Ribonucleases 1 of ruminants

Breukelman, Heleen Jacqueline

Summary

Mammalian ribonucleases 1 form a group of homologous proteins. They are pyrimidine specific and cleave RNA endonucleolytically to yield 3' phosphomono nucleotides and oligonucleotides via 2',3' cyclic intermediates.

Three homologous ribonucleases were found in ox; pancreatic RNase, seminal RNase and brain RNase. Bovine pancreatic RNase (RNase A) consists of a single polypeptide chain of 124 amino acids, containing four disulfide bridges. Only a fraction of the secreted amount of enzyme is N-linked glycosylated at position 34.

Bovine seminal RNase (RNase BS) consists of two identical subunits of 124 amino acid residues, with four disulfide bridges in each subunit. The two subunits are covalently linked by two additional disulfide bridges between cysteine residues at positions 31 and 32, with Cys31 of one subunit binding with Cys32 of the other subunit. No attachment sites for N-linked carbohydrate are found in the protein.

Bovine brain RNase (BRb RNase) shows a high similarity to the other bovine ribonucleases, except that it contains an extension of 17 amino acid residues at the C-terminus. This extension is proline-rich and is O-glycosylated at Thr129 and Ser133. N-linked carbohydrate is attached at position 62.

Until about 15 years ago about 40 ribonuclease sequences were determined at the protein level. Most parsimonious analyses on these sequences indicated that the three different bovine ribonucleases are probably products of two gene duplication events at approximately the origin of the true ruminants, so after the divergence from the other artiodactyls.

To prove this hypothesis, investigations were started on genes coding for ribonucleases in a wide variety of ruminants. Southern blot analyses of genomic DNA of various ruminants and other mammals were performed as described in *Chapter 2*. Several parts of the bovine pancreatic RNase gene: a fragment of the single intervening sequence located in the 5' non-translated region, a fragment of the coding part of the gene and a fragment of the 3' non-translated region, were used as probes. Genomic DNA of most investigated mammals showed only one hybridisation band, while several hybridisation bands could be detected with ruminant genomic DNA. These results suggest that a unique RNase coding sequence is present in the investigated nonruminants, whereas several genomic regions related to the bovine pancreatic RNase gene are present in ruminants. However, recent studies show that also in other mammalian taxa gene duplications have taken place independently from the duplications

within the ruminants.

To demonstrate that the genome of giraffe possess different ribonuclease genes, RNase sequences coding for amino acids 31 to 114 were amplified by the polymerase chain reaction and the nucleotide sequences of the cloned amplification products were determined. Three different RNase sequences were obtained that showed the highest identity to bovine pancreatic RNase, bovine seminal RNase and bovine brain RNase, respectively and may code for active ribonucleases. Our colleagues in Naples meanwhile published the complete sequences of the three ribonuclease genes from giraffe and also the three orthologues from sheep. Most interesting of their results was the demonstration that the 'seminal' ribonucleases of these two species represent pseudogenes with deletions, which disrupt the coding reading frame, in which also active site residues are replaced.

The ribonuclease genes of two deer species, hog deer (subfamily Cervinae) and roe deer (subfamily Odocoileinae), were investigated and the already known ribonuclease sequences of giraffe were extended in order to demonstrate that the existence of the three different ribonuclease 1 genes extends from the bovids and giraffes to the cervids. The sequences of the coding regions of the orthologues of the three bovine ribonucleases 1 in deer species are reported in *Chapter 3* and compared with those of ox, sheep and giraffe. For each species there are three different sequences, encoding pancreatic-, seminal- and brain-type ribonucleases.

The determined pancreatic-type RNase sequences showed a high degree of identity both in the coding region and in the 3' untranslated region with the bovine pancreatic RNase gene. The same was observed for the obtained brain-type RNase sequences compared with the bovine brain sequence.

A high degree of identity could also be observed in the seminal-type sequences of these spieces compared with the bovine seminal RNase gene. The two deer sequences showed a deletion or an insertion in the coding region resulting in frame shifts leading almost directly to a stop codon. When read in the reading frame of the bovine seminal RNase gene many non-synonymous substitutions, at positions coding for amino acid residues essential for structure and function, could be observed. These substitutions combined with a ratio of 2:1 of non-synonymous to synonymous substitutions of the determined nucleotide sequences compared with the bovine seminal RNase gene are typical features of pseudogenes.

Maximum parsimony analyses were performed with the complete nucleotide sequences of the coding regions and the 3' untranslated regions of the pancreatic-,

seminal- and brain-type RNases of ox, sheep, giraffe, hog deer and roe deer. Three resulting trees showed three monophyletic groups formed by pancreatic-, seminal- and brain-type RNases, respectively, well supported by high bootstrap values. Because the topologies in these trees of the three RNase types showed no congruency of giraffe with the other ruminants, the sequences were also analysed by the neighbour-joining and maximum likelihood methods. With these methods trees were obtained with the three different gene-types again in separate groups, but still giraffe occupied non-congruent positions relative to the bovids and cervids.

In *Chapter 4* the ribonuclease sequences of chevrotain, a member of the superfamily Tragulidae, the earliest diverged taxon of the ruminants, are investigated. Southern blot analyses of chevrotain genomic DNA with a RNase 1 probe showed five - eight hybridisation bands. With RNase specific primers five different ribonuclease 1 sequences could be determined by PCR amplification.

