

[19] 中华人民共和国国家知识产权局



# [12] 发明专利说明书

专利号 ZL 200510073393. X

[51] Int. Cl.

*C07K 14/435 (2006.01)*

*C07K 14/47 (2006.01)*

*C07K 19/00 (2006.01)*

*A61K 38/17 (2006.01)*

*A61P 35/00 (2006.01)*

*C12N 15/12 (2006.01)*

[45] 授权公告日 2007 年 11 月 14 日

[11] 授权公告号 CN 100348614C

[22] 申请日 2005.6.3

[21] 申请号 200510073393. X

[73] 专利权人 北京大学

地址 100080 北京市海淀区颐和园 5 号

[72] 发明人 陈慰峰 乔欢 庞学雯 王俞

田 婵

[56] 参考文献

WO 0121644 A 2001.3.29

Large scale identification of human hepatocellular carcinoma - associated antigens by autoantibodies. . Wang, Y. et al. J. Immunol. , Vol. 169 No. 2. 2002

审查员 杨振宇

[74] 专利代理机构 北京泛华伟业知识产权代理有

限公司

代理人 王凤华

权利要求书 1 页 说明书 17 页 附图 11 页

[54] 发明名称

一种肝癌 - 睾丸特异性抗原蛋白质和抗原肽

[57] 摘要

本发明涉及一种肝癌 - 睾丸特异性抗原蛋白质和抗原肽。该肝癌 - 睾丸特异性抗原蛋白质为具有图 1 所示的氨基酸序列的蛋白质及其衍生蛋白质。所述的抗原肽的氨基酸序列为在图 1 所示的序列中起止 AA108 - AA191。本发明提供的肝癌 - 睾丸特异性抗原蛋白质和抗原肽在 pRB/E2F 信号转导中可作为 E2F 活性阻断剂, 对其同源类似物 TFDP - 1 发挥拮抗作用。该蛋白质/肽可作为 CT 抗原或抗原基因而制成疫苗, 这种肿瘤抗原蛋白质通过细胞内分解, 能产生与主要组织相容性复合体 (MHC) - I 类分子相结合的、在结合状态能被 T 细胞所识别的肽片段, 激发免疫反应而达到治疗癌症的目的, 从而可以应用于制备治疗癌症, 特别是肝癌的药物中。

1、一种肝癌-睾丸特异性抗原蛋白质，具有如下氨基酸序列(a)或衍生蛋白质(b)：

所述氨基酸序列(a)具有图1所示的氨基酸序列；

所述衍生蛋白质(b)为将所述氨基酸序列(a)中的氨基酸序列经过一个或几个氨基酸残基的取代，缺失或添加突变所产生的衍生蛋白质，该衍生蛋白质与所述氨基酸序列(a)具有相同或相似的功能。

2、一种权利要求1所述的抗原蛋白质在制备治疗肝癌药物中的应用。

3、一种编码权利要求1所述的肝癌-睾丸特异性抗原蛋白质的基因。

4、一种肝癌-睾丸特异性抗原肽，其具有如下肽段(c)或衍生肽(d)：

所述肽段(c)为上述肝癌-睾丸特异性抗原蛋白质的特异性DNA结合结构域的含84个氨基酸的肽段，该氨基酸的序列为图1所示的序列中起止AA108-AA191；

所述衍生肽(d)为将所述肽段(c)中的氨基酸序列经过一个或几个氨基酸残基的取代，缺失或添加突变所产生的衍生肽，该衍生肽与所述肽段(a)具有相同或相似的功能。

5、一种权利要求4所述的抗原肽在制备治疗肝癌药物中的应用。

6、一种编码权利要求4所述的肝癌-睾丸特异性抗原肽的基因。

## 一种肝癌-睾丸特异性抗原蛋白质和抗原肽

### 技术领域

本发明涉及一种肝癌-睾丸特异性抗原蛋白质和抗原肽。

### 背景技术

肿瘤-睾丸 (cancer-testis, 简称 CT) 抗原蛋白质是目前鉴定的肿瘤抗原中最的一类, 它们编码基因的特点是在很多类型的肿瘤中都有表达, 如在黑色素瘤、肺癌、肉瘤和膀胱癌中都有表达, 但在除睾丸外的正常组织都不表达, 在胎盘、卵巢、胰腺中有一定量的表达。因为睾丸是免疫特许部位, 所以这一类抗原在治疗上被认为是肿瘤特异性的抗原, 是最有希望用作为肿瘤免疫治疗用途的一类抗原。目前试用于临床上的也主要就是这一类抗原, 如 MAGE-1 和 NY-ESO-1 即是在一定类型的肿瘤中有高阳性率的表达, 又有很好的免疫原性的抗原蛋白质, 用于肿瘤的免疫治疗时有很好的前景。

原发性肝细胞癌是人类原发性肝癌中最常见的一种类型, 也是全世界最常发生的恶性肿瘤之一, 尤其在东南亚及非洲, 它的发病率更高。就中国而言, 每年约有 110, 000 人死于肝癌, 占全世界每年肝癌死亡人数的四分之一。这种高死亡率, 主要是由于早期肝癌缺乏特异临床症状, 导致大部分患者就诊迟, 而又缺乏有效治疗手段及术后易复发所致。发病率高、恶性度高、诊断迟、治疗方法少、预后差, 是肝癌的几大特点。尽管国内外科科研人员都对肝癌进行了多方面的研究, 但目前肝癌发生发展的分子机理还不太清楚, 对肝癌的治疗也主要是采取手术治疗及局部或全身化疗, 少数病人辅以肿瘤生物疗法, 但术后生存期仍不尽人意。为此, 对肝癌发生发展机理的研究成为当前的迫切要求, 希望籍此早日阐明肝癌发生发展的分子基础, 或与肝癌发生发展密切相关的分子, 以便随访高危人群, 有效预防、早日诊断、综合治疗、减少复发。

以前的研究表明, 肝癌的发生是一个多阶段、多步骤的过程, 这一过程中伴随着机体许多的细胞学、遗传学的改变, 在细胞内存在的正向调控和负向调控之间的相对平衡被打破, 出现多种基因表达异常, 包括癌基因的激活和/或抑癌基因灭活, 激发异常的信号转导通路, 导致细胞周期及细胞凋亡异常, 从而使得细胞恶性转化, 并随着各种分子异常的累积, 肿瘤出现进展、转移。而且, 肝癌在分子水平的改变是异质性的, 可能存在多个基因, 多个途径交叠作用。同时, 细胞内在的自我防护机制也可能会诱导产生拮抗性分子。所以, 明确这些差异表达的基因, 对于肝癌发生发展机制的阐明和对肝癌的临床诊断、治疗都具有重要意义。

## 发明内容

本发明的目的在于提供一种可用于肝癌治疗的肿瘤睾丸特异性抗原蛋白和肿瘤睾丸特异性抗原肽。

本发明的目的是通过如下的技术方案实现的：

本发明提供一种肝癌-睾丸特异性抗原蛋白质 TFDP-3 (HCA661)，其具有如下氨基酸序列 a 或衍生蛋白质 b：

所述氨基酸序列 a 具有图 1 所示的氨基酸序列；

所述衍生蛋白质 b 为将所述氨基酸序列 a 中的氨基酸序列经过一个或几个氨基酸残基的取代，缺失或添加突变所产生的衍生蛋白质，该衍生蛋白质与所述氨基酸序列 a 具有相同或相似的功能。

本发明提供一种编码上述肝癌-睾丸特异性抗原蛋白质的基因。

本发明所述的基因属于 DP 家族，其具有图 2 所示的核苷酸序列，在基因文库中登录号为：AF219119。

本发明提供一种包含上述肝癌-睾丸特异性抗原蛋白质的融合蛋白。

本发明提供一种肝癌-睾丸特异性抗原肽，其具有如下肽段 c 或衍生肽 d：

所述肽段 c 为上述肝癌-睾丸特异性抗原蛋白质的特异性 DNA 结合结构域的含 84 个氨基酸的肽段，该氨基酸的序列为在图 1 所示的序列中起止 AA108-AA191；

所述衍生肽 d 为将肽段 c 中的氨基酸序列经过一个或几个氨基酸残基的取代，缺失或添加突变所产生的衍生肽，该衍生肽与所述肽段 a 具有相同或相似的功能。

本发明提供一种编码上述肝癌-睾丸特异性抗原肽的基因。

本发明提供一种包含上述肝癌-睾丸特异性抗原肽的融合蛋白。

本发明提供的肝癌-睾丸特异性抗原蛋白质、抗原肽、及其们的融合蛋白在 pRB/E2F 信号转导中可作为 E2F 活性阻断剂，对其同源类似物 TFDP-1 发挥拮抗作用。该蛋白质/肽及编码其的 DNA 可作为 CT 抗原或抗原基因而制成疫苗，这种肿瘤抗原蛋白质通过细胞内分解，能产生与主要组织相容性复合体 (MHC) -I 类分子相结合的、在结合状态能被 T 细胞所识别的肽片段，激发免疫反应而达到治疗癌症的目的，从而可以应用于制备治疗癌症，特别是肝癌的药物中。应用本发明的肿瘤抗原蛋白质可以提供活化抗肿瘤免疫的医药、可提供肿瘤的诊断方法。抗本发明的肿瘤抗原蛋白质的抗体可用于亲和层析、cDNA 文库的筛选、免疫学诊断或医药的制备。

