J Hum Genet (2000) 45:12–17

ORIGINAL ARTICLE

Takao Kinjo · Minoru Isomura · Teruo Iwamasa
Yusuke Nakamura

Molecular cloning and characterization of two novel genes on chromosome 8p21.3

Received: August 30, 1999 / Accepted: October 6, 1999

Abstract Through large-scale sequencing of genomic DNA from human chromosome 8p22–p21.3 we have isolated two novel genes, designated *GK1* and *G5*. Their predicted products showed no significant similarity to any known proteins in public databases. A comparison of *GK1* cDNA sequences, which encode a 1270-amino-acid protein, with corresponding genomic DNA sequences revealed that this gene consists of 15 exons and spans an approximately 113-kb genomic region. Northern blot analysis revealed ubiquitous expression of 7.0- and 4.4-kb transcripts; in addition, we detected a 5.0-kb skeletal muscle-specific transcript and a 4.0-kb transcript specifically expressed in heart and pancreas. Computer and immunocytochemical analyses of a *GK1* Green fluolesent protein (GFP) fused construct indicated that the gene product, which contains putative leucine-zipper domains, was likely to be a mitochondrial protein. The other novel gene, *G5*, expressed four transcripts (4.2, 2.2, 1.7, and 1.0-kb) ubiquitously; the longer three transcripts, which differed only in the 3′-non coding region, encoded identical 397-amino-acid peptides. The *G5* gene consists of 14 exons and spans approximately 52 kb of genomic DNA; its deduced 397-amino acid product appears to contain coiled-coil domains and a proline-rich region, and to be located in cytoplasm.

Key words Large-scale DNA sequencing · In silico cDNA cloning · 8p22–p21.3 · Leucine-zipper domain · Coiled-coil domain

T. Kinjo · M. Isomura · Y. Nakamura (✉)
Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
Tel. +81-3-5449-5372; Fax +81-3-5449-5433
e-mail: yusuke@ims.u-tokyo.ac.jp

T. Kinjo · T. Iwamasa
Second Department of Pathology, Faculty of Medicine, The University of the Ryukyus, Okinawa, Japan

Introduction

Recent progress in the development of large-scale DNA-sequencing technologies and bioinformatics has been contributing to rapid and efficient identification of novel genes. High-throughput automated fluorescent sequencers, as well as high-capacity computation using sophisticated algorithms, have already yielded a huge quantity of DNA sequences from human and other genomes.

Identification of genes from within genomic DNA sequences has become easier since the advent of two computer-assisted in silico cDNA cloning methods. One of them involves searching for homologies between new genomic DNA sequences and the expressed sequence tag (EST) database; more than 800000 ESTs, representing 40000–50000, genes have been archived so far (Rowen et al. 1997), and the number is still growing. The second method applies programs such as GRAIL (Uberbacher and Mural 1991), FEXH (Solovyev et al. 1994), and GENSCAN (Burge and Kerlin 1997) to predict exons from anonymous genomic sequences with reliable sensitivity and accuracy (Claverie et al. 1997; Elkahloun et al. 1997; Ishikawa et al. 1998; Daigo et al. 1999). Thus, sequencing large genomic regions and "trapping" genes within those sequences with the assistance of computer analysis has become a highly efficient and powerful approach for identification of novel genes.

We have been performing large-scale genomic DNA sequencing of the 8p22–p21.3 region, where loss of heterozygosity (LOH) is frequently observed in carcinomas of liver, lung, colorectum (Emi et al. 1992), and prostate (Bergerheim et al. 1991). Computer analysis of these genomic DNA sequences by GRAIL2 (Xu et al. 1994) and GENSCAN identified 11 genes in this region. We report here the isolation and characterization of 2 of them, both novel: *GK1* and *G5*. The product of *GK1* is likely to be located in mitochondria and may be associated with protein-protein interaction. The *G5* product may also be associated with protein-protein interaction, but it appears to be located in cytoplasm.