Two sequences are identified as pancreatic RNase sequences, one encoding the pancreatic enzyme, the other one encoding a pseudogene. The identity of the pancreatic enzyme was confirmed by isolation of the protein and N-terminal sequence analysis. It is the most acidic pancreatic RNase identified so far. Formation of the mature enzyme requires cleavage by signal peptidase of a peptide bond between two glutamic acid residues. This unusual signal peptide cleavage may be due to an increase in the pK value of the first glutamic acid, caused by the vicinity of the second residue and the hydrophobic membrane environment, resulting in an uncharged, small aliphatic residue, in agreement with signal peptide cleavage site requirements. The pseudogene contains three deletions of single nucleotides at different positions in the coding sequence. The substitutions in this sequence are predominantly non-synonymous and the ratio of non-synonymous to synonymous substitutions for this pseudogene compared with the pancreatic RNase gene of chevrotain is 3:1.

The seminal-type RNase sequence of chevrotain shows no deletions or insertions in the coding region and in the deduced amino acid sequence all catalytically active residues are conserved. However, some remarkable substitutions involving cysteine residues can be observed, suggesting that also in chevrotain the seminal-type RNase may not be expressed.

The brain-type RNase sequence of chevrotain shows all the characteristics of a ruminant brain-type RNase gene, like the presence of C-terminal extensions, the presence of several putative glycosylation sites and a high excess of positive charges. The sequence possesses a deletion of the codon for amino acid 19, compared to other

brain-type ribonucleases.

The fifth sequence is a hybrid formed by pancreatic and brain-type RNase sequences of chevrotain. The part of the sequence encoding the signal peptide has the same substitution in the codon for amino acid -1 as the pancreatic RNase gene, and the hybrid is also missing the codon for amino acid 1. In the region encoding amino acids 15 - 18 no similarity with either the pancreatic or the brain-type RNase gene is observed. The rest of this hybrid is almost identical to the brain-type RNase sequence, including the deletion of the codon for amino acid 19. Probably a gene rearrangement or gene conversion has occurred during evolution, with the region encoding amino acids 15 - 20 acting as a hot spot for interchanges.

Phylogenetic analyses of RNase 1 sequences of several ruminant, other artiodactyl and whale species showed again three distinct groups of pancreatic, seminal-type and brain-type RNases. Within each group the chevrotain sequence is the first one to diverge. Average branch length were used to calculate rates. In taxa with duplications of the RNase gene, the gene evolved at least twice as fast as in taxa in which only one gene could be observed. Probably related to new functional adaptations, the two gene duplications in ancestral ruminants caused initially an approximate fourfold increase in evolutionary rate. Later the rate slowed down again and recently diverged bovid and cervid RNases show few substitutions.

Brain-type ribonucleases of ruminants do not show features of pseudogenes. The results described in *Chapter 5* indicate that the genomic sequences of ruminant brain-type RNase code for enzymatically active proteins which are expressed in brain tissue of ox, sheep, roe deer and chevrotain. The bovine enzyme is also expressed in bovine mammary gland but not in milk. However, the bovine pancreatic RNase could be isolated from both sources. This suggests different ways of secretion of the two enzymes probably due to structural differences. The brain-type enzymes may be of functional importance in the species in which they occur with a special role of the C-terminal extensions.

The most striking feature of this sequence of chevrotain brain-type RNase is an 8-fold repeat of six nucleotides with the consensus sequence TCACCC, coding for the dipeptide Ser-Pro. Alignment with the brain-type RNase sequences of bovids, deer, giraffe and pronghorn showed that this repeat occurs five times and is less well conserved in these species. Repetitive Ser-Pro or Thr-Pro sequences are rather frequently observed in protein sequences in which the hydroxy amino acids may be glycosylated and in hinge regions between two separate folding domains in proteins. Similar

structures may occur at the C-termini of ruminant brain-type RNases. The sequence of the brain-type ribonuclease of chevrotain suggests that the C-terminal extensions of ruminant brain-type RNases originate from deletions in the ancestral DNA (including a region with stop codons), followed by insertion of a 5 to 8-fold repeated hexanucleotide sequence, coding for a proline-rich polypeptide.

The ribonuclease sequences of pronghorn, the only representative of the fourth family of true ruminants or pecora, are investigated in order to solve the incongruence of the position of giraffe. The sequences described in *Chapter 6* encode the pancreatic and brain-type genes of pronghorn (*Antilocapra americana*). The determined brain-type sequence does not yet include a small part of the C-terminal extension and the stop codon of this gene. In pronghorn the seminal-type ribonuclease sequence could not be detected, neither by southern blot analysis nor by PCR amplification with primers specific for seminal-type RNases of ruminants. This might be explained by the deletion of the complete seminal-type sequence in pronghorn and not only part of it.

Maximum parsimonious analyses of nucleotide sequences of pancreatic and brain-type ribonucleases of pronghorn and other ruminants with chevrotain as outgroup were performed. No congruency was found between the pancreatic and brain-type ribonuclease subtrees in the obtained four most parsimonious trees. With the maximum parsimony and maximum likelihood method tandemly aligned nucleotide sequences of the two ribonuclease types were investigated. One tree was found with the maximum parsimony method. In this tree the two bovid sequences are not monophyletic and show rather low bootstrap values. The tree obtained with the maximum likelihood method shows giraffe and pronghorn grouping together and these two again with a monophyletic clade of the two bovids, although the bootstrap percentages are rather low.

Phylogenetic analyses of other molecules from ruminant species are also presented in this chapter. In all analyses which include chevrotain this taxon groups with the other ruminants as first diverging clade. Within the pecora the relationships between bovids, cervids, giraffe and pronghorn remains unresolved, although in the majority of analyses pronghorn diverges first, separately or joined with giraffe. The deer species are always found to be monophyletic, but the bovids are sometimes found to be diphyletic. This may be improved with a broader taxon sampling of bovid species, like in earlier studies of amino acid sequences.