Allele specific inhibition. A novel approach to cancer therapy

ten Asbroek, A.L.M.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
We have analyzed the genomic structure of ribonuclease (RNase) H2 loci in the human genome. Human PAC library screening combined with database searches indicated that several loci are present. The transcribed gene is localized on chromosome 2p25. This was confirmed by RNA analysis of a monochromosomal hybrid cell line that expressed human chromosome 2. These data contrast a previous report, as well as the current Human Genome Project annotation, which had placed the gene on chromosome 17p11.2. This location represents a pseudogene. Another highly similar pseudogene is present at a separate locus located more distal on chromosome 17p, while a third pseudogene is localized on chromosome 1q.
INTRODUCTION

Ribonuclease H, RNase H, has a unique function to degrade specifically the RNA moiety of RNA-DNA hybrid duplexes (1,2). This enzyme activity is present in all eukaryotes and prokaryotes (3). Almost all species examined thusfar contain at least two genes encoding RNase H. Only archaeal bacteria contain a single RNase H gene in their genome (4).

The mammalian RNase H enzyme activities have been classified into two groups based on biochemical characteristics (5,6). Class I enzymes have molecular weights of 68-90 kDa, are activated by both Mg\(^{2+}\) and Mn\(^{2+}\), and are active in the presence of SH-blocking agents. Class II enzymes have a molecular mass of 30-45 kDa, are activated only by Mg\(^{2+}\), and inhibited by Mn\(^{2+}\) and SH-blocking agents. RNase H enzymes are involved in removing RNA primers in prokaryotic and eukaryotic DNA synthesis reconstitution experiments in vitro (7,8). However, the physiological role of eukaryotic RNases H remains as yet undetermined, as other enzymes can also fulfill these steps in the DNA replication process (9).

Cloning and sequencing of cDNAs encoding the human enzymes revealed that the human class II enzyme, RNase H2, is the sequence homologue of E.coli RNase H1, the major prokaryotic activity. The human class I enzyme, RNase H1, is the homologue of the minor E.coli RNase H11 activity (10-13).

The RNases H have gained renewed attention since the use of antisense oligodeoxynucleotides as a tool to inhibit gene expression, as RNase H activity is an essential element in that process (14). In a previous study we have tried to dissect the contribution of the human RNases H to the mechanism of antisense inhibition of gene expression (15). The interpretation of the role of RNase H2 was obscured by the fact that the probes derived from the coding region of this enzyme recognize two transcripts in Northern blot analyses. The main 1.2 kb mRNA is present at varying levels in different cell lines, whereas a 5.5 kb mRNA variant is present at a more consistent level. This mRNA species is not cloned yet, and is thought to be a polyadenylated processing variant of the 1.2 kb transcript (13).

Mammalian RNase H2 is the minor enzyme and accounts for only 15% of the total RNase H activity (16). However, because of its homology to the prokaryotic major enzyme, which has been studied in detail, it is most extensively studied. The gene has previously been mapped to chromosome 17p11.2 by fluorescent in situ hybridization (FISH) analysis (10). This location has also been identified in the Human Genome Project as the gene sequence. The Ensemble database sequence of the 17p locus is different from the cDNA sequence as has been identified and cloned by various research groups. In order to obtain more insight into the genomic organization of the gene encoding RNase H2, we have now screened the human genome for putative RNase H2-encoding sequences.
Figure 1: FISH analysis of the human RNase H2 PAC groups
A, PAC 41K9 (in green) localizes to chromosome 1q (centromere 1 labeled in red). B, PAC 80N18 (PAC1) and 61E3 (PAC2) in green near the centromere 17 in red. C, PAC 80N18 in red and 61E3 in green on chromosome 17p and the subtelomeric probe for 17q in red. D, interphase nucleus with PAC 80N18 in red and 61E3 in green. Photographs represent computer-enhanced images.
RESULTS

A PAC library with 5-fold coverage of the human genome was hybridized to a cDNA probe containing the entire coding region of human RNase H2. Twelve PACs were obtained, which could subsequently be divided into three groups based on Southern blot analyses, using various restriction endonucleases and probes for RNase H2 as well as vector sequences (data not shown). The first group (called PAC1) consists of PACs 80N18, 150M8, 198L10, 179K1, 179I2, and 213A7; the second group (called PAC2) consists of PACs 182H15, 17E7, 19A22, 61E3, and 66G8; the third group contains only a single PAC, 41K9 (full names for all PACs are given in the Materials and Methods section). All PACs show signal of similar strengths on Southern blot with an RNase H2 probe.

