**目的**

克隆丙型肝炎病毒(HCV)核心蛋白结合的肝细胞的蛋白基因, 对新发现的基因及编码产物的结构与功能, 及其基因表达的调节机制的结构基础进行生物信息学分析.

**方法**

应用酵母双杂交技术, 以HCV的核心蛋白作为"诱饵(bait)", 筛选鉴定与其结合的肝细胞中蛋白的编码基因.

应用生物信息学(bioinformatics)技术, 对其中筛选得到的人HCV核心蛋白结合蛋白6(HCBP6)全长基因及其编码产物的一级结构序列, 进行生物信息学分析.

获得该基因的基因组序列, 并利用同源基因序列的比对, 确定该基因在染色体上的定位, 并确定人HCBP6基因的内含子序列.

对其编码上游的基因组DNA序列进行分析, 获得该基因启动子序列的一些信息.

同时, 根据同源基因序列的比较, 确定了小鼠HCBP6的基因序列和氨基酸残基序列.

通过对氨基酸残基序列的疏水性特点的分析, 确定了HCBP6蛋白的潜在的抗原结构位点.

通过蛋白质一级结构的在线计算机软件的分析, 对人HCBP6蛋白质一级结构潜在的功能性结构位点进行了计算机辅助分析预测.

**0 引言**

丙型肝炎病毒(HCV)作为输血传播的非甲非乙型病毒性肝炎(NANBH)的主要病原体, 是一种单股正链RNA病毒。HCV与肝细胞之间的相互作用, 可能是HCV感染发病机制的重要部分。

HCV核心蛋白作为一种反式激活蛋白, 对感染的肝细胞中基因表达谱产生影响, 同时, HCV核心蛋白自身结合形成同二聚体结构, 也可以与肝细胞中其他类型的蛋白之间进行结合, 形成异二聚体或多聚体结构, 对肝细胞中的信号转导产生严重干扰. 通过这些生物学作用, 对肝细胞的细胞凋亡、细胞周期进行调节, 从而参与HCV感染的发病机制。

**1 材料和方法**

***1.1 丙型肝炎病毒核心蛋白结合蛋白的酵母双杂交筛选***

应用酵母双杂交技术筛选HCV核心蛋白结合的肝细胞, 蛋白的研究原理、方法、操作步骤等参考基因治疗研究中心相关的研究论文

***1.2 人丙型肝炎病毒核心蛋白结合蛋白6基因及蛋白一级结构的确定***

利用核苷酸序列数据库和基因序列同源性在线分析途径(<http://www.ncbi.nlm.nih.gov/blast>), 获得与酵母双杂交技术筛选结果同源性的核苷酸序列.

***1.3 人丙型肝炎病毒核心蛋白结合蛋白6同源基因序列的搜索和比较***

利用美国国立图书馆国立生物工程信息中心(NCBI)建立的核苷酸数据库(GenBank)以及同源基因序列的在线搜索分析(<http://www.ncbi.nlm.nih.gov/blast>)软件进行分析.

***1.4 人丙型肝炎病毒核心蛋白结合蛋白6基因组结构的确定***

以人HCBP6的cDNA序列为参照, 应用在线分析工具(<http://www.ncbi.nlm.nih.gov/blast>)分析, 确定人HCBP6的基因组DNA的核苷酸序列.

***1.5 人丙型肝炎病毒核心蛋白结合蛋白6基因启动子的序列分析***

首先以人HCBP6的cDNA序列为参照, 对GenBank数据库中与其同源的人基因组DNA序列进行搜寻, 确定人HCBP6基因组DNA序列中编码基因上游约3 000 nt的核苷酸序列, 并且认为人HCBP6基因启动子序列即位于此段序列之中. 然后采用NCBI的GenBank核苷酸数据库以及相关的3种基因启动子序列在线分析软件(<http://www.cbs.dtu.dk/services/promoter/>)、(<http://www.fruitfly.org/cgi-bin/seq_tools/promoter.pl>)和(<http://bimas.dcrt.nih.gov/molbio/proscan/>)版本1.7, 对潜在的启动子序列和可能的结合的转录因子蛋白的结构位点进行分析.

***1.6 人丙型肝炎病毒核心蛋白结合蛋白6疏水位点的分析***

利用核苷酸序列数据库以及在线软件(<http://bioinformatics.weizmann.ac.il/hyd-bin/plot_hydroph.pl>)对人HCBP6蛋白质一级结构序列进行分析, 发现疏水位点, 对HCBP6蛋白质一级结构中潜在的抗原位点进行分析预测.

