Proc. Natl. Acad. Sci. USA Vol. 89, pp. 7856–7860, September 1992 Cell Biology

Isolation and characterization of a full-length cDNA coding for an adipose differentiation-related protein

(obesity/adipocyte/adipose tissue)

HUI-PING JIANG* AND GINETTE SERRERO[†]

W. Alton Jones Cell Science Center, Inc., 10 Old Barn Road, Lake Placid, NY 12946

Communicated by Gordon H. Sato, April 22, 1992

ABSTRACT We have previously isolated from a 1246 adipocyte cDNA library a cDNA clone called 154, corresponding to a mRNA that increases abundantly at a very early time during the differentiation of 1246 adipocytes and in adipocyte precursors in primary culture. We show here that the mRNA encoded by this cDNA is expressed abundantly and preferentially in mouse fat pads. A full-length cDNA for clone 154 was isolated by the RACE (rapid amplification of cDNA ends) protocol. Sequence analysis of this cDNA indicates that it encodes a protein of the 425 amino acids [tentatively named adipose differentiation-related protein (ADRP)] that does not have any similarity with sequences contained in the GenBank DNA and Protein Identification Resource protein data bases. Immunoblot of 1246 cell extracts with an antibody raised against the expressed ADRP shows that the 1246 cells contain a 50-kDa protein, the production of which increases as the cells differentiate. Localization of ADRP in 1246 cells indicates that ADRP is absent from nuclear and cytosolic fractions and is found as a membrane-associated protein. These results demonstrate that adipocyte differentiation is accompanied by early expression of a mRNA encoding a membrane-associated adipose differentiation related protein that is adipose tissue specific in vivo.

Elucidation of molecular events occurring during adipose differentiation has been greatly facilitated with the isolation and characterization of genes that are activated during the process of differentiation (1). Fat cells express proteins and mRNAs that are distinct from other cell types and that determine their ability to carry out their specialized functions (2, 3). cDNA libraries have been established by using mRNA isolated from several adipogenic cell lines and screened either by differential hybridization (4-7) or by subtraction methods (8). The application of these techniques has led to isolation of cDNAs encoding proteins induced during adipose differentiation and involved in lipid metabolism in the adipocytes. Moreover, cDNAs coding for newly discovered adipose tissue-specific proteins were obtained (4-8). The characterization of these genes and elucidation of the function of the proteins they encode will help to better understand the specialized properties of the adipocytes. In our laboratory, we have been studying adipose differentiation by using a C3H mouse teratoma-derived adipogenic cell line called 1246 (9). It was shown that 1246 cells could proliferate and differentiate in defined medium and that insulin was required for both processes (10). Recently, we have constructed a cDNA library of fully differentiated 1246 cells cultivated in defined medium. Among the cDNAs isolated, clone 154 was particularly interesting and was further characterized. Clone 154 contains a 0.8-kilobase (kb) insert and hybridizes to a 1.7-kb mRNA, which is induced at a very early stage during the

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.

onset of adipose differentiation in 1246 cells and in adipocyte precursors in primary culture (11). Our results show that clone 154 mRNA is induced 100-fold by day 1 of differentiation. This induction of clone 154 mRNA occurs 1 day earlier than lipoprotein lipase, which is considered to be an early marker of adipose differentiation. Expression of clone 154 mRNA is enhanced by the addition of stimulators of differentiation such as dexamethasone and isobutylmethylxanthine and is repressed by factors that are potent inhibitors of 1246 cell differentiation, such as transforming growth factor β , tumor necrosis factor, and epidermal growth factor (11). Sequencing of the clone 154 0.8-kb insert revealed an open reading frame whose nucleotide sequence and deduced amino acid sequence had no similarity with sequences in the GenBank and Protein Identification Resource protein data bases. Considering that clone 154 is the earliest marker of adipocyte differentiation reported to date and appears to encode a newly discovered polypeptide, we have pursued the characterization of the 154 gene product referred to as adipose differentiation-related protein (ADRP). We report here the molecular cloning of the entire coding sequence of ADRP by using the rapid amplification of cDNA ends (RACE) technique (12) and the characterization of ADRP in 1246 cells and in adipose tissue.[‡]

MATERIALS AND METHODS

Animals. Three-month-old C57BL/6J mice from The Jackson Laboratory were used for total RNA isolation from various tissues. Male Chinchilla rabbits (4–5 lb) were purchased from Cornell University (Ithaca, NY) and used for raising antisera.

Northern Blot Analysis. Differentiation of 1246 cells in defined medium was carried out as described (11). Total cellular RNA from undifferentiated and differentiated cells was isolated by the method of Chirgwin *et al.* (13). Poly(A)⁺ RNA was twice purified by chromatography on oligo(dT)-cellulose (Collaborative Research) according to Aviv and Leder (14). The method for total RNA isolation from mouse tissues was described by Chomczynski and Sacchi (15). Northern blot analysis of clone 154 mRNA expression was performed as described (11).

DNA Amplification by PCR. The RACE protocol, as described by Frohman *et al.* (12), was used to obtain cDNAs containing the 5' and 3' coding regions of clone 154. First-strand cDNA was synthesized from $1-5 \mu g$ of poly(A)⁺ RNA from fully differentiated 1246 cells by using a red module kit

Abbreviations: ADRP, adipose differentiation-related protein; RACE, rapid amplification of cDNA ends; GST, glutathione S-transferase.

^{*}Present address: Division of Biology, California Institute of Technology, Pasadena, CA 91125.

[†]To whom reprint requests should be addressed.

[‡]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M93275).

(Invitrogen, San Diego). Forty cycles of amplification were carried out for each PCR. Each cycle consisted of a denaturation step (94°C; 1 min), an annealing step (50°C-55°C; 2 min), and an extension step (72°C; 3 min). The final amplification cycle was followed by a 10-min extension reaction at 72°C. The oligonucleotide 5'-AACCAAGGGCCTCAT-TATGGTC-3' (underlined in Fig. 2) was used as a 3' RACE primer. The two primers used for 5' RACE were 5'-ACTC-GATGTGCTCAACACAGTG-3' (5-RT) and 5'-CGTCG-TAGCCGATGCTTCTCTT-3' (5-Amp) (see Fig. 2). The primer 5-RT was used to direct first-strand cDNA synthesis, whereas primer 5-Amp was used as one of the amplification primers as described by Frohman et al. (12). The cDNA fragment containing the entire coding sequence for ADRP was amplified by PCR using two primers each having a Xma I (Sma I) site. The two primers are HP-1 (5'-CTGACCCGG-GATGGCAGCAGCAGTAGTGGATC-3') and HP-2 (5'-CTGACCCGGGTTTACTGAGCTTTGACCTCAGA-3'). The 3' and 5' RACE products of clone 154 were purified by electrophoresis in an agarose gel and then electroeluted. The purified cDNAs were blunt-end ligated into the Sma I site of the pBluescript SK vector (Stratagene) using T4 DNA ligase (GIBCO/BRL). The PCR-amplified entire coding region for ADRP was digested with the restriction enzyme Xma I (New England Biolabs) and purified by gel electroelution. The purified cDNA was then cloned into the Xma I site in Bluescript SK vector. The recombinant plasmids were purified and the inserts were sequenced by the dideoxynucleotide chain-termination method (16).

