Human guanylate kinase (GUK1): cDNA sequence, expression and chromosomal localisation

Jude Fitzgibbon[a], Nicholas Katsanis[b], Dagan Wells[c], Joy Delhanty[c], William Vallins[d], David M. Hunt[a,*]

[a]*Department of Molecular Genetics, Institute of Ophthalmology, University College London, Bath St., London ECIV 9EL, UK*
[b]*Unit of Neurogenetics, Department of Biochemistry and Molecular Genetics, Imperial College School of Medicine at St. Mary's, Norfolk Place, London W2 1PG, UK*
[c]*Department of Biology, University College London, 4 Stephenson Way, London NW1 2HE, UK*
[d]*Pharmacia Biotechnology, 22 Grosvenor Road, St. Albans AL1 3AW, UK*

Received 16 March 1996

Abstract Guanylate kinase (GK) catalyses the conversion of GMP to GTP as part of the cGMP cycle. In mammalian phototransduction, this cycle is essential for the regeneration of cGMP following its hydrolysis by phosphodiesterase. Mutations in different parts of this signalling cascade lead to retinal degeneration in humans. Protein studies have localized a locus for GK to a region of human chromosome 1 that also contains an autosomal recessive form of retinitis pigmentosa (RP12) and Usher's type 11a (USH2A). We report the sequence of this human GK (GUK1) and a further refinement of its localization to 1q32-41, placing it in the same interval as USH2A.

Key words: Guanylate kinase; cGMP cycle; Phototransduction; Usher's type 11a

1. Introduction

An integral component of phototransduction in mammals is the hydrolysis of cGMP by a cGMP phosphodiesterase, resulting in the closure of cGMP-gated channels in the plasma membrane. The cGMP cycle (cGMP-GMP-GDP-GTP-cGMP) [1] is required to regenerate this intracellular messenger so that, in the dark, cGMP concentration returns to its original level, allowing the cGMP-gated channels to re-open [2]. The production of GDP from GMP and ATP in the cycle is catalysed by guanylate kinase (GK, EC 2.7.4.8). Surprisingly, GK has not been considered as a candidate gene for retinal disease even though mutations in other parts of the phototransduction pathway are responsible for both autosomal recessive [3-5] and dominant forms of retinitis pigmentosa [6]. This may reflect the absence of a retinal-specific form of this protein and the housekeeping role of the cGMP cycle in signal transduction and other cellular processes [7].

In mammals, GK activity is encoded by three distinct loci [8]. One of these, GUK1, was initially assigned to human chromosome 1 by isozyme studies of somatic cell hybrids [9]. Using a combination of hybrids generated by X-irradiation and hybrid cell lines that retain different chromosome 1 translocation breakpoints [10,11], this localization was further refined to chromosome 1q31-1q42 and subsequently confirmed by the identification of a gene dosage effect for the GUK1 protein in a patient with a duplication of chromosome 1 [12]. Since loci for an Usher's syndrome (USH2A) and an autosomal recessive retinis pigmentosa (RP12) [13,14] map to this region of chromosome 1, we have evaluated GUK1 as a candidate for retinal disease. In this paper, we present the cDNA sequence of the human gene and the refinement of its location to the distal portion of chromosome 1.

2. Materials and methods

2.1. Database searches

GUK1 ESTs and homologues in other species were identified using the ENTREZ document retrieval system available through the National Centre for Biotechnology Information at the National Institutes of Health, USA.

2.2. DNA sequence analysis

The cDNA clones 172943 and 172478, made available through the Image Consortium, were sequenced in their entirety using an automated sequencer (ALF express, Pharmacia) and the Autoread sequencing kit (Pharmacia). Genomic DNA fragments were subcloned into pBluescript and sequenced using the T7 sequencing kit (Pharmacia). Standard 6% polyacrylamide sequencing gels were used. Partial exon-intron structure was determined by computer alignment of the genomic sequence with the cDNA sequence. Sequences were aligned using ClustalV [15].

