Gene, 146 (1994) 291–295

GENE 08039

Sequence of a putative human housekeeping gene (*HK33*) localized on chromosome 1

(Alternative polyadenylation; four transcripts; ubiquitous expression; chromosomal location)

A. Braun[a,b], S. Kammerer[a], W. Weissenhorn[c], E.H. Weiss[a,c] and H. Cleve[a,†]

[a]*Institute of Anthropology and Human Genetics, University of Munich, D-80333 Munich, Germany;* [b]*Children's Hospital, Laboratory of Molecular Biology, University of Munich, D-80337 Munich, Germany. Tel. (49-89) 5160-3900; and* [c]*Institute of Immunology, University of Munich, D-80336 Munich, Germany. Tel. (49-89) 5996-688*

Received by H. Zachau: 26 October 1993; Revised/Accepted: 10 January 1994; Received at publishers: 21 April 1994

SUMMARY

A gene (*HK33*) localized on human chromosome 1 has been detected by crossreaction of its fusion protein with a monospecific antiserum directed against human vitamin-D-binding protein (hDBP; group-specific component). Its cDNA sequence analysis showed no evident homologies neither to the sequence encoding hDBP nor to any other sequence. The largest cDNA clone of 3.2 kb includes a 897-bp coding region and a large 3′ untranslated region with at least four polyadenylation sites. Further cDNA amplification using PCR demonstrated a total cDNA length of approx. 3.7 kb. Northern blot analysis revealed signals at about 2.2–2.5 kb and 4.0 kb, the shorter transcripts representing mRNAs using one of the two polyadenylation sites at about 2.0 kb. Synthesis of the 299-amino-acid polypeptide (33 kDa) in the bacterial host, with subsequent Western blot analysis, verified the sequence-specific recognition by the hDBP-specific antiserum. The search of protein databanks revealed no homology of HK33 to any known sequence. Since the gene is transcribed in all cells and tissues tested so far, it is a strong candidate for another housekeeping gene.

INTRODUCTION

The gene encoding the human vitamin-D-binding protein (hDBP) is known for its high level expression in the liver (Prunier et al., 1964). However, Mc Leod and Cooke (1989) detected low level transcription of the *DBP* gene of the rat in several other cells and tissues, including kidney, testis and placenta. This observation prompted us to try to identify hDBP in placenta by an immunological screening of a human placenta cDNA expression library in λgt11 with hDBP-specific antibodies. For that purpose we used a polyclonal hDBP-specific rabbit antiserum prepared in our laboratory which was preabsorbed with albumin and purified by hDBP affinity chromatography. But instead of hDBP, we detected the crossreacting fusion protein of a cDNA clone that showed no evident homologies to hDBP. The aim of our studies was to characterize this cDNA.

Correspondence to: Dr. E.H. Weiss, Institute of Anthropology and Human Genetics, University of Munich, Richard Wagner Street 10/I, D-80333 Munich, Germany. Tel. (49-89) 5203-267; Fax (49-89) 5203-286.
†Deceased January 17, 1994.

Abbreviations: aa, amino acid(s); βME, β-mercaptoethanol; bp, base pair(s); cDNA, DNA complementary to RNA; *E.*, *Escherichia*; GST, glutathione *S*-transferase; (h)DBP, (human) vitamin-D-binding protein; *HK33*, putative *h*ou*s*e*k*eeping gene, encoding a 33-kDa polypeptide; IPTG, isopropyl-β-D-thiogalactopyranoside; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; ORF, open reading frame; PAGE, polyacrylamide-gel electrophoresis; PCR, polymerase chain reaction; RT, reverse transcription; SDS, sodium dodecyl sulfate; *UTR*, untranslated region(s).