本发明克隆的新基因是一个肝癌-睾丸特异性表达的、与肝癌发生相关的基因，对于它的功能的研究，将有助于肝癌发生发展机理的阐明。该基因编码的新蛋白 TFDP-3 (HCA661) 是一个新的 CT 抗原，在原发性肝细胞性肝癌中，TFDP-3 (HCA661) 阳性约占 1/3，作为肿瘤疫苗具有应用于肝细胞性肝癌治疗的潜在的临床应用价值。同时，与其它 CT 抗原不同，该蛋白具有明显的功能，是 DP 家族的新成员，该家族可以与

E2F 形成异源二聚体,从而增强 E2F 的转录调节功能,处于 pRB/E2F 信号通路中的核心位置。pRB 蛋白是多种 DNA 肿瘤病毒立早基因产物作用的靶蛋白,如猴病毒 40 大 T 抗原、乳头状瘤病毒 E7 蛋白、腺病毒 E1a 蛋白等。这些癌基因产物可以与 E2F 竞争结合 pRB 蛋白,破坏 E2F 与 pRB 的作用与结合,导致 pRB 对 E2F 活性的抑制被消除,激活 E2F 依赖性靶基因的转录。在多数人类肿瘤中, E2F 的调控机制是失调的,而 E2F 功能的失调往往会伴随着细胞的转化,甚至肿瘤的发生。以前的许多研究都已证明, DP 蛋白或 E2F 蛋白具有原癌基因的特性。TFDP-3 (HCA661) 的结构分析表明,它具有与 TFDP-1 相似的结构,即含有一个 DNA 结合结构域和一个二聚化结构域。功能研究表明, TFDP-3 (HCA661) 可以与 E2F 家族的所有具有二聚化结构域的蛋白形成异二聚体,在蛋白相互作用的水平上, TFDP-3 (HCA661) 具有与 TFDP-1 完全一致的生物学特性。但在蛋白与 DNA 相互作用的水平上, TFDP-3 (HCA661) 与 E2F 蛋白形成的异二聚体极大地削弱了 E2F 结合 DNA 的能力,并可抑制 E2F 的转录活性,亦可抑制 TFDP-1 对 E2F 的别构激活。TFDP-3 (HCA661) 对 E2F 功能影响的独特性质使其可望作为当病毒感染或 pRB 突变导致的 pRB/E2F 信号通路失控时的拮抗剂,从而弥补 pRB 的功能,至少是部分弥补。TFDP-3 (HCA661) 在组织表达的特异性以及它所具有的与 pRB 相似的抑制 E2F 活性的特性,可视为肝癌睾丸特异性的抑癌基因,这个基因和蛋白可能为将来发展肝癌特异性治疗提供一个较好的靶点。

此外,本发明首次将 TFDP-3 (HCA661) 基因鉴定为肿瘤-睾丸特异性抗原基因;在使用免疫方法治疗癌症时,可以用本发明的肿瘤抗原蛋白质诱导机体产生只针对肿瘤细胞的免疫应答反应,从而在不杀伤正常细胞的前提下,特异杀伤体内的肿瘤细胞而达到治疗或者防治肿瘤细胞转移的作用;同时,可以通过对该肿瘤抗原基因或抗体的检测,对肿瘤的发生、发展以及是否有转移进行诊断。

#### 附图说明

图 1 为本发明提供的肝癌-睾丸特异性抗原蛋白质的氨基酸序列;

图 2 为本发明提供的编码肝癌-睾丸特异性抗原蛋白质的核苷酸序列;

图 3 为使用 CLUSTAL W (1.82) 分析程序对 TFDP-3 (HCA661) 蛋白的氨基酸序列与其 DP 家族同源类似物 TFDP-1 的氨基酸序列的对比分析结果;其中,“\*”指序列中的氨基酸残基完全一致;“:”指序列中的氨基酸残基属于保守替换;“.”指序列中的氨基酸残基属于半保守替换;

图 4 为 TFDP-3 (HCA661) mRNA 在肝细胞性肝癌 (HCC) 中的 Northern blot 验证图;其中,Testis 为睾丸组织;Adj 为邻近非癌组织;HCC 为肝细胞性肝癌;Normal liver 为正常肝组织;

图 5 为 TFDP-3 (HCA661) 在 HCC 细胞系中的表达分布;其中, L02 为正常肝细胞

系, Jurkat为急性T淋巴细胞白血病细胞系, 其余为HCC细胞系;

图6 为TFDP-3 (HCA661) 与E2F家族蛋白相互作用的GST Pull-down分析; 其中, A图: E2F的可溶性GST融合蛋白在大肠杆菌 JM109中诱导和表达, 可溶性蛋白的表达水平通过考马斯亮蓝染色评估; 箭头所示为可溶性蛋白的特异性迁移带; 无融合基因的GST蛋白作为阴性参照。B图: 作为阴性参照的GST蛋白和E2F的GST融合蛋白分别与谷胱甘肽-琼脂糖微珠结合, 然后与体外转录、翻译的TFDP-3 (HCA661) 蛋白相互作用, 并洗掉未结合的蛋白, 剩余的与谷胱甘肽-琼脂糖微珠结合的蛋白通过电泳、放射自显影分析。体外转录、翻译的TFDP-3 (HCA661) 蛋白作为阳性参照;

图7 为TFDP-3 (HCA661) 与E2F蛋白相互作用的免疫共沉淀分析; 其中, 上图为采用抗HA多克隆抗体通过免疫印迹分析从共转染的HeLa细胞中提取的免疫共沉淀蛋白; 中图为采用抗HA多克隆抗体通过免疫印迹分析共转染的HeLa细胞裂解物中的E2F蛋白的表达水平; 下图为采用抗FLAG M2单克隆抗体通过免疫印迹分析共转染的HeLa细胞裂解物的TFDP-3 (HCA661) 蛋白的表达水平; 箭头所示为TFDP-3 (HCA661) 或E2F蛋白的特异性迁移带;

图8 为 TFDP-3 (HCA661) 与 E2F 蛋白相互作用形成的异二聚体的 DNA 结合能力分析 (EMSA);

图9 为 TFDP-3 (HCA661) 亚细胞定位分析; 其中, A, E2F 或 DP 蛋白单独转染时的亚细胞定位; B, E2F 和 DP 共转染时的亚细胞定位; 所有图像通过配备了 CCD 的免疫荧光倒置显微镜在放大倍数为 480 倍的视野下摄取;

图10 为 TFDP-3 (HCA661) 对 E2F 转录活性影响的分析; 其中, A 为 TFDP-3 (HCA661) 抑制 E2F 的转录活性结果; B 为 TFDP-3 (HCA661) 抑制 E2F/TFDP-1 介导的转录激活结果; C 为 TFDP-3 (HCA661) 过表达能抑制内源性 E2F 活性结果; 所显示的数据是三次实验的平均值及其标准差;

图11 为 TFDP-3 (HCA661) 突变体的图示; 其中, A 为两个已知的人 DP 家族蛋白 TFDP-1 和 TFDP-2 的 DNA 结合结构域与 TFDP-3 (HCA661) 的 DNA 结合结构域的氨基酸残基的比较; “\*”指序列中的氨基酸残基完全一致; “:”指序列中的氨基酸残基属于保守替换; “.”指序列中的氨基酸残基属于半保守替换; 序列分析所使用的分析程序为 CLUSTAL W (1.82); “Δ”指该氨基酸残基参与了与 DNA 碱基的相互作用; “▲”指该氨基酸残基参与了与 DNA 骨架的相互作用; “■”指该氨基酸残基参与了与蛋白质的异二聚化的相互作用; B 为从全长 TFDP-3 (HCA661) 和 TFDP-1 的氨基酸序列中衍生出的替换突变体的图示; TFDP-3 (HCA661) 和 TFDP-1 根据结构域的组成可以将氨基酸序列分成 4 个部分, 即 N 端未知功能区(UNT)、DNA 结合结构域(DBD)、异二聚化结构域(HD)和 C 端未知功能区(UCT); C 为 TFDP-3 (HCA661) 和 TFDP-1 在 DNA 结合结构域内进行的替换衍生出的突变体的图示; TFDP-3 (HCA661) 的氨基

酸残基用阴影标记；同时在突变体的右侧也简要地标注了该突变体的转录性质；

图 12 为 TFDP-3 (HCA661) 独特功能产生的分子基础分析；其中，A 为 TFDP-3 (HCA661) 和 TFDP-1 突变体的蛋白表达；B 和 C 为 TFDP-3 和 TFDP-1 的替换突变体对 E2F-4 的转录活性的影响的报告基因评估；所显示的数据是三次实验的平均值及其标准差。

## 具体实施方式

### 实施例1. 基因分析

TFDP-3 (HCA661) 是通过SEREX方法从病人的肝癌组织中用血清筛选出的新基因。根据生物信息学分析，TFDP-3 (HCA661) 基因定位于人基因的X染色体上，是一个单外显子组成的基因。其转录本大小为1,680 bp，其中编码框为1,218 bp，起止点为90-1,307，共编码405个氨基酸的蛋白（如图1所示）。该蛋白是转录因子DP家族的新成员，与其DP家族同源类似物TFDP-1有高度的相似性（75.2% 氨基酸一致性；82.0% 氨基酸相似性）。二者的氨基酸序列的对比分析如图3所示，与TFDP-1类似，TFDP-3的氨基酸序列包括一个进化保守的DNA结合结构域（AA108-AA191）和一个进化保守的异二聚化结构域（AA192-AA264），而且，在DNA结合结构域中，也含有一个在E2F家族和DP家族普遍具有的RRXYD DNA识别基序（AA162-AA166）。在已知的E2F和DP家族中，RRXYD DNA识别基序中的第三位氨基酸通常为疏水性的氨基酸（Valine (V) 或Isoleucine (I)），但在TFDP-3 (HCA661) 中，该氨基酸为亲水性的Threonine (T)，这种非保守性的氨基酸替换对TFDP-3 (HCA661) 的DNA结合功能是否有影响尚不清楚。此外，TFDP-3 (HCA661) 的羧基端的20个氨基酸有15个是酸性氨基酸，而TFDP-1和TFDP-2只有13个和11个酸性氨基酸。这些分析结果强烈地提示TFDP-3 (HCA661) 是人的DP家族的第三个新成员。

