J Hum Genet (2001) 46:289–292

SHORT COMMUNICATION

Chisako Iriyama · Satoru Matsuda · Ryu Katsumata
Michinari Hamaguchi

Cloning and sequencing of a novel human gene which encodes a putative hydroxylase

Received: December 15, 2000 / Accepted: January 22, 2001

Abstract Using a conventional two-hybrid technique with MAWD as bait protein, a novel full-length cDNA was isolated and sequenced from a human liver cDNA library. This cDNA consists of 2575 base pairs and has a predicted open reading frame encoding 255 amino acids. Overall, it is similar to the catalytic enzyme PHZF, catalyzing the hydroxylation of phenazine-1-carboxylic acid to 2-hydroxyphenazine-1-carboxylic acid. Polymerase chain reaction-based mapping with both a monochromosomal hybrid panel and radiation hybrid cell panels placed the gene to human chromosome 10q21.1 near the marker D10S210.

Key words Two-hybrid · MAWD · PHZF · Chromosome 10q21 · RH mapping

Introduction

In this article we describe the cloning and characterization of a novel human gene that encodes a predicted protein similar to hydroxylase. We named the molecule *MAWBP* as it seemed to be a MAWD binding protein. MAWD was recently identified as a novel protein containing several WD-40 motifs (van der Voorn and Ploegh 1992), and it was found to promote the anchorage-independent growth of cells (Matsuda et al. 2000a); however, the biochemical and physiological role of MAWD has not yet been clarified. The identification and characterization of a binding protein to MAWD is necessary to elucidate the further precise mechanisms of its cellular function. Accordingly, we tried to identify potential MAWD partners. The use of the yeast two-hybrid system enabled us to isolate MAWBP. Other commonly used protein binding assays have not confirmed the interaction at present. However, as MAWBP had never been described before, this prompted us to further characterize the gene.

C. Iriyama · S. Matsuda (✉) · R. Katsumata · M. Hamaguchi
Department of Molecular Pathogenesis, Nagoya University School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8550, Japan
Tel. +81-52-744-2463; Fax +81-52-744-2464
e-mail: smatsuda@med.nagoya-u.ac.jp

The first two authors contributed equally to this work.
The nucleotide sequence data reported in this article have been deposited with the DDBJ, EMBL, and GenBank data libraries under the accession number AB049758.

Methods, results, and discussion

We used the two WD-40 repeats region (a.a. 214 to 350) of human MAWD for bait plasmids in the yeast two-hybrid screens, and identified a clone from a cDNA library constructed from human liver poly (A)+RNA. The clone cDNA was reconfirmed to be interaction-positive by using an independent yeast clone containing the bait. The positive clones contained insert sequences with the metallothionein gene (accession number, V00594), the S-protein gene (accession number, X03168) and an unknown gene; however, the metallothionein and the S-protein were fused to the GAL4 in a frameshift manner. This prompted us to characterize further the last unknown gene whose product seemed to be fused to GAL4 in a proper frame manner. DNA sequencing of the insert region with the unknown sequence was then performed, using a Sequenase Kit (Amersham, Cleveland, OH, USA) and an automatic sequencer for dideoxy sequencing (Amersham, Cleveland, OH, USA), according to the supplier's instructions. We used the insert nucleotide sequences for further screening to isolate a full-length clone and for searching the gene database to confirm the sequence and the gene structure (Matsuda et al. 1996; Miyazaki et al. 2000). The determined nucleotide sequence and predicted amino acid sequence are shown in Fig. 1A. The cDNA of 2575 bp contains an open reading frame of 765 bp. The MAWBP protein was then predicted to consist of 255 amino acids. There is a potential ATG start codon favored (Kozak 1989) from position 165 (Fig. 1A). A human genomic sequence that is, presumably, identical to the *MAWBP* gene, has already been registered in GenBank (AC016395) and there are two differences, on nucleotide (nt) 359 and nt 2047. At present, we can not rule out the