SSDI 0378-1119(94)00250-V

EXPERIMENTAL AND DISCUSSION

(a) Detection and characterization of a placental *HK33* cDNA clone

To probe a human placenta cDNA expression library in λgt11 (human placenta cDNA library, Clontech Laboratories, Heidelberg, Germany) we used a hDBP-monospecific antiserum. One positive hybridization signal in 200000 λ clones examined could be verified in the rescreen (screening and subsequent purification steps were performed according to Sambrook et al., 1989). DNA was isolated from this cDNA clone λP1.2 and digested with *Eco*RI (Boehringer-Mannheim, Mannheim, Germany), resulting in two insert fragments of 500 and 750 bp, respectively. Sequencing of the 1250-bp cDNA revealed a 200-bp ORF at the 5′ end (Fig. 1).

(b) Sequence of cDNA clones using different polyadenylation sites

Since Northern blot results indicated the presence of at least three transcripts (2.2, 2.5 and 4.0 kb) hybridizing to the λP1.2 insert, various cDNA libraries were screened in order to obtain full-length cDNA clones representative for the three mRNA species. First, the 1250-bp probe was used to screen the human placental cDNA library again.

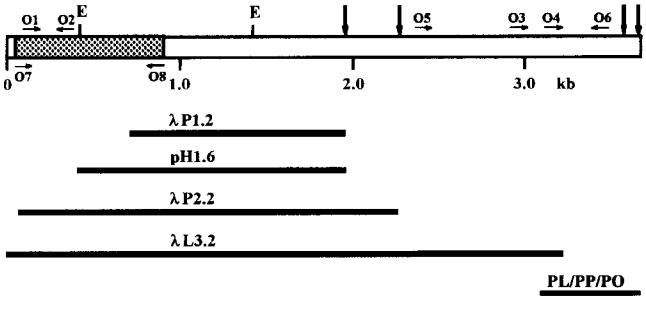

Fig. 1. Scheme of the *HK33* cDNA. The structure has been deduced from the cDNA clones λP1.2, placenta clone (1.2 kb), pH1.6, hepatic clone (1.6 kb), λP2.2, placenta clone (2.2 kb), λL3.2, leukocyte clone (3.2 kb), and the PCR products amplified at the 3′ end, PL, PP an PO. These products were obtained by PCR, using the more upstream primer O3, vector-specific reverse primers of cDNA libraries of different cells and tissues (the leukocyte library in λYES, the placenta library and an ovary library in λgt11, Clontech), and 10^6 phages of each library, respectively. After booster reaction with a nested, e.g., downstream primer O4, products were subcloned in pUC19 and sequenced. The 897-bp coding region is indicated as a shaded box. Alternative poly(A) sites are marked by downward arrows. E, *Eco*RI sites used for subcloning. Sequences of the primers used for amplification of probes and for chromosomal localization (O1-O2, O5-O6), for amplification of the very 3′ products PL, PP and PO (O3-O4) and for amplification of the coding region (O7-O8) are as follows: O1, 5′-CCTGCAGCCTCTGAGAGC; O2, 5′-GGAATTGGAGGAGCTTCCG; O3, 5′-GTTTGACAGC-ACCCAAACTG; O4, 5′-GGCAG TCTTTCTCAAACTATG; O5, 5′-TATGGTATTGGTGAGGAAGG; O6, 5′-CTTTAAAGTTCGAGAG-TCGCA; O7, 5′-AACGCCGCCGCTGAGGAAG; O8, 5′-TCTAG-ACTCACATGATCAGACACTG.

The *Eco*RI digest of the largest clone resulted in three cDNA fragments of about 330, 1050 and 800 bp, respectively. These were subcloned in the *Eco*RI site of pUC19 (Gibco-BRL, Eggenstein, Germany) (Yanisch-Perron et al., 1985) and transformed in *E. coli* MC1061 (Meissner et al., 1987). Sequencing demonstrated a total length of 2225 bp and an ORF up to the 5′ end. Furthermore, this cDNA has a large 3′ *UTR* of 1376 bp which contains two polyadenylation signals.

The first signal at 1960 bp (indicating the first nt of the motif in the sequence submitted to EMBL Data Library) agrees with the consensus sequence 5′-AATAAA (Proudfood and Brownlee, 1976; Proudfood, 1991) and is located 19 bp upstream from the poly(A) tail. This sequence obviously served as poly(A) signal in the 1.6-kb cDNA clone pH1.6 (Fig. 1) that we detected in a human hepatic cDNA library in the plasmid vector CDM8 (construction of the library has been described by Estaller et al., 1991). The motif 5′-AATATA at nt 2231 that served as recognition site in the clone P2.2 differs in one nt from the consensus sequence and precedes the poly(A) site by 35 bp.