2.3. Cosmid identification

Cosmid clones for GUK1 were identified by screening a gridded chromosome 1-specific ICRF reference library kindly provided by the ICRF reference library system (D. Nizectic, unpublished; [16]). Approx. 40000 clones were screened with a probe of about 700 bp that contains the entire coding region of the bovine GUK1 cDNA ([17], kindly provided by Dr. Yakhyaev at the Shemyakin Institute of Bioorganic Chemistry, Moscow). All hybridizations were carried out under standard conditions [18].

2.4. Subcloning of genomic DNA fragments

The cosmid clones ICRFc112I0445 and ICRFc112O0286 were restriction digested with *Pst*I, *Apa*I, and *Apa*I-*Pst*I, and the resulting fragments were subcloned into the appropriate site of pBluescript KS (+). Positive clones were identified by hybridisation screening with a ^{32}P-labelled probe that contained the entire coding region of GUK1.

2.5. Fluorescence in situ hybridisation

The procedures for fluorescence in situ hybridization (FISH) were adapted from the original description by Pinkel et al. [19]. Cosmid DNA was prepared from 10 ml bacterial cultures grown overnight at 37°C, using the Magic Miniprep DNA purification system (Promega). Biotin labelled cosmid DNA was hybridized overnight at 37°C to human chromosome spreads prepared from normal lymphocyte cultures, followed by signal detection with avidin-fluorescein isothiocyanate (avidin-FITC). For R-banding, chromosomes were counterstained with propidium iodide and diamidophenylindole (DAPI). A dual bypass filter allowed the simultaneous visualisation of signals and banded chromosomes. Images were captured and analysed using an MRC600 confocal laser attachment. Approx. 20 metaphase chromosome spreads were analysed for accurate assessment of signal localization.

*Corresponding author. Fax: (44) (171) 608 6863.

PII S0014-5793(96)00365-1

```
ATGTCGGGCCCCAGGCCTGTGGTGCTGAGCGGGCCTTCGGGAGCTGGGAAGAGCACCCTGCTGAAGAGGCTGCTCCAGGAGCACAGCGGC   90
 M  S  G  P  R  P  V  V  L  S  G  P  S  G  A  G  K  S  T  L  L  K  R  L  L  Q  E  H  S  G    30

ATCTTTGGCTTCAGCGTGTCCCATACCACGAGGAACCCGAGGCCCGGCGAGGAGAACGGCAAAGATTACTACTTTGTAACCAGGGAGGTG  180
 I  F  G  F  S  V  S  H  T  T  R  N  P  R  P  G  E  E  N  G  K  D  Y  Y  F  V  T  R  E  V    60

ATGCAGCGTGACATAGCAGCCGGCGACTTCATCGAGCATGCCGAGTTCTCGGGGAACCTGTATGGCACGAGCAAGGTGGCGGTGCAGGCC  270
 M  Q  R  D  I  A  A  G  D  F  I  E  H  A  E  F  S  G  N  L  Y  G  T  S  K  V  A  V  Q  A    90

GTGCAGGCCATGAACCGCATCTGTGTGCTGGACGTGGACCTGCAGGGTGTGCGGAACATCAAGGCCACCGATCTGCGGCCCATCTACATC  360
 V  Q  A  M  N  R  I  C  V  L  D  V  D  L  Q  G  V  R  N  I  K  A  T  D  L  R  P  I  Y  I   120

TCTGTGCAGCCGCCTTCACTGCACGTGCTGGAGCAGCGGCTGCGGCAGCGCAACACTGAAACCGAGGAGAGCCTGGTGAAGCGGCTGGCT  450
 S  V  Q  P  P  S  L  H  V  L  E  Q  R  L  R  Q  R  N  T  E  T  E  E  S  L  V  K  R  L  A   150

GCTGCCCAGGCCGACATGGAGAGCAGCAAGGAGCCCGGCCTGTTTGATGTGGTCATCATTAACGACAGCCTGGACCAGGCCTACGCAGAG  540
 A  A  Q  A  D  M  E  S  S  K  E  P  G  L  F  D  V  V  I  I  N  D  S  L  D  Q  A  Y  A  E   180

CTGAAGGAGGCGCTCTCTGAGGAAATCAAGAAAGCTCAAAGGACCGGCGCCtgaggcttgctgtctgttctcggcaccccgggcccatac  630
 L  K  E  A  L  S  E  E  I  K  K  A  Q  R  T  G  A  *                                       197

aggaccagggcagcagcattgagccacccccttggcaggcgatacggcagctctgtgcccttggccagcatgtggagtggaggagatgct  720

gcccctgtggttggaacatcctgggtgacccccgacccagcctcgctgggctgtcccctgtccctatctctcactctggacccagggctg  810

acatcctaataaaataactgttggattagaaaaaaaa                                                        847
```

