Primary structure of the α-subunit of bovine adenylate cyclase-stimulating G-protein deduced from the cDNA sequence

Toshihide Nukada, Tsutomu Tanabe, Hideo Takahashi, Masaharu Noda, Tadaaki Hirose*, Seiichi Inayama* and Shosaku Numa+

Departments of Medical Chemistry and Molecular Genetics, Kyoto University Faculty of Medicine, Kyoto 606 and **Pharmaceutical Institute, Keio University School of Medicine, Tokyo 160, Japan*

Received 18 November 1985

The primary structure of the α-subunit of the adenylate cyclase-stimulating G-protein (G_s) has been deduced from the nucleotide sequence of cloned DNA complementary to the bovine cerebral mRNA encoding the polypeptide. Comparison of the amino acid sequences of the α-subunits of G_s and transducin reveals that some of the highly conserved regions show sequence homology with elongation factor-Tu and *ras* p21 proteins and correspond to functional regions of guanine nucleotide-binding proteins.

Adenylate cyclase-stimulating G-protein *cDNA cloning* *Nucleotide sequence* *Elongaton factor Tu* ras *protein* *Transducin*

1. INTRODUCTION

A group of membrane-associated G-proteins are essential for transducing signals generated at cell surface receptors into changes in cellular function and metabolism [1]. These proteins are a complex of 3 subunits termed α, β and γ. The α-subunit is responsible for binding guanine nucleotides and is unique to each G-protein. The stimulatory G-protein (G_s) mediates hormonal stimulation of adenylate cyclase [1]. The α-subunit of G_s is the activator of the catalytic moiety of the adenylate cyclase complex. Here, the primary structure of the α-subunit of G_s has been deduced by cloning and sequencing cDNA encoding it. The amino acid sequence homology observed between the α-subunits of G_s and transducin [2–5] is discussed in relation to functional regions of G-proteins.

+ To whom correspondence should be addressed

Abbreviations: G-protein, guanine nucleotide-binding protein; G_s, adenylate cyclase-stimulating G-protein; G_o, a G-protein purified from brain; IAP, islet-activating protein

2. MATERIALS AND METHODS

Total RNA was extracted from adult bovine cerebral cortex as in [6], and poly(A)$^+$ RNA was isolated as in [7]. A library of cDNA clones was constructed by the method of Okayama and Berg [8], using 7.8 μg poly(A)$^+$ RNA and 4.2 μg of the vector-primer DNA. The procedures for transformation and screening were as in [9,10]. Two oligodeoxyribonucleotide probes, 5′-AT^T_CTTCA-T^T_CTG^T_CTT-3′ (probe a) and 5′-TG$^A_{G,T}$ATCCA^T_CT-T^T_CTT-3′ (probe b), were synthesized by the triester method [11]. Probes a and b, labelled with ^{32}P at the 5′-end, were used for hybridization at 34 and 35°C, respectively. The 225-base-pair *Bst*NI-*Acc*I fragment excised from the 5′-terminal region of the cDNA insert of clone pGα28 (see text) was labelled by nick-translation [12] with [α-^{32}P]dCTP and used for hybridization at 60°C. DNA se-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

quencing was carried out according to Maxam and Gilbert [13].

3. RESULTS AND DISCUSSION

A cDNA library derived from bovine cerebral cortex poly(A)$^+$ RNA was screened by hybridization with oligodeoxyribonucleotide probes (probes a and b; see section 2). Probes a and b were synthesized on the basis of the pentapeptide sequences Lys-Gln-Met-Lys-Ile and Lys-Lys-Trp-Ile-His, respectively, which are known to be contained in the α-subunits of both transducin [2–5] and G_o [14], a G-protein purified from brain. From about 1.2×10^5 transformants, one clone (pGα28) hybridizing with both the probes was isolated. Nucleotide sequence analysis of clone pGα28 showed that its cDNA insert encodes an amino acid sequence that is homologous both with the sequence of transducin [2] and with the known partial sequence of G_o [14], but not identical with either. To isolate an upstream cDNA sequence of clone pGα28, we then screened the same cDNA library using a hybridization probe excised from the 5′-terminal region of the cDNA insert of this clone. From about 1.4×10^5 transformants, 4 positive clones were isolated. These clones differed in restriction pattern from clone pGα28. One of them, clone pGSα7, was subjected to nucleotide sequence analysis.