As explained later, these two cDNA clones represent the lower Northern blot signal at approx. 2.2–2.5 kb as the poly(A) tail extends over 200–300 bp in vivo (Manley, 1988).

Using a 290-bp 5′ probe (described in Fig. 1) we screened a randomly primed human leukocyte cDNA library in λYES (human leukocyte cDNA stretch library, Clontech Laboratories) to search for larger clones that could account for the 4.0-kb signal. Two clones with a 3.2 and a 2.2-kb insertion, respectively, were obtained.

The 5′ fragment of the 3.2-kb clone λL3.2 extends the previous cDNA of λP2.2 by 58 bp and reveals a start codon at nt 11 which was in frame with the ORF of P2.2. It was impossible to obtain cDNA clones with a larger 5′ end, because all clones analysed stopped in the region of nt 1–10. But it seems probable that the coding region starts at nt 11, since the sequence flanking the ATG agrees in the most highly conserved positions with the consensus sequence determined by Kozak (1987).

The resulting ORF of 897 bp (Fig. 1) determines a polypeptide of 299 aa (32.8 kDa). The nt sequence of this region together with the deduced aa sequence is presented in Fig. 2. The coding sequence is identical in all cDNA clones analysed.

Sequence analysis of the 3′ *Eco*RI 1.7-kb fragment revealed a 3′ *UTR* that was 0.9-kb longer than that in the clone λP2.2, but contained no poly(A) tail. To determine the very 3′ end of the larger transcript, we analysed different PCR products (Fig. 1). All showed a poly(A) tail, but the PCR product derived from one tissue (ovary cDNA library) was 65-bp longer than those of the leuko-

```
                                                    5' AGGTAGCAAG  10
ATGGCCGCCGCTGAGGAAGGCTGTAGTGTCGGGGCCGAAGCGGACAGGGAATTGGAGGAG  70
MetAlaAlaAlaGluGluGlyCysSerValGlyAlaGluAlaAspArgGluLeuGluGlu  20

CTTCTGGAAAGTGCTCTTGATGATTTCGATAAAGCCAAACCCTCCCCAGCACCCCCTTCT 130
LeuLeuGluSerAlaLeuAspAspPheAspLysAlaLysProSerProAlaProProSer  40

ACCACCACGGCCCCTGATGCTTCGGGGCCCCAGAAGAGATCGCCAGGAGACACTGCCAAA 190
ThrThrThrAlaProAspAlaSerGlyProGlnLysArgSerProGlyAspThrAlaLys  60

GATGCCCTCTTCGCTTCCCAAGAGAAGTTTTTCCAGGAACTATTCGACAGTGAACTGGCT 250
AspAlaLeuPheAlaSerGlnGluLysPhePheGlnGluLeuPheAspSerGluLeuAla  80

TCCCAAGCCACTGCGGAGTTCGAGAAGGCAATGAAGGAGTTGGCTGAGGAAGAACCCCAC 310
SerGlnAlaThrAlaGluPheGluLysAlaMetLysGluLeuAlaGluGluGluProHis 100

CTGGTGGAGCAGTTCCAAAAGCTCTCAGAGGCTGCAGGGAGAGTGGGCAGTGATATGACC 370
LeuValGluGlnPheGlnLysLeuSerGluAlaAlaGlyArgValGlySerAspMetThr 120

TCCCAACAAGAATTCACTTCTTGCCTAAAGGAAACACTAAGTGGATTAGCCAAAAATGCC 430
SerGlnGlnGluPheThrSerCysLeuLysGluThrLeuSerGlyLeuAlaLysAsnAla 140

ACTGACCTTCAGAACTCCAGCATGTCGGAAGAAGAGCTGACCAAGGCCATGGAGGGGCTA 490
ThrAspLeuGlnAsnSerSerMetSerGluGluGluLeuThrLysAlaMetGluGlyLeu 160

GGCATGGACGAAGGGGATGGGGAAGGGAACATCCTCCCCATCATGCAGAGTATTATGCAG 550
GlyMetAspGluGlyAspGlyGluGlyAsnIleLeuProIleMetGlnSerIleMetGln 180

AACCTACTCTCCAAGGATGTGCTGTACCCATCACTGAAGGAGATCACAGAAAAGTATCCA 610
AsnLeuLeuSerLysAspValLeuTyrProSerLeuLysGluIleThrGluLysTyrPro 200

GAATGGTTGCAGAGTCATCGGGAATCTCTACCTCCAGAGCAGTTTGAAAAATATCAGGAG 670
GluTrpLeuGlnSerHisArgGluSerLeuProProGluGlnPheGluLysTyrGlnGlu 220

CAGCACAGCGTCATGTGCAAAATATGTGAGCAGTTTGAGGCAGAGACCCCCACAGACAGT 730
GlnHisSerValMetCysLysIleCysGluGlnPheGluAlaGluThrProThrAspSer 240

GAAACCACTCAAAAGGCTCGTTTTGAGATGGTGCTGGATCTTATGCAGCAGCTACAAGAT 790
GluThrThrGlnLysAlaArgPheGluMetValLeuAspLeuMetGlnGlnLeuGlnAsp 260

TTAGGCCATCCTCCAAAAGAGCTGGCTGGAGAGATGCCTCCTGGCCTCAACTTTGACCTG 850
LeuGlyHisProProLysGluLeuAlaGlyGluMetProProGlyLeuAsnPheAspLeu 280

GATGCCCTCAATCTTTCGGGCCCACCAGGTGCCAGTGGTGAACAGTGTCTGATCATGTGA 910
AspAlaLeuAsnLeuSerGlyProProGlyAlaSerGlyGluGlnCysLeuIleMet*** 299
```