Fig. 1. Nucleotide and deduced amino acid sequence of the human GUK1 cDNA (GenBank accession no. L76200). The poly(A) addition signal and poly(A) tail are indicated by single and double underlining, respectively. A 19 bp region that is absent in clones H56620 and T08090 is indicated by a bar above the relevant sequence.

3. Results and discussion

3.1. Isolation of the human GUK1 cDNA

A number of human expressed sequence tag sites (ESTs) showing homology to the porcine GK gene were identified as part of the Washington-Merck sequencing project (see section 2). Two clones, 172943 (accession no. H20470) and 172478 (accession no. H20158), made available through the Image Consortium, were sequenced in their entirety. The consensus nucleotide sequence of human GUK1 is shown in Fig. 1. Comparison of the protein sequence with other mammalian GKs suggests that these clones contain the entire coding sequence (Fig. 2). The predicted amino acid sequence is 197 residues in length, and the resultant polypeptide has an estimated M_r of approx. 22 kDa. Amino acid alignment with the bovine [17], porcine [20], yeast (*Saccharomyces cerevisiae*) [21], and bacterial (*E. coli*) [22] enzymes give amino acid identity of 91, 89, 64 and 54%, respectively. All five sequences show considerable homology at regions implicated in specific interactions in yeast GK. The putative ATP-binding site consisting of the sequence GxxGxGKS [23] is identical across all five species, and with the exception of conservative changes in the yeast and bacterial proteins, the Mg^{2+} motif [24] and the sites of interaction with the guanine ring and phosphate groups of GMP [25] are also conserved. There is also a single potential N-linked glycosylation site at residue 171 that is conserved in all three mammalian proteins.

We have shown that all available GUK1-related EST clones share a common 3′ untranslated region (unpublished observation). In contrast, the original protein isozyme studies indicated that GK activity is encoded by three distinct loci that code for proteins with M_r of approx. 22, 27 and 32 kDa, respectively [8]. This latter study used enzymic staining to follow the production of GDP and ADP so it is possible that the other proteins with GK activity are specified by additional GK loci that are not closely related to the GUK1 gene. Two such loci are the genes encoding the p55 membrane protein [26] and DLG2 [27], a mammalian homologue of the *Drosophila* tumor suppressor gene *dlg-A*; both contain a GK domain that has many of the conserved features of GUK1, and it been proposed that these genes may encode proteins with GK activity [27-29]. However, since the M_r values of these proteins do not correspond to either the 27 or 32 kDa sizes of the GK proteins identified from protein isozyme analysis [8], it is unlikely that either of the above genes encodes the other isozymic variants of GK.

