·基础论著·

乙型肝炎病毒表面抗原中蛋白结合蛋白1基因的克隆及生物信息学分析

李志群 马英骥 成军 王琳 刘妍 洪源 伦永志 李晓东

【摘要】 **目的** 克隆乙型肝炎病毒表面抗原中蛋白(MHBs)结合蛋白1(MBP1)基因,构建其真核表达载体并应用生物信息学技术初步探讨其结构及功能。**方法** 应用PCR技术从HepG2细胞提取的cDNA扩增MBP1基因,选用pGEM-T载体进行TA克隆,通过PCR及限制性酶切分析及测序进行鉴定,并应用生物信息学初步分析其物理化学性质、蛋白质结构域、功能及染色体定位。**结果** 成功从HepG2细胞提取并逆转录的cDNA扩增出MBP1基因,其编码区为201个核苷酸(nt),编码产物由66个氨基酸残基(aa)组成,进行GenBank同源序列搜寻发现其与已知基因序列和蛋白序列之间无显著同源性,属于未知功能的新基因,并成功进行TA克隆,经酶切和测序验证正确后亚克隆至pcDNA™3.1/myc-His A真核表达载体,生物信息学分析此基因位于第12号染色体,其编码产物分子量为16645.7,理论pI为5.33,半衰期为4.4 h(体外哺乳类网状细胞),属于不稳定蛋白,疏水指数较高,含潜在的2个螺旋区域和1个β折叠,预测其可能包含2个蛋白激酶C磷酸化作用位点、2个酪蛋白激酶Ⅱ磷酸化作用位点和2个N-豆蔻酰化位点并推测其可能具有不紧密的球蛋白结构,具有信号肽及2个跨膜结构域。**结论** 发现并成功克隆了乙型肝炎病毒表面抗原中蛋白结合蛋白1(MBP1)基因,为以后研究此基因的生物学功能及其在HBV损害肝细胞等方面的具体作用提供了新线索。

【关键词】 乙型肝炎病毒表面抗原中蛋白;克隆;生物信息学分析

Cloning and bioinformatic analysis of gene of hepatocyte protein 1 binding with MHBs

LI Zhi-qun, MA Ying-ji, CHENG Jun, WANG Lin, LIU Yan, HONG Yuan, LUN Yong-zhi, LI Xiao-dong. Institute of Infectious Diseases, Beijing Ditan Hospital, Beijing 100011, China

Corresponding author: CHENG Jun, Email: cj@genetherapy.com.cn

【**Abstract**】 **Objective** To clone and identify a new gene MHBs-binding protein 1(MBP1) and construct its recombinant eukaryotic expression vector to explore its function and structure by bioinformatic analysis. **Methods** MBP1 gene was amplified by reverse transcription-polymerase chain reaction (RT-PCR) technique using HepG2

作者单位:100011 北京市,北京地坛医院传染病研究所(李志群、马英骥、成军、王琳、刘妍、洪源、伦永志、李晓东)

通讯作者:成军 Email: cj@genetherapy.com.cn

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn

cDNA as template and was inserted into vector pGEM-T by TA cloning. Bioinformatic analysis was used to explore its possible physical and chemical characteristics, structure, function and chromosome location. **Results** MBP1 gene was successfully amplified by RT-PCR and it consists of 201 nucleotides which code for a protein of 66 amino acid. After searching in GenBank, we found the new gene had no significant homology with the known genes. The DNA fragment has been inserted into pGEM-T vector by TA cloning. Bioinformatic analysis showed that the new gene was located at chromosome 12 which code for protein with molecular weight 16645.7 and theoretical pI as 5.33, and its estimated half-life was about 4.4 hours (mammalian reticulocytes in vitro). The protein was unstable with its aliphatic index very high and secondary structure predicted it had two helix regions and one β folding region, motif searching showed that it may have two protein kinase C (PKC) phosphorylation sites and two casein kinase Ⅱ phosphorylation site and two N-myristoylation sites. The new protein was predicted to be globular but not as compact as a domain and it may have signal peptide and two transmembrane regions. **Conclusions** MBP1 gene had been recognized and successfully cloned which brought some new clues for studying the biological functions of this new gene and its pathogenesis during HBV inducing liver damage.