Fig.1 shows the 1599-nucleotide sequence (excluding the poly(dA) tract) of the cDNA insert of clone pGSα7. The sequence of nucleotides 82–183 corresponds precisely to the sequence of 34 amino acids predicted from a cDNA clone that has recently been identified as that encoding the α-subunit of bovine G_s [17]. This indicates that clone pGSα7 carries a cDNA sequence for the G_s α-subunit. The primary structure of this polypeptide was deduced from the cDNA sequence by using the reading frame corresponding to the 34-amino-acid sequence (fig.1). The assignment of the translational initiation site to the methionine codon composed of nucleotides 1–3 is based on the alignment of the deduced amino acid sequence with the sequence of the transducin α-subunit [2] (fig.2). This assignment is supported by the fact that the nucleotide sequence surrounding this ATG triplet agrees with the favoured sequence that flanks functional initiation codons in eukaryotic mRNAs, i.e. CC$\overset{A}{G}$CCAUG(G) [19]. The possibility that the initiating methionine is located upstream of the 5′-end of the cDNA insert of clone pGSα7 cannot be excluded. A translational termination codon (TAA) occurs in frame after the 394th codon specifying leucine. Thus, the α-subunit of bovine G_s consists of 394 amino acid residues (including the initiating methionine) and has a calculated M_r of 45706, which agrees with the reported value [20,21]; it is to be noted that rabbit liver contains an additional G_s polypeptide of M_r 52000 [22].

