Meiotic human sperm cells express a leucine-rich homologue of Caenorhabditis elegans early embryogenesis gene, Zyg-11.

Chloé Féral, Yuan-Qing Wu, André Pawlak, Georges Guellaën

Meiotic human germ cells express a leucine-rich homologue of

Caenorhabditis elegans early embryo genesis gene, Zyg-11

Chloé Féral, Yuan-Qing Wu, André Pawlak and Georges Guellaën

Unité INSERM 99, Hôpital Henri Mondor, 94010 Créteil, France

running title: Human testis homologue of *C.elegans* Zyg-11.

Georges Guellaën
Unité INSERM 99
Hôpital Henri Mondor
94010 Créteil (France)
Tel: 33 1 49 81 35 30
Fax: 33 1 48 98 09 08
E-mail: guellaen@im3.inserm.fr

Abstract

We cloned a human protein (Hzyg) homologue to the *Caenorhabditis elegans* Zyg-11, an essential protein for the cell division at the initial developmental stages of this species, and a *Drosophila melanogaster* gene (*Mei-1*) likely to be involved in meiosis. Hzyg mRNA encodes a protein of 766 amino acids (88 kDa) containing 14 % of leucine residues, some of them being arranged in four leucine rich repeat motives usually involved in protein-protein interactions. Hzyg is encoded by a single gene, located on chromosome 9q31-q32.4, and transcribed as two mRNA: i) a 5 kb transcript heavily expressed in testis and skeletal muscle, and barely detectable in other human tissues; ii) a 3.1 kb abundant mRNA detected only in the testis. By using in situ hybridisation and immunohistology, we clearly established the presence of Hzyg in pachytene spermatocytes (stage V) and spermatids (stage I and/or two) surrounding meiosis. The cell specific expression of Hzyg transcripts in testis, and the conservation of this gene among distant species, endowed this protein with a putative important role during meiosis.

Introduction

We produced a large series of human testis Expressed Sequence Tags (EST), (Pawlak et al. 1995). During the course of their characterization, the clone A11107 (EST Hzyg) exhibited a significant homology to the *Caenorhabditis elegans* Zyg-11 gene, a gene which contributes to the cytoplasm organization and spindle orientation at one cell stage in this worm (Kemphues et al. 1986). Actually, temperature-sensitive Zyg-11 defective mutants of *Caenorhabditis elegans* embryos exhibit development retardation, arrest of meiosis at metaphase II, a delay in the formation of the pronuclei and an incorrect placement of the first cleavage furrow. All these effects can be rescued by injection of Zyg-11 cDNA (Carter et al. 1990).

EST Hzyg also displayed a significant homology to a *Drosophila melanogaster* gene located in region 62A1-62A2 of the chromosome 3L. Mutations in this region disrupt male meiotic division I (Ivy 1981) and are supposed to target a yet unidentified gene named *Mei-1*. This gene is not related to the gene *mei-1* from *Caenorhabditis elegans* (Clak-Maguire et al. 1994). The analysis of this genomic region revealed several genes and the gene homologous to EST Hzyg seems the sole likely candidate for *Mei-1*.

Based on the similarities of EST Hzyg with these *Caenorhabditis elegans* and *Drosophila melanogaster* genes potentially important for meiosis, we investigated the expression of the Hzyg mRNA and the associated protein in human testis. Hzyg was found expressed specifically at meiotic stages of spermatogenesis. The conservation of this gene among distant species endowed this protein with a putative important role during human male meiosis.

Materials and Methods

Tissues, DNA and RNA preparation

Testicular samples, genomic DNA and mRNA, were obtained and prepared as previously reported (Pawlak et al. 1995; Levy et al. 1996).

cDNA cloning and sequencing

The initial Hzyg cDNA (1 kb-clone A11107) is part of a series of human testis cDNA clones already described (Pawlak et al. 1995). The 5' end region of the open reading frame was obtained by extending this 1 kb following two successive primer extensions. First, a 28 bp oligonucleotide (nt 341-369 of clone A11107) was hybridised to 1 μg of human testis mRNA and was elongated by using the Marathon cDNA amplification kit (Clontech, Palo Alto, USA). The elongated products were ligated at their 5' end with an adaptator (Clontech, Palo Alto, USA) and PCR-amplified using two oligonucleotides complementary to the adaptator and to the position 314-341 of clone A11107 respectively. A fragment of 1.2 kb was obtained. A second primer extension was done using the same protocol and two oligonucleotides located on the 1.2 kb fragment at position 400-423 and 384-406 nt for reverse transcription and amplification steps respectively. This second primer extension generated a fragment of 900 bp. These two overlapping fragments were cloned into the pSPORT2 plasmid. The A11107, 1.2kb and 0.9 kb cDNA clones were sequenced using the Erase-a-base system (Promega, Madisson, USA) as reported (Giuili et al. 1992). The fragments were sequenced at least five times on each strand as previously described (Pawlak et al. 1995).

The nucleotide and protein sequences were analysed on the server Infobiogen (www.infobiogen.fr) using several programs including Protparam (Expasy), SAPS (ISREC), Blocks, Psort and NetOglyc WWW server.

Southern and northern blot analysis

Southern blotting of human genomic DNA was performed as previously reported (Levy et al. 1996), using 20 μg per lane of human genomic DNA digested with the restriction enzyme HindIII, PstI, PvuII, EcoRI and KpnI (New England Biolabs, Hertfordshire, UK). Northern blots of human tissue RNAs were obtained from Clontech, Palo Alto, USA. The blots were hybridised to the insert of clone A11107, [^{32}P] labelled by random priming using the Megaprime protocol (Promega, Madisson, USA). The Southern blot was washed at 68 °C in 0.1 x SSC (1 x SSC: 150 mM NaCl, 15 mM sodium citrate pH 7.0), 0.1% SDS and exposed overnight on a ß-max X-ray film (Amersham France, Les Ulis, F) with one intensifying screen. The northern blots were washed at 65 °C in 0.05 x SSC, 0.1% SDS and exposed as for the Southern blots except for 48 hours.