Materials and methods

DNA sequencing

Two BAC clones (15504 and 54f05) and three cosmids (A266, 3040, and A260) corresponding to genomic DNA on human chromosome 8p21.3 were sequenced by a combination of shotgun and primer-walking methods (Daigo et al. 1999). DNA sequences were assembled by means of ABI "Auto Assembler" computer software (Norwalk, CT, USA). Direct-cosmid sequencing, using primers designed from the end-sequences of assembled segments, filled gaps between segments.

Isolation of cDNA

We used GRAIL2 and GENSCAN to analyze genomic DNA sequences from the target region, and the Repeat Masker program (Washington University, Seattle, Washington, USA) to screen DNA sequences for interspersed repeats such as Alu and L1. The BLAST algorithm (Altschul et al. 1997) was used to compare similarities and identities with known genes and EST sequences in the GenBank and EMBL databases. To investigate whether predicted candidate exons were actually transcribed, exon-connection experiments by reverse-transcriptase polymerase chain reaction (RT-PCR) were performed as described previously (Horii et al. 1993). Connected cDNA fragments were used as probes to screen cDNA libraries constructed from human umbilical-vein endothelial cells (HUVEC) and K562 (a cell line that originated from chronic myeloblastic leukemia).

Northern blot analysis

Human multiple-tissue blots (Clontech, Palo Alto, CA, USA) were hybridized with cDNA fragments labeled by random-oligonucleotide priming. Prehybridization, hybridization, and washing were performed according to the supplier's recommendations. The blots were auto-radiographed with intensifying screens for 72 h at −80°C.

Green fluoresent protein (GFP)-fused protein assay

To investigate the intracellular locations of *GK1* and *G5* proteins, we constructed plasmids to express GFP-*GK1* or GFP-*G5* constructs. The coding sequences of each gene, including the translation-initiation site, were amplified by Pyrobest DNA polymerase (TaKaRa, Otsu, Shiga, Japan) using human uterus cDNA as template, and inserted into plasmid pEGFP-N1 (Clontech, Palo Alto, CA, USA). The constructed vectors were verified by sequencing on an ABI PRISM 377 instrument and purified by equilibrium centrifugation in CsCl-ethidium bromide gradients.

HepG2 (a cell line that originated from hepatocellular carcinam) cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, in 35-mm dishes. At 40%–50% confluence, the cultures were transfected by lipofection with 2 μg of one or the other constructed plasmid. After a 4-h incubation, the cells were washed with fresh medium and incubated for an additional 24 or 48 h. After one or the other of these incubation times, cells were stained with 4′, 6-diamidino-2-phenylindole (DAPI) and examined by fluorescence microscopy.

Immunocytochemical analysis

After transfected HepG2 cells were fixed for 1 h in 4% paraformaldehyde/phosphate-buffered saline (PBS), their membranes were rendered permeable by incubation in 0.2% Triton-X100 in PBS. The cells were incubated overnight in blocking serum (2.5% normal goat serum in PBS), then for 1 h at room temperature in the presence of polyclonal murine antibody to human mitochondria, for 1 h in PBS containing rhodamine-conjugated goat anti-mouse secondary antibody, and for 30 min for DAPI staining. Images were acquired by fluorescence microscopy.

Results

Isolation of *GK1* cDNA clone

Computer analysis of genomic DNA sequences corresponding to two BAC clones, 15504 and 54f05, and a cosmid clone, A266, predicted 63 possible exons (38 fragments by GRAIL2 with an excellent score and 42 fragments by GENSCAN; 17 were predicted as exons by both programs). Of these exon candidates, 5 (all predicted by GENSCAN) were identical to parts of a single EST sequence (GenBank Accession No. AA459720). To learn whether these exon-like fragments were in fact transcribed in human tissues, we synthesized oligonucleotides corresponding to the candidate regions and performed exon-connection experiments (Fearon et al. 1990). RT-PCR experiments and subsequent DNA sequencing confirmed that all 5 exon-candidate segments were parts of the same transcript. To define the 5′- and 3′- ends of the transcript we screened two human cDNA libraries, one derived from umbilical-vein endothelial cells and the other from K562 cells, and obtained 16 positive clones. The assembled cDNA sequences consisted of 4252 nucleotides, including an open reading frame of 1308 nucleotides. Northern blot analysis had detected transcripts of 7.0 and 4.4 kb in all human tissues examined; we also detected a 5.0-kb skeletal muscle-specific transcript and a 4.0-kb transcript specifically expressed in heart and pancreas (data not shown); this cDNA was considered to correspond to the 4.4-kb transcript. To obtain the cDNA corresponding to the 7.0-kb transcript, we performed additional exon-connection experiments and obtained a cDNA of 6362 nucleotides, including an open reading frame of 3810 nucleotides. Comparison of cDNA and genomic DNA sequences defined 15 exons spanning approximately 113 kb of genomic DNA (Fig. 1). Table 1 documents the nucleotide sequences around the exon-intron boundaries.

