrated directly.

Although the conservation of the GTP-binding sites has been described (15–17), the results shown in Table 1 suggest that no significant relationship exists between members of the G protein and ras families when the whole protein is scored. Another point emerging from this analysis is that the *rho*, *ypt*, *ral*, and R-*ras* genes code for proteins that align better than ARF with the ras family of proteins but not as well as the three ras members do among themselves.

ARF-Related Sequences. Analysis of several computer data bases for ARF-related sequences failed to produce more related sequences than those already discussed. However, a review of the literature located a processed pseudogene from chicken termed CPS-1 (30). Bovine ARF and CPS-1 share 74.1% amino acid identity over the entire length of the sequences with only one gap near the amino terminus (Fig. 3). Thus, CPS-1 is much more closely related to ARF than to either the α subunits of the G proteins or ras proteins. It is probable that CPS-1 is a processed ARF pseudogene. Since the CPS-1 sequence retains a similar degree of identity throughout the length of the sequence, it is likely that CPS-1 represents a close homologue of the expressed chicken ARF.

FIG. 1 (*on opposite page*). Restriction maps and sequences of bovine and yeast ARF. The restriction maps of pARF1 (*A*) and yARF1 (*C*) are shown. Arrows indicate the direction and lengths of DNA sequences determined. kb, Kilobase; X, *Xba* I; B, *Bam*HI; P, *Pst* I; N, *Nar* I; H, *Hinc*II; S, *Sau*3A; Pv, *Pvu* II; Bg, *Bgl* II; A, *Acc* I; E, *Eco*RI. The coding regions are shown with amino and carboxyl termini indicated. The nucleic acid sequences and translated coding regions of pARF1 (*B*) and yARF1 (*D*) are also shown.

Table 1. Comparisons of the amino acid sequences of GTP-binding proteins in terms of their alignment scores by using the ALIGN program

	bARF	yARF	G_s	G_i	G_o	G_t	rho	ypt	ral	R-ras	yRAS	Ha-ras	Ki-ras
bARF		84.50	9.26	14.15	10.34	10.83	6.80	8.99	6.42	8.31	6.06	7.14	8.27
yARF	84.50		10.98	11.84	11.92	9.15	6.45	11.02	8.23	5.58	7.16	6.51	5.82
G_s	9.26	10.98		55.86	62.92	72.55	0.91	1.72	0.21	1.90	0.77	2.24	2.27
G_i	14.15	11.84	55.86		137.47	125.84	0.27	0.18	−0.06	3.30	2.42	2.03	2.09
G_o	10.34	11.92	62.92	137.47		114.74	3.24	−0.32	1.87	2.79	1.59	3.26	3.28
G_t	10.83	9.15	72.55	125.84	114.74		0.81	1.07	2.04	3.22	1.19	1.96	1.62
rho	6.80	6.45	0.91	0.27	3.24	0.81		21.74	19.14	24.84	22.61	20.15	21.03
ypt	8.99	11.02	1.72	0.18	−0.32	1.07	21.74		29.98	28.11	29.91	24.19	37.58
ral	6.42	8.23	0.21	−0.06	1.87	2.04	19.14	29.98		53.41	47.28	66.98	48.50
R-ras	8.31	5.58	1.90	3.30	2.79	3.22	24.84	28.11	53.41		63.80	58.39	61.05
yRAS	6.06	7.16	0.77	2.42	1.59	1.19	22.61	29.91	47.28	63.80		61.95	57.67
Ha-ras	7.14	6.51	2.24	2.03	3.26	1.96	20.15	24.19	66.98	58.39	61.95		98.69
Ki-ras	8.27	5.82	2.27	2.09	3.28	1.62	21.03	37.58	48.50	61.05	57.67	98.69	

The ALIGN program first determines the optimal alignment and match score by using a version of the algorithm of Needleman and Wunsch (22). Numbers shown are alignment scores in standard deviation units as calculated by taking the difference between the match score and the average match score for 100 random permutations of the two sequences being compared and dividing by the standard deviation of the random scores. A gap penalty of 25 was used with a matrix bias parameter of 0. Boxes are drawn around clusters of similar scores to facilitate viewing only. It has been empirically determined that scores below three are generally considered not significant, whereas those above six are probably significant. bARF, bovine ARF; yARF, yeast ARF; yRAS, yeast RAS.