【**Key words**】 MHBs protein; Cloning; Bioinformatic analysis

我国学者于1984年报道了中国大陆流行的adr血清型的乙型肝炎病毒(HBV)的DNA全基因序列[1]。HBV感染与肝细胞癌的发生密切相关,病毒基因组编码的蛋白与宿主肝细胞间的相互作用可能是病毒致癌作用的重要分子机制。本研究应用PCR技术从我国流行株adr亚型的HBV DNA全基因克隆出MHBs基因,在酵母双杂交系统筛选人肝细胞cDNA文库过程中发现了一个未知功能的基因并将其命名为HBV表面抗原中蛋白(MHBs)结合蛋白1(MBP1)。

材料和方法

一、材料和试剂

肝母细胞瘤细胞系HepG2及大肠埃希菌DH5α为本室保存,Taq酶(鼎国生物),T4 DNA连接酶、pGEM-T载体、IPTG及X-α-gal购于Promega公司,玻璃奶回收试剂盒购自博大泰克公司,引物合成及DNA序列测定均由Invitrogen公司完成,限制性内切酶*Eco* RⅠ和*Bam* HⅠ购自TaKaRa生物公司。

二、方法

1. MBP1基因序列的确定:本实验室通过酵母双杂交系统筛选人肝细胞cDNA文库中与HBV表面抗原中蛋白(MHBs)结合的肝细胞蛋白基因。挑选克隆后测序并与GenBank数据库进行初步比较,结果发现一个新基因。根据基因起始密码子的Kozak规则和终止密码子下游保守的多聚腺苷酸信号序列,经电子拼接后推定该基因的开放读码框架并获得其相应的全长编码基因,将该未知功能

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn

基因命名为 MBP1,发现其开放读码框长 201 bp,编码产物由 66 个氨基酸残基组成。

2. 多聚酶链反应扩增 MBP1 基因:根据 MBP1 的全长编码基因设计引物,有义链引物 5'-GAATTCATGAATTTCCTGACAGG-3',引入 *Eco* R Ⅰ酶切位点,反义链引物 5'- GGATCCGGGAGAAAGTAAGT-3',引入 *Bam* H Ⅰ酶切位点,提取 HepG2 细胞的总 RNA 并以反转录产物 cDNA 为模板进行 PCR 扩增,扩增条件如下:95℃预变性 5 min,94℃变性 45 s,55℃退火 45 s,72℃延伸 60 s,共进行 35 个循环,72℃终延伸 10 min。

3. 目的片段的克隆及鉴定:PCR 产物经 1% 琼脂糖凝胶电泳后切取目的片段并以玻璃奶法回收,与 pGEM-T 载体连接,转化 DH5α 感受态细菌,在含氨苄青霉素的 LB/X-gal/IPTG 培养板上,37℃培养 12 ~ 16 h 后挑取阳性菌落,增菌后提取质粒进行限制性酶切分析鉴定并将经鉴定的克隆送测序。

4. 新基因的生物信息学分析:将该基因序列与 NCBI 进行 BLASTn 比对后确定其染色体定位,应用 ProtParam 工具(www. expasy. ch/tools/protparam. html)对 MBP1 蛋白质一级结构的生物信息学进行预测,包括蛋白质的相对分子量、理论 pI 值、消光系数、半衰期、不稳定系数以及总平均亲水性等,应用 nnpredict (www. cmpharm. ucsf. edu/ ~ nomi/nnpredict. html)预测氨基酸序列中各个氨基酸对应的二级结构;应用 PredictProtein(www. embl-heidelberg. de/ predictprotein/predictprotein. html)对蛋白质结构的分析;应用 ProtScale (www. expasy. org /tools/protscale. html)进行蛋白质的疏水性分析;应用 SignalP-信号肽预测工具(http://www. cbs. dtu. dk/services/SignalP/)进行蛋白质结构信号肽分析,这有助于蛋白质功能域的区分及蛋白质细胞定位;应用 Tmpred-跨膜结构分析(http://www. ch. embnet. org/software/ TMPRED_form. html)对于正确认识蛋白质结构和功能具有重要的意义,Coils 蛋白质卷曲区域的预测(http://www. ch. embnet. org/software/COILS_form. html)用于预测蛋白质中卷曲螺旋。

结　　果

一、MBP1 全长编码序列的确定

新基因开放读码框长 201 bp,编码 66 个氨基酸残基(图 1),该序列被美国核苷酸序列数据库 GenBank 收录,收录号码为 DQ307498。

二、MBP1 基因的 PCR 扩增及其重组载体的构建

以 HepG2 细胞 cDNA 文库为模板,PCR 扩增后经 1% 琼脂糖凝胶电泳,可见长度为 201 bp 左右的电泳条带(图 2)。纯化后的 PCR 产物 TA 克隆至 pGEM-T vector System Ⅰ载体并进一步将 pGEM-T- MBP1 进行 *Eco* R Ⅰ和 *Bam* H Ⅰ双酶切得到两条带(图 3),回收目的片段经测序验证为正确的克隆,命名为 pGEM-T-MBP1,表明 pGEM-T-MBP1 构建成功。