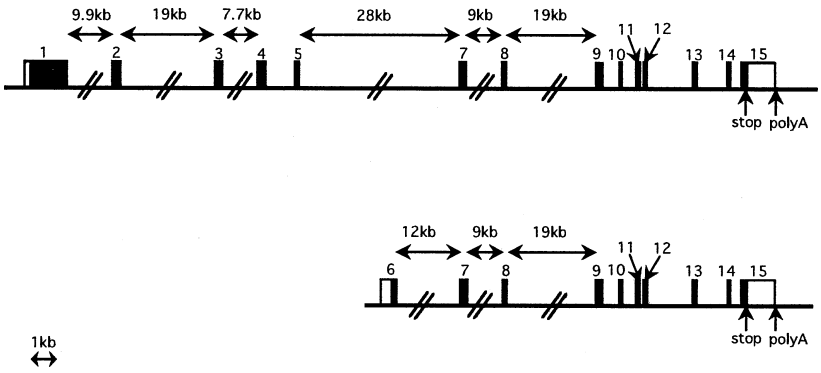

Fig. 1. Genomic structure of *GK1*, showing locations of exons in each of the transcripts generated by alternative splicing. Coding regions are indicated as *filled boxes*

Table 1. Exon — intron boundary sequences of the *GK1* gene

Exon number	Exon length (bp)	cDNA position	Splice acceptor	Splice donor	Intron number	Intron length (bp)
1	196	1–196	attttcac**ag** GGCTCTGCTT	TCCAGTTCTG **gt**aagaataa	1	19,775
2	162	197–358	aaaccccca**ag** GAAAGCCTAC	AGCAACAATT **gt**aagtcgac	2	7,769
3	135	359–493	tttattac**ag** CTGGTAATGC	CCTCCAAAAG **gt**acagtcct	3	2,513
4	39	494–532	tatatctt**ag** GTCCTTCGAG	CTTAATGCAG **gt**gagtgacc	4	15,549
5	410	533–942	gcatgcac**ag** TGTGATGTGG	TTCCACACAG **gt**ctgtctcg	5	12,704
6	215	943–1157	cgttttgc**ag** TTGAAAAGAG	GCTGTCTGAG **gt**aaggaggt	6	9,041
7	67	1158–1224	ttttccgc**ag** CGGGAGGAAG	GGAGAGCTAG **gt**aagagctt	7	19,176
8	203	1225–1427	ctgttttc**ag** TCACTGCTTC	GCAAGAGCAG **gt**ctgtgctt	8	1,198
9	106	1428–1533	ttgcttac**ag** TTTGACAACT	TCCCTTTCAG gcaaggatgc	9	1,062
10	74	1534–1607	attttccc**ag** AAATTAAGAA	ATCGCTAGAG **gt**gggtgaat	10	143
11	96	1608–1703	tgtgttttt**ag** AAGCAAATCA	AGCAAATTTG **gt**aagttgtg	11	3,224
12	117	1704–1820	tcctttcc**ag** AAAAATCCTC	GGAGAAACTG **gt**atgtttcc	12	2,767
13	98	1821–1918	ttatgcca**ag** GTGGACTTCA	CAATCTCAAG **gt**aaaaataa	13	842
14	2,334	1919–4252	tttcttca**ag** GCAGCTTTCC	GAATACTCAA gaattgcatc	14	

Characterization of the *GK1* gene product

A search for homology of the predicted product of the 4.4-kb *GK1* transcript (436 amino acids), using the FASTA program (GenomeNet, Human Genome Center, University of Tokyo, Tokyo, Japan) revealed 24% identity in amino acids (98 identical in 409 amino acids) with Rat synaptonemal complex protein 1 (Meuwissen et al. 1992). The PSORT sequence analysis program (Horton and Nakai 1996) predicted that *GK1* is likely to be a mitochondrial protein which has two coiled-coil structures with a leucine-zipper pattern within the coiled-coil conformation (Fig. 2). The N-terminus of *GK1* is rich in basic amino acids.