Fig. 1A,B. Representation of the *MAWBP* cDNA and sequences. **A** Nucleotide and predicted amino acid sequences of the *MAWBP* gene. The complete sequence of the cDNA was determined and confirmed by the primer walking strategy, using dideoxy sequencing. The nucleotide sequence data have been submitted to the GenBank/EMBL data libraries under the accession number AB049758. Numbering of the nucleotide and amino acid sequences is shown on the *right*. In the 3′-noncoding DNA sequence, the polyadenylation signal (AATAAA) exists (nucleotide [nt] 2429 to 2434). The original sequence of the two-hybrid positive clone started from nt 114 and ended at 2575. **B** Sequence comparison of *MAWBP* and T04A11.2. Sequence homology searches were conducted with the protein database at the National Center for Biotechnology Information (National Institutes of Health, Bethesda, MD, USA), using the BLASTP and CLUSTALW sequence database search tools. Amino acid sequence comparisons between the predicted human MAWBP (amino acids 1-255) and *Caenorhabditis elegans* T04A11.2 (amino acids 39-316) are shown. Residues of the proteins, indicated by *plus* (+), represent similar amino acid residues, and gaps are indicated by *dashes*. The one-letter amino acid code is used

```
AAGTGCCTTGAATTCCTTGGTAATTATTTCCTTTTAAGCGTGTAACTCTTATTATTGCTAAACTGCTTTTTAAGTTAATGGTAGTTGTTG     90
TCCAAAGGAAGTCTGATCTGTGTTTTTGAAACCTCAATGATCGTCTGTTTCTAATTAAGGAATTGGATGAAGACATGCATCAGAAAATTG    180
                                                                    M  H  Q  K  I  A      6
CAAGGGAGATGAACCTCTCTGAAACTGCTTTTATCCGAAAACTGCACCCGACAGACAACTTTGCACAAAGTTCCTGCTTTGGACTGAGAT    270
   R  E  M  N  L  S  E  T  A  F  I  R  K  L  H  P  T  D  N  F  A  Q  S  S  C  F  G  L  R  W     36
GGTTTACACCAGCGAGTGAGGTCCCACTCTGTGGCCATGCCACCCTGGCTTCTGCAGCTGTGCTGTTTCACAAAATAAAAAACATGAACA    360
   F  T  P  A  S  E  V  P  L  C  G  H  A  T  L  A  S  A  A  V  L  F  H  K  I  K  N  M  N  S     66
GCACGCTCACGTTTGTCACTCTGAGTGGAGAACTAAGGGCCAGACGAGCAGAGGACGGCATCGTCCTGGACTTGCCTCTTTATCCAGCCC    450
   T  L  T  F  V  T  L  S  G  E  L  R  A  R  R  A  E  D  G  I  V  L  D  L  P  L  Y  P  A  H     96
ACCCCCAGGACTTCCATGAAGTAGAGGACTTGATAAAGACTGCCATAGGCAACACACTGGTCCAGGACATCTGTTATTCTCCAGATACCC    540
   P  Q  D  F  H  E  V  E  D  L  I  K  T  A  I  G  N  T  L  V  Q  D  I  C  Y  S  P  D  T  Q    126
AAAAGCTCCTCGTCCGCCTCAGTGACGTTTACAACAGGTCGTTTCTGGAGAACCTGAAAGTGAACACGGAGAATCTGCTGCAAGTTGAAA    630
   K  L  L  V  R  L  S  D  V  Y  N  R  S  F  L  E  N  L  K  V  N  T  E  N  L  L  Q  V  E  N    156
ACACAGGGAAGGTGAAAGGGCTTATTCTTACCCTTAAAGGAGAGCCTGGTGGGCAGACCCAAGCATTTGACTTTTACTCAAGATATTTTG    720
   T  G  K  V  K  G  L  I  L  T  L  K  G  E  P  G  G  Q  T  Q  A  F  D  F  Y  S  R  Y  F  A    186
CACCGTGGGTTGGTGTGGCTGAAGACCCAGTGACAGGGTCTGCACACGCTGTTCTCAGCAGCTACTGGTCCCAGCATCTGGGGAAGAAAG    810
   P  W  V  G  V  A  E  D  P  V  T  G  S  A  H  A  V  L  S  S  Y  W  S  Q  H  L  G  K  K  E    216
AAATGCATGCTTTTCAGTGTTCCCACCGAGGAGGAGAGCTGGGAATTTCCCTTCGTCCAGACGGAAGGGTTGACATTAGAGGAGGTGCAG    900
   M  H  A  F  Q  C  S  H  R  G  G  E  L  G  I  S  L  R  P  D  G  R  V  D  I  R  G  G  A  A    246
CTGTTGTTTTAGAGGGCACACTGACAGCCTAGAGGTGGTTATGCTGTGACGCTGCTGTCTCTAACCACCAAGTATTTTCTGCTTAAAAAG    990
   V  V  L  E  G  T  L  T  A  *                                                               255
AAATGTAAGGGGCTGCCTTTAGCAAATGTGCGTAGTAGTCTACTTAATCCTCATGTTAAAAATCGAAAAATGGGCCAGGCGCAGTGGCTC   1080
ATGCCTGTAATCGTAGCACTTTGAGAGGCCAAGGTGGGTGGATCACCTGAGGTCAGGGGTTCGACACCAGCCTGGCCAACATGGTGAAAC   1170
CTCGTCTCTACTACAAATACAAAAATTAGCTGGGTGTGGTGCCACATGCCTGTAATTCCAGCTACTTGGGAGGCTGAGGCATGGAGAATC   1260
GCTTGAACCCAGGAGGCAGAAGTTGCAGTGAGCAGAGATCACACCACTGCACTCCAGCCTGGGCAACAGAGTAAGACTCCATCCACACAC   1350
ACACACACACACAAAATTGAAAAGTGAAGACATTTTAATGGAGATTTAATAGTGCTTCCAGACTAATGAACTAATGGAGTTTTGGCTC   1440
CACTCATGAGTGTATTTGAAATGTAAGTAACCAGCTACAAAGAATAATGTCACTTCATTTGATTATGACTACCAATCAAGAGAAGGAGGA   1530
ATACATTTCTGAGGAGTGATACTAAACCATTTGAGCTTAAATGAGTACCTGATTTTGCAGCCATTAAAATGAGTCAATAACTATGTGGAA   1620
ATATAGAAAAATATTTTATAATATAATACTAAGAAAGAGGACAGTAGGCCATAAAGTGTTAGCAGTGTTATGCTATGTCTATATAAATAA   1710
GCATGGGCCATTAAAAATAGCAAACTGAAAACAGTTCATGTGTGGGGGGCAAGTTTAGAAGTAATTTTAATGTATCTAAGAAGTGATTTC   1800
CTTTATTGTTGTACTCAGGTACGAAGACTAAAAATCAGGATGAATCCAGAAAATTAACACAAAAAAGTATGTATTAGGCTACTTCCTGTC   1890
ATAGTGAAATTAAACACTCAAAGTAATGTAATCTCAGCTGGGCGTGGCGGCTCATGCCTGTAATCCCAGCACTTTGGCAGGACAAGGCAG   1980
GAGGATCACTTGAACCTAGGAGTTCGAGACCAGCCTGGGCAACACAGCGAGAACTTGTCTCTATTAAAAAAAAAAAAAAGGAATGTAATT   2070
TCAATCTTTTTCTTGATATATGGCTTGAGAATGATAATGTAAAAGGAATTCTTCTCTTACTTCAATAAAATGGGTTTTAACATAACTTTA   2160
AATTCAGTTAAATATACAATATTGAATACCTATAGTTGACTTTGGGATGGGGACTTTTTCAAGTCATTAAGAGTGTTTGTTTAAGGTGAT   2250
CTCATTGATGGTAGTTCTCAGCCGTCTCAAAAACTGCAAGCTAATCAGTCAGACATTCTTTAATGACCCCAATTTTTTCACTTTAATTGT   2340
TACCATGTTTTCTATTTTTACTGATTTTTGCTAAAGCATGTAAGAGTGAATTTATTATAGCAGTAATCTTGTGTTTCTCCTGATGTGCAA   2430
TAAATCAATAATTCACCAGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA   2520
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA                               2575
```