The amino acid sequences of the α-subunits of G_s and transducin show 42% homology (fig.2); gaps have been counted as one substitution regardless of their length. Some of the highly conserved regions exhibit sequence homology with elongation factor-Tu and *ras* p21 proteins and correspond to functional regions of G-proteins [23–26]. The segment comprising positions 42–60 in the aligned sequences (fig.2) is homologous with the region of elongation factor-Tu and *ras* proteins that is proposed as being involved in interaction with the phosphate groups of the GDP ligand [24–26]. The side chain of the lysine of these proteins corresponding to the lysine at position 53 is considered to contribute to the charge neutralization of one of the phosphate groups. Furthermore, it is suggested that this region of *ras* proteins is involved in GTPase activity [27]. The segment comprising positions 171–175 is homologous with the region of elongation factor Tu and *ras* proteins including the aspartic acid (corresponding to that at position 173) that may form a salt bridge with an Mg^{2+} located close to the β-phosphate group of the GDP ligand [24–26]. This region of *ras* proteins is also thought to be involved in GTPase activity [28,29]. The segment comprising positions 287–300 is homologous with the region of elongation factor Tu and *ras* proteins that is implicated in interaction with the guanine ring [24–26]. The aspartic acid corresponding to that at position 295 may form a hydrogen bond to the amino group, and the asparagine corresponding to that at position 292 to the keto group of the guanine ring [24–26]. In our previous report [2], the regions of *ras* proteins corresponding to positions 171–175 and 287–300 were aligned improperly with the transducin α-subunit because no information concerning structural details of the guanine nucleotide-binding site of elongation factor Tu was available.

```
                                                                                                                    5'-----TCTCGGCCCGCGTGAGGCCGCCCGCGCCGGCCCCGCCGCCGCCGCCGCC   -1
  1                                        10                                        20                                        30
Met Gly Cys Leu Gly Asn Ser Lys Thr Glu Asp Gln Arg Asn Glu Glu Lys Ala Gln Arg Glu Ala Asn Lys Lys Ile Glu Lys Gln Leu
ATG GGC TGT CTC GGA AAC AGC AAG ACC GAG GAC CAG CGC AAC GAG GAG AAG GCG CAG CGC GAG GCC AAC AAG AAG ATC GAG AAG CAG CTG    90

                                        40                                        50                                        60
Gln Lys Asp Lys Gln Val Tyr Arg Ala Thr His Arg Leu Leu Leu Leu Gly Ala Gly Glu Ser Gly Lys Ser Thr Ile Val Lys Gln Met
CAG AAG GAC AAG CAG GTC TAC CGG GCC ACG CAC CGT CTG CTG CTG CTG GGT GCT GGA GAA TCT GGT AAA AGC ACC ATT GTG AAG CAA ATG   180

                                        70                                        80                                        90
Arg Ile Leu His Val Asn Gly Phe Asn Gly Glu Gly Gly Glu Glu Asp Pro Gln Ala Ala Arg Ser Asn Ser Asp Gly Glu Lys Ala Thr
AGG ATC CTG CAT GTT AAT GGG TTT AAT GGA GAG GGC GGC GAA GAG GAC CCG CAG GCT GCA AGG AGC AAC AGC GAT GGT GAG AAG GCC ACC   270

                                       100                                       110                                       120
Lys Val Gln Asp Ile Lys Asn Asn Leu Lys Glu Ala Ile Glu Thr Ile Val Ala Ala Met Ser Asn Leu Val Pro Pro Val Glu Leu Ala
AAA GTG CAG GAC ATC AAA AAC AAC CTG AAA GAG GCC ATT GAA ACC ATC GTG GCC GCC ATG AGC AAC CTG GTG CCC CCT GTG GAG CTG GCC   360

                                       130                                       140                                       150