### 实施例2. TFDP-3 (HCA661) mRNA Northern blot 分析

为执行Northern blot 分析，从HCC标本和对应的非癌组织标本以及睾丸组织中提取总RNA。RNA的完整性通过电泳分析。电泳时，每孔总RNA的上样量为30 mg，然后转膜。用特异性<sup>32</sup>P标记的cDNA探针与硝酸纤维素膜在65℃杂交过夜。用0.1xSSC/0.1% SDS溶液洗涤三次，每次30分钟，在-70℃放射自显影。全长mRNA的大小通过与28S和18S的核糖体RNA的迁移程度对比得出。

由于核苷酸序的高度相似性，Northern blot杂交结果（如图4所示）显示，在睾丸组织的RNA中出现了两条强带，分别为~1.7 kb和2.1 kb。在HCC组织的RNA中只出现~1.7 kb的带，而非癌组织以及正常肝组织只检测到了2.1 kb的带。DNA序列分析表明，2.1 kb的基因是TFDP-1，1.7 kb的基因是TFDP-3 (HCA661)。

### 实施例3. TFDP-3 (HCA661) 在HCC和邻近非癌组织标本中的表达分布

TFDP-3 (HCA661) mRNA的表达谱通过RT-PCR分析, 结果列于表1。组织包括从Clontech购买的16种正常组织cDNA: 脑、心、肾、肝、肺、胰腺、胎盘、骨骼肌、结肠、卵巢、外周血白细胞、前列腺、小肠、脾、睾丸和胸腺, 从HCC组织和与之配对的邻近非癌组织提取的总RNA逆转录成cDNA。使用特异性引物对扩增TFDP-3 (HCA661) 片段。Glyceraldehyde 3-phosphate dehydrogenase (G3PDH) mRNA作为内参。所用引物如下: TFDP-3, forward, 5'-TAC ACT CGG CCT GGA AGA ATT G-3'; reverse, 5'-TCT TCC TCC TCG ACT GCT G-3', size, 1,244 base pairs (bp)。

在16种正常组织中, 只有睾丸检测到了TFDP-3 (HCA661) mRNA的高水平表达, 在正常胰腺组织中表达很弱, 其余各种正常组织未检测到表达。在HCC组织中, 17份HCC组织标本中阳性标本有5份, 与17份HCC组织标本相对应的邻近非癌组织中仅有2份标本有微弱的表达, 9份正常肝组织无一表达, 是一个典型的癌-睾丸 (CT) 抗原。

表1、TFDP-3 (HCA661) 在HCC和邻近非癌组织标本中表达的RT-PCR分析

	HCC	ADJ	Cirrhosis	Normal liver
Case	17	17	5	9
+ve	5	2	0	0

### 实施例4. TFDP-3 (HCA661) mRNA在HCC细胞系中的表达

采用TRIZOL (GIBCO BRL) 从L02 (人正常肝细胞系), CH-Hep-1, CH-Hep-2, CH-Hep-3, CH-Hep-4, CH-Hep-5, CH-Hep-6, Hep 3B, Bel-7402, Bel-7405, SSMC7721, Hep G2和HLE HCC细胞系, SK-Mel-37和Mel-526 黑色素瘤细胞系, Jurkat (人急性T淋巴细胞白血病细胞系) 细胞提取总RNA。总RNA逆转录并使用特异性引物对扩增TFDP-1, TFDP-2和TFDP-3 (HCA661) 片段, G3PDH mRNA作为内参。使用下列参数执行RT-PCR: TFDP-3 (HCA661), 35个循环, 94°C 30 秒; 60°C 30秒; 72°C 90秒, 72°C 延伸10分钟; TFDP-1, 35个循环, 94°C 30秒; 55°C 30秒; 72°C 90秒, 72°C 延伸10分钟; TFDP-2, 35个循环, 94°C 30秒; 55°C 30秒; 72°C 60秒, 72°C 延伸10分钟; G3PDH, 25个循环, 94°C 15秒; 64°C 20秒; 72°C 20秒, 72°C 延伸10分钟。TFDP-3 (HCA661), TFDP-1, TFDP-2和G3PDH 引物序列和PCR产物片段长度如下: TFDP-3, forward, 5'-TAC ACT CGG CCT GGA AGA ATT G-3'; reverse, 5'-TCT TCC TCC TCG ACT GCT G-3', size, 1,244 base pairs (bp); TFDP-1, forward, 5'-ATG GCA AAA GAT GCC GGT CTA ATT G-3'; reverse 5'-TCG TCC TCG TCA TTC TCG TTG-3', size, 1,229 bp; TFDP-2, forward, 5'-CCC TGC ACC AGC AAT GGT TAC TCA G-3; reverse, 5'-ACT GCT GGA CTG GTG ACT GTT TGG G-3', size, 997 bp; G3PDH, forward, 5'-ACC ACA



GTC CAT GCC ATC AC-3, reverse, 5'-TCC ACC ACC CTG TTG CTG TA-3', size, 452 bp.

我们将DP家族各个成员TFDP-1、TFDP-2和TFDP-3 (HCA661) 在肿瘤细胞系尤其在HCC细胞系中的表达水平通过RT-PCR进行分析。使用G3PDH作为内参, 评估mRNA的质量。我们检测了TFDP-3 (HCA661) 在肝癌组织中的表达, 在这个实验中, 我们设计了如上所示的特异性引物对, 分别扩增TFDP-3 (HCA661), TFDP-1和TFDP-2基因的片段, 检测DP家族各成员在肿瘤细胞系尤其在HCC细胞系中的表达分布。TFDP-1和TFDP-2 mRNA表达于所有细胞系的细胞中, 而TFDP-3 (HCA661) mRNA的表达在不同的细胞系中有很大的差异。在SMMC-7721中表达相对较高的水平, Hep G2, HLE和Jurkat中则表达水平较低, 其余细胞系未检测到TFDP-3 (HCA661) mRNA的表达 (如图5中所示), 与TFDP-3 (HCA661) 在肝癌组织中的表达分布相一致。与TFDP-1和TFDP-2的广泛表达不同, TFDP-3 (HCA661) 的表达具有明显的组织细胞限制性。因此, TFDP-3 (HCA661) 具有不同于DP家族其它成员的表达谱。

#### 实施例5. TFDP-3 (HCA661) 蛋白与E2F蛋白在体外的相互作用

为了产生E2F-1~6的GST融合蛋白以利于蛋白纯化和后续的GST Pull-down分析, E2F基因构建入pGEX-4T2, 然后转化Escherichia coli (JM109), 通过传统方法获取和纯化大量的GST-E2F融合蛋白 (见图6A)。TFDP-3 (HCA661) 蛋白通过Promega的TNT<sup>®</sup> Quick Coupled Transcription/Translation system 试剂盒于50  $\mu$ l 反应体系中在放射性<sup>35</sup>S-methionine存在下体外转录和翻译来获取。体外结合反应中, 适宜量的纯化过的GST或GST-E2F融合蛋白先与谷胱苷肽-琼脂糖微珠反应, 然后加入含50 mM Tris (pH 8.0), 150 mM NaCl, 10 mg/ml lysozyme, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 50 mg/ml leupeptin, 50 mg/ml protease inhibitor, 50 mg/ml aprotinin和50 mM dithiothreitol的TFDP-3 (HCA661) 裂解缓冲液中。于4 $^{\circ}$ C作用2.5小时后, 离心并收集沉淀物, 用裂解缓冲液洗涤四次, 去除未结合的蛋白。加入SDS 上样缓冲液, 在12.5%的聚丙烯酰胺凝胶上电泳、干胶, 放射自显影。体外转录和翻译的TFDP-3 (HCA661) 蛋白作为阳性参照, 非融合的GST蛋白作为阴性参照。结果显示, TFDP-3 (HCA661) 能与E2F-1~E2F-6相互作用 (见图6B, lanes 3 to 8)。TFDP-3 (HCA661) 与E2F蛋白之间的相互作用是高效率的而且是特异性的, 因为TFDP-3 (HCA661) 与阴性参照GST蛋白之间无相互作用 (见图6B, lane 2)。这个体外结合数据支持TFDP-3 (HCA661) 是DP家族的一个新成员。