A

```
MAWBP   : 4   KIAREMNLSETAFIRKLHPTDNFAQSSCFGLRWFTPASEVPLCGHATLASAAVLFHKIKN 63
              KIA E NLSETAF   +  TD F   S F LRWFTP +EVPLCGHATLA++ VLF+++ N
T04A11.2: 39  KIAAEFNLSETAFPVPIGSTD-FKTCSQFSLRWFTPKTEVPLCGHATLATSHVLFNEVGN 97

MAWBP   : 64  MNSTLTFVTLSGELRARRAEDG-IVLDLPLYPAHPQDFHEVE----------------D 105
              +N  + F T SG L A+R E G + ++ P Y      F++                  +
T04A11.2: 98  VNKEIKFDTQSGVLIAKRDEKGNVEMNFPEYDLASVKFNDTPNPLQGIFSEFEAPSFLFE 157

MAWBP   : 106 LIKTAI-GNTLVQDICYSPDTQKLLVRLSDVYNRSFLENLKVNTENLLQVENTGKVKGLI 164
              +IK  +    +++ + YS  ++KL+V +     +  LE ++++   +L++ +   V+GL
T04A11.2: 158 VIKCVVPTEMIIESVVYSSKSRKLIVIVDPETTKFELEAVRIDCSKMLEIHDGSFVRGLA 217

MAWBP   : 165 LTLK-------GEPGGQTQAFDFYSRYFAPWVGVAEDPVTGSAHAVLSSYWSQHLGKKEM 217
              ++L+       G      + +D+  RYFAPWVG+ EDP TGSA  V+  +WS+ LGK E+
T04A11.2: 218 ISLRPSNPMAQGFVDSSDEPYDYACRYFAPWVGIDEDPATGSAQCVMGPFWSKMLGKNEL 277

MAWBP   : 218 HAFQC-SHRGGELGISLRPDGRVDIRGGAAVVLEGTLTA- 255
              +AFQ    RG +  I L+ D RV + G +  VL G L
T04A11.2: 278 YAFQAFPTRGAQFRIRLQ-DDRVILNGPSVTVLRGELALN 316
```