In situ hybridisation

An EcoRI-HindIII fragment of 555 bp (nt 228-783 of clone A11107) was cloned into the corresponding restriction sites of pSPORT1 (Life technologies, Gibco BRL). The resulting plasmid, linearised either using the restriction enzymes BamHI or RsrII, was used as template for sense and antisense single strand RNA probes, using Sp6 or T7 RNA polymerase respectively. The labelling reactions were done as described previously (Matsuoka et al.

1992), using [^{33}P] UTP instead of [^{35}S] UTP.

Testis tissue samples from a normal adult male were frozen in OCT-compound and sectioned at 10 μm. The sections were fixed and hybridised as previously reported (Matsuoka et al. 1992) except that the probe (6.10^5 cpm) was added to the tissue sections in 30 μl. The sections were rinsed twice in 4 x SSC and 2 x SSC for 5 min and in 1 x SSC at room temperature for 10 min. The non-specific signal was removed by treatment with a solution containing 20 μg/ml RNAse A, 20 U/ml RNAse T1, 10 mM Tris pH 8, 0.5 M NaCl and 1 mM EDTA for 45 min at 37 °C. The sections were then washed at 42 °C in 1 x SSC for 30 min, 0.1 x SSC for 15 min and at 60 °C in 0.05 x SSC three times for 30 min and finally at room temperature in 0.05 x SSC for 10 min. The sections were dehydrated in ethanol in the presence of 0.3 M ammonium acetate, air dried and exposed to ß-max film (Amersham France, Les Ulis, F) for 4 days. The sections were then coated with LM-1 emulsion (Amersham France, Les Ulis, F) exposed for 3 weeks, developed in Kodak reactives and counterstained with O-toluidine. Controls were carried out on adjacent sections hybridised to the sense probe.

Antibodies

A 13 amino acid peptide (CSNFKEENMDTSR), corresponding to the carboxy terminal part of Hzyg, was coupled to keyhole limpet hemocyanin (KLH) as immunogen (Syst:em, Nimes, F). Hzyg antibodies against this antigen were raised in rabbits and were purified on NHS-activated column (Amersham France, Les Ulis, F) coupled to the specific peptide following manufacturer's instructions.

Western blot analysis

Human testis and kidney tissues were homogenized in 25 mM Tris-HCl buffer pH 6.8 containing 10% glycerol, 2% SDS, 0.7 M ß-mercaptoethanol, and 125 μg/ml bromophenol blue. After centrifugation, the supernatants were boiled for 5 min. Proteins were separated on 10% SDS-PAGE, blotted onto a PVDF transfer membrane (Millipore, St Quentin Yvelines, F), and probed with the anti-Hzyg serum. Treatment of the membrane was carried out according to the enhanced chemifluorescence protocol (ECF, Amersham France, Les Ulis, F). Sheep anti-rabbit IgG coupled to horseradish peroxidase (1/10000) was used as secondary antibody. The membrane was incubated for 7 min in ECF solution and fluorescent signals were visualized using a phosphofluoroimager STORM 840 (Molecular Dynamics, Sunnyvale, USA).

Immunohistofluorescence

Normal human testis paraffin sections (6 μm), picked up on polylysine slides, were deparaffinised and heated in a microwave oven for 3 x 5 for antigen retrieval. Slides were incubated in 1x PBS containing 10% sheep serum and 0.3% Triton X100 for 30 min, and then

incubated with affinity-purified anti-Hzyg antibodies at room temperature for 45 min. The sections were then washed and overlaid with CY3-conjugated anti-rabbit IgG (Amersham France, Les Ulis, F) at room temperature for 30 min. DNA was stained using 4,6-diamidino-2-phenylindole (DAPI).

Results

Cloning of Hzyg

In a precedent study, we identified a large series of human testis EST (Pawlak et al. 1995). One of them, translated in amino-acid sequence, exhibited a significant homology (blast score = 124; P: 4 e^{-27}) with amino acids 552 to 619 of the *Caenorhabditis elegans* Zyg-11 protein. The corresponding cDNA clone (clone A11107 - 1 kb), elongated by two successive 5'primer extensions, gave two overlapping fragments of 1.2 kb and 900 bp respectively. From the sequences of the initial clone and these two fragments we generated a 2548 bp unique nucleotide sequence (Accession number: X99802).

The ATG, at nt 39, in a correct context for the initiation of translation (Kozak 1991), opens a reading frame (2298 bp) susceptible to encode a protein (Hzyg) of 766 amino acids (figure 1). This sequence, in its 5' end, perfectly matches with two human EST (Acc n° AL036600 and AI903478), which extend 160 nt upstream of Hzyg cDNA, and contains several stop codons in frame with the ATG initiation codon at nt 39. Four EST (Acc n°: AA025815, AA687640, AA431332 and AW204619) extended the 3'Hzyg nucleotide sequence (212bp) for 46 nt. Thus, the 3' UTR mRNA is 258 nt long; it contains an AACAAA sequence as a non-canonical polyadenylation signal located 10 bp upstream the polyA tail. Such an unusual polyadenylation signal has already been observed in several human genes including hRPS27 (Tsui et al. 1996), ARF3 (Tsai et al. 1991), factor XI (Fujikawa et al. 1986) as well as in the *Drosophila melanogaster* tropomyosin gene (Boardman et al. 1985) and the chicken type II procollagen gene (Sandell et al. 1984).