Production of Polyclonal Antibodies Against ADRP. pGEX-2T (17) (Pharmacia) was used as a vector to express ADRP in the form of a fusion protein with the 27-kDa glutathione S-transferase (GST; RX: glutathione R-transferase, EC 2.5.1.18) in bacteria as described by Smith and Johnson (17). pGEX-2T vector was first digested with BamHI, then treated with Klenow fragment, and religated. The vector DNA was then digested with Xma I and ligated with the Xma I-treated cDNA fragment containing the entire coding region for ADRP as described above, therefore making ADRP in-frame with GST. The orientation of the cDNA encoding ADRP in the plasmid was determined by Pst I mapping. The GST-ADRP fusion protein was purified by glutathione Sepharose-4B (Pharmacia) affinity chromatography according to the manufacturer's procedure. Affinitypurified GST-ADRP fusion protein (150 μ g of protein per injection) was used as an antigen to raise antisera in chinchilla rabbits using the Ribi adjuvant system (Ribi Immunochem). The GST-ADRP fusion protein was denatured by either 0.5% SDS or 8 M urea to expose the entire primary structure of this protein before the injection. ADRP antiserum was immunoaffinity-purified by GST Sepharose chromatography before being used.

Preparation of Cell Lysate and Subcellular Fractions from 1246 Cells for Immunoblot. Differentiation of 1246 cells in defined medium was carried out as described (11). At each time point, cells were washed with phosphate-buffered saline (PBS) and lysed in 62.5 mM Tris·HCl (pH 6.8) containing 10% glycerol and 2% SDS. The final volume of this cell lysate was measured and the protein concentration was determined by using a micro-BCA protein assay reagent kit (Pierce). At equivalent time points, the cell number from duplicate plates of 1246 cells was determined in order to use cell lysates containing the same number of cells for Western blot analysis. For preparation of subcellular fractions, differentiated 1246 cells were collected on day 8. Briefly, 1246 cells were washed with PBS twice, gently scraped off the culture dishes, and pelleted by low-speed centrifugation. Cells were resuspended in TSM buffer (10 mM Tris HCl, pH 7.4/10 mM NaCl/3 mM MgCl₂) containing EDTA (5 mM), aprotinin (10 μ g/ml), leupeptin (10 μ g/ml), pepstatin A (10 μ g/ml), and

phenylmethylsulfonyl fluoride (0.2 mM) as protease inhibitors and homogenized with 10 strokes at \approx 800 rpm. The homogenate was centrifuged at 600 × g for 5 min. The particulate and cytosol fractions were separated by ultracentrifugation of the supernatant at 100,000 × g for 60 min. The nuclear fraction was obtained by resuspending the 600 × g pellet in sucrose solution (0.25 M sucrose/10 mM Tris·HCl, pH 7.4/2 mM MgCl₂) and recentrifuged at 600 × g. Equal amounts of protein from each fraction were applied for immunoblotting.

For analysis of 154 protein expression in cell and tissue samples by immunoblot, electrophoresis on a SDS/10% polyacrylamide gel was performed (18). Proteins were electrophoretically transferred to $0.2-\mu m$ poly(vinylidene difluoride) membranes (Millipore). The membranes were blocked with TBS (20 mM Tris/500 mM NaCl, pH 7.5) containing 5% nonfat milk for 1 hr at room temperature and rinsed twice with TBS. The membranes were then incubated with a 1:200 dilution of ADRP antiserum for 1–2 hr in TBS/1% bovine serum albumin at room temperature followed by several washes with TTBS (TBS/0.05% Tween 20) and incubated with alkaline phosphatase-conjugated goat anti-rabbit IgG for 1 hr. Bound antibodies were visualized by color development with substrates NBT (nitroblue tetrazolium) and BCIP (5bromo-4-chloro-3-indolyl phosphate) (Bio-Rad).

Immunofluorescence Microscopy. 1246 cells were cultured in chamber slides (Nunc) at a density of 0.2×10^4 cells per cm² and differentiation of the 1246 cells was carried out. At day 8, the 1246 cells were washed with PBS, fixed, and permeabilized by cold methanol as described by Osborn and Weber (19); then they were incubated with either nonimmune serum or anti-ADRP antiserum at room temperature for 1 hr, washed with PBS, and incubated with fluorescein isothiocyanate-conjugated goat anti-rabbit IgG for 1 hr. The bound antibodies were visualized by immunofluorescence microscopy.

Sequence Analysis. Software and data bases for sequence analysis were provided by IntelliGenetics. Sequence similarity searches were performed with the GenBank (release 69) and Swiss-Prot (release 19) data bases.

RESULTS

Tissue-Specific Expression of Clone 154 mRNA *in Vivo*. We have previously shown that clone 154 mRNA is expressed abundantly during differentiation of 1246 cells as well as in adipocyte precursors in primary culture (11). Here we determined the tissue distribution of 154 mRNA *in vivo*. As shown in Fig. 1, clone 154 mRNA was predominantly expressed in mouse inguinal and epididymal fat pads and was virtually undetected in liver, kidney, heart, and brain. The size of 154



FIG. 1. Tissue distribution of clone 154 mRNA. Total RNA from mouse kidney, liver, heart, brain, and inguinal fat pad (FP) was isolated and 15 μ g of RNA from each tissue was loaded on each lane of a denaturing agarose gel for Northern blotting analysis.

mRNA in adipose tissue was similar to that found in 1246 cells and differentiated adipocyte precursors in primary culture.