Asn Pro Glu Asn Gln Phe Arg Val Asp Tyr Ile Leu Ser Val Met Asn Val Pro Asp Phe Asp Phe Pro Pro Glu Phe Tyr Glu His Ala
AAC CCA GAG AAC CAG TTC AGA GTG GAT TAC ATT CTG AGC GTG ATG AAC GTG CCG GAC TTT GAT TTC CCT CCC GAA TTC TAC GAG CAT GCC   450

                                       160                                       170                                       180
Lys Ala Leu Trp Glu Asp Glu Gly Val Arg Ala Cys Tyr Glu Arg Ser Asn Glu Tyr Gln Leu Ile Asp Cys Ala Gln Tyr Phe Leu Asp
AAG GCT CTC TGG GAG GAT GAA GGG GTG CGT GCC TGC TAT GAG CGC TCC AAC GAG TAC CAG CTG ATT GAC TGC GCC CAG TAC TTC CTG GAC   540

                                       190                                       200                                       210
Lys Ile Asp Val Ile Lys Gln Asp Asp Tyr Val Pro Ser Asp Gln Asp Leu Leu Arg Cys Arg Val Leu Thr Ser Gly Ile Phe Glu Thr
AAG ATT GAT GTC ATC AAG CAG GAT GAC TAC GTG CCC AGC GAC CAG GAT CTG CTC CGC TGC CGT GTC CTG ACT TCT GGA ATC TTT GAG ACC   630

                                       220                                       230                                       240
Lys Phe Gln Val Asp Lys Val Asn Phe His Met Phe Asp Val Gly Gly Gln Arg Asp Glu Arg Arg Lys Trp Ile Gln Cys Phe Asn Asp
AAG TTC CAG GTG GAC AAA GTC AAC TTC CAC ATG TTT GAC GTG GGC GGC CAG CGC GAT GAA CGC CGC AAA TGG ATC CAA TGC TTC AAT GAT   720

                                       250                                       260                                       270
Val Thr Ala Ile Ile Phe Val Val Ala Ser Ser Ser Tyr Asn Met Val Ile Arg Glu Asp Asn Gln Thr Asn Arg Leu Gln Glu Ala Leu
GTG ACT GCC ATC ATC TTC GTG GTT GCC AGC AGC AGC TAC AAC ATG GTC ATT CGG GAG GAC AAC CAG ACC AAC CGC CTG CAG GAG GCT CTG   810

                                       280                                       290                                       300
Asn Leu Phe Lys Ser Ile Trp Asn Asn Arg Trp Leu Arg Thr Ile Ser Val Ile Leu Phe Leu Asn Lys Gln Asp Leu Leu Ala Glu Lys
AAC CTC TTC AAG AGC ATC TGG AAT AAC AGA TGG CTG CGC ACC ATC TCT GTG ATT CTG TTC CTC AAC AAG CAA GAT CTG CTG GCT GAG AAA   900

                                       310                                       320                                       330
Val Leu Ala Gly Lys Ser Lys Ile Glu Asp Tyr Phe Pro Glu Phe Ala Arg Tyr Thr Thr Pro Glu Asp Ala Thr Pro Glu Pro Gly Glu
GTC CTT GCT GGA AAA TCG AAG ATT GAG GAC TAC TTT CCA GAA TTT GCT CGC TAC ACT ACT CCT GAG GAT GCG ACT CCC GAG CCC GGA GAG   990

                                       340                                       350                                       360
Asp Pro Arg Val Thr Arg Ala Lys Tyr Phe Ile Arg Asp Glu Phe Leu Arg Ile Ser Thr Ala Ser Gly Asp Gly Arg His Tyr Cys Tyr
GAC CCA CGC GTG ACC CGG GCC AAG TAC TTC ATT CGA GAT GAA TTT CTG AGA ATC AGC ACT GCT AGT GGA GAC GGG CGC CAC TAC TGC TAC  1080

                                       370                                       380                                       390
Pro His Phe Thr Cys Ala Val Asp Thr Glu Asn Ile Arg Arg Val Phe Asn Asp Cys Arg Asp Ile Ile Gln Arg Met His Leu Arg Gln
CCT CAC TTC ACC TGC GCT GTG GAC ACC GAG AAC ATC CGC CGT GTG TTC AAC GAC TGC CGT GAC ATC ATC CAG CGC ATG CAC CTC CGT CAG  1170

Tyr Glu Leu Leu
TAT GAG CTG CTC TAA GAAGGGAACCTCCAGATTTAATTAAGGCCTTAAGCGCAATTAATTAAAAGTAAGATATAATTGTACACGCAGTTGATCACCCACCATAGGGCATGATTAA  1285

CAAAGCAACCTTTCCTTTCCCTCAAGTGATTTTGCGAAACCCCCTTCACCTTACAGCTTGCTTAAATATTCCAAATTTAGAAAGCTTAAGGCAGCCTATAGATTAAGATTAAGAAAAAAA  1405

GGCCACAAATGTTCCCTTCTCTCTTTAAGTAACTACAATAATAGCAGCAAACAGAAATAAATAAATGAAATAAATGAAGCAAATGAAATAAATATCGTGTTGTGCAGCATTAAAAAAATC  1525

AAAATAAAAATTAAATGTGAGC-----3'
```