Fluorescent in situ hybridization (FISH) was performed to map the three groups on the human genome (Figure 1). The single PAC of group 3 hybridizes to chromosome 1q (Fig. 1A), while PAC1 and PAC2 hybridize to chromosome 17p (Fig 1B). Co-hybridization of these PACs indicates that they map to different loci on chromosome 17p (Fig. 1C and 1D). FISH analysis with additional markers on this chromosomal region indicated that the locus of PAC1 is at or near band 17p11.2, while PAC2 maps at or near band 17p13.1 (data not shown). The fact that each PAC recognizes only its own locus in the FISH analyses indicates that the hybridization signal is mainly derived from the flanking sequences and not from the RNase H2 coding sequence that was used to screen the PAC library.

The single PAC comprising group 3 maps to a genomic region that has been shown in the public Human Genome Project (HGP) database to contain a pseudogene of RNase H2 (in contig NT_021877.7 at chromosome 1q32.1-41). With that sequence available we could PCR amplify and sequence this region of PAC 41K9, which confirmed the presence of the RNase H2 pseudogene locus on this PAC. The RNase H2 loci of PAC1 and PAC2 were both contained in a single HindIII restriction fragment of approximately 10 kb. Cloning of these fragments into a plasmid, as a whole or in a shotgun approach, was not or only partly possible, due to recombination during growth of the plasmid-containing bacteria. In order to obtain the sequences of these RNase H2 loci we used a PCR-based PAC-walk sequencing method (described in the Materials and Methods section) that utilizes ligated fragments as template in a PCR reaction, thus avoiding the need for growth in bacteria. The overlapping fragments that were generated were verified to represent a continuous DNA segment by direct amplification of the entire RNase H2 sequence-containing region on the PAC DNAs. The sequences from PAC1 and PAC2 revealed that both loci represent also pseudogenes of RNase H2, which cannot lead to the generation of functional protein. Figure 2 shows the comparison of the three pseudogene sequences with the RNase H2 cDNA sequence of GenBank accession AF039652. We have confirmed the cDNA sequence of RNase H2 as presented by AF039652 in various human cell lines following RT-PCR (15).
Figure 2: Sequence comparison of the RNase H2 pseudogene loci and the cDNA sequence

The sequences obtained for the PAC1 and PAC2 groups and PAC 41K9 are aligned with the cDNA sequence of GenBank ID AF039652, using ClustalW of the EMBL European Bioinformatics Institute (www.ebi.ac.uk). The start codon and stop codon in the cDNA are given in bold, and the exons indicated by alternating gray and white blocks. An asterisk indicates an identical residue present in all sequences. Gaps are included by the algorithm to allow optimal alignment, and are indicated by dashes.
The sequence identity of all PACs within each group with respect to the RNase H2 region was addressed with single-strand conformation polymorphism analyses (data not shown). This indicated that all PACs within each group have the same RNase H2-like sequence. Only PAC 182H15 in the PAC2 contig did not contain the entire RNase H2-like sequence, since it was located on the border of the PAC. This was confirmed by sequencing.

The PAC library screen thus only produced pseudogenes for human RNase H2. The PAC1 position is the gene locus assignment of the HGP, as well as the locus reported earlier in literature (10). To identify the real gene that encodes this sequence we screened the public HGP database and the commercial Celera (CDS) database for homology with the coding region of RNase H2 (releases of December 2001). Both databases contain two RNase H2-like sequences in contigs mapped to chromosome 17p11.2, plus the pseudogene locus on chromosome 1q32-41. The proximal PAC1 locus on 17p is similar to HGP contig NT_024893.7 and CDS scaffold GA_x5J8B7Q55UD. The more distal PAC2 locus resembles HGP contig NT_030157.2 and CDS scaffold GA_x5J8B7Q55UB, as well as the Ensemble sequence ENST0207138. The CDS database contained a fourth locus positioned on chromosome 2p25 (hCG1784271). This locus presents a gene structure of 8 exons that, once joined, yields exactly the RNase H2 cDNA sequence of GenBank AF039652. The exons are indicated in the sequence of AF039652 in Figure 2. The locus size is approximately 10 kb. The exact size cannot be given, because some regions, among which the start of exon 3, still contain stretches of unknown sequence. Comparison to the CDS mouse database shows that the mouse gene, mCG19128 localized on chromosome 12A1.2 and there contained in 9.9 kb, has the same exon structure and uses the same splice site positions as the human gene.