Molecular Cloning of the Entire Coding Sequence of Clone 154 by Using the RACE Protocol. The 823-base-pair (bp) clone 154 isolated from the 1246 differentiated cDNA library (11) was sequenced by the dideoxynucleotide chain-termination method (16). Computer-assisted analysis of the clone 154 cDNA sequence (823 bp) indicated that clone 154 cDNA contained an open reading frame and had no similarity to DNA sequences in the GenBank data base. Moreover, the putative protein encoded by the major open reading frame of clone 154 cDNA also showed no similarity to protein sequences in the protein data base. This suggested that the clone 154 mRNA encoded a newly discovered protein. Thus, it was of interest to isolate and determine the complete nucleotide sequence of the full-length transcript corresponding to clone 154 and the amino acid sequence of its encoded protein. To obtain the sequence information of the 5' and 3' regions flanking the original 823-bp cDNA, we used the RACE technique (12). The cDNA strand that contained the major open reading frame was assumed to be the (+)-strand. The 3' RACE reaction was performed by using as primer an oligonucleotide complementary to the (-)-strand as shown underlined in Fig. 2. The 3' RACE reaction yielded a 0.3-kb fragment, which hybridized strongly to ³²P-labeled clone 154 cDNA (data not shown). This 0.3-kb fragment was cloned into the *Sma* I site in pBluescript SK vector by blunt-end ligation and sequenced. Three clones from different PCRs containing the 0.3-kb 3' RACE fragment were sequenced in order to eliminate possible PCR errors. Our results showed that this 0.3-kb fragment covered most of the 3' sequence of the original clone 154 with several additional nucleotides followed by a poly(A) tail (Fig. 2). In addition, the 3' RACE results confirmed that our assumed direction of clone 154 is correct.

The 5' RACE reaction was carried out using clone 154specific oligonucleotides as primers. The sequences of both primers, referred to as 5-RT and 5-Amp in Fig. 2, are complementary to the (+)-strand. The primer 5-RT was used to direct first-strand cDNA synthesis, while 5-Amp was used