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn

图 1 MHBs 蛋白结合蛋白 1(MBP1)基因序列及推定蛋白(GenBank DQ307498)

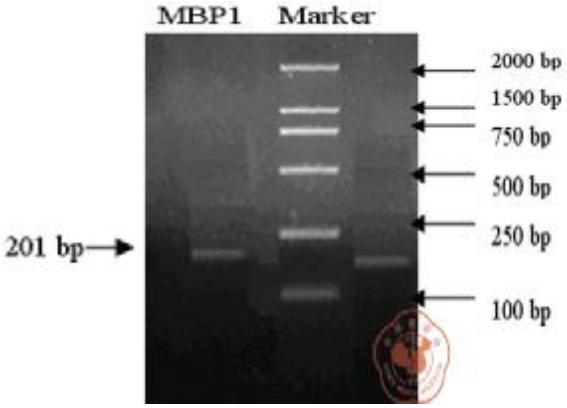

图 2 MBP1 基因的 PCR 产物琼脂糖凝胶电泳

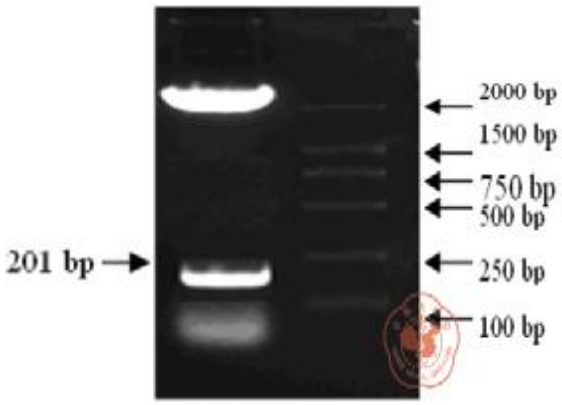

图 3 重组质粒 pGEM-T-MBP1 的酶切鉴定

三、MBP1 基因生物信息学分析

将基因测序结果与 NCBI 进行 BLASTn 比对分析发现 MBP1 基因位于人染色体 12 号,ProtParam 分析其编码产物的分子量为 16645.7,理论 pI 为5.33,消光系数3250M-1 cm-1(280 nm)是在6.0 mmol/L 盐酸胍,0.02 mmol/L 磷酸盐缓冲液,pH 6.5 以半胱氨酸进行计算,半衰期为 4.4 h(体外哺乳类网状细胞),不稳定系数为 59.97,故认为该蛋白是不稳定的,疏水性指数为 19.90,总平均亲水性为 0.713。nnpredict 预测的二级结构为:----------HHHHHHHHHH-------HHHH----------EEEE----------(H 表示螺旋,E 表示 β 折叠,“ - ”表示转角结构),PredictProtein 预测包含潜在的蛋白激酶 C 磷酸化位点(5/TGSE 及 19/SALD),酪蛋白激酶Ⅱ磷酸化位点(5/TGSE 及 19/SALD),N-豆蔻酰化位点(43/GLAGGA 及50/GQGSSF),预测的蛋白质区域结构见图 4。PROF 预测提示包含提交蛋白质的二级结构含量与疾病分类、蛋白质亲水表面统计、残基组成、预测的二级结构等相关。OBS-sec 表示二级结构的观测值;PROF-sec 表示预测的二级结构;Rel-sec 为预测二级结构的可信度,数值越大表示越可信;SUB-secPROF 表示预测的二级结构的集合,所有区域平均精确度大于 82%,H 为螺旋,E 为折叠,L 为随机卷曲;O-3-acc 表示相对亲水表面的观测值,共有 3 种状态,b = 0 ~ 9%,I = 9% ~ 36%,e = 36% ~ 100%;P-3-acc 表示预测的相对亲水表面;Rel-acc