B

possibility that either a polymorphism or a mutation may exist at this site.

Searching the protein database (SWALL and PIR) revealed that MAWBP had high homology to several other proteins. The amino acid sequence of MAWBP was most homologous to that of T04A11.2 (37% identical and 55% similar) of *Caenorhabditis elegans*. The amino acid sequences in these proteins were compared (Fig. 1B). How-

ever, the alignment of the deduced amino acid sequence of MAWBP with the conceptual translation of genes from various species showed that the MAWBP seemed to be highly conserved throughout evolution from single-cell to higher eukaryotes, such as CAB77771, BAB04002, AAG06158, AAG18908, and AAB82044 (data not shown). It is amazing that the biochemical functions of all these proteins basically remain elusive. Because one of these homologous proteins is the catalytic enzyme PHZF (accession number; AAD15343; 37% identical and 50% similar to overall MAWBP), which is suspected to catalyze the hydroxylation of phenazine-1-carboxylic acid to 2-hydroxyphenazine-1-carboxylic acid, we speculate that MAWBP may serve as a similar catalyzing enzyme. In order to understand the biological functions of MAWBP, it is important to confirm the enzymatic activity and to identify the MAWBP binding protein, excepting MAWD.

The tissue distribution of *MAWBP* transcripts in various tissues was examined by using cycle-limited reverse transcription-coupled polymerase chain reaction (RT-PCR), as described previously (Seki et al. 1999; Matsuda et al. 2000b). Primers used for RT-PCR correspond to the coding region of the gene (the expected product corresponds to nt 803 to 979). Template-cDNAs from the human tissues were purchased from Clontech (Palo Alto, CA, USA), and ten nanograms of each cDNA was used in the experiment. As shown in Fig. 2, expression of *MAWBP* mRNA was detected in brain, heart, lung, liver (most obvious), pancreas (obvious), kidney (obvious), and placenta, but not in skeletal muscle, by using cycle-limited RT-PCR (28 cycle-amplified), suggesting that *MAWBP* may be expressed ubiquitously, except for skeletal muscle, although the level of expression varied. In this experiment, genomic DNA may not have been amplified, because primers designed from the noncoding region of *MAWBP* could not amplify (data not shown). Considering its expression profile, MAWBP may be involved in the basic functions of many cells.