#### 实施例 6. TFDP-3 (HCA661) 蛋白与 E2F 蛋白在体内的相互作用

人宫颈癌细胞系 HeLa 细胞于 37 $^{\circ}$ C, 5% CO<sub>2</sub> 的细胞培养箱中培养, 所用培养基为

含 100 units/ml penicillin、100  $\mu\text{g/ml}$  streptomycin 和 10%新生小牛血清的 DMEM 培养基。转染前, HeLa 细胞用胰酶消化, 用无双抗的含 10%新生小牛血清的新鲜 DMEM 培养基将细胞稀释于 6 孔板中, 培养至细胞汇合度达到转染所需的 90-95%。转染时, 1.4  $\mu\text{g}$  的 HA-tagged E2F 质粒 DNA 和 1.4  $\mu\text{g}$  FLAG-tagged DP 质粒 DNA, 或 1.4  $\mu\text{g}$  的空载体 DNA, DNA 总量加至 2.8  $\mu\text{g}$ 。LipofectAMINE 2000 reagent (Invitrogen) 用来将 DNA 带入细胞内。HeLa 细胞转染 24-36 小时后, 按常规方法制备非变性细胞裂解物。这些蛋白的表达水平通过抗 HA 兔多克隆抗体或抗 FLAG M2 小鼠单克隆抗体的免疫印迹分析细胞裂解物来监控 E2F 蛋白 (如图 7 中图) 或 TFDP-3 (HCA661) 蛋白 (如图 7 下图) 的表达。为了执行免疫共沉淀分析, 转染的 HeLa 细胞先用 1x PBS 洗涤二次, 再加入适量的含 20 mM Tris (pH 7.5), 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 5  $\mu\text{g/ml}$  aprotinin, 5  $\mu\text{g/ml}$  leupeptin, and 2 mM PMSF 的裂解缓冲液裂解细胞, 制备非变性细胞裂解物。通过 Bradford 分析总蛋白的浓度。免疫共沉淀时, 细胞裂解物在 4 $^{\circ}\text{C}$  含 2  $\mu\text{g/ml}$  抗 FLAG 抗体和 25  $\mu\text{l}$  的 protein A-agarose 的细胞裂解缓冲液中至少作用 2 小时。离心后的沉淀用含 500mM NaCl 的细胞裂解缓冲液洗涤三次, 去除未结合的蛋白。洗涤离心后沉淀物加入上样缓冲液, 在 12.5%的聚丙烯酰胺凝胶上电泳、转膜。电转后的硝酸纤维膜先用含 5%脱脂奶粉的 TBS 封闭缓冲液封闭 1 小时, 然后用含 1  $\mu\text{g/ml}$  抗 HA 抗体的封闭液温育 1 小时, 最后用 HRP 偶联的抗兔 IgG 的二抗反应, ECL 显色。E2F 蛋白在所有转染了 E2F DNA 的免疫共沉淀物中检测到, 但未在只转染空载体 DNA 或 TFDP-3 (HCA661) DNA 免疫共沉淀物中检测出。这些数据表明 TFDP-3 (HCA661) 能有效率地与 E2F 家族中含有异二聚化结构域的成员相互作用并形成异二聚体, 进一步支持 TFDP-3 (HCA661) 是 DP 家族新成员的推测。

实施例7. TFDP-3 (HCA661) 与E2F蛋白相互作用形成的异二聚体的DNA结合能力分析 (EMSA, electrophoretic mobility shift assay)。

人宫颈癌细胞系HeLa细胞于37 $^{\circ}\text{C}$ , 5% CO<sub>2</sub>的细胞培养箱中培养, 所用培养基为含 100 units/ml penicillin、100  $\mu\text{g/ml}$  streptomycin和10%新生小牛血清的DMEM培养基。转染前, HeLa细胞用胰酶消化, 用无双抗的含10%新生小牛血清的新鲜DMEM培养基将细胞稀释于6孔板中, 培养至细胞汇合度达到转染所需的90-95%。转染时, 1.4  $\mu\text{g}$  的 E2F DNA和1.4  $\mu\text{g}$  DP DNA, 或1.4  $\mu\text{g}$ 的空载体DNA, DNA总量加至2.8  $\mu\text{g}$ 。

ipofectAMINE 2000 reagent (Invitrogen) 用来将DNA带入细胞内。HeLa细胞转染 24-36小时后, 按常规方法制备非变性细胞蛋白提取物, 等量的HeLa细胞蛋白提取物加入含有10 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES; pH7.9), 375 mM KCl, 12.5 mM MgCl<sub>2</sub>, 0.5 mM EDTA, 5 mM DTT, 15% Ficoll, 10 mg/ml BSA, and 1 mg/ml polydI/dC的 EMSA缓冲液中, 同时加入0.1 ng的放射性<sup>32</sup>P 同位素标记的含有特

异性野生型E2F结合位点的寡核苷酸探针(ATTTAAGTTTCGCGCCCTTTCCAA)在室温温育10~15分钟。反应混合物然后注入5%非变性聚丙烯酰胺凝胶在4℃恒压180 V电泳2.25小时,干胶,在-70℃放射自显影分析4-16小时。为了检测蛋白结合野生型DNA探针的特异性,反应液中加入20倍于野生型探针量的含有E2F突变位点的未标记的双链寡核苷酸探针作为竞争分子(ATTTAAGTTTCGatCCCTTTCTCAA)。

如图8所示,因为有内源性E2F基础水平的存在,单个E2F或DP转染时,它们的DNA结合能力仍处在基础水平(显影带2、3、6、9与1比较),表明单独转染E2F或DP并不能提高其DNA结合活性。当TFDP-1与E2F共转染时,E2F/TFDP-1异二聚体结合DNA的能力与单独转染E2F相比显著地增强了(显影带5与3、8与6、11与9相比)。相反,TFDP-3(HCA661)与E2F共转时,E2F/TFDP-3(HCA661)异二聚体的DNA结合能力并没有得到显著地提高(显影带4与3、7与6、10与9相比)。这表明,TFDP-3(HCA661)蛋白中可能存在着破坏DNA结合力的尚未鉴定的基序,从而导致E2F/TFDP-3异二聚体丧失结合E2F DNA识别位点。

#### 实施例 8. TFDP-3 (HCA661) 的亚细胞定位

SV40转化的绿猴肾细胞系 COS-7 细胞于 37℃, 5% CO<sub>2</sub> 的细胞培养箱中培养, 所用培养基为含 100 units/ml penicillin、100 μg/ml streptomycin 和 10%新生小牛血清的 DMEM 培养基。转染前, COS-7 细胞用胰酶消化, 用无双抗的含 10%新生小牛血清的新鲜 DMEM 培养基将细胞稀释至  $8 \times 10^4$  cells/ml, 各加入 1ml 于 24 孔板中, 培养至细胞汇合度达到转染所需的 90-95%。单独转染时, 280 ng 的 EGFP-tagged E2F 质粒 DNA 或 280 ng FLAG-tagged DP 质粒 DNA 与 280 ng 的空载体 DNA, DNA 总量加至 560 ng, 同时, pEGFP-N1 作为阴性参照。共同转染时, 280 ng 的 EGFP-tagged E2F 质粒 DNA 和 280 ng FLAG-tagged DP 质粒 DNA, DNA 总量加至 560 ng。LipofectAMINE 2000 reagent (Invitrogen) 用来将 DNA 带入 COS-7 细胞内。转染后 24 小时用 1xPBS 稍加洗涤, 然后加入-70℃预冷的甲醇在-20℃固定 20 分钟, 加入含 0.2% Triton X-100 的 1x PBS 通透细胞膜 4 分钟。用含 1% BSA 的 1x PBS 封闭液封闭非特异性抗体结合位点后, 固定的 COS-7 细胞在室温于含 1 μg/ml 抗 FLAG M2 单克隆抗体和 1% BSA 的 1x PBS 中温育 1 小时, 用 1x PBS 洗涤三次后, 加入含 1:100 稀释的 tetramethylrhodamine isothiocyanate (TRITC)偶联的山羊抗小鼠 IgG 的二抗和 1% BSA 的 1x PBS 于室温反应 1 小时。COS-7 细胞核用 10 μg/ml Hoechst 33342 逆染, E2F 采用 EGFP 标记和检测。通过配备了 CCD 的荧光显微镜采集免疫荧光图像, 所有图像通过 Adobe Photoshop 7.0.1 版处理。

实验结果表明, 当单独转染时, TFDP-3 (HCA661) 和 TFDP-1 均定位于 COS-7 细胞的细胞浆内, E2F-1 定位于细胞核内, 而 E2F-4 则定位于细胞浆内 (见图 9A)。

E2F-2、E2F-3 也定位于细胞核内，而 E2F-5 定位于细胞浆内（未显示）。

要发挥功能，DP 家族成员必须与 E2F 家族成员形成异二聚体。因此，我们分析了蛋白与蛋白之间的相互作用对 TFDP-3 (HCA661) 的亚细胞定位的影响。正如所推测的一样，TFDP-3 (HCA661) 与 E2F-1 (见图 9B)、E2F-2 或 E2F-3 (未显示) 共同转染时，与 TFDP-1 类似，均定位于细胞核内。与 E2F 的第一个亚家族的定位相反，当与 E2F-4 (见图 9B) 或 E2F-5 (未显示) 共同转染时，TFDP-3 (HCA661) 和 TFDP-1 仍然定位于细胞浆内。这些数据表明，TFDP-3 (HCA661) 的亚细胞定位与 TFDP-1 相一致。

#### 实施例 9. TFDP-3 (HCA661) 的报告基因分析

人正常肝细胞系 L02 于 37°C, 5% CO<sub>2</sub> 的细胞培养箱中培养，所用培养基为含 100 units/ml penicillin、100 µg/ml streptomycin 和 10% 新生小牛血清的 DMEM 培养基。转染前，L02 细胞用胰酶消化，用无双抗的含 10% 新生小牛血清的新鲜 DMEM 培养基将细胞稀释至  $1 \times 10^5$  cells/ml 于 24 孔板中各加入 1ml，并培养细胞使之汇合度达到转染所需的 90-95%。报告基因分析中，DNA 总量加至 560 ng，其中含 140 ng 的报告基因 6x E2F-firefly luciferase。执行 TFDP-3 (HCA661) 对 E2F/TFDP-1 介导的转录激活的抑制效应实验中，DNA 总量加至 1,120 ng。执行 TFDP-3 (HCA661) 或 TFDP-1 对内源性 E2F 活性的影响时，DNA 总量加至 840 ng。LipofectAMINE 2000 reagent (Invitrogen) 用来将 DNA 带入细胞内。L02 细胞转染 24 小时后先用 1xPBS 稍加洗涤，将残留的培养基去除，然后加入被动裂解缓冲液 (PLB)，室温裂解 15 分钟后，收获裂解液并离心，除去细胞碎片，收获上清，按照厂商提供的操作手册使用双特异性 luciferase 分析系统分析产物。为了消除转染效率的差异对报告基因 6x E2F-firefly luciferase 值的影响，每次转染都加入 140 ng 的 pRL-SV40 Renilla luciferase 质粒作为内参。