| | TAGTGGTGAT | CTGGACCGTG | CGGACTTGCT | CGTCCCTCAG | CTCTCCTGTT | AGGCGTCTCT | 60 |
|-------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
| | TTTCTCCAGG | AGGAAAAAAA | GGCAGCAGCA | GTAGTGGATC | CGCAACAGAG | CGTGGTGATG | 120 |
| 14 | | м | AAA | VVDP | QQS | VVM | |
| | AGAGTGGCCA | ACCTGCCCTT | GGTGAGCTCT | ACCTACGACC | TTGTGTCCTC | CGCTTATGTC | 180 |
| 34 | R V A N | LPL | VSS | TYDL | VSS | AYV | |
| | AGTACAAAGG | ATCAGTACCC | GTATTTGAGA | TCCGTGTGTG | AGATGGCCGA | GAAGGGCGTG | 240 |
| 54 | STKD | QYP | YLR | SVCE | MAE | KGV | |
| | AAGACCGTGA | CCTCTGCGGC | CATGACAAGT | GCCCTGCCCA | TCATCCAGAA | GCTGGAGCCA | 300 |
| 74 | ктут | SAA | MTS | ALPI | ΙΩΚ | LEP | |
| | CAAATTGCGG | TTGCCAATAC | CTATGCCTGC | AAGGGGCTAG | ACAGGATGGA | GGAAAGACTG | 360 |
| 94 | QIAV | ANT | YAC | KGLD | RME | ERL | |
| | CCTATTCTGA | ACCAGCCAAC | GTCCGAGATT | GTTGCCAGTG | CCAGAGGTGC | CGTAACTGGG | 420 |
| 114 | PILN | QPT | SEI | VASA | RGA | V T G | |
| | GCGAAGGATG | TGGTGACGAC | TACCATGGCT | GGAGCCAAGG | ATTCTGTAGC | CAGCACAGTC | 480 |
| 134 | AKDV | V T T | TMA | GAKD | S V A | S T V | E 4 0 |
| | TCAGGGGTGG | TGGATAAGAC | CAAAGGAGCA | GTGACTGGCA | GCGTGGAAAG | GACCAAGICI | 540 |
| 154 | SGVV | DKT | K G A | V T G S | VER | T K S | 600 |
| | GTGGTCAATG | GCAGCATCAA | TACAGTTTTG | GGGATGGTGC | AGTTCATGAA | CAGTGGAGTA | 600 |
| 174 | V V N G | SIN | TVL | GMVQ | F M N | SGV | ~~~ |
| | GATAATGCCA | TCACCAAGTC | GGAGATGCTG | GTAGACCAGT | ACTTCCCTCT | CACTCAGGAG | 660 |
| 194 | DNAI | TKS | EML | VDQY | FPL | TQE | |
| | GAGCTGGAGA | TGGAAGCAAA | AAAGGTGGAA | GGATTTGATA | TGGTTCAGAA | GCCGAGCAAC | 720 |
| 214 | ELEM | EAK | KVĒ | GFDM | νοκ | PSN | |
| | TATGAACGGC | TGGAGTCCCT | GTCTACCAAG | CTCTGCTCTC | GGGCTTATCA | CCAGGCTCTC | 780 |
| 234 | YERL | ESL | S T K | LCSR | а у н | QAL | |
| | AGCAGGGTTA | AAGAGGCCAA | ACAAAAGAGC | CAGGAGACCA | TTTCTCAGCT | CCACTCCACT | 840 |
| 254 | SRVK | EAK | QKS | QETI | SQL | нст | |
| | | | | | | | |
| | GTCCACCTCA | | CACCAACAAT | ATCCACACTC | CCARCCAGAA | AATTCACCCT | 900 |
| 274 | GTCCACCTGA | TTGAATTCGC | CAGGAAGAAT | ATGCACAGTG | CCAACCAGAA | AATTCAGGGT | 900 |
| 274 | GTCCACCTGA V H L I | ITGAATTCGC E | CAGGAAGAAT R K N | ATGCACAGTG M H S A | CCAACCAGAA N Q K 5-AMP | AATTCAGGGT I Q G | 900 |
| 274 | GTCCACCTGA V H L I | ► TTGAATTCGC E F A | CAGGAAGAAT R K N | ATGCACAGTG M H S A | CCAACCAGAA N Q K 5-AMP | AATTCAGGGT I Q G | 900 |
| 274 | GTCCACCTGA V H L I GCTCAGGATA | TTGAATTCGC E F A AGCTCTATGT | CAGGAAGAAT R K N CTCGTGGGTG | ATGCACAGTG M H S A GAGTGGAAGA | CCAACCAGAA N Q K 5-AMP GAAGCATCGG | AATTCAGGGT I Q G CTACGACGAC | 900 960 |
| 274 294 | GTCCACCTGA V H L I GCTCAGGATA A Q D K | TTGAATTCGC E F A AGCTCTATGT L Y V | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-PT | ATGCACAGTG M H S A GAGTGGAAGA E W K R | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G | AATTCAGGGT I Q G CTACGACGAC Y D D | 900 960 |
| 274 294 | GTCCACCTGA V H L I GCTCAGGATA A Q D K | TTGAATTCGC E F A AGCTCTATGT L Y V | CAGGAAGAAT R K N CTCGTGGGGTG S W V 5-RT | ATGCACAGTG M H S A GAGTGGAAGA E W K R | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G | AATTCAGGGT I Q G CTACGACGAC Y D D | 900 960 |
| 274 294 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT | TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC F H T | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT | AATTCAGGGT I Q G CTACGACGAC Y D D CGCCCGCAAC | 900 960 1020 |
| 274 294 314 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC E H I TACATCCCAC | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGCTAT L A I | AATTCAGGGT IQGG CTACGACGAC YDD CGCCCGCAAC ARN AGGGTTACCA | 900 960 1020 |
| 274 294 314 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC | CAGGAAGAAT R K N CTCGTGGGGTG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C C | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTG | CCAACCACAAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCAA | AATTCAGGGT I Q G CTACGACGAC Y D D CGCCCGCAAC A R N AGGGTTACCA | 900 960 1020 1080 |
| 274 294 314 334 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CACABCTCC | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T ACCTCACCC | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG T V L V | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCCA N A Q CACCCCCACAT | AATTCAGGGT IQG CTACGACGAC YDD CGCCCGCAAC ARN AGGGTTACCA GLP CTACTCCTA | 900 960 1020 1080 |
| 274 294 314 334 | $ \begin{array}{c} {\rm GTCCACCTGA} \\ {\rm V} & {\rm H} & {\rm L} & {\rm I} \\ {\rm GCTCAGGATA} \\ {\rm A} & {\rm Q} & {\rm D} & {\rm K} \\ {\rm ACCGATGAGT} \\ {\rm T} & {\rm D} & {\rm E} & {\rm S} \\ {\rm CTGACCCAGC} \\ {\rm L} & {\rm T} & {\rm Q} & {\rm Q} \\ {\rm CAGAACATTC} \end{array} $ | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG T V L V GGGGTGATGG | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGCTAT L A I TCAACCCCA N A Q CAGGCGACAT | AATTCAGGGT IQGC CTACGACGAC YDD CGCCCGCAAC ARN AGGGTTACCA GLP CTACTCCGTA | 900 960 1020 1080 1140 |
| 274 294 314 334 354 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q Q | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A | CAGGAAGAAT R K N CTCGTGGGGG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG G V M A GGGGTGATGG G V M A | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCCA N A Q CAGGCGACAT G D I | AATTCAGGGT I Q G CTACGACGAC Y D D CGCCCGCAAC A R N AGGGTTACCA G L P CTACTCCGTA Y S V | 900 960 1020 1080 1140 |
| 274 294 314 334 354 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGGTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC E H I TACATGCACA T C Q CAAACACTTG K H L TAAGAAGTG | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG T V L V GGGGTGATGG G V M A TCCGATGGCG | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCA N A Q CAGGCGACAT G D I TCCTCACATC | AATTCAGGGT IQGCCGCAC YDDD CGCCCGCAAC ARN AGGGTTACCA GLP CTACTCCGTA YSV TAGCAAGGGG | 900 960 1020 1080 1140 1200 |
| 274 294 314 334 354 374 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT A S F | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L TAAGGAAGTG K E V | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG T V L V GGGGTGATGG G V M A TCCGATGGCG S D G V | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCCA N A Q CAGGCGACAT G D I TCCTCACATC L T S | AATTCAGGGT IQGC CTACGACGAC YDD CGCCCGCAAC ARN AGGGTTACCA GLP CTACTCCGTA YSV TAGCAAGGGG SKG | 900 960 1020 1080 1140 1200 |
| 274 294 314 334 354 374 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A CAGCTGCGGAGA | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT A S F AAATGAAGAA | CAGGAAGAAT R K N CTCGTGGGGG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L TAAGGAAGTG K E V ATCCTTAGAT | ATGCACAGTG M H S A GAGTCGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG T V L V GGGGTGATGG G V M A TCCGATGGCG S D G V GAAGTTATGG | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGCTAT L A I TCAACCCCA N A Q CAGGCGACAT G D I TCCTCCACTC L T S ATTACTTTGT | AATTCAGGGT I Q G CTACGACGAC Y D D CGCCCCCCAAC A R N AGGGTTACCA G L P CTACTCCGTA Y S V TAGCAAGGGG S K G TAACAACAG | 900 960 1020 1080 1140 1200 1260 |
| 274 294 314 334 354 374 394 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A CAGCTGCAGA Q L Q K | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCT A S F AATGAAGGA M K E | CAGGAAGAAT R K N CTCGTGGGGGG S W V 5-RT TGAGCACACTC E H I TACATGCCAC T C Q CAAACACCTG K H L TAAGGAAGG K E V ATCCTTAGAT S L D | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG G V M A TCCGATGGCG S D G V GAAGTTATGG E V M D | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCA N A Q CAGGCGACAT G D I TCCTCACATC L T S ATTACTTTGT Y F V | AATTCAGGGT IQGCCGCGAC YDDD CGCCCGCAAC ARN AGGGTTACCA GLP CTACTCCGTA YSV TAGCAAGGG SKG TAACAACACG NNT | 900 960 1020 1080 1140 1200 1260 |
| 274 294 314 334 354 374 394 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A CAGCTGCAGA Q L Q K CCTCTCAACT | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT A S F AAATGAAGGA M K E GGCTGGTAGG | CAGGAAGAAT R K N CTCGTGGGGG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L TAAGGAAGTG K E V ATCCTTAGAT S L D TCCCTTTAT | ATGCACAGAGA M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG T V L V GGGGTGATGG G V M A TCCGATGGCG S D G V GAAGTTATGG E V M D CCTCAGTCTA | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCA N A Q CAGCCGACAT G D I TCCTCACATC L T S ATTACTTTGT Y F V CCGAGGTGAA | AATTCAGGGT IQGCCGCAC YDDD CGCCCGCAAC ARN AGGGTTACCA GLP CTACTCCGTA YSV TAGCAAGGGG SKG TAACAACAGG NNT CAAGGCCAGC | 900 960 1020 1080 1140 1200 1260 1320 |