To determine the chromosomal localization of the *MAWBP* gene, PCR-based monochromosomal somatic cell hybrid mapping (Quantum, Montreal, Quebec, Canada) was performed with a set of 3′ untranslated region (UTR) primers (5′ > TGACAGCCTAGAGGT < 3′ and 5′ > TGCTACGATTACAGGCAT < 3′, corresponding to the nt from 911 to 1098). This study indicated that the *MAWBP* gene is located on human chromosome 10 (Fig. 3A). To

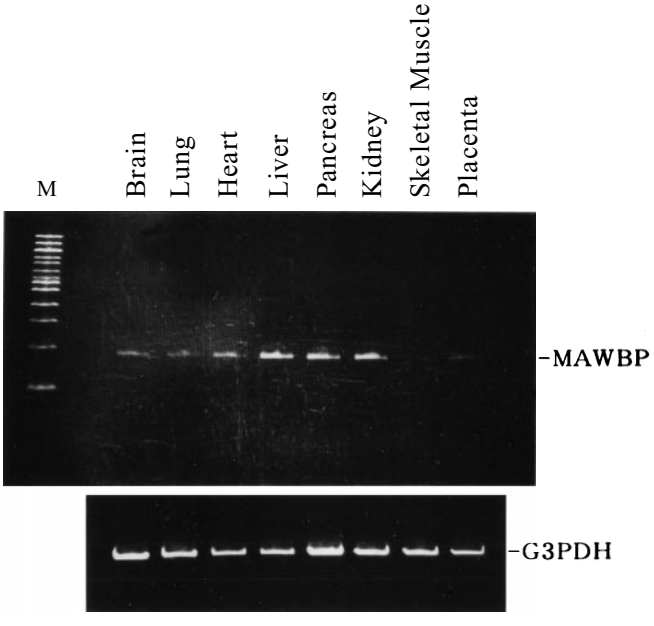

Fig. 2. Expression of the *MAWBP* gene in various normal human tissues. Reverse transcription and amplification by polymerase chain reaction (PCR) with the specific primers for the *MAWBP* gene were carried out to analyze *MAWBP* expression (*top panel*). The eight tissues examined are indicated *above each lane*. The template cDNAs for brain (*lane 1*), lung (*lane 2*), heart (*lane 3*), liver (*lane 4*), pancreas (*lane 5*), kidney (*lane 6*), skeletal muscle (*lane 7*), and placenta (*lane 8*) of the normal human tissues were purchased from Clontech (Palo Alto, CA, USA). *Lane m*, marker. The *bottom panel* shows the expression of the glycerol-3-phosphate dyhydrogenase (G3PDH) that was analyzed as a control

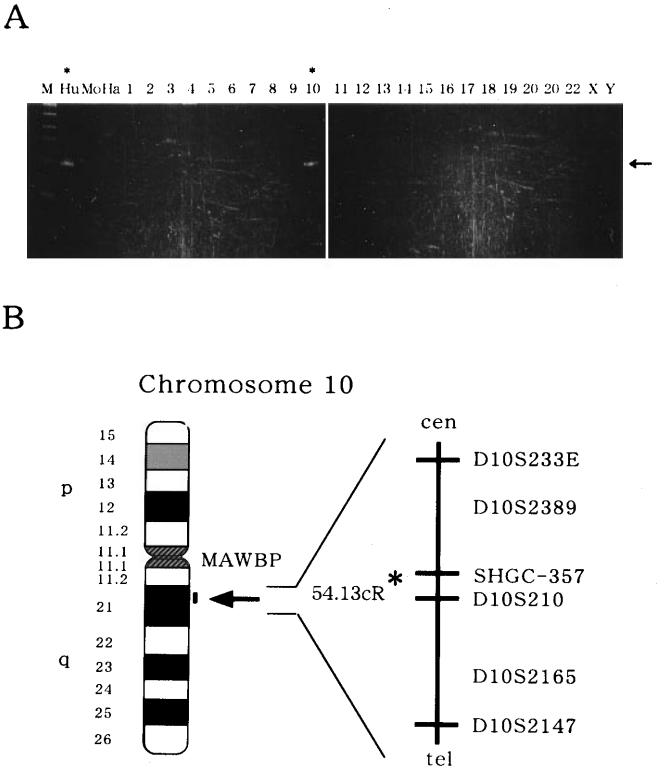

Fig. 3A,B. Chromosomal assignment of the *MAWBP* gene. **A** Representative PCR-based monochromosomal somatic cell hybrid mapping of *MAWBP*. Primers from the 3′-untranslated region (UTR) of the *MAWBP* cDNA were used. *Lanes 1–22* and *lanes X and Y* represent human chromosomes. *Lanes Hu*, *Mo*, *and Ha* represent cell hybrids containing genomic DNA from human, mouse, and hamster, respectively. A single product of the expected size was generated from chromosome 10 and human genomic DNA (indicated by the *asterisk*). *Arrow* indicates expected size of PCR product. **B** Chromosomal localization of the *MAWBP* gene in a schematic ideogram of human chromosome 10. The approximate corresponding cytogenetic location of the gene on chromosome 10q21.1, determined by the results of radiation hybrid (RH) panel mapping, is shown. The *asterisk* shows the putative localization of the *MAWBP* gene. *Cen*, Centromere; *tel*, telomere