Fig.1. Nucleotide sequence of the cDNA encoding the α-subunit of bovine G_s. Nucleotide residues are numbered in the 5′ to 3′ direction, beginning with the first residue of the ATG triplet encoding the initiating methionine, and the nucleotides on the 5′-side of residue 1 are indicated by negative numbers; the number of the nucleotide residue at the right-hand end of each line is given. The deduced amino acid sequence of the G_s α-subunit is shown above the nucleotide sequence, and amino acid residues are numbered beginning with the initiating methionine. The 5′-terminal sequence presented does not extend to the 5′-end of the mRNA. The 3′-terminal sequence shown is followed by a poly(dA) tract connected with the vector DNA sequence [8]. The 3′-noncoding region contains 5 and 3 copies of the polyadenylation signals AATAAA [15] (nucleotides 1461–1466, 1465–1470, 1474–1479, 1492–1497 and 1528–1533) and ATTAAA [16] (nucleotides 1228–1233, 1514–1519 and 1535–1540), respectively.

The hydropathy profile [30] and the predicted secondary structures [31] of the G_s α-subunit are generally similar to those of the transducin α-subunit [2]. The region comprising positions 241–253 of both the α-subunits represents a highly hydrophobic segment with predicted secondary structure. This region corresponds to one of the β-strands proposed as being located in the vicinity of the guanine nucleotide-binding site of elongation factor Tu and *ras* proteins [25]. It is also possible that this region is involved in hydrophobic interaction with other subunits of the G-proteins, recep-

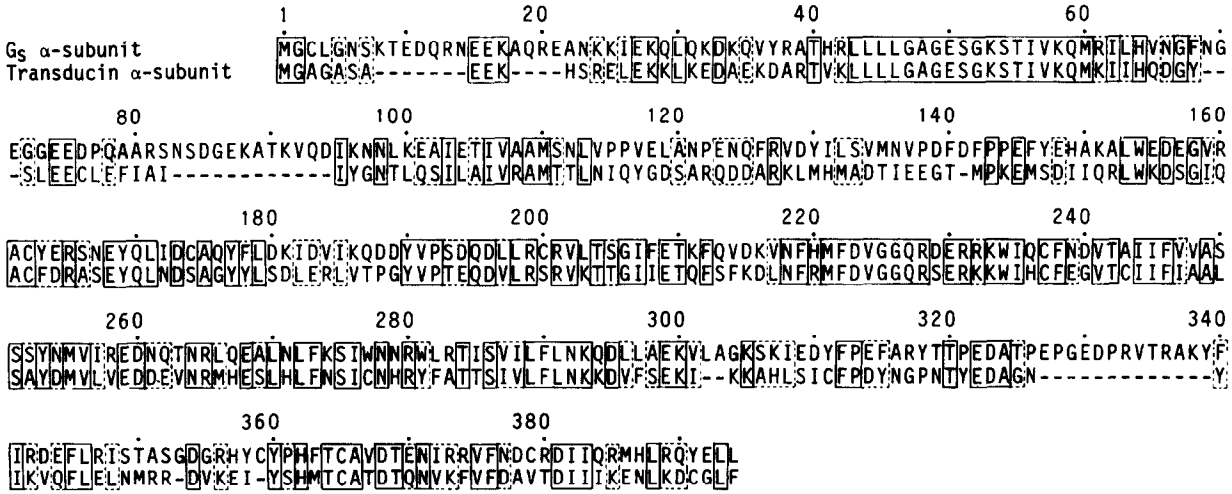

Fig.2. Alignment of the amino acid sequences of the α-subunits of bovine G_s (top) and transducin (bottom). The one-letter amino acid notation is used. The sequence data for the transducin α-subunit have been taken from [2]. Sets of identical residues are enclosed with solid lines, and sets of conservative residues with dashed lines. Conservative amino acid substitutions are defined as pairs of residues belonging to one of the following groups: S, T, P, A and G; N, D, E and Q; H, R and K; M, I, L and V; F, Y and W [18]. Gaps (−) have been inserted to achieve maximum homology. Position numbers in the aligned sequences coincide with amino acid numbers of the G_s α-subunit.

tor or effector proteins or the plasma membrane.

The tetrapeptide sequence Ser-Arg-Val-Lys of the transducin α-subunit (positions 200–203) has been identified as the site that is ADP-ribosylated by cholera toxin [32]. The ADP-ribose is linked to the guanidinium group of the arginine. The G_s α-subunit, which is also ADP-ribosylated by cholera toxin [22], contains an arginine at the corresponding position, and the region surrounding it (positions 190–212) is highly conserved. On the other hand, the carboxy-terminal nonapeptide sequence of the transducin α-subunit, identified as the site of ADP-ribosylation by IAP [14,33], is not well conserved in the G_s α-subunit, which is not ADP-ribosylated by IAP [34].

ACKNOWLEDGEMENTS

We thank Dr Paul Berg and Dr Hiroto Okayama for providing their high-efficiency cloning system and Dr Takashi Miyata and Mr Hidenori Hayashida for computer analysis. This investigation was supported in part by research grants from the Ministry of Education, Science and Culture of Japan, the Mitsubishi Foundation and the Japanese Foundation of Metabolism and Diseases.