| 274 294 314 334 354 374 394 414 | | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT A S F AAATGAAGGC M K E GGCTGGTAGG L V G | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L TAAGGAAGTG K E V ATCCCTTAGAT S L D TCCCTTTAT P F Y | ATGCACAGAGA M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG T V L V GGGGTGATGG G V M A TCCGATGGCG S D G V GAAGTTATGG E V M D CCTCAGTCTA P Q S T | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCCA N A Q CAGGCGACAT G D I TCCTCACATC L T S ATTACTTGT Y F V CCGAGGTGAA E V N | AATTCAGGGT I Q G CTACGACGAC Y D D CGCCCGCAAC A R N AGGGTTACCA G L P CTACTCCGTA Y S V TAGCAAGGGG S K G TAACAACACG N N T CAAGGCCAGC K A S | 900 960 1020 1080 1140 1200 1260 1320 |
| 274 294 314 334 354 374 394 414 | $ \begin{array}{cccc} GTCCACCTGA \\ V & H & L & I \\ GCTCAGGATA \\ A & Q & D & K \\ \\ ACCGATGAGT \\ T & D & E & S \\ CTGACCCAGC \\ L & T & Q & Q \\ CAGAACATTC \\ Q & N & I & Q \\ CTCCGCAATG \\ F & R & N & A \\ CAGCTGCAGA \\ Q & L & Q & K \\ CCTCTCAACT \\ P & L & N & W \\ CTGAAGGTCC \\ \end{array} $ | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT A S F AAATGAAGGA M K E GGCTGGTAGG L V G AGCAGTCTGA | CAGGAAGAAT R K N CTCGTGGGGTG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L TAAGGAAGTG K H L TAAGGAAGTG K E V ATCCTTAGAT S L D TCCCTTTAT P F Y GGTCAAAGCT | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG T V L V GGGGTGATGG G V M A TCCGATGGCG S D G V GAAGTTATGG E V M D CCTCAGTCTA P Q S T CAGTAAACCC | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCCA N A Q CAGGCGACAT G D I TCCTCCACATC L T S ATTACTTTGT Y F V CCGAGGTGAA E V N CTCCTTGTCA | AATTCAGGGT I Q G CTACGACGAC Y D D CGCCCCCCAAC A R N AGGGTTACCA G L P CTACTCCGTA Y S V TAGCAAGGGG S K G TAACAACACG N N T CAAGGCCAGC K A S CCAGAGCATG | 900 960 1020 1080 1140 1200 1260 1320 1380 |
| 274 294 314 334 354 374 394 414 425 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A CAGCTGCAGA Q L Q K CCTCTCCACT P L N W CTGAAGGTCC L K V Q | H→ TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCT A S F AATGAAGGA M K E GGCTGGTAGG L V G AGCAGTCTGA Q S E | $\begin{array}{c c} CAGGCAAGAAT R & K & N \\ \hline R & K & N \\ \hline CTCGTGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGG$ | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG G V M A TCCGATGGCG G V M A TCCGATGGCG E V M D CCTCAGTCTA P Q S T CAGTAAACCC Q * | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCA N A Q CAGGCGACAT G D I TCCTCCACAT L T S ATTACTTTGT Y F V CCGAGGTGAA E V N CTCCTTGTCA | AATTCAGGGT IQGCCGCACGAC YDDD CGCCCGCAAC ARN AGGGTTACCA GLP CTACTCCGTA YSV TAGCAAGGG SKG TAACAACACG NNT CAAGGCCAGC KASS CCAGAGCATG | 900 960 1020 1080 1140 1200 1260 1320 1380 |
| 274 294 314 334 354 374 394 414 425 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A CAGCTGCAGA Q L Q K CCTCTCAACT P L N W CTGAAGGTCC L K V Q ATGTTGCTGG | H► TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGGTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT A S F ANATGAAGGA M K E GGCTGGTAGG L V G AGCAGTCTGA Q S E CCAGATGACC | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L TAAGGAAGTG K E V ATCCTTTAGAT S L D TCCCTTTAT P F Y GGTCAAAGCT V K A CCTTTTGCTG | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG T V L V GGGGTGATGG G V M A TCCGATGGCG S D G V GAAGTTATGG E V M D CCTCAGTCTA P Q S T TCAGTAAACCC Q * TATTGAAATT | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCA N A Q CAGCCGACAT G D I TCCTCACATC L T S ATTACTTTGT Y F V CCGAGGTGAA E V N CTCCTTGTCA AACTTGGTAG | AATTCAGGGT I Q G CTACGACGAC Y D D CGCCCGCAAC A R N AGGGTTACCA G L P CTACTCCGTA Y S V TAGCAAGGGG S K G TAACAACACG N N T CAAGGCCAGC K A S CCAGAGCATG ATGGCTTTAG | 900 960 1020 1080 1140 1260 1320 1320 1380 1440 |
| 274 294 314 334 354 374 394 414 425 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A CAGCTGCAGA Q L Q K CCTCTCAACT P L N W CTGAAGGTCC L K V Q | H► TTGGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT A S F AAATGAAGGA M K E GGCTGGTAGG L V G AGCAGTCTGA Q S E CCAGATGACC * | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L TAAGGAAGTG K E V ATCCTTAGAT S L D TCCCTTTTAT P F Y GGTCAAAGCT V K A CCTTTTGCTG | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG G V L V GGGGTGATGG G V M A TCCGATGGCG S D G V GAAGTTATGG E V M D CCTCAGTCTA P Q S T CAGTAAACCC Q * | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCA A Q CAGCCGACAT G D I TCCTCACATC L T S ATTACTTGT Y F V CCGAGGTGAA E V N CTCCTTGTCA | AATTCAGGGT IQGC CTACGACGAC YDD CGCCCGCAAC ARN AGGGTTACCA GLP CTACTCCGTA YSV TAGCAAGGG SKG TAACAACAG NNT CAAGGCCAGG KAS CCAGAGCATG | 900 960 1020 1080 1140 1260 1320 1380 1440 |
| 274 294 314 334 354 374 394 414 425 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A CAGCTGCAGA Q L Q K CCTCTCAACT P L N W CTGAAGGTCC L K V Q ATGTTGCTGG | H→ TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT A S F AAATGAAGGA M K E GGCTGGTAGG L V G AGCAGTCTGA Q S E CCACTCTTA | CAGGAAGAAT R K N CTCGTGGGGTG S W V S-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L TAAGGAAGTG K E V ATCCTTAGAT S L D TCCCTTTAT P F Y GGTCAAAGCT V K A CCTATGCTG | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG T V L V GGGGTGATGG G V M A TCCGATGGCG S D G V GAAGTTATGG E V M D CCTCAGTCTA P Q S T CAGTAAACCC Q * TATTGAAATT CCTCATTATG | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCCA N A Q CAGGCGACAT G D I TCCTCCACATC L T S ATTACTTTGT Y F V CCGAGGTGAA E V N CTCCTTGTCA AACTTGGTAG | AATTCAGGGT IQQGC CTACGACGAC YDDD CGCCCCGCAAC ARN AGGGTTACCA GLP CTACTCCGTA YSV TAGCAAGGGG SKG TAACAACACG NNT CAAGGCCAGC KAS CCAGAGCATG ATGGCTTTAG GCTCAGTTAT | 900 960 1020 1080 1140 1200 1260 1320 1380 1440 |
| 274 294 314 334 354 374 394 414 425 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A CAGCTGCAGA Q L Q K CCTCTCAACT P L N W CTGAAGGTCC L K V Q ATGTTGCTGG GGTCTTGCCC | H→ TTGAATTCGC E F A AGCTCTATGT L Y V CCCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTA A S F AAATGAAGGA M K E GGCTGGTAGG L V G AGCAGTCTGA Q S E CCAGATCTTA CAGCTCTCTA CAGCTCTCTA | CAGGAAGAAT R K N CTCGTGGGGGG S W V 5-RT TGAGCACATC E H I TACATGCAC TACTGCACACTG K H L TAAGGAAGG K E V ATCCTTAGAT S L D TCCCTTTAT P F Y GGTCAAAGCT V K A CCTTTGCTG GACCAAGGG | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG G V M A TCCGATGGCG G V M A TCCGATAGCCTA P Q S T CAGTAAACCC Q * TATTGAAATT CCTCATTATG GTTCCTTAC | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCCA N A Q CAGGCGACAT G D I TCCTCACATC L T S ATTACTTTGT Y F V CCGAGGTGAA E V N CTCCTTGTCA AACTTGGTAG G <u>GTC</u> ACTCACA | AATTCAGGGT I Q G CTACGACGAC Y D D CGCCCCCCAAC A R N AGGGTTACCA G L P CTACTCCGTA Y S V TAGCAAGGG S K G TAACAACACG N N T CAAGGCCAGC K A S CCAGAGCATG ATGGCCTTTAG GCTCAGTTAT AGTGGCCTGT | 900 960 1020 1140 1200 1260 1320 1320 1320 1340 1550 |
| 274 294 314 334 354 374 394 414 425 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A CAGCTGCAGA Q L Q K CCTCTCAACT P L N W CTGAAGGTCCTGG GGTCTTGCCCC GTTAGTCTGCCC | H→ TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGGTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT A S F ANATGAAGGA M K E GGCTGGTAGG L V G AGCATCTGA Q S E CCAGATGACCC TGAGGTCCTA CAGCTCTCTA | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L TAAGGAAGTG K E V ATCCTTTAGAT S L D TCCCTTTAT P F Y GGTCAAAGCT V K A CCTTTTGCTG GAACCAAGGG TGGCACAGGA | ATGCACAGTG M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG G V L V GGGGTGATGG G V M A TCCGATGGCG S D G V CCTCAGTCTA P Q S T TATTGAAATT CCTCATTATG GTTCTCTTAC AAAGCTCAGA | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGGCTAT L A I TCAACGCCA A A Q CAGGCGACAT G D I TCCTCACATC L T S ATTACTTTGT Y F V CCGAGGTGAA E V N CTCCTTGTCA AACTTGGTAG GTCACTCACA | AATTCAGGGT I Q G CTACGACGAC Y D D CGCCCGCAAC A R N AGGGTACCA G L P CTACTCCGTA Y S V TAGCAAGGGG S K G TACAACACG N N T CAAGGCCAGC K A S CCAGAGCATG ATGGCTTTAG GCTCAGTTAT AGTGGCCTGT | 900 960 1020 1080 1140 1200 1320 1320 1320 1380 1440 1550 1560 |
| 274 294 314 334 354 374 394 414 425 | GTCCACCTGA V H L I GCTCAGGATA A Q D K ACCGATGAGT T D E S CTGACCCAGC L T Q Q CAGAACATTC Q N I Q TTCCGCAATG F R N A CAGCTGCAGA Q L Q K CCTCTCAACT P L N W CTGAAGGTCC L K V Q ATGTTGCTGG GTTAGTTGCTCG GTTAGTCTTG GGTCTTTGCCCC | H→ TTGAATTCGC E F A AGCTCTATGT L Y V CCCACTGTGT H C V AGCTCCAGAC L Q T AAGATCAGGC D Q A CTGCCTCCTT A S F AAATGAAGGA M K E GGCTGGTAGG L V G AGCTTCTAA Q S E CCAGATGACC * CAGCTCGGCCCT ATGTTATTG | CAGGAAGAAT R K N CTCGTGGGTG S W V 5-RT TGAGCACATC E H I TACATGCCAG T C Q CAAACACTTG K H L TAAGGAAGTG K E V ATCCTTAGAT S L D TCCCTTTAT P F Y GGTCAAAGCT Q GAGCACAGGG GAGCACAGGA GCTGTAAACG | ATGCACAGAGA M H S A GAGTGGAAGA E W K R GAGTCACGTA E S R T ACTGTCCTGG GGGTGAGG G V M A TCCGATGGCG S D G V GAAGTTATGG E V M D CCTCAGTCTA P Q S T CAGTAAACCC Q * TATTGAAATT CCCTCATTATG GTTCTCTAGA | CCAACCAGAA N Q K 5-AMP GAAGCATCGG S I G CTCTGCTAT L A I TCAACCCCA N A Q CAGGCGACAT G D I TCCTCACATC L T S ATTACTTGT Y F V CCGAGGTGAA E V N CTCCTTGTCA AACTTGGTAG GGCACTCACA CTGCCTGGTG TGCACTTACA | AATTCAGGGT I Q G CTACGACGAC Y D D CGCCCGCAAC A R N AGGGTTACCA G L P CTACTCCGTA Y S V TACCAAGGGG S K G TAACAACAGG N N T CAAGGCCAGC K A S CCAGGCCAGC ATGGCTTAG GCTCAGTTAT AGTGGCCTGT GTCTTGTCTG ATTCATTCAC | 900 960 1020 1080 1140 1260 1320 1380 1380 1440 1550 1620 1620 |