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn

表示预测亲水表面的可信度;SUB-acc 表示预测的亲水表面的集合(图 5),并预测该蛋白可能为球蛋白,但结构并非特别紧密,蛋白质疏水性预测最大值为 1.189,最小值为 -1.744,在 22 ~ 24 aa,36 ~ 46 aa,50 ~ 64 aa 区域的氨基酸具有很强的疏水性,其次为 32 ~ 36 aa 区域具有一定的疏水性(图 6)。用 SignalP-信号肽预测工具预测该蛋白信号肽概率为 0.971,锚定蛋白概率为 0.015,最大分隔位点概率为 0.085,位于 57 与 58 氨基酸间(图 7)。跨膜结构分析该蛋白包括一个强跨膜螺旋(从 33 ~ 56 aa,跨膜方向由内向外)(图 8),无明显的卷曲螺旋结构。

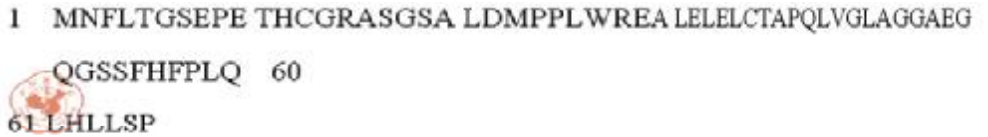

图 4　预测 MBP1 蛋白质的区域结构

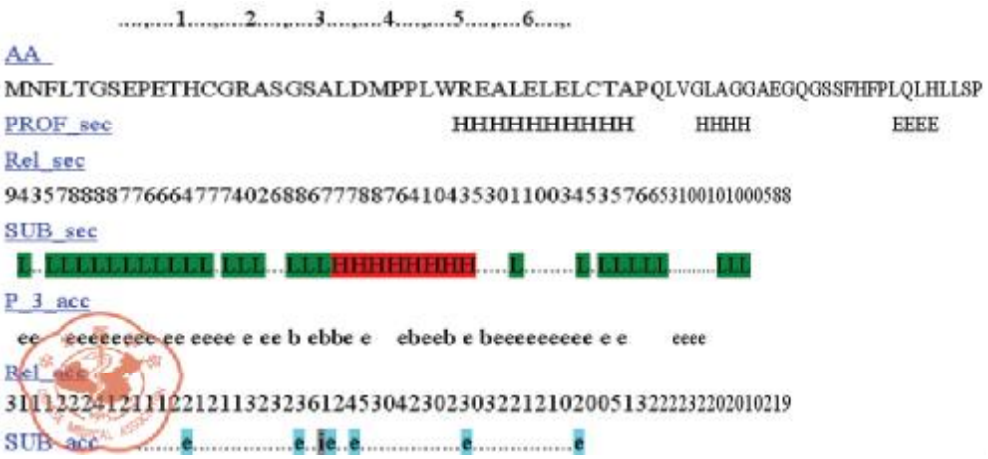

图 5　MBP1 序列的 PROF 预测

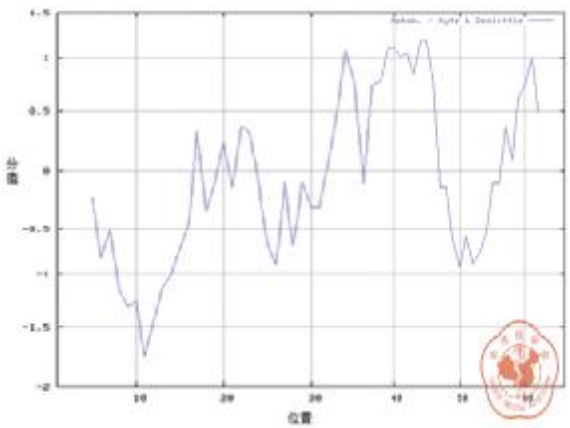

图 6　MBP1 蛋白的疏水性分析

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn

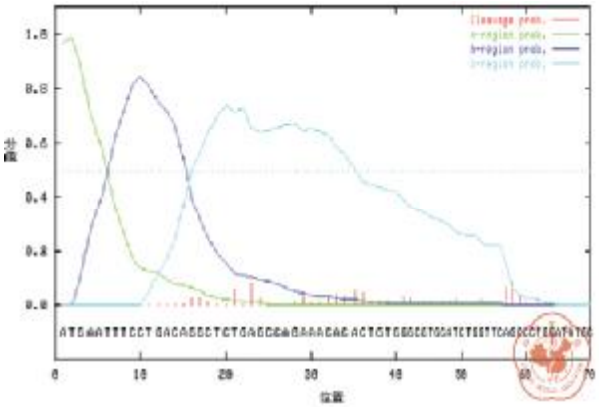

图7 MBP1蛋白的信号肽分析

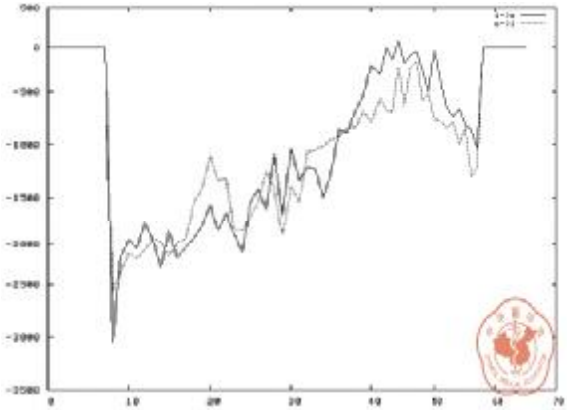

图8 MBP1蛋白的跨膜结构域分析
注:0表示跨膜方向由外向内

讨 论

本研究应用酵母双杂交系统筛选人肝细胞cDNA文库中与MHBs相结合的肝细胞蛋白基因。经克隆分析发现新基因与GenBank中注册的已知功能基因序列无明显同源性。经电子拼接推定新基因的开放读码框架并获得相应的全长编码基因,该序列被美国核苷酸序列数据库GenBank收录,编号为DQ307498,并命名为HBV表面抗原中蛋白(MHBs)结合蛋白1(MBP1)。根据该序列设计引物,以人肝癌细胞cDNA为模板经PCR扩增新基因序列,产物经测序完全符合计算机分析结果,表明我们顺利得到了MBP1基因的全长编码序列。

MBP1基因为一种功能未知的新基因,对于该基因结构与功能的研究目前尚为空白,MBP1基因开放读码框长201 bp,编码66个氨基酸残基,将基因测序结果与NCBI进行BLASTn比对分析发现MBP1基因位于人第12号染色体,ProtParam分析其分子量为16645.7,理论pI为5.33,半衰期为4.4 h(体外哺乳类网状细胞),大于20 h(酵母体内)及10 h(埃希氏菌属大肠埃希菌体内),不稳定系数为59.97,属于不稳定蛋白,疏水指数较高19.90,总平均亲水性为0.713,在22~24 