Hzyg ORF encodes a 766 amino acid sequence for a total molecular weight of 88 kDa. The sequence of this protein, analysed by using Psort and Protparam programs, revealed that Hzyg is likely to be cytoplasmic and relatively unstable. It contains 103 acidic residues (Asp + Glu) and 83 basic residues (Arg + Lys) not arranged in clusters. There are three potential N-glycosylation sites (Asn 294, 464, 553) and one O-glycosylation site (Ser 308). Hzyg is very rich in Leucine with 108 residues (14.1 % of the total residues). At positions 105, 130, 297 and 354 (figure 1), several of these Leu are part of consensus motives for leucine-rich repeats (LRR) (Kobe and Deisenhofer 1994; Kobe and Deisenhofer 1995). LRR contain 20 to 29 amino acids; the 10 amino terminal ones fit the consensus motif LXXLXLXX(N/T/C), in which X is any amino acid; the carboxy terminal region of the LRR, variable in length, contains several aliphatic amino acids such as A, V, L, I, F, Y or M.

Hzyg exhibited 24% identity and 45% similarity with *Caenorhabditis elegans* Zyg-11 (figure 2), and matched also with several unknown genes in this species, indicating that Zyg-11 belongs to a *Caenorhabditis elegans* multigene family. In the *Drosophila melanogaster* genome, Hzyg matched with only a single gene product (Acc. n°: AAF47500) (figure 2). This gene, located on chromosome 3L, codes for a putative yet uncharacterised protein, which exhibited 38% identical residues and 54% similarities with Hzyg. Zyg-11 and the *Drosophila melanogaster* gene exhibited 2 and 3 LRR respectively.

Southern blot analysis

A Southern blot of human genomic DNA, digested by using KpnI, EcoRI, PvuII, PstI and HindIII restriction enzymes, and hybridised to Hzyg cDNA, revealed 1 to 4 bands per

lane (figure 3). Such a simple pattern argues for the presence of a single Hzyg gene in the human genome.

Chromosomal localization

Four sequence tag sites (STS) corresponding to Hzyg nucleotide sequences (Acc N°: A005A01; stSG4875; stSG4266; A007G01) have been mapped on chromosome 9 between D9S282 and D9S260 (136-141 cM) (Unigene database - Accession number: Hs.29285). In addition, the comparison of Hzyg nucleotide sequence with the human genome sequence (Acc. N° AL359678) confirmed the location of the Hzyg gene on Chromosome 9. Thus, it clearly appears that Hzyg is encoded by a single gene located on the long arm of chromosome 9 in q32-q34.1.

Hzyg mRNA expression

Northern blot analysis of Hzyg mRNA expression in several tissues revealed, under high stringency conditions, a strong signal at 3.1 kb, detected only in human testis. An abundant 5 kb transcript was also detected in testis and skeletal muscle and to a lower level in prostate, ovary, small intestine, heart, brain and pancreas (figure 4). A strong signal at 3.1 kb, was detected only in human testis.

The cellular localization of Hzyg mRNA expression was determined by in situ hybridization on human testis sections. Dark field examination at low magnification revealed a distribution of the grains inside the tubules (figure 5 A and B). At higher magnification, the signal was mainly located on cells with large nucleus corresponding to pachytene

spermatocytes (figure 5, E, F) as well as on smaller cells located in the center of the tubule

and identified as spermatides (figure 5, G, H). Under identical conditions the sense probe did

not give any signal (figure 5 C and D). A very low signal was observed in the intersitial

parenchyma. A careful examination of the grain distribution clearly showed that the signal

was concentrated in the cytoplasm of round cells at the central and adluminal compartments

of seminiferous tubules, suggesting that the spermatocytes and spermatids were the main

source of Hzyg transcripts. The signal was low near the boundaries of tubules, thus excluding

a major contribution of Sertoli cells and spermatogonia, which are located along the basement

membrane. Therefore, in the human testis, Hzyg is expressed mainly in germ cells that

undergo the meiotic division.

Hzyg protein expression

Antibodies raised against the carboxy-terminal part of Hzyg revealed a single band at

the expected size (88 kDa), on human testis extracts (figure 6). By using these antibodies on

human testis sections, we observed a very specific expression of Hzyg in the cytoplasm of late

pachytene spermatocytes (mainly stage V), and in the cytoplasm of round spermatids (stage I

and/or II) (figure 7).

Discussion

We cloned from a human testis library a new human protein (Hzyg), which is expressed at meiotic stages of human spermatogenesis.

In these cells, Hzyg mRNA is transcribed from a single gene as two major transcripts (5 and 3.1 kb), whereas a single 5 kb band is detected in skeletal muscle and to a lower extent in prostate, ovary, small intestine, heart, brain and pancreas. Such a specific expression pattern of a single gene in testis as compared to somatic tissues is often observed in germinal cells. For example, this is the case for the farnesyl pyrophosphate synthetase transcript which displays a longer 5'UTR in testis as compared to the transcript expressed in somatic tissues (Teruya et al. 1990), and c-abl (Meijer et al. 1987) or ß1-galactosyl transferase mRNA (Shaper, 1990) which are polyadenylated at a different site in testis as compared to other tissues. The structure of the two Hzyg transcripts detected in testis is not known yet, however they may differ in their 3'UTR sequences as indicated from the Expressed Sequences Tags present in databases. Nevertheless, if both Hzyg transcripts are functional, the detection of a single band on western blot of testis extracts suggests that either they code for the same protein or that post-transcriptional events modify the epitope reacting with the antibody.