FIG. 2. Nucleotide sequence and deduced amino acid sequence for ADRP. The ADRP nucleotide sequence was obtained by direct sequencing of the cDNA of original clone 154 (11) and cDNAs of 3' and 5' RACE clones. The stop codons prior to the initial codon are boxed and asterisks represent stop codons within the open reading frame. The poly(A) polymerase recognition site is underlined, and arrows indicate the cDNA sequence of original clone 154. The 3' RACE primer sequence is underlined, whereas the 5' RACE primers are labeled 5-RT and 5-Amp. The sequences obtained by RACE protocols were from at least three clones from different PCRs to eliminate possible PCR errors. as one of the amplification primers as described by Frohman et al. (12). The 5' RACE reaction produced an \approx 1.1-kb weak and rather broad band after electrophoresis in a 1.0% agarose gel, which hybridized strongly with ³²P-labeled clone 154 cDNA (data not shown). This DNA fragment was cloned into the Sma I site in Bluescript SK vector and sequenced. Five clones containing the 1.1-kb 5' RACE product obtained from three different PCRs were sequenced to eliminate possible PCR errors. This yielded an additional 850 bp of sequence information to the original 154 cDNA sequence. By combining the sequence of original clone 154 cDNA together with the ones obtained by 3' and 5' RACE reactions, we thereby obtained a cDNA sequence of 1.7 kb. A major open reading frame of 1375 bp was found in this 1.7-kb cDNA, which encodes a protein of 425 amino acids with a molecular mass of 47 kDa called ADRP. In addition, this open reading frame was the same as the one found in the original clone 154. As shown in Fig. 2, the first methionine codon of the open reading frame (box) is at nucleotide 79; two upstream inframe stop codons (underlined) indicate that this represents the initiation methionine codon. There are two stop codons at nucleotides 1354 and 1396, respectively, followed by a stretch of 3' noncoding sequence. A single poly(A) polymerase recognition sequence 5'-AATAAA-3' is located 14 nucleotides upstream of the start of the poly(A) tail.