Immunocytochemical analysis, using a GFP-fused *GK1* protein, revealed granular bright green signals in the cytoplasm of HepG2 cells (Fig. 3a). The red signals that resulted from staining with antibody to human mitochondria (Fig. 3b) overlapped the green signals of the fusion product (Fig. 3c), indicating that *GK1* was indeed a mitochondrial protein, as the computer analysis had predicted.

Isolation of a *G5* cDNA clone

Computer analysis of genomic DNA sequences present in two cosmid clones (3040 and A260) predicted 25 possible exons (8 by GRAIL2, 8 by GENSCAN, and 9 by both). Four of them (3 predicted by GENSCAN, 1 by both programs) were identical to one EST sequence (GenBank Accession No. N34601). Using this EST sequence as a probe, we searched the EST database and obtained an overlapping EST (GenBank Accession No. H02721) which contained DNA sequences identical to 3 additional computer-predicted exons (1 by GENSCAN, 2 by both programs). RT-PCR and subsequent DNA sequencing revealed that all 7 exon-candidate segments were, in fact, parts of the same transcript. We then screened human testis and umbilical-vein endothelial-cell (HUVEC) cDNA libraries, and obtained ten positive clones. The assembled cDNA sequences consisted of 3296 nucleotides, with an open reading frame of 1191 nucleotides. The comparison of cDNA and genomic DNA sequences finally defined 12 exons spanning approximately 52 kb of genomic DNA (Fig. 4). Table 2 summarizes the nucleotide sequences around the exon-intron boundaries of the *G5* gene.

Northern blot analysis revealed that transcripts of four different sizes (4.2, 2.2, 1.7 and 1.0 kb), corresponding to 4 types of cDNA, were expressed in all human tissues examined. However, skeletal muscle, heart, and testis seemed to express the 2.2-kb transcript more abundantly than other tissues (data not shown).

Characterization of the *G5* gene product

A search of protein databases, using the FASTA program to detect homologies with the predicted 397-amino-acid protein, revealed that the amino acid sequence of *G5* pos-

sessed 26% identity to the tyrosine phosphatase 99A precursor of *Drosophila* (Rubin et al. 1991). The homology was limited to a region containing casein kinase 2 phosphorylation sites and N-myristoylation sites of *Drosophila* tyrosine phosphatase; no significant homology was seen in the catalytic tyrosine-phosphatase domain.