Fig. 2. The nt and deduced aa sequence of the coding region of the human *HK33* gene. Sequencing of double-stranded plasmid template of the clones showed in Fig. 1 was performed by the dideoxy chain-termination method of Sanger et al. (1977) using [α-^{32}P]dATP (Amersham, Braunschweig, Germany), Sequenase Version 2.0 (Tabor and Richardson, 1987; US Biochemical, Bad Homburg, Germany) and universal pUC primers (M13/pUC 'sequencing' and 'reverse sequencing' primer, Boehringer-Mannheim). This sequence and the complete 3′ *UTR* sequence (Fig. 1) has been submitted to the EMBL Data Library (accession No. X75535).

cyte and the placenta library, demonstrating additional alternative polyadenylation in this region of the cDNA (Fig. 1).

The poly(A) signal, 5′-AATTAA, at 3566 bp, precedes the poly(A) tail by 17 bp, while the signal at nt 3612/3624 showed two successive motifs of this sequence and was followed by a poly(A) tail 36/24 nt downstream. The fact that these motifs differed from the consensus sequence might explain the reduced amount of the 4.0-kb mRNA in some cells and tissues analysed (Fig. 3A; Sheets et al., 1990).

(c) Northern blot analysis

To analyse whether the Northern blot signals only differ in the extent of the 3′ *UTR*, further Northern blots were performed (Fig. 3). Hybridization was done with a 290-bp 5′ probe which detected all transcripts of 2.2, 2.5 and 4.0 kb, as expected (Fig. 3A), and with a 1060-bp 3′ probe to confirm that the 4.0-kb transcripts were due to a larger 3′ end caused by alternative polyadenylation (Fig. 3B). Indeed this probe revealed exclusively a single band at 4.0 kb. Different mRNAs are present in varying ratios in all tissues and cells analysed.