3.2. Expression analysis of GUK1

GUK1 has been shown to code for the e, f, and g isozymes of guanylate kinase with M_r values of 22, 22.5 and 22.5 kDa, respectively [8]. Isozymes e and f are present in almost all tissues while isozyme g is restricted to red cells [8]. In red cells, there is a reduction with age in the amount of the e isozyme present, with an associated increase in the amounts of iso-

```
                                                        G  P  P  P
Human     MSGPRPVVLSGPSGAGKSTLLKRLLQEHSGI-PGFSVSHTTRNPRPGEEN   49
Bovine    ......................K.....GS.-..........D.......   49
Porcine   .GS.....................K......S.-..........D.......   49
Yeast     ..---.I.I.....T.......K.FA.YPDS-......S...T..A..V.   47
Bacterial .AQGTLYIV.A.......S.IQA..KTQPLYDTQV.......Q.....VH   50

          P                  G          P G
Human     GKDYYFVTREVMQRDIAAGDFIEHAEFSGNLYGTSKVAVQAVQAMNRICV   99
Bovine    ....................................A..R..........   99
Porcine   ....................................A..R..........   99
Yeast     ....N..SVDEFKSM.KNNE...W.Q....Y..STVAS.KQ.SKSGKT.I   97
Bacterial .EH.F..NHDEFKEM.SRDA.L...EVF..Y....RE.IEQ.L.TGVDVF  100

          PGG
Human     LDVDLQGVRNIKAT-DLRPIYISVQPPSLHVLEQRLRQRNTETEESLVKR  148
Bovine    ............K.-........F.....................A..   148
Porcine   ............K.-...Q....F.....D.................A..   148
Yeast     ..I.M...KSV..IPELNARFLFIA...VED.KK..EG.G......IN..   147
Bacterial ..I.W..AQQ.RQKMP-HA.S.F.L...KIE.DR...G.GQDS..VIA..   149

                                    *
Human     LAAAQADMESSKEPGLFDVVIINDSLDQAYAELKEALSEEIKKAQRTG--   196
Bovine    ....R.............LI.......K..WA.............G..Q-   197
Porcine   ....K.............LI.......K..WA.............A..H-   197
Yeast     .S....ELA-YA.T.AH.K..V..D..K..K...DFIFA.----------   186
Bacterial M.Q.V.E.SHYA.---Y.YL.V..DF.T.LTD..TIIRA.RLRMS.QKQR   196

Human     ---------A-                                         197
Bovine    ---------S-                                         198
Porcine   ---------S-                                         198
Yeast     ----------K                                         197
Bacterial HDALISKLL.D                                         207
```

Fig. 2. Amino acid alignment of guanylate kinases. The ATP and Mg^{2+} binding motifs are indicated by an overscore and underscore respectively. The sites of interaction with the guanine ring and phosphate groups of GMP are indicated by G and P, respectively. The single glycosylation site in the three mammalian proteins is identified by an asterisk.