Hzyg contains 766 amino acids (88 kDa) and the main feature of its structure resides in the presence of 14 % leucine residues. Several of them gathered in four consensus motives for leucine-rich repeats (LRR) (Kobe and Deisenhofer 1994; Kobe and Deisenhofer 1995). LRR motives are generally involved in protein-protein interactions, and in 50 % of the cases in signal transduction pathways (Kobe and Deisenhofer 1994; Kobe and Deisenhofer 1995).

This structural feature will help for future investigations in order to decipher the role of Hzyg in human testis.

The most interesting point is a possible function of Hzyg in meiosis. In fact, we observed a specific expression of Hzyg in the cytoplasm of late pachytene spermatocytes (mainly stage V), as well as in the cytoplasm of the round spermatids (stage I and/or II), at two stages corresponding to meiotic division. The observation of Hzyg transcripts in somatic tissues is not contradictory to a specific role of Hzyg during meiosis. As an example proteins such as SPO11, which initiates the homologous recombination during meiosis (Romanienko and Camerini-Otero 1999) are also expressed in somatic tissues. In addition, Hzyg exhibits significant homologies with the *Caenorhabditis elegans* Zyg-11 protein, a protein which a determinant role during meiosis, as shown by the meiosis arrest at metaphase II and an incorrect placement of the first cleavage furrow in Zyg-11 deficient worms. Finally, Hzyg is the orthologue of a *Drosophila melanogaster* gene located in region 62A1-62A2 of the chromosome 3L likely to correspond to the gene *Mei-1* and mutations at this locus promote a non-disjunction of the chromosome at first meiotic division. The second phase of meiosis appears normal except for the non-haploid complements resulting from metaphase I chromosome misbehaviour (Ivy 1981).

Apparently this protein is not indispensable for somatic mitosis since it was not detected in all human somatic tissues undergoing cell division and we did not detect this protein in rapidly dividing tissues such as human testis tumours (data not shown). In summary Hzyg is likely to play a role during male germ cell division. Already two animal models with knock

out genes related to Hzyg are available. In fact, major alterations of meiosis have been reported in *Drosophila melanogaster* and *Caenorhabditis elegans* deficient in Hzyg homologous gene. Further investigations are needed to decipher the role of Hzyg in human germ cells.

Acknowledgments

Literature cited

Boardman, M., Basi G.S., and Storti R.V. (1985) Multiple polyadenylation sites in a *Drosophila melanogaster* tropomyosin gene are used to generate functional mRNAs. *Nucleic Acids Res* **13**: 1763-76.

Carter, P.W., Roos J.M., and Kemphues K.J. (1990) Molecular analysis of zyg-11, a maternal-effect gene required for early embryogenesis of Caenorhabditis elegans. *Mol Gen Genet* 22 **1**: 72-80.

Clak-Maguire, S., and Mains P.E. (1994) *mei-1*, a gene required for meiotic spindle formation in *Caenorhabditis elegans,* is a member of a family of ATP-ases. *Genetics* **136**: 533-46.

Fujikawa, K., Chung D.W., Hendrickson L.E. *et al.* (1986) Amino acid sequence of human factor XI, a blood coagulation factor with four tandem repeats that are highly homologous with plasma prekallikrein. *Biochemistry* **25**: 2417-24.

Giuili, G., Scholl U., Bulle F. *et al.* (1992) Molecular cloning of the cDNAs coding for the two subunits of soluble guanylyl cyclase from human brain. *FEBS Lett* **304**: 83-8.

Ivy, J. (1981) Mutations that disrupt meiosis in males of drosophila melanogaster. (Thesis) University of California, San Diego, California, USA.

Kemphues, K.J., Wolf N., Wood W.B. *et al.* (1986) Two loci required for cytoplasmic organization in early embryos of Caenorhabditis elegans. *Dev Biol* **113**: 449-60.

Kobe, B. and Deisenhofer J. (1994) The leucine-rich repeat: a versatile binding motif. *Trends Biochem Sci* **19**: 415-21.

Kobe, B. and Deisenhofer J. (1995) Proteins with leucine-rich repeats. *Curr Opin Struct Biol* **5**: 409-16.

Kozak, M. (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. *J Cell Biol* **115**: 887-903.

Levy, I., Wu Y.Q., Roeckel N. *et al.* (1996) Human testis specifically expresses a homologue of the rodent T lymphocytes RT6 mRNA. *FEBS Lett* **382**: 276-80.

Matsuoka, I., Giuili, G., Poyard M. *et al.* (1992) Localization of adenylyl and guanylul cyclase in rat brain by in situ hybridization: comparison with calmodulin mRNA distribution. *J Neurosci* **12**, 3350-60.

Meijer, D., Hermans A., von Lindern M. *et al.* (1987) Molecular characterization of the testis specific c-abl mRNA in mouse. *Embo J* **6**: 4041-8.

Pawlak, A., Toussaint C., Levy I. *et al.* (1995) Characterization of a large population of mRNAs from human testis. *Genomics* **26**: 151-8.

Romanienko, P.J. and R.D. Camerini-Otero R.D. (1999) Cloning, characterization, and localization of mouse and human SPO11. *Genomics* **61**: 156-69.

Sandell, L.J., Prentice H.L., Kravis D. *et al.* (1984) Structure and sequence of the chicken type II procollagen gene. Characterization of the region encoding the carboxyl-terminal telopeptide and propeptide. *J Biol Chem* **259**: 7826-34.

Teruya, J.H., Kutsunai S.Y., Spear D.H. *et al.* (1990) Testis-specific transcription initiation sites of rat farnesyl pyrophosphate synthetase mRNA. *Mol Cell Biol* **10**: 2315-26.