Figure 3: Northern blot analysis of monochromosomal hybrid cell lines

RNA from human tumor cell lines MiaPacaII and 15PC3 and monochromosomal hybrid cell lines containing human chromosome #1, #2, or #17 is hybridized to the human RNase H2 coding region (A) and the mouse RNase H2 coding region (B). The mouse probe hybridizes with less intensity to the hamster mRNA of hybrid #2, and shows some cross-hybridization to the human transcript. C, 28S rRNA band visible after ethidium staining of the gel prior to blotting. The RNA samples are indicated on top of the lanes; the arrows point at the transcripts as indicated.
To confirm that human chromosome 2 contains the gene encoding RNase H2 and that chromosomes 1 and 17 do not, we performed Northern blot analysis using monochromosomal hybrid cell lines that contain a single human chromosome in a murine background (Figure 3). The hybrid cell line containing human chromosome 2 indeed expresses the 1.2 kb RNase H2 mRNA (lane mono #2 in Fig. 3A), while those containing chromosome 1 (mono #1) or chromosome 17 (mono #17) only express the 1.4 kb mouse mRNA (Fig. 3B). To demonstrate that these cell lines also express genes located on their human chromosome, RT-PCR was performed with primers for human CTP synthetase and RNA polymerase II, localized respectively on chromosome 1 and 17 (data not shown). The 5.5 kb human RNase H2 mRNA was not detected in the hybrid cell line with human chromosome 2. The integrity of the hybrid cell lines with respect to the presence of the human chromosomes was analyzed by FISH with unique probes (centromeric and subtelomeric) for the chromosomes. For each cell line 100 interphase nuclei were counted using standard criteria and presuming that one spot represents one chromosome\(^{(25)}\). Hybrid #1 nuclei displayed 90% single spots, 5% double spots and 5% no spots; hybrid #2 nuclei had 65% single, 15% double, and 20% no spots; hybrid # 17 nuclei had 75% single, 24% double and 1% no spots. FISH analysis with the PACs from the three groups confirmed that the three cell lines only contained human chromosome 1, 2 or 17 (data not shown).

DISCUSSION

The Human Genome Project has placed the gene encoding human RNase H2 on chromosome 17p11.2, a locus that had previously also been identified by FISH analysis\(^{(10)}\). The sequence that is given in the database (Ensemble ENST0207138) differs from the cDNA sequence, nor leads to functional RNase H protein, however. We screened a human PAC library with a probe containing the entire coding region of RNase H2. This has resulted in the isolation of twelve PACs, which represented 3 different loci as demonstrated by FISH analyses. Two different loci were identified on chromosome 17p and sequence analysis showed that these were both pseudogenes. These loci had DNA sequences devoid of introns and contained several stop codons in all reading frames requiring a significant number of frameshifts during translation to produce a protein containing RNase H. These two loci, at or near position 17p11.2 and 17p13.1 on the chromosome, are highly similar to each other (>90% identity). The public HGP database and the commercial CDS database contain both pseudogene sequences in separate contigs or scaffolds, but place them both at chromosome 17p11.2. The RNase H2 sequences that we have obtained were determined for two individual PACs in each group (6 PACs were obtained for the 17p11.2 locus and 5 PACs for the 17p13.1 locus). They
were subsequently shown to be the same for all the PACs within a locus-group by comparative analyses with single-strand conformation analysis. The third PAC group is only represented by a single PAC in our screen, and presents a pseudogene on chromosome 1q32-41 that is known in the HGP as well as the CDS databases. The PAC library screen using the RNase H2 cDNA sequence thus only provided pseudogenes, while the protein-encoding gene remained undetermined. The pseudogenes have in common that they contain an RNase H2-like sequence without introns, and are all contained in chromosomal regions that are characterized by highly repetitive DNA. This is probably the reason for the lack of contiguous sequence in the genome databases in these regions. We were not able to extend the continuous DNA sequence of the chromosome 17 loci any further by our PAC-walk sequencing approach, because we had arrived in Alu-rich repetitive regions flanking the RNase H2 sequence.

The gene encoding RNase H2 could only be obtained by searching the CDS databases for the mouse and the human genome. The HGP database (release December 2001) does not contain this sequence. The gene was found positioned at chromosome 2p25 in human and chromosome 12A1.2 in mouse. The mouse locus had been reported on 12A3 by FISH analysis\(^ {10}\). The mouse genome database search did not display pseudogenes or other sequences that contain RNase H2-like sequences.

