Genomic structure of the gene for mouse germ-cell nuclear factor (GCNF). II. Comparison with the genomic structure of the human GCNF gene

ABSTRACT
By homology with the mouse gene, 11 GCNF coding exons can be defined on human chromosome 9. All human GCNF cDNAs identified so far are, however, derived from mRNAs generated by splicing the fourth to the second exon. Although the genomic sequence is highly conserved, the analysis suggests that alternative splicing generates a higher complexity of human GCNF isoforms compared with the situation in the mouse.

INTRODUCTION
Background The nuclear receptors comprise a family of transcriptional regulators involved in a wide variety of biological processes such as embryonic development, differentiation and homeostasis. The family includes ligand-dependent zinc-finger transcription factors for steroid hormones, estrogens, thyroid hormones, retinoids, vitamin D and other hydrophobic molecules. In addition, several family members are 'orphan receptors' for which ligands have yet to be identified. Nuclear receptors have been assigned to six subfamilies on the basis of evolutionary studies. As the first member of the sixth subfamily, GCNF is also known by its systematic name NR6A1. On the basis of homology and expression profile, the receptor has been given the alternative name of retinoic acid receptor-related testis-associated receptor (RTR). GCNF lacks known ligands and is therefore referred to as an orphan receptor. The gene has been mapped to chromosome 9q33-q34.1. Transfection experiments reveal that GCNF can act as a constitutive repressor when binding as a homodimer to promoters containing a direct repeat DNA element 5'-AGGTCAAGGTCA-3' (DRo). Gene targeting in the mouse shows that GCNF has essential functions during embryogenesis. The mouse receptor (mGCNF) is highly expressed in the developing embryonic nervous system and the labyrinthine layer of the placenta. In the adult, high transcript levels are restricted to the developing germ cells. Northern analysis reveals a transcript of 7.5 kilobases (kb) in somatic cells and an additional message of approximately 2.4 kb in male germ cells. This size difference is at least partially due to different polyadenylation sites, and it is therefore assumed that both transcripts code for identical proteins of 495 amino acids. The protein sequence is encoded by 11 exons. When differentiation of P19 embryonal carcinoma cells is triggered by retinoic acid, the transcript and the protein are temporarily upregulated and then downregulated. Isolation of a human cDNA coding for a protein (hGCNF) with an identity to the mouse protein of 98.7%, similar regulation in mouse P19 cells and in the human embryonal carcinoma cell line NT2/D1, together with the presence of two mRNAs of approximately 7.5 and 2.2 kb in human testis, suggested similar functions for mouse and human GCNF. The cloning of human cDNAs that give rise to different hGCNF isoforms, however, suggests a higher complexity in humans. Currently, four different hGCNF cDNAs have been isolated that code for isoforms ranging in size from 454 to 480 amino acids. We have investigated the genomic structure of mammalian GCNF to determine how the different GCNF isoforms are generated. Here we compare the exon/intron structure of the previously characterized mouse gene with the human ortholog. Our study shows that alternative splicing generates at least three of the different GCNF isoforms.

CONCLUSION

In summary, our analysis reveals a conserved structure for GCNF, allows the verification and systematic analysis of splice variants, and may be the basis of a better understanding of GCNF. The human GCNF gene consists of at least 10 exons. The conservation of the intron-exon boundaries is consistent with the extremely high degree of amino-acid conservation between the human and the mouse proteins. The generation of the proteins hGCNF-1, hGCNF-2a and hGCNF-2b can be explained by alternative splicing of the RNA. The sequence of the third coding mouse exon, including the splice sites, is highly conserved; however, at present no human cDNA has been isolated containing this putative exon. Alternative splicing provides a plausible means for generating diversity and may contribute to a higher instructive complexity in human GCNF.