refine the further subchromosomal location of the gene, the radiation hybrid (RH) panels (Stanford G3 and Genebridge 4) were utilized. Both of the linkage analyses of the PCR results (the data vector for *MAWBP* of the Stanford G3 and the Genebridge 4 RH panels: 11000 00010 00001 00010 00000 00000 00010 00001 00111 00100 00101 11100 01000 10010 00000 00010 101 and 00011 10010 00000 00010 11101 01110 11001 11000 01010 00100 00000 10000 01101 00100 01010 00000 11100 01000 111, respectively) showed consistently that the *MAWBP* gene was linked to several markers adjacent to the marker D10S210 in chromosome 10q21.1, with lod score values higher than 5.6. The most likely order of the refined loci is shown in Fig. 3B. Other genes that have been mapped to chromosome 10q21.1 near the *MAWBP* gene include the myxoid chondrosarcoma translocation gene (Shen et al. 1990) and the Charcot-Marie-Tooth disease type 1D (*CMT1D*) gene (De Jonghe et al. 1999). Although there is no evidence of MAWBP involvement in the genesis of these diseases, our precise chromosomal positioning data could contribute toward positional candidate approaches for these disease genes linked to this locus. Future studies will address the biological role of MAWBP in this regard, with reference to MAWD-binding.

Acknowledgments This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas and for COE Research from the Ministry of Education, Science, and Culture of Japan, a Grant under the Monbusho International Scientific Research Program, and a Grant from the Aichi Cancer Research Foundation. We thank Yu-ki Iwata for her excellent technical assistance.

References

De Jonghe P, Timmerman V, Nelis E, De Vriendt E, Lofgren A, Ceuterick C, Martin JJ, Van Broeckhoven C (1999) A novel type of hereditary motor and sensory neuropathy characterized by a mild phenotype. Arch Neurol 56:1283–1288

Kozak M (1989) The scanning model for translation: an update. J Cell Biol 108:229–241

Matsuda S, K-Tsuzuku J, Ohsugi M, Yoshida M, Emi M, Nakamura Y, Onda M, Yoshida Y, Nishiyama A, Yamamoto T (1996) Tob, a novel protein that interacts with p185erbB2, is associated with anti-proliferative activity. Oncogene 12:705–713

Matsuda S, Katsumata R, Okuda T, Yamamoto T, Miyazaki K, Senga T, Machida K, Thant AA, Nakatsugawa S, Hamaguchi M (2000a) Molecular cloning and characterization of human MAWD, a novel protein containing WD-40 repeats frequently overexpressed in breast cancer. Cancer Res 60:13–17

Matsuda S, Miyazaki K, Ichigotani Y, Kurata H, Takenouchi Y, Yamamoto T, Nimura Y, Irimura T, Nakatsugawa S, Hamaguchi M (2000b) Molecular cloning and characterization of a novel human gene (*NESCA*) which encodes a putative adapter protein containing SH3. Biochim Biophys Acta 1491:321–326

Miyazaki K, Matsuda S, Ichigotani Y, Takenouchi Y, Hayashi K, Fukuda Y, Nimura Y, Hamaguchi M (2000) Isolation and characterization of a novel human gene (*NESH*) which encodes a putative signaling molecule similar to e3B1 protein. Biochim Biophys Acta 1493:237–241

Seki N, Hattori A, Hayashi A, Kozuma S, Ohira M, Hori T, Saito T (1999) Structure, expression profile and chromosomal location of an isolog of DNA-PKcs interacting protein (*KIP*) gene. Biochim Biophys Acta 1444:143–147

Shen WP, Young RF, Walter BN, Choi BH, Smith M, Katz J (1990) Molecular analysis of a myxoid chondrosarcoma with rearrangements of chromosomes 10 and 22. Cancer Genet Cytogenet 45:207–215

van der Voorn L, Ploegh HL (1992) The WD-40 repeat. FEBS Lett 307:131–134