报告基因分析结果如图 10A 所示，表明 E2F 基因和 DP 基因单独转染时，140 ng 的 TFDP-1 对报告基因 6x E2F-firefly luciferase 产生了大约 2 倍的转录激活，同样，20 ng 的 E2F-1 或 E2F-3 也分别产生了 10 倍或 30 倍的转录激活，70 ng 的 E2F-4 或 140 ng E2F-5 也产生了 17 倍或 11 倍的转染激活。当 E2F 与 TFDP-1 共转染时，别构激活了报告基因 6x E2F-firefly luciferase，达到了 30 至 70 倍。然而，与 TFDP-1 形成鲜明对比的是，当 E2F 与 TFDP-3 (HCA661) 共转染时，观察到了 TFDP-3 (HCA661) 对 E2F 介导的转录激活产生了显著的抑制效应。同一实验至少重复三次 ( $P < 0.01$ )。

TFDP-3 (HCA661) 具有抑制 E2F 介导的转录激活能力，因此，我们想进一步分析 TFDP-3 (HCA661) 与 E2F 的相互作用是否能有效地抑制 E2F/TFDP-1 异二聚体介导的转录激活。实验中，E2F/TFDP-1 的各种组合与 0、140、280 或 560 ng 的 TFDP-3 (HCA661) 共转染 L02 细胞，采用 6x E2F-firefly luciferase 报告基因分析系统分析

TFDP-3 (HCA661) 对 E2F/TFDP-1 异二聚体的作用。

如上所述, 所有 E2F/TFDP-1 的组合显著地提高了 6x E2F-firefly luciferase 报告基因的转录活性。当 E2F/TFDP-1 组合在 TFDP-3 (HCA661) 存在时, E2F/TFDP-1 的转录活性显著地被抑制了, 而且是剂量依赖性的。甚至能完全阻断 TFDP-1 对 E2F 的别构激活 (如图 10B 所示)。这个发现表明, 尽管 TFDP-3 (HCA661) 是 DP 家族的一个新成员, 但它是作为一个抑制因子发挥功能的, 抑制 E2F 依赖性靶基因的转录调节。

最后, 我们也评估了 TFDP-3 (HCA661) 对内源性 E2F 活性的影响。如期所愿, FLAG-tagged TFDP-1 增强了 6x E2F-firefly luciferase 报告基因的转录活性, 而 FLAG-tagged TFDP-3 (HCA661) 则抑制了 6x E2F-firefly luciferase 报告基因的转录活性, 而且是剂量依赖性的 (见图 10C), 这表明, TFDP-3 (HCA661) 有能力阻断内源性 E2F 复合物对 6x E2F-firefly luciferase 报告基因的转录激活。而且, 我们也从用 E2F 的靶基因 Cyclin A 天然调控序列构建的 Cyclin A-firefly luciferase 报告基因中得到了相似的结果 (未显示)。

TFDP-3 (HCA661) 对 E2F 介导的转录的抑制能力与 TFDP-3 (HCA661) /E2F 异二聚体丧失结合特异性 E2F DNA 结合序列的能力相一致。而且, TFDP-3 (HCA661) 的抑制效应也表明, TFDP-3 (HCA661) 可能在 TFDP-3 (HCA661) 和 TFDP-1 共存时扮演着拮抗 TFDP-1 的角色。

总之, 所有上述结果表明, TFDP-3 (HCA661) 和 TFDP-1 之间的功能差异可能源于 E2F/TFDP-3 (HCA661) 异二聚体的 DNA 结合力的丧失, 抑制作用的分子基础应该存在于 TFDP-3 (HCA661) 分子中, 因为对某一个特定的 E2F 蛋白而言, 其转录活性的高低取决于由 E2F 蛋白和 DP 蛋白共同组成的完整的 DNA 结合结构域的 DNA 结合能力的强弱。

#### 实施例 10. TFDP-3 (HCA661) 独特功能产生的分子基础的探索

为了探索 TFDP-3 (HCA661) 与 E2F 蛋白形成的异二聚体的 DNA 结合能力的丧失以及 TFDP-3 (HCA661) 对 E2F 转录活性的抑制的分子基础, 我们构建了 TFDP-3 (HCA661) 和 TFDP-1 的替换突变体, 采用 6x E2F-luciferase 报告基因分析系统分析其突变体的转录性质, 间接评估 TFDP-3 (HCA661) 与 E2F 蛋白形成的异二聚体的 DNA 结合能力。TFDP-3 (HCA661) 和 TFDP-1 的替换突变体以相应质粒为模板, 通过 PCR 方法构建而成。先在替换位点分别设计正向突变引物和反向突变引物, 引物内含有相应的需要替换的碱基, 分别与 TFDP-3(HCA661)或 TFDP-1 ORF 的反向引物 (DP3CF-OA 或 DP1CF-OA)和正向引物(DP3CF-OS 或 DP1CF-OS)配对, 通过 PCR 合成两条小片段, 然后再通过连接 PCR 将两条小片段连接成一个完整的基因。

下列突变体以 pCDNA3-TFDP-3-FLAG and pCDNA3-TFDP-1-FLAG 为模板, 所用

的突变引物如下所示：TFDP-3M, forward, 5'-CGG CGC gtc TAC GAT GCC TTA AAC GTG-3'; reverse, 5'-ATC GTA gac GCG CCG TTT TAT GTT TTT C-3'; TFDP-1M, forward, 5'-CGG CGC acc TAC GAT GCC TTA AAC GTG-3'; reverse, 5'-ATC GTA ggt GCG CCG TCT TAT GTT TTT C-3'; TFDP-3MR, forward, 5'-aag ggc cta cgg cat ttc tcc atg aag gtc tgc gag aag GTG CAG AGG -3'; reverse, 5'-ctt ctc gca gac ctt cat gga gaa atg ccg tag gcc ctt GCC ATT CTT C-3'; TFDP-1MR, forward, 5'-atg ggc ctg tgc cgt ctt tcc atg aag gtc tgg gag acg GTG CAG AGG-3'; reverse, 5'-cgt ctc cca gac ctt cat gga aag acg gca cag gcc cat GCC ATT CTT C-3'. 3D-1H, forward, 5'-ggt ctg acc acc AAC TCG GCT CAG-3'; reverse, 5'-ctg agc cga gtt GGT GGT CAG ACC-3'; 1D-3H, forward, 5'-ggt ctg ccc acc AAC TCG GCT CAG-3'; reverse, 5'-ctg agc cga gtt GGT GGG CAG ACC-3'.

TFDP-3ΔH 和 TFDP-3ΔD、TFDP-1ΔH 和 TFDP-1ΔD 的结构域替换突变体以 3D-1H and 1D-3H 为模板, PCR 所用的突变引物如下: TFDP-3ΔH, forward, 5'-acc agc aag aag ACC GTC ATC AAC-3'; reverse, 5'-ggt gat gac ggt CTT CTT GCT GGT G-3'; TFDP-1ΔH, forward, 5'-agt agc aag aag ACG GTC ATC GAC-3'; reverse, 5'-gtc gat gac cgt CTT CTT GCT AC-3'; TFDP-3ΔD, forward, 5'-gga gag aag aat GGC AAG GGC CTA C-3'; reverse, 5'-tag gcc ctt gcc ATT CTT CTC TCC-3'; TFDP-1ΔD, forward, 5'-gga gag aag aat GGC ATG GGC CTG-3'; reverse, 5'-cag gcc cat gcc ATT CTT CTC TCC-3'.

TFDP-3ΔDH 以 TFDP-3ΔD 和 TFDP-3ΔH 为模板, 突变引物如下: forward, 5'-ggt ctg ccc acc AAC TCG GCT CAG-3'; reverse, 5'-ctg agc cga gtt GGT GGG CAG ACC-3'. TFDP-1ΔDH 以 TFDP-1ΔD 和 TFDP-1ΔH 为模板, 突变引物如下: forward, 5'-ggt ctg acc acc AAC TCG GCT CAG-3; reverse, 5'-ctg agc cga gtt GGT GGT CAG ACC-3'.

TFDP-3MR2~TFDP-3MR5 替换突变体以 pCDNA3-TFDP-3-FLAG 质粒为模板, 突变引物如下: TFDP-3MR2, forward, 5'-tac aac gaa gtc gca gac gag ctg gtt gcg gag ttc agt gct gcc gac AACC-3'; reverse, 5'-gtc ggc agc act gaa ctc cgc aac cag ctc gtc tgc cac ttc gtt gta GGAAG-3'; TFDP-3MR3, forward, 5'-atc tta cca aac gag tca gct tat gac cag aaa aac ata aga CGGCGC-3'; reverse, 5'-tct tat gtt ttt ctg gtc ata agc tga ctc gtt tgg taa gat GTGGTTG-3'; TFDP-3MR4, forward, 5'-aag gag aag aag gag atc aag tgg att ggt ctg ccc ACCAACTCG-3'; reverse, 5'-ggg cag acc aat cca ctt gat ctc ctt ctt ctc ctt GGAGATG-3'; TFDP-3MR5, forward, 5'-gca gac gag ctg gtt gcg gag ttc agt GCT GCC AGC-3'; reverse, 5'-act gaa ctc cgc aac cag ctc gtc CAC TTC CTG-3'.