Both nucleotide sequence and amino acid sequence of the 1.7-kb cDNA were analyzed by a similarity search. No significant similarity could be found with sequences contained in either the GenBank or the protein data base. In addition, computer-assisted analysis of functional motifs such as kinase domains (20), zinc fingers (21), nuclear localization signals (22, 23), signal peptide sequence (24), and leucine zipper sequences (25, 26) indicated that no such consensus sequences are present in ADRP.

Expression of ADRP in 1246 Cells Undergoing Differentiation. Polyclonal antibody was raised against ADRP fused to GST. This antibody was used to follow the expression of ADRP in 1246 cells undergoing differentiation by Western blot analysis as described. Cell lysates of the 1246 cells were collected on day 4, when the cells reach confluency and are still undifferentiated; on day 5, when the cells start to differentiate; and on days 6, 8, and 12, when the cells are undergoing differentiation. As shown in Fig. 3A, a single band with a molecular mass of 50 kDa was visualized by Western blotting with anti-ADRP antibody. Expression of this 50-kDa protein increased as the cells became differentiated. The ADRP could be detected at a very low level on day 4, when 1246 cells reached confluency, it started to increase on day 5, and it reached a maximum level on day 8. This increased expression of the 50-kDa protein is in agreement with the increased expression of clone 154 mRNA observed in 1246 cells undergoing differentiation (11). The 80-kDa band observed in all lanes of the Western blot corresponds to a nonspecific staining as it also appears after incubation with nonimmune serum. The results of Western blot analysis using a nonimmune rabbit serum as primary antibody did not detect the 50-kDa protein (data not shown). Moreover, if anti-ADRP antiserum was first incubated with excess fusion protein before being used in the Western analysis, no 50-kDa band was detected in the differentiated 1246 cells (Fig. 3B).

Localization of ADRP in 1246 Cells. The distribution of 154 protein in 1246 cells was investigated by immunofluorescence staining with anti-ADRP antiserum. Experiments were performed on day 8, when the level of ADRP in 1246 cells reaches a maximum. As shown in Fig. 4, 1246 cells stained positive with anti-ADRP antiserum. Immunofluorescence was mainly localized in the cytoplasm in the vicinity of the plasma membrane. No nuclear localization was observed.

Localization of ADRP was further investigated by immunoblotting 1246 subcellular fractions with anti-ADRP anti-



FIG. 3. Identification of ADRP in 1246 cells by Western blotting analysis. (A) Differentiation of 1246 cells was carried out in defined medium. Whole-cell lysates were collected at various time points as indicated. Cell lysates equivalent to 5×10^4 cells for each time point were used for Western blotting. Samples were electrophoresed on a 10% polyacrylamide gel according to the method of Laemmli (18) and transferred to a filter. The antiserum raised against ADRP-GST fusion protein was used for immunoblot analysis. (B) Differentiated 1246 cell extracts were subjected to SDS/PAGE as described in A. After transfer to the filter, the extract was subjected to immunoblot analysis with anti-154-GST antibody in the absence (lane 1) or presence (lane 2) of excess 154-GST fusion protein.

body (Fig. 5). The three subcellular fractions were nuclear fraction (lane c), cytosol fraction (lane d), and particulate fraction (lane b). Since there was no conventional signal peptide sequence in the amino acid sequence of clone 154encoded protein, indicating that ADRP was not secreted, conditioned medium of fully differentiated 1246 cells (lane a) was included as a negative control. The results of Fig. 5 revealed that ADRP was detected predominantly in the particulate fraction, whereas no detectable amount of this protein was found in the cytosolic and nuclear fractions.

DISCUSSION

Using the RACE technique (12), we cloned a full-length cDNA from 1246 adipocytes that encodes a newly identified protein called ADRP, which is adipose tissue specific *in vivo*. The full-length cDNA for ADRP is 1.7 kb and contains an open reading frame of 1375 bp, which encodes a protein of 425 amino acids. The sequence of ADRP cDNA was confirmed



FIG. 4. Localization of ADRP in 1246 cells by immunofluorescence. Differentiated 1246 cells were cultivated in chamber slides as described. The 1246 cells were used on day 8. Cells were fixed and permeabilized with cold methanol, stained with anti-ADRP antiserum or control serum, and then stained with fluorescein isothiocyanate-conjugated goat anti-rabbit IgG.