(d) Test of different cells and tissues for transcription of *HK33*

Many other cells and tissues, including lung, kidney, uterus, colon, ovary, testis, peripheral blood leukocytes, fibroblasts, two hepatoma cell lines (H4, Hep G2) and a retina cDNA library (human fetal retina library in λUni-ZAP XR vector, Stratagene, Heidelberg, Germany), were tested for the presence of *HK33* mRNA using RT-PCR. In all these tissues a PCR product of the expected size could be amplified (data not shown). Therefore, the *HK33* mRNA seems to be ubiquitously expressed. The protein synthesis will be analysed quantitatively with a mono-

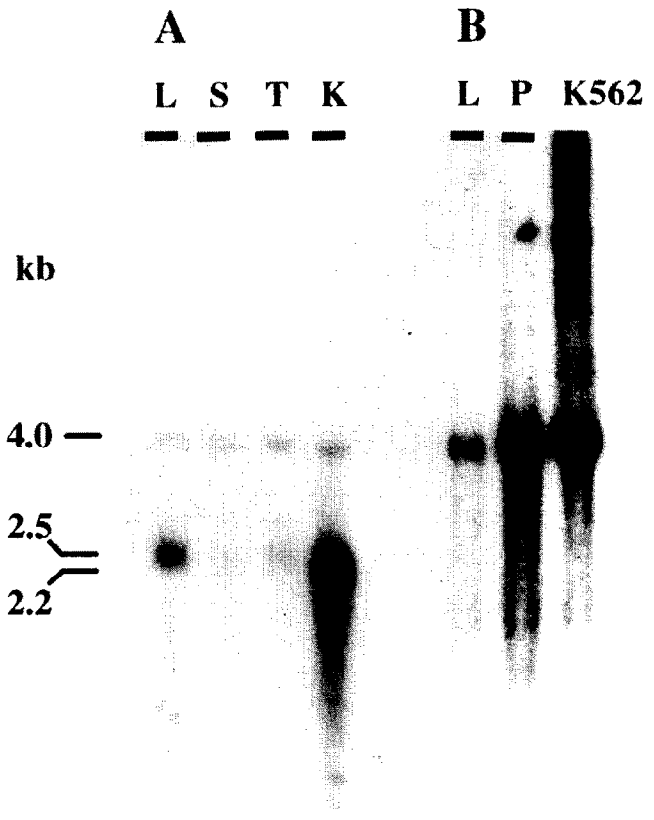

Fig. 3. Northern blot analysis of *HK33* transcripts in different cells and tissues. (**A**) Tissues analysed: **L**, liver; **S**, spleen; **T**, thymus; **K**, kidney. Hybridization with a 290-bp 5′ probe. The probe was amplified from the 5′ region of λP2.2 using 10 ng of the purified DNA and two insert-specific primers (O1-O2, Fig. 1) and radiolabelled by PCR (*Taq* polymerase supplied by Promega, Heidelberg, Germany; primers synthesized by MWG Biotech, Ebersberg, Germany) with [α-^{32}P]dATP (Amersham). Hybridization was performed overnight by a modified method of Church and Gilbert (1984) at 65°C in a buffer containing 1% bovine serum albumin (BSA)/7% SDS/0.25 M Na_2HPO_4/NaH_2PO_4 pH 7.2. Filters were washed three times for 10 min at 65 C with 40 mM Na_2HPO_4/NaH_2PO_4 pH 7.2/1% SDS. (**B**) Tissues and cell line tested: **L**, liver; **P**, placenta; **K562**, chronic myelogenous leukemia cell line, and hybridization with a 1060-bp 3′ probe. The probe was amplified by PCR using 200 ng genomic DNA and specific primers (O5-O6, Fig. 1) and radiolabelled by the random priming method (Feinberg and Vogelstein, 1983) using a hexamer mixture, Klenow large fragment of *E. coli* DNA polymerase I (both Boehringer-Mannheim) and [α-^{32}P]dATP (Amersham). Hybridization was performed as in **A**. RNA was prepared by the procedure of Chirgwin et al. (1979). Total RNA (10 to 20 μg) was separated on 2.2 M formaldehyde/1.2% agarose gels and transferred to nylon membrane (Hybond-N, Amersham).

specific HK33-antiserum (S.K., A.B. and H.C., data not shown).