aa,36~46 aa,50~64 aa区域的氨基酸具有很强的疏水性,含有2个螺旋区域,1个β折叠区,2个潜在的蛋白激酶C磷酸化位点(5/TGSE和19/SALD),2个酪蛋白激酶Ⅱ磷酸化位点(5/TGSE和19/SALD)和2个N-豆蔻酰化位点(43/GLAGGA和50/GQGSSF),其中蛋白激酶C磷酸化位点与介导胞外分泌有关[2],酪蛋白激酶Ⅱ磷酸化位点与细胞凋亡和细胞周期的调节有关[3],N-豆蔻酰化位点可能与蛋白定位于内质网或线粒体外膜有关[4]。用SignalP-信号肽预测工具预测该蛋白信号肽概率为0.971,锚定蛋白概率为0.015,跨膜结构分析其包括两个强跨膜螺旋(从33~56 aa,跨膜方向由内向外;0表示跨膜方向由外向内)。本研究结果为进一步研究新基因结构和功能奠定基础,同时也为研究MBP1蛋白的生物学功能及HBV的致病机制指明了新方向。

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn

参 考 文 献

1 贺光, 孙开来. 生物信息学在蛋白质研究中的应用. 国外医学遗传学分册,2002,25:156-158.

2 Morgan A, Burgoyne RD, Barclay JW, et al. Regulation of exocytosis by protein kinase C. Biochem Soc Trans,2005,33:1341-1344.

3 Yamane K, Kinsella TJ. Casein kinase 2 regulates both apoptosis and the cell cycle following DNA damage induced by 6-thioguanine. Clin Cancer Res,2005,11:2355-2363.

4 Colombo S, Longhi R, Alcaro S, et al. N-myristoylation determines dual targeting of mammalian NADH-cytochrome b5 reductase to ER and mitochondrial outer membranes by a mechanism of kinetic partitioning. J Cell Biol,2005,168:735-745.

(收稿日期:2008-04-27)

(本文编辑:温少芳)

李志群,马英骥,成军,等. 乙型肝炎病毒表面抗原中蛋白结合蛋白 1 基因的克隆及生物信息学分析[J/CD]. 中华实验和临床感染病杂志:电子版, 2008,2(3):162-168.

PDF 文件使用 "pdfFactory Pro" 试用版本创建 www.fineprint.com.cn