The PSORT sequence analysis program (Horton and Nakai 1996) predicted *G5* to be a nuclear protein. However, immunocyochemical analysis of GFP-*G5* fused protein detected bright green signals in the cytoplasm of HepG2 cells (Fig. 6), indicating that the gene product is more likely to be a cytoplasmic protein. *G5* protein appeared to have two coiled-coil structures in which three or four heptad repeats of hydrophobic residues, including valine, were present in the coiled-coil conformation. In addition, two proline-rich regions (18% and 21% in codons 70–163 and 191–232, respectively) lie between the coiled-coil domains (Fig. 5).

```
1    MTDDNSDDKIEDELQTFFTSDKDGNTHAYNPKSPPTQNSSASSVNWNSANPDDMVVDYET   60
61   DPAVVTGENISLSLQGVEVFGHEKSSSDFISKQVLDMHKDSICQCPALVGTEKPKYLQHS  120
121  CHSLEAVEGQSVEPSLPFVWKPNDNLNRAGYCDALELNQTFDMTVDKVNCTFISHHAIGK  180
181  SQSFHTAGSLPPTGRRSGSTSSLSYSTWTSSHSDKTHARETTYDRESFENPQVTPSEAQD  240
241  MTYTAFSDVVMQSEVFVSDIGNQCACSSGKVTSEYTDGSQQRLVGEKETQALTPVSDGME  300
301  VPNDSALQEFFCLSHDESNSEPHSQSSYRHKEMGQNLRETVSYCLIDDECPLMVPAFDKS  360
361  EAQVLNPEHKVTETEDTQMVSKGKDLGTQNHTSELILSSPPGQKVGSSFGLTWDANDMVI  420
421  STDKTMCMSTPVLEPTKVTFSVSPIEATEKCKKVEKGNRGLKNIPDSKEAPVNLCKPSLG  480
481  KSTIKTNTPIGCKVRKTEIISYPRPNFKNVKAKVMSRAVLQPKDAALSKVTPRPQQTSAS  540
541  SPSSVNSRQQTVLSRTPRSDLNADKKAEILINKTRKQQFNKLITSQAVHVTTHSKNASHR  600
601  VPRTTSAVKSNQEDVDKASSSNSACETGSVSALFQKIKGILPVKMESAECLEMTYVPNID  660
661  RISPEKKGEKENGTSMEKQELKQEIMNETFEYGSLFLGSASKTTTTSGRNISKPDSCGLR  720
721  QIAAPKAKVGPPVSCLRRNSDNRNPSADRAVSPQRIRRVSSSGKPTSLKTAQSSWVNLPR  780
781  PLPKSKASLKSPALRRTGSTPSIASTHSELSTYSNNSGNAAVIKYEEKPPKPAFQNGSSG  840
841  SFYLKPLVSRAHVHLMKTPPKGPSRKNLFTALNAVEKSRQKNPRSLCIQPQTAPDALPPE  900
901  KTLELTQYKTKCENQSGFILQLKQLLACGNTKFEALTVVIQHLLSEREEALKQHKTLSQE  960
961  LVNLRGELVTASTTCEKLEKARNELQTVYEAFVQQHQAEKTERENRLKEFYTREYEKLRD 1020
1021 TYIEEAEKYKMQLQEQFDNLNAAHETSKLEIEASHSEKLELLKKAYEASLSEIKKGHEIE 1080
1081 KKSLEDLLSEKQESLEKQINDLKSENDALNEKLKSEEQKRRAREKANLKNPQIMYLEQEL 1140
1141 ESLKAVLEIKNEKLHQQDIKLMKMEKLVDNNTALVDKLKRFQQENEELKARMDKHMAISR 1200
1201 QLSTEQAVLQESLEKESKVNKRLSMENEELLWKLHNGDLCSPKRSPTSSAIPLQSPRNSG 1260
1261 SFPSPSISPR                                                    1270
```

Fig. 2. Deduced amino acid sequence of the GK1 product. Coiled-coil regions are indicated by *bold underlining*, and leucine-zipper regions by *dotted underlining*

Discussion

By combining genomic DNA sequencing with a computer-assisted in silico exon-identification method, we have isolated and characterized two novel human genes, termed *GK1* and *G5*, on chromosome 8p21.3. *GK1* appears to be a mitochondrial protein, whose N-terminal region is rich in basic amino acids, and contains two putative leucine-zipper motifs and 27 phosphorylation sites. Some known mitochondrial proteins also contain the leucine-zipper motif. For example, human mitochondrial transcription-termination factor, mTERF, binds to DNA as a monomer, although multiple intramolecular leucine-zipper interactions are required to bring the two basic domains into close register with the mTERF-target DNA sequence (Fernandez-Silva et al. 1997). Tim23, a key factor for importation of proteins across the inner mitochondrial membrane, forms dimers through its leucine-zipper motif in a membrane voltage-dependent manner (Bauer et al. 1996). The leucine-zipper domain of *GK1* may also be important for maintaining this protein's biological function.

The second gene reported here, *G5*, showed partial similarity to the tyrosine phosphatase 99A precursor of *Drosophila* (Rubin et al. 1991). Its encoded amino acid sequence contains two coiled-coil motifs and two proline-rich regions.