(e) Chromosomal localization

To determine the chromosome encoding the putative housekeeping gene *HK33*, we performed PCR with a panel of 18 human/rodent hybridoma cell lines. A human cell line and two rodent cell lines (CHO and mouse) were used as controls. PCR was done using 200 ng DNA of each cell line and gene-specific primers (O5-O6, Fig. 1). The PCR products were examined by Southern blot analysis using the 1060-bp 3′ probe described in Fig. 3B. Four hybridoma cell lines and the human cell line were positive, the other lines including the two rodent cell lines were negative. The positive cell lines contained the human chromosome 1 which was absent in the negative lines. The gene *HK33* is, thus, localized on human chromosome 1 (data not shown).

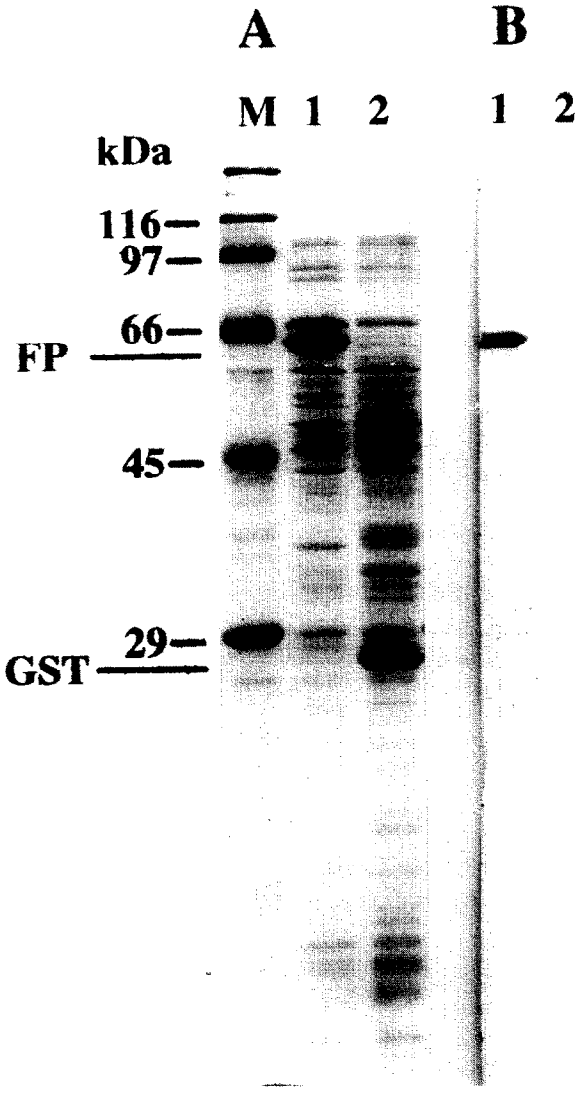

Fig. 4. Expression of HK33 from the prokaryotic vector pGEX-3X (Smith and Johnson, 1988; Johnson et al., 1989; Pharmacia, Freiburg, Germany). (**A**) After induction with IPTG (Promega) bacterial growth was continued for 8–10 h. Lysis with a reducing SDS loading buffer containing βME and separation by a 0.1% SDS-12% PAGE were followed by staining with Coomassie brilliant blue R-250 (Serva, Heidelberg, Germany). **M**, Markers (MW-SDS-200, Sigma, Deisenhofen, Germany); lane **1**, pGEX-3X with *HK33* insert producing the 64-kDa fusion protein (FP); lane **2**, pGEX-3X without insert producing merely the 26-kDa GST portion (GST). (**B**) Subsequent Western blot analysis was performed as described in Hoffmann et al. (1990) using the hDBP-monospecific polyclonal antiserum described in the introduction (lane **1** and **2** as in **A**).

(f) Synthesis of HK33 in a prokaryotic system

The coding region of the gene was specifically amplified by PCR using hepatic cDNA and oligos flanking this region (O7-O8, Fig. 1), subcloned in the *Sma*I site of the prokaryotic expression vector pGEX-3X and transformed into *E. coli* MC1061. The IPTG-induced fusion protein of about 60 kDa consisted of a GST portion of approx. 27 kDa and a HK33 portion of about 33 kDa. Subsequent Western blot analysis confirmed the sequence-specific recognition of the protein HK33 by the hDBP-specific antiserum (Fig. 4).