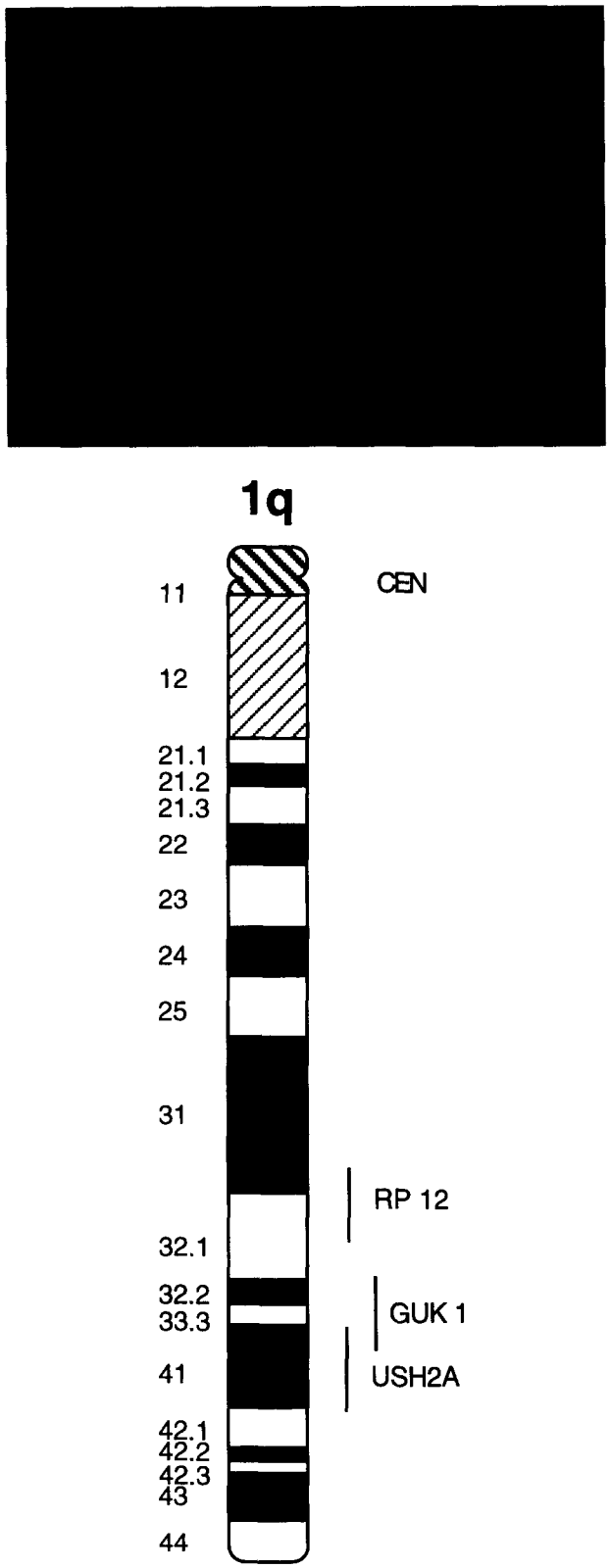

Fig. 3. (a) Chromosomal localization of GUK1 to human chromosome 1 by FISH. The hybridization signal was assigned to chromosome 1q32-41 (arrowed). Enlargement of both copies of chromosome 1 are also shown. (b) Idiogram of human chromosome 1q showing the position of GUK1 FISH hybridization signals and the physical locations of the RP12 and USH2A loci.

zymes f and g, indicating that e is the primary form, and that it is modified in some manner to produce the two secondary isozymes. This posttranslational modification may involve glycosylation, as a single glycosylation site is conserved in all mammalian GUK1 proteins (Fig. 2).

The broad tissue-expression pattern of the GUK1 protein is consistent with the identification of ESTs for GUK1 in retina (H95538, H92316), infant and adult brain (H17287, H17288, T08090, H19929, H20252, H20158, H40996, H43002), pineal gland (H56620), lung (T34435), breast (R54941; R54752), foetal liver-spleen (H38217) and melanocytes (H83699, H97119). We have also noted that two ESTs, H56620 and T08090, lack a 19 bp region between nucleotides 157 and 176 of the human GUK1 cDNA (Fig. 1); the absence of this region results in the formation of a truncated protein of 53 amino acids that is unlikely to be functionally important.

3.3. Mapping of human GUK1 to chromosome 1

A gene coding for a GK has previously been assigned by protein studies [9,11,12] to the distal region of the long arm of chromosome 1 at q31-42. A full length bovine GK cDNA clone [17] was used to screen a human chromosome 1 cosmid library [16] and two overlapping postive clones, ICRFc112I0445 and ICRFc112O0286, were identified. These cosmids were shown by partial sequence analysis of several subclones to contain the GUK1 gene and from this, we have been able to define the exon-intron structure of the 3′ end of the gene. Two introns were identified after nucleotides 390 and 476, although it is possible that more 5′ introns are also present.

The genes for an autosomal recessive RP (RP12) and USH2A have been localized to chromosome 1q31-q32.1 and 1q41, respectively (Fig. 3b). To determine whether GUK1 is a candidate for either of these retinal diseases, the positions of ICRFc112I0445 and ICRFc112O0286 cosmid clones were refined by fluorescence in situ hybridization to chromosome 1q32-41 (Fig. 3); this localization is consistent with previous mapping studies using the GUK1 protein and places the gene distal to the region containing the RP12 locus. Further studies will be required to evaluate GUK1 as a candidate for USH2A.

Acknowledgements: The work was supported by a grant from the Wellcome Trust. The confocal imaging equipment was provided by the Medical Research Council as part of the Human Genome mapping project. N.K. is funded by the Birth Defects and Leventis Foundation. We are also indebted to the Reference Library Database ICRF and Image Consortium for availability of clones. In particular we are grateful to Gaidarov and colleagues at the Shemyakin Institute Moscow for suggesting in their paper that GK would be worth considering as a candidate gene for retinal disease. We also thank the following for helpful comments on the manuscript: Dr. Alex Morris, Dr. Helen Cadiou, Mr. Neil Ebenezer, Dr. Arthur Bergen and Prof. David Hopkinson.

References

[1] Hall, S.W. and Kuhn, H. (1986) Eur. J. Biochem. 161, 551–556.
[2] Farber, D. (1995) Int. Ophthalmol. Vis. Sci. 36, 263–275.
[3] McLaughlin, M.E., Sandberg, M.A., Berson, E.L. and Dryja, T.P. (1993) Nat. Genet. 4, 130–134.
[4] Kumaramanickavel, G., Maw, M., Denton, M.J., John, S., Srikumari, C., Orth, U., Oehlmann, R. and Gal, A. (1994) Nat. Genet. 8, 10–11.