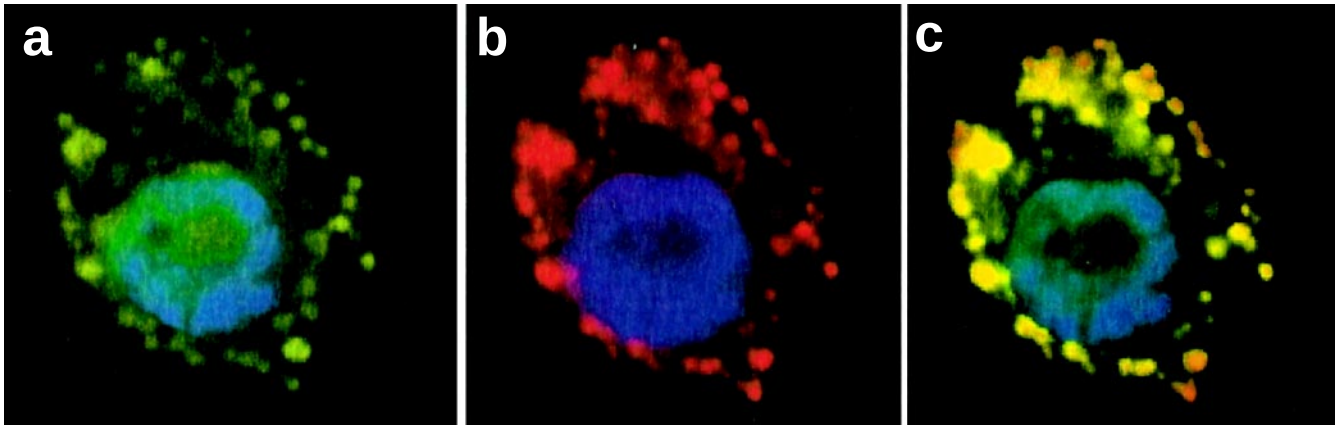

Fig. 3. **a** Green fluoresent protein (GFP)-fused protein assay. *Bright green signals* indicate the location of GFP. **b,c** Immunocytochemistry. *Red signals* indicate the location of mitochondria, and *yellow signals* indicate the overlap of GFP-fused *GK1* with the mitochondria

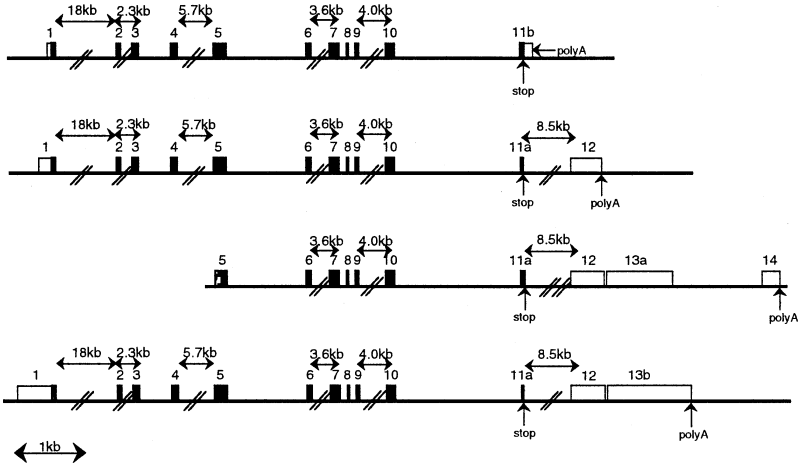

Fig. 4. Genomic structure of the *G5* gene. Locations of exons are shown in each of the four transcripts generated by alternative splicing; coding regions are indicated as *filled boxes*. The 5′ non-coding region of the transcript that includes exon 14 has not yet been found