(g) Conclusions

(*1*) The gene *HK33* shows no evident homologies to any known nt or aa sequence.

(*2*) Comparison of the aa sequences of HK33 and hDBP did not show motifs which could be considered as candidates for the immunological cross reaction.

(*3*) Analysis of different cDNA clones resulted in a full-length cDNA of about 3.7 kb with at least four alternative poly(A) sites that are all used in vivo.

(*4*) Northern blot hybridizations revealed at least two bands which are obviously due to the alternative polyadenylation. Successive sequencing demonstrated a minimum of four transcripts.

(*5*) Chromosomal localization using a panel of different human/rodent hybridoma cell lines demonstrated *HK33* to be located on chromosome 1.

(*6*) Transcription of the gene *HK33* in all cells and tissues tested so far makes it a candidate for a further housekeeping gene. It is our forthcoming project to determine the exact biological function of this gene.

ACKNOWLEDGEMENTS

This work was supported by grants of the Deutsche Forschungsgemeinschaft (DFG) to H.C. (Cl 27/14-2) and of the Bundesministerium für Forschung und Technologie (BMFT) to E.H.W. (Genzentrum München).

REFERENCES

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J.: Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18 (1979) 5294–5299.

Church, G.M. and Gilbert, W.: Genomic sequencing. Proc. Natl. Acad. Sci. USA 81 (1984) 1991–1995.

Estaller, C., Koistinen, V., Schwaeble, W., Dierich, M.P. and Weiss, E.H.: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel member of the short consensus repeat family related to the C-terminal of the classical 150-kDa molecule. J. Immunol. 146 (1991) 3190–3196.

Feinberg, A.P. and Vogelstein, B.: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132 (1983) 6–13.

Hoffmann, R., Braun, A. and Cleve, H.: A monoclonal antibody against human vitamin-D-binding protein for the analysis of genetic variation in the group-specific component system (Gc). Human Genetics 84 (1990) 137–146.

Johnson, K.S., Harrison, G.B.L., Lightowlers, M.W., O'Hoy, K.L., Cougle, W.G., Dempster, R.P., Lawrence, S.B., Vinton, J.G., Heath, D.D. and Rickard, M.D.: Vaccination against ovine cysticercosis using a defined recombinant antigen. Nature 338 (1989) 585–587.

Kozak, M.: An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15 (1987) 8125–8148.

Manley, J.L.: Polyadenylation of mRNA precursors. Biochim. Biophys. Acta 950 (1988) 1–12.

McLeod, J.F. and Cooke, N.E.: The vitamin D-binding protein, α-fetoprotein, albumin multigene family: detection of transcripts in multiple tissues. J. Biol. Chem. 264 (1989) 21760–21769.

Meissner, P.S, Sisk, W.P. and Berman, M.L.: Bacteriophage cloning system for the construction of directional cDNA libraries. Proc. Natl. Acad. Sci. USA 84 (1987) 4171 4175.

Proudfood, N.J. and Brownlee, G.G.: 3' non-coding region sequences in eukaryotic mRNA. Nature 263 (1976) 211–214.

Proudfood, N.J.: Poly(A) signals. Cell 64 (1991) 671–674.

Prunier, J.H., Bearn, A.G. and Cleve, H.: Site of formation of the group-specific component and certain other serum proteins. Proc. Soc. Exp. Biol. N.Y. 115 (1964) 1005–1007.

Sambrook, J., Fritsch, E.F. and Maniatis, T.: Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Sheets, M.D., Ogg, S.C. and Wickens, M.P.: Point mutations in AAUAAA and the poly(A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. Nucleic Acids Res. 18 (1990) 5799–5805.

Smith, D.B. and Johnson, K.S.: Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. Gene 67 (1988) 31–40.

Tabor, S. and Richardson, C.C.: DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84 (1987) 4767–4771.

Yanisch-Perron, C., Vieira, J. and Messing, J.: Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene 33 (1985) 103–119.