Tsai, S.C., Haun R.S., Tsuchiya M., *et al.* (1991) Isolation and characterization of the human gene for ADP-ribosylation factor 3, a 20-kDa guanine nucleotide-binding protein activator of cholera toxin. *J Biol Chem* **266**: 23053-9.

Tsui, S.K., Lee S.M., Fung K.P. *et al.* (1996) Primary structures and sequence analysis of human ribosomal proteins L39 and S27. *Biochem Mol Biol Int* **40**: 611-6.

Legend to figures

Figure 1

Nucleotide and amino acid sequences of Hzyg. The non-coding nucleotide sequences are indicated in low cap. The sequences deduced from alignments with EST are represented in italic and are located outside the bracket. The different stop codons present in the 5'untranslated region are underlined. The polyadenylation signal is underlined. Double arrow (⋈) map the 4 LRR and the XLXXLXLXX(N/T/C) motives at their amino terminal parts are underlined.

Figure 2

Alignment of protein sequences of Hzyg, the *Drosophila melanogaster* gene *Mei-1* (Mei-1?), and *Caenorhabditis elegans* Zyg-11. The black boxes represent identical residues between the human sequence and the sequences of the two other species.

Figure 3

Southern blot, of human genomic DNA digested with different restriction enzymes (K: KpnI, E: EcoRI, Pv: PvuII, Ps: PstI, H: HindIII), hybridised with the insert of clone A11107. The blot was washed under high stringency conditions. The size of the markers is indicated on the right side.

Figure 4

Northern blots of human tissue mRNA hybridised with the insert of the clone A11107 and washed under high stringency conditions. The blots were exposed for two nights. The size of the markers is indicated on the right side.

Figure 5

In situ hybridisation to human testis sections of single strand RNA antisense (A, B, E-H) or sense (C, D) probes specific for Hzyg. A, B, D, F, H and C, E, G photographs were obtained under dark field and bright field respectively. Magnification for sections are as follows: 25 (A), 100 (B), 630 (C-H). Scale bare: (A) = 200 µ; (B) = 25 µ; (C-H) = 10 µ

Figure 6

Western blot of human testis protein extracts incubated with affinity-purified IgG raised against a 13 amino-acids specific carboxy-terminal Hzyg peptide. A human kidney extract was used as negative control, since no mRNA signal was observed on northern blot. Sizes of the molecular weight standards (kDa) are reported on the left.