**2 结果**

***2.2 人丙型肝炎病毒核心蛋白结合蛋白6基因序列的确定***

利用NCBI建立的核苷酸数据库同源基因序列的在线分析(<http://www.ncbi.nlm.nih.gov/blast>), 在核苷酸序列数据库中检索, 与我们发现的HCBP6同源的核苷酸序列共有6个. 通过核苷酸序列同源性分析发现两个显著的特点: (1)人和小鼠组织中, 都存在HCBP6不同的基因类型, 这种基因类型的生物学意义非常重要, 目前还不清楚; (2)人HCBP6和利用同样技术筛选鉴定的人HCV核心蛋白结合蛋白1(HCBP1)两种基因序列的同源性在44 %以上, 提示这两种基因可能组成一个基因超家族, 也很可能还有其他超家族的成员有待于进一步的研究发现.

***2.3 人HCBP6的基因组DNA结构***

以人HCBP6的cDNA序列为参照, 应用GenBank数据库分析和在线分析工具(<http://www.ncbi.nlm.nih.gov/blast>)分析, 确定人HCBP6的cDNA与人22号染色体的基因组核苷酸序列100%同源, 所以将人的HCBP6基因定位于第22号染色体. 人HCBP6基因组序列与其cDNA序列完全同源, 未发现内含子(intron)序列.

***2.4 人HCBP6疏水性位点的分析***

利用在线软件(<http://bioinformatics.weizmann.ac.il/hyd-bin/plot_hydroph.pl>)对人HCBP6蛋白质一级结构序列进行分析, 发现疏水位点在蛋白质一级结构的第80-110氨基酸残基序列之间. 可以根据此序列人工合成抗原多肽, 进行动物免疫获得抗体或经过噬菌体表面展示技术筛选人源化单链可变区抗体(scFv), 进行免疫组织化学染色, 探讨正常肝组织中或在急性、慢性病毒性肝炎、肝硬化、肝细胞癌的肝组织中HCBP6蛋白表达的特点及变化规律.

***2.5 人HCBP6的编码区上游的基因序列分析***

***2.6 小鼠HCBP6的cDNA及其编码产物的序列***

根据核苷酸序列同源序列的搜索分析, 确定了小鼠的HCBP6基因以及蛋白质一级结构的序列, 如图[4](https://www.wjgnet.com/1009-3079/full/v11/i4/378.htm?s=qc" \l "F4)所示. 小鼠的HCBP6核苷酸序列由555 nt组成, 其氨基酸残基序列由184 aa组成.

**3 讨论**

研究新基因的结构与功能是目前分子生物学研究领域中最为具有挑战性的工作之一.我们在肝炎病毒致病的分子生物学机制研究中, 经常会遇到研究蛋白和蛋白间相互结合、作用, 病毒基因表达产物对肝细胞表达基因谱的影响[[28](https://www.wjgnet.com/1009-3079/full/v11/i4/378.htm?s=qc#B28)-[30](https://www.wjgnet.com/1009-3079/full/v11/i4/378.htm?s=qc#B30)], 以及肝炎病毒蛋白的抗体[[31](https://www.wjgnet.com/1009-3079/full/v11/i4/378.htm?s=qc#B31)-[44](https://www.wjgnet.com/1009-3079/full/v11/i4/378.htm?s=qc#B44)]、抗原（配体）、结合蛋白、模拟表位[[45](https://www.wjgnet.com/1009-3079/full/v11/i4/378.htm?s=qc#B45)-[47](https://www.wjgnet.com/1009-3079/full/v11/i4/378.htm?s=qc#B47)]、DNA/RNA-结合蛋白的筛选等方面的问题, 其中现代的分子生物学技术如酵母双杂交技术(yeast two-hybrid)[[48](https://www.wjgnet.com/1009-3079/full/v11/i4/378.htm?s=qc#B48)]、抑制性消减杂交技术(SSH)[[49](https://www.wjgnet.com/1009-3079/full/v11/i4/378.htm?s=qc#B49)]、噬菌体表面展示(phage display)技术[[50](https://www.wjgnet.com/1009-3079/full/v11/i4/378.htm?s=qc#B50)]等的应用, 都是研究蛋白质的功能与基因的克隆化相耦联的技术途径.

而现在只需要以此序列为参照, 利用GenBank的在线同源基因搜索, 就可以获得这部分序列, 快速有效, 十分简便. 如果我们得到的是人或小鼠等种属的基因类型, 利用同样的技术, 就可以得到其他种属的同源性的基因. 如在本研究中, 我们应用人HCBP6的cDNA序列作为参照, 进行GenBank数据库搜寻时, 就可以获得小鼠、大鼠、牛等种属生物的HCBP6的基因序列. 但是预测软件还有许多不完善.

但是, 对蛋白质结构的疏水性位点的分析结果, 还是相对可靠的. 我们已经利用这些分析结果获得了人HCBP6蛋白的抗原位点, 而且通过噬菌体表面展示技术对半合成的抗体基因文库的筛选获得了相应的人源化单链可变区抗体.

生物信息学技术不仅可以用于结构基因的功能预测, 而且还可以用于调节基因的预测. 例如, 我们应用不同的生物信息学技术对人HCBP6基因编码区上游的核苷酸序列进行分析, 初步获得了HCBP6基因启动子的序列结构特点.