下列突变体以 pCDNA3-TFDP-3-FLAG 质粒为模板, 突变引物如下: TFDP-3CY, forward, 5'-G ACC ACT TCC tac CAG GAA GTG-3'; reverse, 5'-CAC TTC CTG gta GGAAGT GGT C-3'; TFDP-3QN, forward, 5'-C ACT TCC TGC aac GAA GTG GTG-3'; reverse, 5'-CAC CAC TTC gtt GCA GGA AGT G-3'; TFDP-3VA, forward, 5'-CAG GAA

GTG gca GGC GAG CTG-3'; reverse, 5'-CAG CTC GCC tgc CAC TTC CTG-3';  
 TFDP-3GD, forward, 5'-GAA GTG GTG gac GAG CTG GTC-3'; reverse, 5'-GAC CAG  
 CTC gtc CAC CAC TTC-3'; TFDP-3KE, forward 5'-CTG GTC GCC gag TTC AGA  
 GC-3'; reverse, 5'-GC TCT GAA ctc GGC GAC CAG-3'; TFDP-3RS, forward, 5'-C AAG  
 TTC agt GCT GCC AGC AAC-3'; reverse, 5'-GTT GCT GGC AGC act GAA CTT G-3';  
 TFDP-3SD, forward, 5'-C AGA GCT GCC gac AAC CAC GC-3'; reverse, 5'-GC GTG  
 GTT gtc GGC AGC TCT G-3'.

所有突变体测序校正后, 经 Hind III-BamHI 限制性内切酶消化后, 克隆进 pCDNA3-FLAG 表达载体中, 图 11 显示了突变体的氨基酸序列。图 12A 显示了各个突变体的蛋白表达。通过 6xE2F-luciferase 报告基因分析系统分析其突变体对 E2F-4 转录活性的影响, 间接评估 TFDP-3 (HCA661) 与 E2F 蛋白形成的异二聚体的 DNA 结合能力。具体实验程序与实施例 9 中所述方法相同。

首先, 我们根据以前的研究结果分析了 TFDP-3 (HCA661) 与 TFDP-1 的 DNA 结合结构域的氨基酸组成的差异, 在 RRXYD DNA 识别模序中对第三个氨基酸进行了单个氨基酸替换, 产生两个替换突变体 (TFDP-3M: aa164T→V; TFDP-1M: aa169V→T), 分析该位氨基酸的非保守性替换是否是导致 TFDP-3 (HCA661) 的独特功能的原因。6xE2F-firefly luciferase 报告基因分析表明, 该位氨基酸的替换与 TFDP-3 (HCA661) 的功能无关。此外, 根据晶体结构, 在 DNA 结合结构域内, 另有 13 个氨基酸参与了与蛋白的异二聚化和 DNA 磷酸骨架的相互作用。其中, 在 TFDP-3 (HCA661) 中, 有三个氨基酸与 TFDP-1 和 TFDP-2 不同, 为此, 我们设计了一个 13 个氨基酸的替换 (TFDP-3MR: aa109-121 MGLCRLSMK~~V~~WET→KGLRHFSM~~K~~VCEK; TFDP-1MR: aa114-126 KGLRHFSM~~K~~VCEK→MGLCRLSMK~~V~~WET)。然而, 这些替换仍然不能逆转 TFDP-3 (HCA661) 的功能。

运用类似的序列替换分析, 我们分析了 TFDP-3 (HCA661) 和 TFDP-1 的全长氨基酸序列, 根据结构域的组成, 将其分成 4 个组成部分, 即氨基端未知功能区 (UNT)、DNA 结合结构域 (DBD)、异二聚化结构域 (HD) 和羧基端未知功能区 (UCT), 并构建了 3D-1H (含 TFDP-3 (HCA661) 的 UNT 和 DBD 和 TFDP-1 的 HD 和 UCT) 和 1D-3H (含 TFDP-1 的 UNT 和 DBD 和 TFDP-3 (HCA661) 的 HD 和 UCT), 报告基因分析表明, 1D-3H 产生的 E2F 介导的转录激活水平相当于野生型 TFDP-1 产生的水平, 不受 TFDP-3 (HCA661) 的 AA192-405 的影响。相反, 3D-1H 表现出对 E2F 介导的转录激活明显的抑制效应。这表明, TFDP-3 (HCA661) 独特功能产生的分子基础应该位于 TFDP-3 (HCA661) 的前半部分 (AA1-191)。

为了进一步分析功能产生的分子基础, 我们通过 PCR 方法从 3D-1H and 1D-3H 中衍生出了如下的替换突变体: TFDP-3ΔD (含 TFDP-1 的 DBD)、TFDP-1ΔD (含 TFDP-3

(HCA661)的DBD)、TFDP-3ΔH(含TFDP-1的HD)、TFDP-1ΔH(含TFDP-3(HCA661)的HD)、TFDP-3ΔDH(含TFDP-1的DBD和HD)、TFDP-1ΔDH(含TFDP-3(HCA661)的DBD和HD)。6xE2F-firefly luciferase 报告基因结果表明,TFDP-3ΔH和TFDP-1ΔH产生的转录抑制或激活水平分别与它们相应的野生型蛋白相似,相反,随着DNA结合结构域的替换,TFDP-3ΔDH和TFDP-3ΔD产生了类似TFDP-1的对E2F转录的别构激活,而TFDP-1ΔDH和TFDP-1ΔD则产生了类似TFDP-3(HCA661)的对E2F转录的抑制效应。这些数据表明,TFDP-3(HCA661)高效率地抑制E2F介导的转录激活的功能区位于TFDP-3(HCA661)的DNA结合结构域(见图12B)。

接下来,我们采用与TFDP-3MR和TFDP-3M构建方法相同的策略,将TFDP-3(HCA661)的DNA结合结构域中与TFDP-1有差异的其余13个氨基酸划分成三部分,分别用TFDP-1相应氨基酸序列替换TFDP-3(HCA661)序列,产生TFDP-3MR2(AA130-145 CQEVVGELVAKFRAAS→YNEVADELVAEFSAAD),TFDP-3MR3(AA148-161 ASPNESAYDVKNIK→ILPNESAYDQKNIR)和TFDP-3MR4(AA179-190 REKKKIKWIGLT→KEKKEIKWIGLP)。报告基因分析显示,TFDP-3MR3和TFDP-3MR4替换突变体没有提高E2F的转录活性,产生了与野生型TFDP-3(HCA661)相似的抑制效应。相反,TFDP-3MR2产生的转录激活水平与单独E2F转染时所产生的转染激活相似。该突变体也具有抑制E2F/TFDP-1转染激活的能力(未显示)。另一个在TFDP-3MR2内部衍生出的替换突变体TFDP-3MR5(AA134-142 VGELVAKFR→ADELVAEFS)对E2F介导的转录激活具有较弱的抑制作用。因此,TFDP-3(HCA661)DNA结合结构域的AA130-145序列是产生抑制效应所必需的(见图12C)。此外,为了寻找关键性氨基酸,我们在该区域内产生了另一组单个氨基酸替换突变体TFDP-3CY(AA130 C→Y),TFDP-3QN(AA131 Q→N),TFDP-3VA(AA134 V→A),TFDP-3GD(AA135 G→D),TFDP-3KE(AA140 K→E),TFDP-3RS(AA142 R→S)和TFDP-3SD(AA145 S→D),但结果表明,TFDP-3(HCA661)功能的产生可能是多个氨基酸共同作用的结果。

总之,报告基因分析结果表明,TFDP-3(HCA661)的DNA结合结构域(AA108-AA191)是导致TFDP-3(HCA661)与E2F蛋白形成的异二聚体丧失DNA结合能力的根本原因,从而抑制E2F的转录活性,与蛋白的其余部分的空间结构无关,包括异二聚化结构域。对DNA结合结构域的进一步分析表明,AA130-AA145(CQEVV GELVAK FRAAS)是TFDP-3(HCA661)对E2F的DNA结合产生抑制作用的必要条件。



sequence-listing.txt  
SEQUENCE LISTING

<110> 北京大学

<120> 一种肝癌-睾丸特异性抗原蛋白质和抗原肽

<130> FPI05135

<160> 2

<170> PatentIn version 3.1

<210> 1

<211> 405

<212> PRT

<213> Homo sapiens

<400> 1

Met Ala Lys Tyr Val Ser Leu Thr Glu Ala Asn Glu Glu Leu Lys Val  
1                   5                   10                   15

Leu Met Asp Glu Asn Gln Thr Ser Arg Pro Val Ala Val His Thr Ser  
20                   25                   30

Thr Val Asn Pro Leu Gly Lys Gln Leu Leu Pro Lys Thr Phe Gly Gln  
35                   40                   45

Ser Ser Val Asn Ile Asp Gln Gln Val Val Ile Gly Met Pro Gln Arg  
50                   55                   60

Pro Ala Ala Ser Asn Ile Pro Val Val Gly Ser Pro Asn Pro Pro Ser  
65                   70                   75                   80

Thr His Phe Ala Ser Gln Asn Gln His Ser Tyr Ser Ser Pro Pro Trp  
85                   90                   95

Ala Gly Gln His Asn Arg Lys Gly Glu Lys Asn Gly Met Gly Leu Cys  
100                   105                   110

Arg Leu Ser Met Lys Val Trp Glu Thr Val Gln Arg Lys Gly Thr Thr  
115                   120                   125

Ser Cys Gln Glu Val Val Gly Glu Leu Val Ala Lys Phe Arg Ala Ala  
130                   135                   140