Note Added in Proof. A second ARF gene, ARF2, has recently been cloned from yeast. The amino acid sequence, the size of the message, and the restriction map are well conserved between ARF1 and ARF2.

We acknowledge Michael Brownstein, Lee Eiden, and Anna Iacangelo and other members of the Laboratory of Cell Biology for their generous contributions of advice and reagents. We are also in debt to Sanford Silverman for providing RNA and advice relating to yeast.

1. Kahn, R. A., Katada, T., Bokoch, G., Northup, J. K. & Gilman, A. G. (1983) in *Post-translational Modifications of Proteins for Function*, ed. Johnson, B. C. (Academic, New York), pp. 373–395.
2. Schleifer, L. S., Kahn, R. A., Hanski, E., Northup, J. K., Sternweis, P. C. & Gilman, A. G. (1982) *J. Biol. Chem.* **257**, 20–23.
3. Enomoto, K. & Gill, D. M. (1979) *J. Supramol. Struct.* **10**, 51–60.
4. Kahn, R. A. & Gilman, A. G. (1986) *J. Biol. Chem.* **261**, 7906–7911.
5. Stryer, L. & Bourne, H. R. (1986) *Annu. Rev. Cell Biol.* **2**, 391–419.
6. Gilman, A. G. (1987) *Annu. Rev. Biochem.* **56**, 615–649.
7. Gilman, A. G. (1984) *Cell* **36**, 577–579.
8. Gibbs, J. B., Sigal, I. S. & Scolnick, E. M. (1985) *Trends Biochem. Sci.* **10**, 350–353.
9. Finkel, T. & Cooper, G. M. (1984) *Cell* **136**, 1115–1120.
10. Toda, T., Uno, I., Ishikawa, T., Powers, S., Kataoka, T., Broek, D., Cameron, S., Broach, J., Matsumoto, K. & Wigler, M. (1985) *Cell* **40**, 27–33.
11. Madaule, P. & Axel, R. (1985) *Cell* **41**, 31–40.
12. Gallwitz, D., Donath, C. & Sander, C. (1983) *Nature (London)* **306**, 704–707.
13. Chardin, P. & Tavitian, A. (1986) *EMBO J.* **5**, 2203–2208.
14. Lowe, D. G., Capon, D. J., Delwart, E., Sakaguchi, A. Y., Naylor, S. L. & Goeddel, D. V. (1987) *Cell* **48**, 137–146.
15. Lochrie, M. A., Hurley, J. B. & Simon, M. I. (1985) *Science* **228**, 96–99.
16. Tanabe, T., Nukuda, T., Nishikawa, Y., Sugimoto, K., Suzuki, H., Takahashi, H., Noda, M., Haga, T., Ichiyama, A., Kangawa, K., Minamino, N., Matsuo, H. & Numa, S. (1985) *Nature (London)* **315**, 242–245.
17. Robishaw, J. D., Russell, D. W., Harris, B. A., Smigel, M. D. & Gilman, A. G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1251–1255.
18. Rose, M. D., Novick, P., Thomas, J. H., Botstein, D. & Fink, G. R. (1988) *Gene*, in press.
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
20. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
21. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
22. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48**, 443–453.
23. Halliday, K. R. (1983) *J. Cyclic Nucleotide Protein Phosphorylation Res.* **9**, 435–448.
24. Gibbs, J. B., Sigal, I. S., Poe, M. & Scolnick, E. M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5704–5708.
25. Towler, D. A., Eubanks, S. R., Towery, D. S., Adams, S. P. & Glaser, L. (1987) *J. Biol. Chem.* **262**, 1030–1036.
26. Buss, J. E., Mumby, S. M., Casey, P. J., Gilman, A. G. & Sefton, B. M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7493–7497.
27. Schultz, A. M., Tsai, S., Kung, H.-F., Oroszlan, S., Moss, J. & Vaughan, M. (1987) *Biochem. Biophys. Res. Commun.* **147**, 1234–1239.
28. Garber, E. A., Cross, F. R. & Hanafusa, H. (1985) *Mol. Cell. Biol.* **5**, 2781–2788.
29. Kamps, M. P., Buss, J. E. & Sefton, B. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4625–4628.
30. Alsip, G. R. & Konkel, D. A. (1986) *Nucleic Acids Res.* **14**, 2123–2130.
31. Kahn, R. A., Goddard, C. & Newkirk, M. (1988) *J. Biol. Chem.*, in press.