FIG. 5. Localization of ADRP in 1246 cells by immunoblotting. Subcellular fractions of 1246 adipocytes were obtained as described. Twenty micrograms of protein from conditioned medium of 1246 adipocytes (lane a), microsomal fraction (lane b), nuclear fraction (lane c), or cytosol fraction (lane d) was used for Western blot analysis. A cell lysate from 1246 adipocytes (lane e) was used as a positive control.

by sequencing exons of a genomic clone (D. Eisinger and G.S., unpublished results). Polyclonal antibodies were raised against the entire ADRP protein expressed as a fusion protein with GST. By immunoblotting, the polyclonal antibody specifically recognizes a 50-kDa protein, which correlates with the calculated molecular mass of ≈ 47 kDa based on the deduced amino acid sequence. We have previously shown that mRNA corresponding to ADRP is induced in differentiated 1246 cells 1 day after the cells reach confluency (day 5) (11). In this paper, we have shown by immunoblotting that the level of the 50-kDa ADRP increases as the cells differentiate, which is in agreement with the increased expression of its mRNA.

Localization of ADRP protein in 1246 cells by immunoblotting indicates that this protein is found not in the cytosol or nuclear fractions but in the particulate fraction. This suggests that it is a membrane-associated protein. Sequence analysis of the protein showed no typical leader sequence or transmembrane domain, which is generally associated with either incorporation into membranes or secretion into the extracellular space. Moreover, analysis of hydropathic characteristics of clone 154 amino acid sequence did not provide any clue with regard to the location of a structure domain involved in membrane interaction (data not shown). It is possible that the membrane association of ADRP protein could occur during subcellular fractionation of the cells. However, immunofluorescence staining of fixed 1246 cells supports the possibility that this protein is preferentially localized to the inner cell wall. It is not known whether ADRP is always membrane associated or whether this association is temporary, as in the case of translocated proteins such as protein kinase C and phospholipase A₂ (27, 28).

Computer-assisted analysis of nucleotide sequence and deduced amino acid sequence for ADRP revealed no similarity to sequences in DNA and protein data bases. Thus, the function of this protein is at present not known. Northern blot analysis of RNA extracted from various tissues failed to detect mRNA corresponding to this protein in nonadipose tissues, indicating that ADRP mRNA is preferentially expressed in adipose tissue and may be developmentally regulated *in vivo*. Interestingly, we have shown that the levels of ADRP mRNA and protein are elevated in the fat pads of *ob/ob* mice when compared to their normal counterparts (data not shown).

On the basis of these results it is interesting to speculate that ADRP may play a role in allowing adipocytes to perform their specialized functions in energy storage and energy balance. Further studies are necessary to investigate this possibility and to elucidate the function of ADRP in adipocytes. Moreover, as ADRP is a newly identified protein, its characterization will provide new insight into the biochemistry of adipocytes.

We wish to thank Drs. Stephen Harris, Jun Hayashi, and Daniel Sussman for helpful advice and critical reading of the manuscript; Carol Baine and Marina LaDuke for the illustrations; and Valerie Oliver for preparation of the manuscript. This work was supported by Grant DK38639 from the National Institutes of Health and Grant 2003 from the Council for Tobacco Research.

- 1. Spiegelman, B. M. (1988) Trends Genet. 4, 203-207.
- 2. Sidhu, R. S. (1979) J. Biol. Chem. 254, 11111-11118.
- Spiegelman, B. M. & Green, H. (1980) J. Biol. Chem. 255, 8811-8818.
- Chapman, A. B., Knight, D. M., Dieckmann, B. S. & Ringold, G. M. (1984) J. Biol. Chem. 259, 15548-15555.
- Spiegelman, B. M., Frank, M. & Green, H. (1983) J. Biol. Chem. 258, 10083–10089.
- Bernlohr, D. A., Angus, W., Lane, M. D., Bolanowski, M. & Kelly, T. J., Jr. (1984) Proc. Natl. Acad. Sci. USA 81, 5468– 5472.
- Dani, C., Doglio, A., Amri, E. Z., Bardon, S., Fort, P., Bertrand, B., Grimaldi, P. & Ailhaud, G. (1989) J. Biol. Chem. 264, 10119-10125.
- Smith, P. J., Wise, L. S., Berkowitz, R., Wan, C. & Rubin, C. S. (1988) J. Biol. Chem. 263, 9402-9408.
- 9. Serrero, G. & Khoo, J. (1982) Anal. Biochem. 120, 351-359.
- Serrero, G. & Sato, G. (1982) in Growth of Cells in Hormonally Defined Media, Cold Spring Harbor Conference on Cell Proliferation, eds. Sato, G., Pardee, G. & Sirbasku, D. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), Vol. 9, pp. 943–955.
- 11. Jiang, H. P., Harris, S. E. & Serrero, G. (1992) Cell Growth Differ. 3, 21-30.
- Frohman, M. A., Duch, M. K. & Martin, G. R. (1988) Proc. Natl. Acad. Sci. USA 85, 8998-9002.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. S. (1979) *Biochemistry* 18, 5294-5299.
- Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
- 15. Chomczynski, P. & Sacchi, M. (1987) Anal. Biochem. 162, 156–159.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 17. Smith, D. B. & Johnson, K. S. (1988) Gene 67, 31-40.
- 18. Laemmli, U. K. (1970) Nature (London) 277, 680-685.
- 19. Osborn, M. & Weber, K. (1977) Cell 12, 561-571.
- 20. Hanks, S. K., Quinn, A. M. & Hunter, T. (1988) Science 241, 42–52.
- Rauscher, F. J., III, Morris, J. F., Tournay, O. E. & Curran, T. (1990) Science 250, 1259–1262.
- 22. Lee, B. A., Maher, D. W., Hannink, M. & Donoghue, D. J. (1987) Mol. Cell. Biol. 7, 3527-3537.
- Goldfarb, D. S., Gariepy, J., Schoolnik, G. & Kornberg, R. D. (1986) Nature (London) 322, 641-644.
- 24. Heijne, G. V. (1986) Nucleic Acids Res. 14, 4683-4690.
- Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) Science 240, 1759-1764.
- Liou, H.-C., Boothby, M. R., Finn, P. W., Davidson, R., Nabavi, N., Zelenik-Le, N. J., Ting, J. P.-Y. & Glimcher, L. H. (1990) Science 247, 1581–1584.
- Parker, P. J., Coussens, L., Totty, N., Rhee, L., Young, S., Chen, E., Stabel, S., Waterfield, M. D. & Ullrich, A. (1986) *Science* 233, 853-859.
- Clark, J. D., Lin, L.-L., Kriz, R. W., Ramesha, C. S., Sultzman, L. A., Lin, A. Y., Milona, N. & Knopf, J. L. (1991) Cell 65, 1043–1051.