Ser Asn His Ala Ser Pro Asn Glu Ser Ala Tyr Asp Val Lys Asn Ile  
145                   150                   155                   160

Lys Arg Arg Thr Tyr Asp Ala Leu Asn Val Leu Met Ala Met Asn Ile  
165                   170                   175

Ile Ser Arg Glu Lys Lys Lys Ile Lys Trp Ile Gly Leu Thr Thr Asn  
180                   185                   190

Ser Ala Gln Asn Cys Gln Asn Leu Arg Val Glu Arg Gln Lys Arg Leu

```

sequence—listing.txt
195                200                205
Glu Arg Ile Lys Gln Lys Gln Ser Glu Leu Gln Gln Leu Ile Leu Gln
210                215                220
Gln Ile Ala Phe Lys Asn Leu Val Leu Arg Asn Gln Tyr Val Glu Glu
225                230                235                240
Gln Val Ser Gln Arg Pro Leu Pro Asn Ser Val Ile His Val Pro Phe
245                250                255
Ile Ile Ile Ser Ser Ser Lys Lys Thr Val Ile Asn Cys Ser Ile Ser
260                265                270
Asp Asp Lys Ser Glu Tyr Leu Phe Lys Phe Asn Ser Ser Phe Glu Ile
275                280                285
His Asp Asp Thr Glu Val Leu Met Trp Met Gly Met Thr Phe Gly Leu
290                295                300
Glu Ser Gly Ser Cys Ser Ala Glu Asp Leu Lys Met Ala Arg Asn Leu
305                310                315                320
Val Pro Lys Ala Leu Glu Pro Tyr Val Thr Glu Met Ala Gln Gly Thr
325                330                335
Phe Gly Gly Val Phe Thr Thr Ala Gly Ser Arg Ser Asn Gly Thr Trp
340                345                350
Leu Ser Ala Ser Asp Leu Thr Asn Ile Ala Ile Gly Met Leu Ala Thr
355                360                365
Ser Ser Gly Gly Ser Gln Tyr Ser Gly Ser Arg Val Glu Thr Pro Ala
370                375                380
Val Glu Glu Glu Glu Glu Glu Asp Asn Asn Asp Asp Asp Leu Ser Glu
385                390                395                400
Asn Asp Glu Asp Asp
405
<210> 2
<211> 1218
<212> DNA
<213> Homo sapiens
<400> 2
atggcaaat atgtcagtct cactgaagct aacgaagaac tcaaggtctt aatggacgag 60
aaccagacca gccgccccgt ggccgttcac acctccaccg tgaacccgct cgggaagcag 120
ctcttgccga aaacctttgg acagtccagt gtcaacattg accagcaagt ggtaattggt 180
atgcctcaga gaccagcagc atcaaacatc cctgtggtag gaagcccaaa cccaccagc 240

```

## sequence—listing.txt

```

actcactttg cctctcagaa ccagcattcc tactcctcac ctctttgggc cgggcagcac 300
aacaggaaaag gagagaagaa tggcatgggc ctgtgccgtc tttccatgaa ggtctgggag 360
acggtgcaga gaaaggagc cacttcctgc caggaagtgg tgggcgagct ggtcgccaag 420
ttcagagctg ccagcaacca cgctcacca aacgagtcag cttatgacgt gaaaaacata 480
aaacggcgca cctacgatgc cttaaacgtg ctgatggcca tgaatatcat ctccagggag 540
aaaaagaaga tcaagtggat tggcttgacc accaactcgg ctcagaactg tcagaactta 600
cgggtggaaa gacagaagag acttgaaaga ataaagcaga aacagtctga acttcaacaa 660
cttattctac agcaaattgc tttcaagaac ctggtgctga gaaaccagta tgtggaggag 720
caggtcagcc agcggccgct gcccaactca gtcacccag tgcccttcat catcatcagc 780
agtagcaaga agaccgtcat caactgcagc atctccgacg acaaatcaga atatctgttt 840
aagttaaca gtcctttga aatccacgat gacacagaag tgctgatgtg gatgggcatg 900
acttttgggc tagagtccgg gagctgctct gccgaagacc ttaaaatgac cagaaatttg 960
gtcccaaagg ctctggagcc gtacgtgaca gaaatggctc agggaaacttt tggaggtgtg 1020
ttcacgacgg caggttccag gtctaattggc acgtggcttt ctgccagtga cctgaccaac 1080
attgcgattg ggatgctggc cacaagctcc ggtggatctc agtacagtgg ctccagggtg 1140
gagacccag cagtcgagga ggaagaggag gaggacaaca acgatgacga cctcagtgag 1200
aatgacgagg atgactga 1218

```

Met Ala Lys Tyr Val Ser Leu Thr Glu Ala Asn Glu Glu Leu Lys Val  
 1                    5                    10                    15  
 Leu Met Asp Glu Asn Gln Thr Ser Arg Pro Val Ala Val His Thr Ser  
                   20                    25                    30  
 Thr Val Asn Pro Leu Gly Lys Gln Leu Leu Pro Lys Thr Phe Gly Gln  
                   35                    40                    45  
 Ser Ser Val Asn Ile Asp Gln Gln Val Val Ile Gly Met Pro Gln Arg  
                   50                    55                    60  
 Pro Ala Ala Ser Asn Ile Pro Val Val Gly Ser Pro Asn Pro Pro Ser  
 65                    70                    75                    80  
 Thr His Phe Ala Ser Gln Asn Gln His Ser Tyr Ser Ser Pro Pro Trp  
                   85                    90                    95  
 Ala Gly Gln His Asn Arg Lys Gly Glu Lys Asn Gly Met Gly Leu Cys  
                   100                    105                    110  
 Arg Leu Ser Met Lys Val Trp Glu Thr Val Gln Arg Lys Gly Thr Thr  
                   115                    120                    125  
 Ser Cys Gln Glu Val Val Gly Glu Leu Val Ala Lys Phe Arg Ala Ala  
                   130                    135                    140  
 Ser Asn His Ala Ser Pro Asn Glu Ser Ala Tyr Asp Val Lys Asn Ile  
 145                    150                    155                    160  
 Lys Arg Arg Thr Tyr Asp Ala Leu Asn Val Leu Met Ala Met Asn Ile  
                   165                    170                    175  
 Ile Ser Arg Glu Lys Lys Lys Ile Lys Trp Ile Gly Leu Thr Thr Asn  
                   180                    185                    190  
 Ser Ala Gln Asn Cys Gln Asn Leu Arg Val Glu Arg Gln Lys Arg Leu  
                   195                    200                    205  
 Glu Arg Ile Lys Gln Lys Gln Ser Glu Leu Gln Gln Leu Ile Leu Gln  
                   210                    215                    220  
 Gln Ile Ala Phe Lys Asn Leu Val Leu Arg Asn Gln Tyr Val Glu Glu  
 225                    230                    235                    240  
 Gln Val Ser Gln Arg Pro Leu Pro Asn Ser Val Ile His Val Pro Phe  
                   245                    250                    255

Ile Ile Ile Ser Ser Ser Lys Lys Thr Val Ile Asn Cys Ser Ile Ser  
                   260                  265                  270  
 Asp Asp Lys Ser Glu Tyr Leu Phe Lys Phe Asn Ser Ser Phe Glu Ile  
                   275                  280                  285  
 His Asp Asp Thr Glu Val Leu Met Trp Met Gly Met Thr Phe Gly Leu  
                   290                  295                  300  
 Glu Ser Gly Ser Cys Ser Ala Glu Asp Leu Lys Met Ala Arg Asn Leu  
                   305                  310                  315                  320  
 Val Pro Lys Ala Leu Glu Pro Tyr Val Thr Glu Met Ala Gln Gly Thr  
                   325                  330                  335  
 Phe Gly Gly Val Phe Thr Thr Ala Gly Ser Arg Ser Asn Gly Thr Trp  
                   340                  345                  350  
 Leu Ser Ala Ser Asp Leu Thr Asn Ile Ala Ile Gly Met Leu Ala Thr  
                   355                  360                  365  
 Ser Ser Gly Gly Ser Gln Tyr Ser Gly Ser Arg Val Glu Thr Pro Ala  
                   370                  375                  380  
 Val Glu Glu Glu Glu Glu Glu Asp Asn Asn Asp Asp Asp Leu Ser Glu  
                   385                  390                  395                  400  
 Asn Asp Glu Asp Asp  
                   405

图 1

```

5-atggcaaat atgtcagtct cactgaagct aacgaagaac tcaaggtctt aatggacgag 60
aaccagacca gccgccccgt ggccgttcac acctccaccg tgaacccgct cgggaagcag 120
ctcttgccga aaacctttgg acagtccagt gtcaacattg accagcaagt ggtaattggg 180
atgcctcaga gaccagcagc atcaaacatc cctgtggtag gaagcccaa cccaccagc 240
actcactttg cctctcagaa ccagcattcc tactcctcac ctccctgggc cgggcagcac 300
aacaggaaag gagagaagaa tggcatgggc ctgtgccgtc ttccatgaa ggtctgggag 360
acgggtcaga gaaagggac cacttcctgc caggaagtgg tgggcgagct ggtcgccaag 420
ttcagagctg ccagcaacca cgctcacca aacgagtcag cttatgacgt gaaaaacata 480
aaacggcgca cctacgatgc cttaaacgtg ctgatggcca tgaatatcat ctccagggag 540
aaaaagaaga tcaagtggat tggctgacc accaactcgg ctcagaactg tcagaactta 600
cgggtggaaa gacagaagag acttgaaga ataaagcaga aacagtctga acttcaaaa 660
cttattctac agcaattgc tttcaagaac ctggtgctga gaaaccagta tgtggaggag 720
caggtcagcc agcggccgct gcccaactca gcatccacg tgcccttcat catcatcagc 780
agtagcaaga agaccgtcat caactgcagc atctccgacg acaaatcaga atatctgttt 840
aagtttaaca gtcctttga aatccacgat gacacagaag tgctgatgtg gatgggcatg 900
acttttgggc tagagtccgg gagctgctct gccgaagacc ttaaatggc cagaaattg 960
gtcccaaagg ctctggagcc gtacgtgaca gaaatggctc agggaacttt tggaggtgtg 1020
ttcacgacgg caggttccag gtctaattgg acgtggcttt ctgccagtga cctgaccaac 1080
attgcgattg ggatgctggc cacaagctcc ggtggatctc agtacagtgg ctccagggtg 1140
gagaccccag cagtcgagga ggaagaggag gaggacaaca acgatgacga cctcagttag 1200
aatgacgagg atgactga-3