Table 2. Exon — intron boundary sequences of the *G5* gene

Exon number	Exon length (bp)	cDNA position	Splice acceptor	Splice donor	Intron number	Intron length (bp)
1	439	1–439		CACACTCCAG **gt**gactggtc	1	18,538
2	75	440–514	tctcctgc**ag** TATAGCCGAA	ACATTAATAT **gt**gagtagaa	2	2,280
3	115	515–629	cctctacc**ag** ATTGCTTCCT	AGTAAACAAT **gt**atgtatat	3	483
4	101	630–730	ttactttc**ag** TTTACAATGC	CATTTCCTTA **gt**aagtatat	4	5,776
5	226	731–956	ttcttttc**ag** TCTATACAGT	TTCAATACCG **gt**tggtatcg	5	1,438
6	71	957–1027	ttttcttt**ag** ACAAGCCAAA	CAGAACTAAG **gt**aaacctgg	6	3,560
7	128	1028–1155	gtgtttct**ag** TGTGTCACAA	GAACTAGCAA **gt**atgttttc	7	83
8	59	1156–1214	tatcttac**ag** GAAAAAATCT	TTTAGATAAG **gt**gagttagt	8	76
9	69	1215–1283	tctctttt**ag** TATGAATTAC	ACTTAGTGAG **gt**aagactgt	9	4,029
10	144	1284–1427	tcttttc**ag** AGCTGTAGTG	AAAGAGAACA **gt**atgtaata	10	1,730
11a	81	1428–1508	tattttct**ag** ATTTGCCACT	TCCACTATAG **gt**aaattgta	11a	8,559
11b	261	1428–1688	tattttct**ag** ATTTGCCACT	ATAACTAGTA gaaatctttt	11b	8,390
12	572	1509–2080	ttctttct**ag** ATTTTGCTGG	GAATAAACCA gctgttctta	12	49
13a	1,182	2081–3262	tgcattta**ag** AAGCAATACA	CATAATCACA tgagtagttc	13a	1,277
13b	805	2081–2885	tgcattta**ag** AAGCAATACA	GATAACTGAG **gt**aagagtgt	13b	1,654
14	501	3263–3763	gtttttat**ag** ACATACACTG	GATATTCAAA atattgatgt	14	

```
1   MSWLFPLTKSASSSAAGSPGGLTSLQQQKQRLIESLRNSHSSIAEIQKDVEYRLPFTINN  60
61  LTININILLPPQFPQEKPVISVYPPIRHHLMDKQGVYVTSPLVNNFTMHSDLGKIIQSLL 120
121 DEFWKNPPVLAPTSTAFPYLYSNPSGMSPYASQGFPFLPPYPPQEANRSITSLSVADTVS 180
181 SSTTSHTTAKPAAPSFGVLSNLPLPIPTVDASIPTSQNGFGYKMPDVPDAFPELSELSVS 240
241 QLTDMNEQEEVLLEQFLTLPQLKQIITDKDDLVKSIEELARKNLLLEPSLEAKRQTVLDK 300
301 YELLTQMKSTFEKKMQRQHELSESCSASALQARLKVAAHEAEEESDNIAEDFLEGKMEID 360
361 DFLSSFMEKRTICHCRRAKEEKLQQAIAMHSQFHAPL                         397
```

Fig. 5. Deduced amino acid sequence of the G5 product. Coiled coil regions are indicated by *heavy underlining*, and proline-rich regions by *lighter underlines*

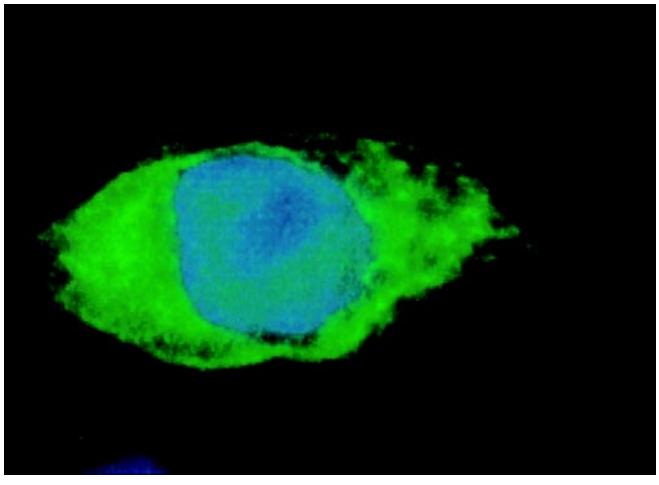

Fig. 6. GFP-fused protein assay for the intracellular location of G5. The photomicrograph was made 24h after transfection. 4′, 6-diamidino-2-phenylindole (DAPI) has stained the nucleus *blue*; *bright green* signal indicates the location of the fusion protein