The loci that we have identified in this study, on chromosome 1, 2 and 17, were analyzed for RNA expression using monochromosomal human/murine hybrid cell lines. Only hybrid cells containing human chromosome 2 produced the 1.2 kb transcript of RNase H2. The 5.5 kb human mRNA is present in all human cells at a relatively constant level even when the 1.2 kb mRNA is present in minute amount\(^ {13,15}\). In contrast, Northern blot analysis of different mouse tissues of various strains only detected the 1.4-kb mRNA (our unpublished results). Apparently, the ability to express the 5.5 kb human RNase H2 transcript requires a factor that is lacking in mouse cells, or synthesis of the 5.5 kb mRNA is prematurely terminated or destabilized in murine cells.

**MATERIALS AND METHODS**

**PAC library**

Human PAC library RPCI-6, constructed by Pieter de Jong from the Roswell Park Cancer Institute, was obtained from the German Human Genome Project Resource Center. The PACs identified and positive for RNase H2 were: LLNLPN1880Q2 (80N18), LLNLPM08150Q2 (150M8), LLNLPL10198Q2 (198L10), LLNLPK01179Q2 (179K1), LLNLPI02179Q2 (179I2), LLNLPA07213Q2 (213A7), LLNLPH15182Q2 (182H15), LLNLPE0717Q2 (17E7),
RNA analysis and Probes
Northern blot analysis of RNA was as described (18). Hybridized probe was visualized on a PhosphorImager (Fuji BAS Imager and the AIDA software). cDNA fragments to be used as probe were generated by RT-PCR and subsequent cloning into the pGEM-T Easy vector (Promega). Probes used contained the coding region of human RNase H2, (GenBank accession number AF039652 position 82-942), and mouse RNase H2 (GenBank accession number NM_011275 position 49-906).

FISH analysis
Fluorescent in situ hybridization was performed according to standard protocols, as described earlier (19). Centromere probes used were pUC1.77 for chromosome 1 (20), CEP2 Spectrum Orange for chromosome 2 (Vysis Inc. # 32-110002), and p17H8 for chromosome 17 (21). Subtelomeric probes were GS-62-1.8 for chromosome 1p, GS-160-H23 for 1q, GA-892-G20 for 2p, GS-1011-O17 for 2q, and GS-362-K4 for 17q (22). Additional probes on chromosome 17 were cCI17-500 for 17p12, cCI17-484 for 17p13.1 (23), and c197-9 for 17p13.3 (24).

Cell lines
Human cell lines 15PC3 (prostate cancer) and MiaPacII (pancreatic carcinoma) were gifts from colleagues. Cells were maintained by serial passage in Dulbecco’s modified Eagle’s medium (DMEM), supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin.
Monochromosomal hybrid cell lines were generous gifts from Prof. Dr. A. Geurts van Kessel (Department of Anthropogenetics, University of Nijmegen, The Netherlands). Monochromosomal hybrid # 1 (GM13139) was maintained in DMEM, supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin, and 250 μg/ml G418 (Roche). Monochromosomal hybrid #2 (GM10826) was cultured in MEM-alpha medium supplemented with 10% FCS, 1 mM L-glutamine, and ribonucleosides. Monochromosomal hybrid #17 (PCT-BA18) was maintained in DMEM plus F10 (1:1), supplemented with 10% FCS, 2 mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin, and HAT selection. Monochromosomal hybrids #1 and #17 contain the human chromosome in a mouse background, while hybrid #2 has a hamster background.

PAC-walk sequencing
This method is derived from a gene-walking method described earlier (25). In separate reactions, the PAC DNA was digested using fine-cutting restriction endonucleases *Alu*I,
**Ddel, Hinfl or NlaIII.** After purification, the digests were incubated for 1 h at 72°C with AmpliTaq polymerase (Perkin Elmer) in MgCl₂-containing PCR buffer provided by the manufacturer, in the presence of dNTPs (50 µM final concentration). This generates an A-overhang that allows subsequent ligation into the pGEM-T Easy vector (Promega). The ligation is used as template in a PCR reaction with one primer specific for the vector (SP6 or T7 primer) and the other specific for the gene sequence that is the starting point for the walking exercise. After purification and sequencing of the generated fragment(s) a next round of PCR plus sequencing can be done, using a gene-specific primer designed on the newly generated sequence. When templates from different digest are used in the successive rounds, overlapping fragments are generated. The PCR cycling used for amplification was 35 times (30 s 94°C, 30 s 55°C, and 3 min 72°C).

**Acknowledgements**

We thank Ad Geurs van Kessel (Anthropogenetics Department, University of Nijmegen, The Netherlands) for the generous gift of the monochromosomal hybrid cell lines. We thank Roos de Jonge for performing the initial FISH experiments.

**REFERENCES**