REFERENCES

[1] Gilman, A.G. (1984) Cell 36, 577–579.
[2] Tanabe, T., Nukada, T., Nishikawa, Y., Sugimoto, K., Suzuki, H., Takahashi, H., Noda, M., Haga, T., Ichiyama, A., Kangawa, K., Minamino, N., Matsuo, H. and Numa, S. (1985) Nature 315, 242–245.
[3] Lochrie, M.A., Hurley, J.B. and Simon, M.I. (1985) Science 228, 96–99.
[4] Medynski, D.C., Sullivan, K., Smith, D., Van Dop, C., Chang, F.-H., Fung, B.K.-K., Seeburg, P.H. and Bourne, H.R. (1985) Proc. Natl. Acad. Sci. USA 82, 4311–4315.
[5] Yatsunami, K. and Khorana, H.G. (1985) Proc. Natl. Acad. Sci. USA 82, 4316–4320.
[6] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[7] Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408–1412.
[8] Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. 2, 161–170.

[9] Hanahan, D. and Meselson, M. (1980) Gene 10, 63–67.
[10] Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Hirose, T., Inayama, S., Nakanishi, S. and Numa, S. (1982) Nature 295, 202–206.
[11] Ito, H., Ike, Y., Ikuta, S. and Itakura, K. (1982) Nucleic Acids Res. 10, 1755–1769.
[12] Weinstock, R., Sweet, R., Weiss, M., Cedar, H. and Axel, R. (1978) Proc. Natl. Acad. Sci. USA 75, 1299–1303.
[13] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[14] Hurley, J.B., Simon, M.I., Teplow, D.B., Robishaw, J.D. and Gilman, A.G. (1984) Science 226, 860–862.
[15] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[16] Goeddel, D.V., Leung, D.W., Dull, T.J., Gross, M., Lawn, R.M., McCandliss, R., Seeburg, P.H., Ullrich, A., Yelverton, E. and Gray, P.W. (1981) Nature 290, 20–26.
[17] Harris, B.A., Robishaw, J.D., Mumby, S.M. and Gilman, A.G. (1985) Science 229, 1274–1277.
[18] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure, vol.5, suppl.3 (Dayhoff, M.O. ed.) pp.345–352, National Biomedical Research Foundation, Silver Springs, MD.
[19] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[20] Hanski, E., Sternweis, P.C., Northup, J.K., Dromerick, A.W. and Gilman, A.G. (1981) J. Biol. Chem. 256, 12911–12919.
[21] Hanski, E. and Gilman, A.G. (1982) J. Cyclic Nucleotide Res. 8, 323–336.
[22] Northup, J.K., Sternweis, P.C., Smigel, M.D., Schleifer, L.S., Ross, E.M. and Gilman, A.G. (1980) Proc. Natl. Acad. Sci. USA 77, 6516–6520.
[23] Halliday, K.R. (1984) J. Cyclic Nucleotide Protein Phosphorylation Res. 9, 435–448.
[24] La Cour, T.F.M., Nyborg, J., Thirup, S. and Clark, B.F.C. (1985) EMBO J. 4, 2385–2388.
[25] Jurnak, F. (1985) Science 230, 32–36.
[26] McCormick, F., Clark, B.F.C., La Cour, T.F.M., Kjeldgaard, M., Norskov-Lauritsen, L. and Nyborg, J. (1985) Science 230, 78–82.
[27] McGrath, J.P., Capon, D.J., Goeddel, D.V. and Levinson, A.D. (1984) Nature 310, 644–649.
[28] Gibbs, J.B., Sigal, I.S., Poe, M. and Scolnick, E.M. (1984) Proc. Natl. Acad. Sci. USA 81, 5704–5708.
[29] Temeles, G.L., Gibbs, J.B., D'Alonzo, J.S., Sigal, I.S. and Scolnick, E.M. (1985) Nature 313, 700–703.
[30] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[31] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–276.
[32] Van Dop, C., Tsubokawa, M., Bourne, H.R. and Ramachandran, J. (1984) J. Biol. Chem. 259, 696–698.
[33] Manning, D.R., Fraser, B.A., Kahn, R.A. and Gilman, A.G. (1984) J. Biol. Chem. 259, 749–756.
[34] Bokoch, G.M., Katada, T., Northup, J.K., Hewlett, E.L. and Gilman, A.G. (1983) J. Biol. Chem. 258, 2072–2075.