Figure 7

Immunofluorescent staining of paraffin-embedded human testis sections incubated: (A) without primary antibodies, (B and D) with Hzyg purified IgG, (C) with purified IgG preabsorbed on Hzyg specific peptide. Sections were overlaid with CY3-conjugated anti-rabbit IgG. DNA was counterstained with DAPI. Original magnification x630. Scale bare = 10 µ.

```
gggtttgttgacagcggaggcggcggcggctgcaaggctccgagccgtaggagccggatcggggggaggggccgggcccaggagcctcagccccgccggcagccctaagggca   112

                                                                                          M   A   S   D     4
Aggcagtggctctgccaatgatcctgtgagtattcaggaatcactgttgcccctgggaatccttgtcctggagtggcccacctgcttgcccccagc ATG GCG TCC GAC   220

 T   P   E   S   L   M   A   L   C   T   D   F   C   L   R   N   L   D   G   T   L   G   Y   L   L   D   K   E    32
ACT CCC GAG TCG CTG ATG GCC CTC TGT ACT GAC TTC TGC TTG CGC AAC CTG GAT GGC ACC CTG GGC TAC CTG CTG GAC AAG GAG   304

 T   L   R   L   H   P   D   I   F   L   P   S   E   I   C   D   R   L   V   N   E   Y   V   E   L   V   N   A    60
ACC CTG CGG CTA CAT CCG GAC ATC TTC TTG CCC AGC GAG ATC TGT GAC CGG CTC GTC AAT GAG TAT GTG GAG CTG GTG AAC GCT   388

 A   C   N   F   E   P   H   E   S   F   F   S   L   F   S   D   P   R   S   T   R   L   T   R   I   H   L   R    88
GCC TGT AAC TTC GAG CCA CAC GAG AGC TTC TTC AGC CTC TTT TCG GAC CCC CGC AGC ACC CGC CTC ACG CGG ATC CAC CTC CGT   472

 E   D   L   V   Q   D   Q   D   L   E   A   I   R   K   Q   D   L   V   E   L   Y   L   T   N   C   E   K   L   116
GAG GAC CTG GTG CAG GAC CAG GAC CTG GAG GCC ATC CGC AAG CAG GAC CTG GTG GAG CTG TAC CTG ACT AAC TGC GAG AAG CTG   556

 S   A   K   S   L   Q   T   L   R   S   F   S   H   T   L   V   S   L   S   L   F   G   C   T   N   I   F   Y   144
TCC GCC AAG AGC CTG CAG ACA CTG AGG AGC TTC AGC CAC ACC CTG GTG TCC TTG AGC CTC TTC GGC TGT ACA AAC ATT TTC TAT   640

 E   E   E   N   P   G   G   C   E   D   E   Y   L   V   N   P   T   C   Q   V   L   V   K   D   F   T   F   E   172
GAG GAG GAG AAC CCA GGG GGC TGT GAA GAT GAG TAC CTC GTC AAC CCC ACC TGC CAG GTG CTG GTT AAG GAT TTC ACC TTC GAG   724

 G   F   S   R   L   R   F   L   N   L   G   R   M   I   D   W   V   P   V   E   S   L   L   R   P   L   N   S   200
GGC TTC AGC CGC CTC CGC TTC CTC AAC TTG GGC CGC ATG ATT GAT TGG GTC CCT GTG GAG TCC CTG CTG CGG CCG CTT AAC TCC   808

 L   A   A   L   D   L   S   G   I   Q   T   S   D   A   A   F   L   T   Q   W   K   D   S   L   V   S   L   V   228
CTG GCT GCC TTG GAC CTC TCA GGC ATT CAG ACG AGC GAC GCC TTC CTC ACC CAG TGG AAA GAC AGC CTG GTG TCC CTC GTC   892

 L   Y   N   M   D   L   S   D   D   H   I   R   V   I   V   Q   L   H   K   L   G   H   L   D   I   S   R   D   256
CTC TAC AAC ATG GAC CTG TCC GAC GAC CAC ATC CGG GTC ATC GTG CAG CTG CAC AAG CTG GGA CAC CTG GAC ATC TCC CGA GAC   976

 R   L   S   S   Y   Y   K   F   K   L   T   R   E   V   L   S   L   F   V   Q   K   L   G   N   L   M   S   L   284
CGC CTC TCC AGC TAC TAC AAG TTC AAG CTG ACT CGG GAG GTG CTG AGC CTC TTT GTG CAG AAG CTG GGG AAC CTA ATG TCC CTG   1060

 D   I   S   G   H   M   I   L   E   N   C   S   I   S   K   M   E   E   E   A   G   Q   T   S   I   E   P   S   312
GAC ATC TCT GGC CAC ATG ATC CTG GAG AAC TGC AGC ATC TCC AAG ATG GAA GAG GAA GCG GGG CAG ACC AGC ATT GAG CCT TCC   1144

 K   S   S   I   I   P   F   R   A   L   K   R   P   L   Q   F   L   G   L   F   E   N   S   L   C   R   L   T   340
AAG AGC AGC ATC ATA CCT TTC CGG GCT CTG AAG AGG CCG CTG CAG TTC CTC GGG CTC TTT GAG AAC TCT CTG TGC CGC CTC ACG   1228

 H   I   P   A   Y   K   V   S   G   D   K   N   E   E   Q   V   L   N   A   I   E   A   Y   T   E   H   R   P   368
CAC ATT CCA GCC TAC AAA GTA AGT GGT GAC AAA AAC GAA GAG CAG GTG CTG AAT GCC ATC GAG GCC TAC ACG GAG CAC CGG CCT   1312