G5 protein appears to be located in cyto-plasm, suggesting that it may be associated with a variety of cellular functions through its coiled-coil domains in which heptad repeats of hydrophobic residues produce amphipathic α-helical segments (Lupas et al. 1991). Although both *G5* and *GK1* have been partially characterized by computational analysis and expression profiles, their precise physiological and pathophysiological roles remain to be clarified.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bauer FB, Sirrenberg C, Neupert W, Brunner M (1996) Role of Tim23 voltage sensor and presequence receptor in protein import into mitochondria. Cell 87:33–41

Bergerheim USR, Kunimi K, Collins VP, Ekman P (1991) Deletion mapping of chromosome 8, 10, and 16 in human prostate carcinoma. Genes Chromosom Cancer 3:215–220

Burge C, Kerlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94

Claverie LM, Poirot O, Lopez F (1997) The difficulty of identifying genes in anonymous vertebrate sequences. Comput Chem 21:203–214

Daigo Y, Isomura M, Nishiwaki T, Tamari M, Ishikawa S, Kai M, Murata Y, Takeuchi K, Yamane Y, Hayashi R, Minami M, Fujino MA, Hojo Y, Uchiyama I, Takagi T, Nakamura Y (1999) Characterization of a 1200-kb genomic segment of chromosome 3p22–p21.3. DNA Res 6:1–8

Elkahloun AG, Kritzman DB, Wang Z, Hofmann TA, Roe B, Melzer PS (1997) Transcript mapping in a 46-kb sequenced region at the core of 12q13.3 amplification in human cancers. Genomics 42:295–301

Emi M, Fujiwara Y, Nakajima T, Tsuchiya E, Tsuda H, Hirohashi S, Maeda Y, Tsuruta K, Miyaki M, Nakamura Y (1992) Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer. Cancer Res 52:5368–5372

Fearon ER, Cho KR, Nigro JM, Kern SE, Simonds JW, Ruppert JM, Hamilton SR, Preisinger AC, Thomas G, Kinzlar KW, Vogelstein B (1990) Identification of a chromosome 18q gene that is altered in colorectal cancers. Science 247:49–56

Fernandez-Silva P, Martinez-Azorin F, Micol V, Attardi G (1997) The human mitochondrial transcription termination factor (mTERF) is a multizipper protein but binds to DNA as a monomer, with evidence pointing to intramolecullar leucine zipper interactions. EMBO J 16:1066–1079

Horii A, Nakatsuru S, Ichii S, Nagase H, Nakamura Y (1993) Multiple forms of the APC transcripts and their tissue-specific expression. Hum Mol Genet 2:283–287

Horton P, Nakai K (1996) A probabilistic classification system for predicting the cellular localization sites of proteins. Intelligent System Mol Biol 4:109–115

Ishikawa S, Kai M, Murata, Tamari M, Daigo Y, Murano T, Ogawa M, Nakamura Y (1998) Genomic organization and mapping of the human activin receptor type IIB(hActR-IIB)gene. J Hum Genet 43:132–143

Lupas A, van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252:1162–1164

Meuwissen RL, Offenberg HH, Dietrich AJ, Riesewijk A, van Iersel M, Heyting C (1992) A coiled-coil related protein specific for synapsed regions of meiotic prophase chromosomes. EMBO J 13:5091–5100

Rowen L, Mahairas G, Hood L (1997) Sequencing the human genome. Science 278:605–607

Rubin GM, Cuang P-T, Hariharan IK (1991) Cloning and characterization of receptor-class phosphotyrosine phosphatase gene expressed on central nervous system axons in *Drosophila melanogaster*. Proc Natl Acad Sci USA 88:11266–11270

Solovyev VV, Salamov AA, Lawrence CB (1994) Predicting internal exons by oligonucleotide composition and discriminate analysis of spliceable open reading frames. Nucleic Acids Res 22:5156–5163

Uberbacher EC, Mural RJ (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc Natl Acad Sci USA 88:11261–11265

Xu Y, Mural RJ, Uberbacher EC (1994) Constructing gene models from accurately predicted exons; an application of dynamic programming. Comput Appl Biosci 10:613–623