[5] Dryja, T.P., Finn, J.T., Peng, Y., McGee, T.L. and Berson, E. (1995) Proc. Natl. Acad. Sci. USA 92, 10177–10181.
[6] Dryja, T.P., McGee, T.L., Reichel, E., Hahn, L.B., Cowley, G.S., Yandell, D.W., Sandberg, M.A. and Berson, E.L. (1990) Nature 342, 364–366.
[7] Bourne, H.R., Sanders, D.A. and McCormick, F. (1990) Nature 348, 125–132.
[8] Jamil, T., Fisher, R. and Harris, H. (1975) Hum. Hered. 25, 402–413.
[9] Meera-Khan, P., Doppert, B.A., Hagemeijer, A. and Westerveld, A. (1974) Cytogenet. Cell. Genet. 13, 130–131.
[10] Burgerhout, W.G., Leupe-de-Smit, S. and Jongsma, A.P.M. (1977) Cytogenet. Cell. Genet. 18, 267–283.
[11] Jongsma, A.P.M. and Burgerhout, W.G. (1977) Cytogenet. Cell. Genet. 18, 255–266.
[12] Dallapiccola, B., Lungarotti, M.S., Falorni, M. and Dacha, M. (1980) Ann. Genet. 23, 83–85.
[13] Van Soest, S., Van den Born, L., Gal, A., Farrar, G.J., Bleeker-Wagemakers, L.M., Westerveld, A., Humphries, P., Sandkuijl, L.A. and Bergen, A.A.B. (1994) Genomics 22, 499–504.
[14] Van Soest, S., Te Nijenhuis, S., Van den Born, L., Bleeker-Wagemakers, L.M., Sharp, E., Sandkuijl, L.A., Westerveld, A. and Bergen, A.A.B. (1996) Cytogenet. Cell Genet., in press.
[15] Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) CABIOS 8, 189–191.
[16] Lehrach, H., Drmanac, R., Hoheisel, J., Larin, Z., Lennon, G., Monaco, A.P., Nizetic, D., Zchetner, G. and Pouska, A. (1990) in: Genome Analysis Volume (Davies, K.E. and Tilghman, S.M. eds.) pp. 39–81, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[17] Gaidarov, I.O., Suslov, O.N. and Abdulaev, N.G. (1993) FEBS Lett. 335, 81–84.
[18] Yulug, I.G., Egan, S.E., See, C.G. and Fisher, E.M.C. (1994) Genomics 22, 313–318.
[19] Pinkel, D., Straume, T. and Gray, J.M. (1986) Proc. Natl. Acad. Sci. USA 83, 2934–2938.
[20] Zschocke, P.D., Schiltz, E. and Schultz, G.E. (1993) Eur. J. Biochem. 213, 263–269.
[21] Berger, A., Schiltz, E. and Schulz, G.E. (1989) Eur. J. Biochem. 184, 433–443.
[22] Gentry, D., Bengra, C., Ikehara, K. and Cashel, M. (1993) J. Biol. Chem. 268, 14316–14321.
[23] Saraste, M., Sibbald, P.R. and Wittinghofer, A. (1990) Trends Biochem. Sci. 15, 430–434.
[24] Yan, H. and Tsai, M. (1991) Biochemistry 30, 5539–5546.
[25] Stehle, T. and Schulz, G.E. (1990) J. Mol. Biol. 211, 249–254.
[26] Ruff, P., Speicher, D.W. and Husain-Chishti, A. (1991) Proc. Natl. Acad. Sci. USA 88, 6595–6599.
[27] Mazoyer, S., Gayther, S.A., Nagai, M.A., Smith, S.A., Dunning, A., Van Rensburg, E.J., Albertson, H., White, R. and Ponder, B.A. (1995) Genomics 28, 25–31.
[28] Woods, D.F. and Bryant, P.J. (1991) Cell 66, 451–464.
[29] Koonin, E.V., Woods, D.F. and Bryant, P.J. (1992) Nat. Genet. 2, 256–257.