 E   I   T   S   R   A   I   N   L   L   F   D   I   A   R   I   E   R   C   N   Q   L   L   R   A   L   K   L   396
GAG ATC ACC TCG CGG GCC ATC AAC TTG CTT TTT GAC ATC GCC CGC GAC CGT TGC AAC CAG CTG CTG CGG CGC CTG AAG CTG   1396

 V   I   T   A   L   K   C   H   K   Y   D   R   N   I   Q   V   T   G   S   A   L   L   F   Y   L   T   N   S   424
GTC ATC ACG GCC CTC AAG TGC CAC AAA TAT GAC AGG AAC ATT CAA GTG ACA GGC AGC GGG CTT CTC TTC TAC CTA ACA AAT TCC   1480

 E   Y   R   S   E   Q   S   V   K   L   R   R   Q   V   I   Q   V   V   L   N   G   M   E   S   Y   Q   E   V   452
GAG TAC CGC TCA GAG CAG AGT GTG AAG CTG CGC CGG CAG GTT ATC CAG GTG GTG CTG AAT GGC ATG GAA TCC TAC CAG GAG GTG   1564

 T   V   Q   R   N   C   C   L   T   L   C   N   F   S   I   P   E   E   L   E   F   Q   Y   R   R   V   N   E   480
ACG GTG CAG CGG AAC TGC TGC CTG ACG CTC TGC AAC TTC AGC ATC CCC GAG GAG CTG GAA TTC CAG TAC CGC CGG GTC AAC GAG   1648

 L   L   L   S   I   L   N   P   T   R   Q   D   E   S   I   Q   R   I   A   V   H   L   C   N   A   L   V   C   508
CTC CTG CTC AGC ATC CTC AAC CCC ACG CAG GAC GAG ATC TCT ATC CAG CGC ATC GCC GTG CAC CTG TGC AAT GCC CTG GTC TGC   1732

 Q   V   D   N   D   H   K   E   A   V   G   K   M   G   F   V   V   T   M   L   K   L   I   Q   K   K   L   L   536
CAG GTA GAC AAC GAC CAC AAG GAG GCC GTG GGC AAG ATG GGC TTT GTC GTG ACC ATG CTG AAG CTG ATT CAG AAG AAG CTG CTG   1816

 D   K   T   C   D   Q   V   M   E   F   S   W   S   A   L   W   N   I   T   D   E   T   P   D   N   C   E   M   564
GAC AAG ACA TGT GAC CAG GTC ATG GAG TTC TCC TGG AGT GCC CTG TGG AAC ATC ACA GAT GAA ACT CCT GAC AAC TGC GAG ATG   1900

 F   L   N   F   N   G   M   K   L   F   L   D   C   L   K   E   F   P   E   K   Q   E   L   I   R   N   M   L   592
TTC CTC AAT TTC AAC GGC ATG AAG CTC TTC CTG GAC TGC CTG AAG GAA TTC CCA GAG AAG CAG GAA CTC ATT AGG AAT ATG CTA   1984

 G   L   L   G   N   V   A   E   V   K   E   L   R   P   Q   L   M   T   S   Q   F   I   S   V   F   S   N   L   620
GGA CTT TTG GGG AAT GTG GCA GAA GTG AAG GAG CTG AGG CCT CAA CTA ATG ACT TCC CAG TTC ATC AGC GTC TTC AGC AAC CTG   2068

 L   E   S   K   A   D   G   I   E   V   S   Y   N   A   C   G   V   L   S   H   I   M   F   D   G   P   E   A   648
TTG GAG AGC AAG GCC GAT GGG ATC GAG GTT TCC TAC AAT GCC TGC GGC GTC CTC TCC CAC ATC ATG TTT GAT GGA CCC GAG GCC   2152

 W   G   V   C   K   P   Q   R   E   E   V   E   E   R   M   W   A   A   I   Q   S   W   D   I   N   S   R   R   676
TGG GGC GTC TGT AAG CCC CAG CGT GAG GAG GTG GAG GAA CGC ATG TGG GCT GCC ATC CAG AGC TGG GAC ATA AAC TCT CGG AGA   2236

 N   I   N   Y   R   S   F   E   P   I   L   R   L   L   P   Q   G   I   S   P   V   S   Q   H   W   A   T   W   704
AAC ATC AAT TAC AGG TCA TTT GAA CCA ATT CTC CGC CTC CTT CCC CAG GGA ATC TCT CCT GTC AGC CAC TGG GCA ACC TGG   2320

 A   L   Y   N   L   V   S   V   Y   P   D   K   Y   C   P   L   L   I   K   E   G   G   M   P   L   L   R   D   732
GCC CTG TAT AAC CTC GTG TCT GTC TAC CCG GAC AAG TAC TGC CCT CTG CTG ATC AAA GAA GGG GGG ATG CCC CTT CTG AGG GAC   2404

 I   I   K   M   A   T   A   R   Q   E   T   K   E   M   A   R   K   V   I   E   H   C   S   N   F   K   E   E   760
ATA ATT AAG ATG GCG ACC GCA CGG CAG GAG ACC AAG GAA ATG GCC CGC AAG GTG ATT GAG CAC TGC AGT AAC TTT AAA GAG GAG   2488

 N   M   D   T   S   R   *                                                                                    767
AAC ATG GAC ACG TCT AGA TAG  aggcctccgtccccatggccgccaccgctctggaccacaggcggggaggaagcatgctcaagcagcccagcggggggcccctt   2592

ccgaggggagcctcccacggagtaaggagacatgggggactttttgcacaaccgacgcttttccttaatgttagtgagatatatatatattatatatatatattttttttttgg   2704

ttaggaagtgtgaagtttgtgtgtatgatttctgtgcaaaaacaaaagcaacactcctg                                                    2764
```

```
          . . . . .10 . . . . .20 . . . . .30 . . . . .40 . . . . .50 . . . . .60 . . . . .70 . . . . .80
Hzyg   ....MASDTEESLMALCTDFCLRNLLGTLGYLLDKETLRLHEDLFLPSELCDRLVNEYVELVNAACNFEPHESFFSLLSD: 76
Mei-1? MEDGSLDEEELTLKEIAYQKLCNNLLIISSHRPDGORG.LNEGLVLENELCDGFLENYORFN.....RPLDDLVIRLLED: 74
Zyg11  ...MDHNAALTTCSTSAIPRLAFLATERIAFLIENDALPEFDHKVIAECSNDIFEVLRQKKK.......LNHQTLQTMCS: 70

          . . . . .90 . . . .100 . . . .110 . . . .120 . . . .130 . . . .140 . . . .150 . . . .160
Hzyg   PRSTRLTRHLREDLVQDODLEAIRKODLVELYLTNCEKLSAKSLOTLRSFSHLTLVSLSL.FGCTNIFYBEENPGGCEDE:155
Mei-1? THRLSLKIVNLRNSTLSSIGLETLMRHKLFALSLWYCDMISVGSHHLLAHYGDSLRSLELGISSHLLQMAEFNEKEPVDF:154
Zyg11  TSRFQINKLDLMGNKVELODLELCSNONLAVSFRLGDIDYFPNTNKIOVADVLDTALNRTS...ROQLRHLDLSGRHRITS:147

          . . . .170 . . . .180 . . . .190 . . . .200 . . . .210 . . . .220 . . . .230 . . . .240
Hzyg   YLVNPTCOVLVKDFTFEGFSRLRFLNLGRMLDWVPVESLLRPLNSLAALDLSGLOLSDAAFLTOWKDSLVSLVLYNMDLS:235
Mei-1? QLTOPHLRRLVLN.GVVMHHRLOEAHLHDLGHLDLTSCVLANFSLEALGSLPNLHLL..............LLLBNVWPI:218
Zyg11  NWPTYVARKLPRLESLAFANRSTANETLSQLGASLLNLRFLDISSTCVTLISCLSSLRN........LEVLLIMYNLNIL:218

          . . . .250 . . . .260 . . . .270 . . . .280 . . . .290 . . . .300 . . . .310 . . . .320
Hzyg   D.LHIRVLVOLHKLRHLDLSRDRLSS..YYKFKLTREVLSLFVOKLGNLMSLDISGHMILENCSISKMEEEAGOTSIELSK:313
Mei-1? A.NOLHALCOLRRLCTLDLSISSSGNGHGTYDLLPDQTLEMLMDNLRHLTHLDISGTNLAGNGVATKESTTTSGMQOSEKM:297
Zyg11  KGLVTETLSNLTKLRVLDISR.KVNTDYLOETSODAHLLDLALGIYNRSVEAIESGTATPWAELRAIDMSGLSIVQFGTDR:297

          . . . .330 . . . .340 . . . .350 . . . .360 . . . .370 . . . .380 . . . .390 . . . .400
Hzyg   SSIIPFRALK.......RPLQFLGLFENS..LCRLTHLF..AYKVSGLKNEEQVLNALEALTEH.RFEITSRLINLLFDI:381
Mei-1? EQHFALTDIPGLASRTQRPLOFLGLYHTAHWACKRHDLL..ALEVAGLANEQCLLTAARYLHD..RFVLLTRVLNLLYHL:373
Zyg11  ALAFVEKIIEAHP.....KLEQISLLATP.LDSSLIELLNRNLQVINTVSRRSIIFALSHYANLDREFAITHLLHSVYYQ:371

          . . . .410 . . . .420 . . . .430 . . . .440 . . . .450 . . . .460 . . . .470 . . . .480
Hzyg   AR..IERCN..QLLRALKLVITALKGHKYDRNICVTGSAALFYLTNSEYRSEQSVKLRRQVLQVVLNGMES.YQEVLVOR:456
Mei-1? FR..FENCK..DIHTALDVLSAMDRHLKFKHMCISGSATLFYIVKGRDRSKFGALLRNHLLRTLLNGMEMHITDDLMIR:449
Zyg11  LQSGYDKFSQEELKECLRLVCISMQQGLNTLPVCIACGSACLYHLCKMKRIKRLSVKEVNNCLERSLDAALQYRSMTQICK:451

          . . . .490 . . . .500 . . . .510 . . . .520 . . . .530 . . . .540 . . . .550 . . . .560
Hzyg   NCOLTLCNLSLEEELEFOYRRVNELLLSILNPLROLESLIQRLAVHLCNALVCQVDL.NDHKEAVGKMGFVVVTMLKLLQKKL:535
Mei-1? NGYLTLTOLHMFVDVLEEYERLIKILLHGVSKLEOEGFVORLTLIYLLNTLACQVDL.GROKLFLGELGVVSLMFLLLKDRL:528
Zyg11  NVWLTLCNDYLLHLDELDFYRTCKVALDTMLLNF.DASVERMTIAIVSIVTPKMRPSEARLILTTETKYVYHLVRIMNDYL:530

          . . . .570 . . . .580 . . . .590 . . . .600 . . . .610 . . . .620 . . . .630 . . . .640
Hzyg   LDKTLCLQ...........VMEFSWSALWNLTDETEDNCEMFLNFNCMKLFLDCLFEFPEKQELHRNMLGLLGNVAEVK:602
Mei-1? TRSVFLD............VMEFVAWSTMWNVTDETAINCKRFLDGRCMEYFLKCLHTFLDRDELLRNMLGLIGNVAEVK:595
Zyg11  EAYTREHRVGHERDNENALYTLKETISALWNLLTDECLATCKAFLDAGCVQIAFRILKAFDYHGNVQTKVLGLLNNLAEVE:610

          . . . .650 . . . .660 . . . .670 . . . .680 . . . .690 . . . .700 . . . .710 . . . .720
Hzyg   ELRP.OLMLSQFLSVFSNLLES.........KADGLEVSYNACGVLSHIMFDGPEAWGVCEPOREEVEERWAAICSWDL:672
Mei-1? WLRP.KLMLQEFLEVFARLLDS.........LSDGLEVSYNALGVLALTLASDGADAWTIKDPSREHMLERWVAAICRWNL:665
Zyg11  ELHLGOLCKNEYLSVLISCLDGSFNEVDSKGRYREVDRSMFFAGILANLLTN.TDGWESEN.QRDEAOEKLLELLEOYPT:688

          . . . .730 . . . .740 . . . .750 . . . .760 . . . .770 . . . .780 . . . .790 . . . .800
Hzyg   NSRRNINYRSFEPILRLLPQGISEVSOHWATWALYNLVSVYPDKYCPLLIKEGCMPLLRDTLKMATAROELKEMARKVLE:752
Mei-1? KSERNINYRSFEPILSLVRCYELOQOHWALWALANLTQVYPBKYCKLVEOENCIQILNELLEHESPYCELKRLARLVLE:745
Zyg11  LPSAMVSYKSELPLFSRLIVRESNENGAIMWCLWGVHHVLQHREKNKAPTYEKG....YIEMFMDSGLSPIVEDVSHGHSK:763

          . . . .810 . . .
Hzyg   HCSNFKEENMDTSR:766
Mei-1? CCDSGSERMVVDG.:758
Zyg11  EIHNLDI.......:770
```
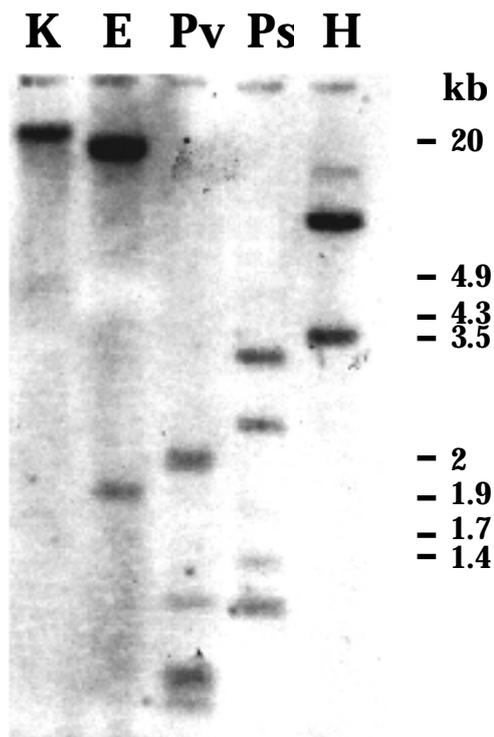
Figure 3