```

1218

图 2

```

TFDF-1 MAKDAGLIRANGELKIVFDQNLSPGKGVSLVAVHPTVWPLGKQLLPKIFGQSSVHIAQ 60
TFDF-3 MAKYVSLIRAMKELKVLMDKQTSRP----VAVHTSTVWPLGKQLLPKIFGQSSVHIDQ 55
    ** . * ** * : * * : .      ***,***** ** *

TFDF-1 QVVIGTPQRPAASNTLVVGSPIHTPSTHWASQNPSSPWSAGKRRKREGEKNGGLEHNS 120
TFDF-3 QVVIGMPQRPAASNTLVVGSPIHTPSTHWASQNPSTSSPPWAGQRRKREGEKNGGLCRLS 115
    ***** ***** : ***** * ** * : ***** ** : *

    _____E2F DNA-binding domain_____
TFDF-1 MKVCEKVKRNGTTSYNEVADELVAKPSAADHGLPMSATDCKIKRRRYTRALNVLMAHH 180
TFDF-3 MKVWETVQRNGTTSQYVVGELVAKFRASHWASPTMSATDVKIKRRTYTRALNVLMAHH 175
    ** * ,***** ;* ,****,* ** * ***** ** :*,*****

    _____Heterodimerization domain_____
TFDF-1 IISREKKKLEWIGLPTMSAGCQNLVREKQKLEKRIKIQSGLQELIQQLAFKELVQRH 240
TFDF-3 IISREKKKLEWIGLPTMSAGCQNLVREKQKLEKRIKIQSGLQELIQQLAFKELVLRH 235
    ** : ** : ***** ,***** : ** * ,***** : ** : ***** **

TFDF-1 RVARQQASRPPPPSVINLPTIIVNTSMKTVIDCSISNMFYLPFDFNPFKIHDDIEVL 300
TFDF-3 QYVDEQVSRPLPISVINLPTIISSTMTVINCSISDNKSEYLPKFNSSPFKIHDDIEVL 295
    : : * , * : * ***** : ***** : ***** : ** * : * : ***** **

TFDF-1 KRGNACGLSESGCSAEDLEHARSLVPHALFPYVTEAAGTVGGVFTTAGSTSNGRFS 360
TFDF-3 MWGNTFGLSESGCSAEDLEHARMLVPHALFPYVTEAAGTFGGVFTT-AGSRSNGTWLS 354
    ** : ***** ,***** ,***** ** * ** * ** * : *

TFDF-1 ASDLTNGADGMLATSSNGSQYSGSRVETP-VSYVGHDDDDDDFHEKDEED 410
TFDF-3 ASDLTNIAIGMLATSSGGSQYSGSRVETPAVEEKEEDDDDDLSEKDEED 405
    ***** * ***** ,***** * , * : * : *****

```

图 3

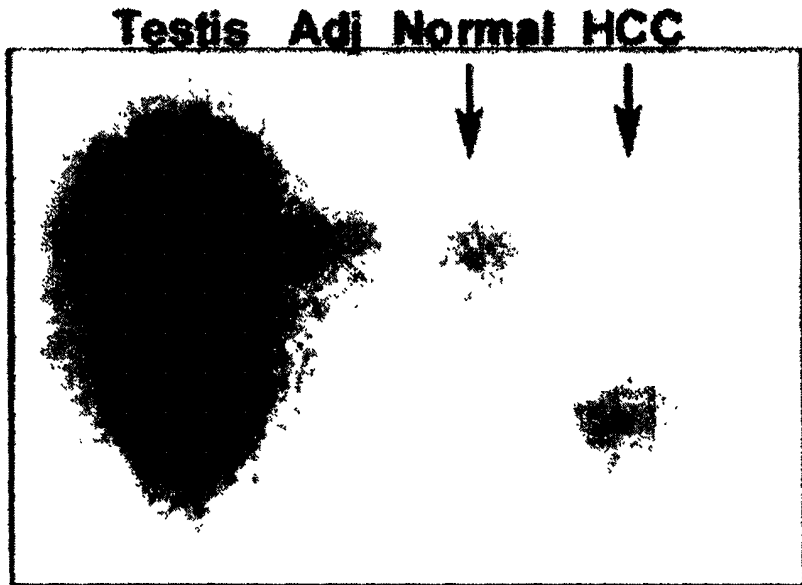


图 4



图 5

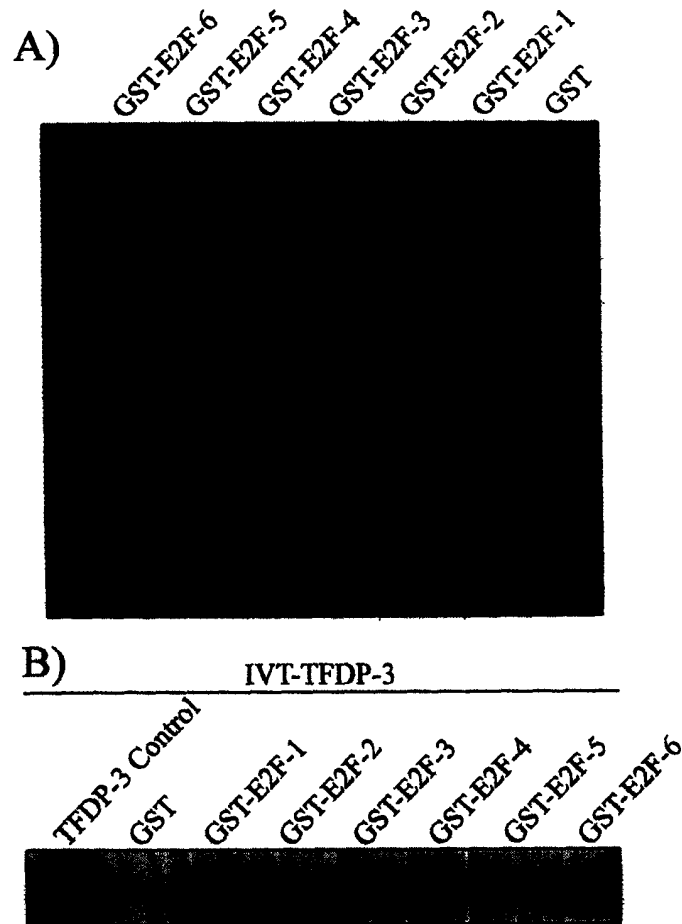


图 6



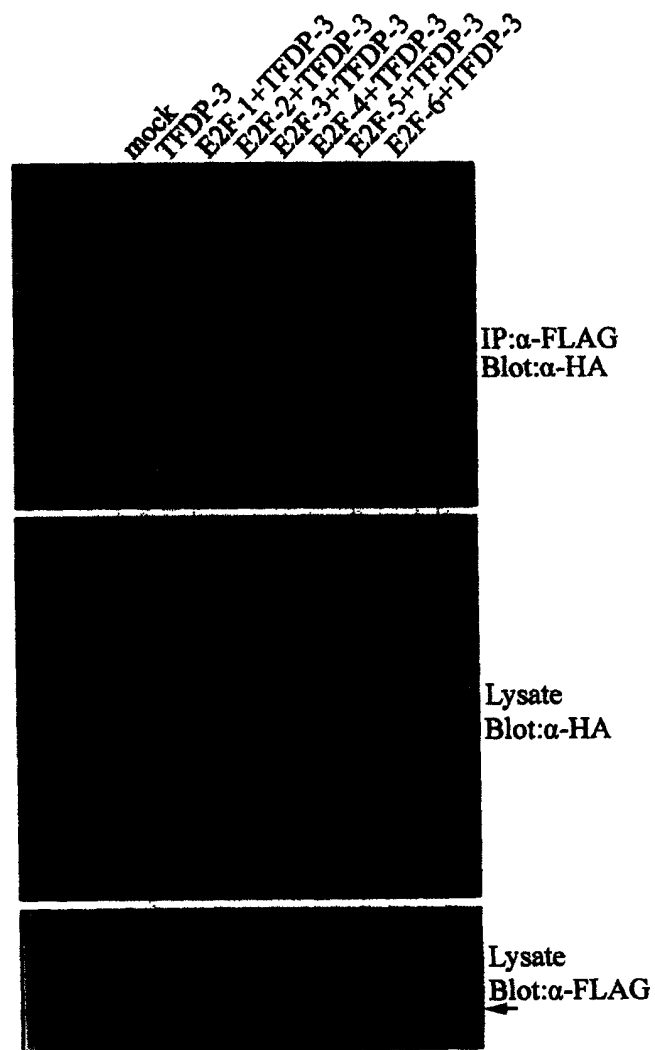


图 7



图 8

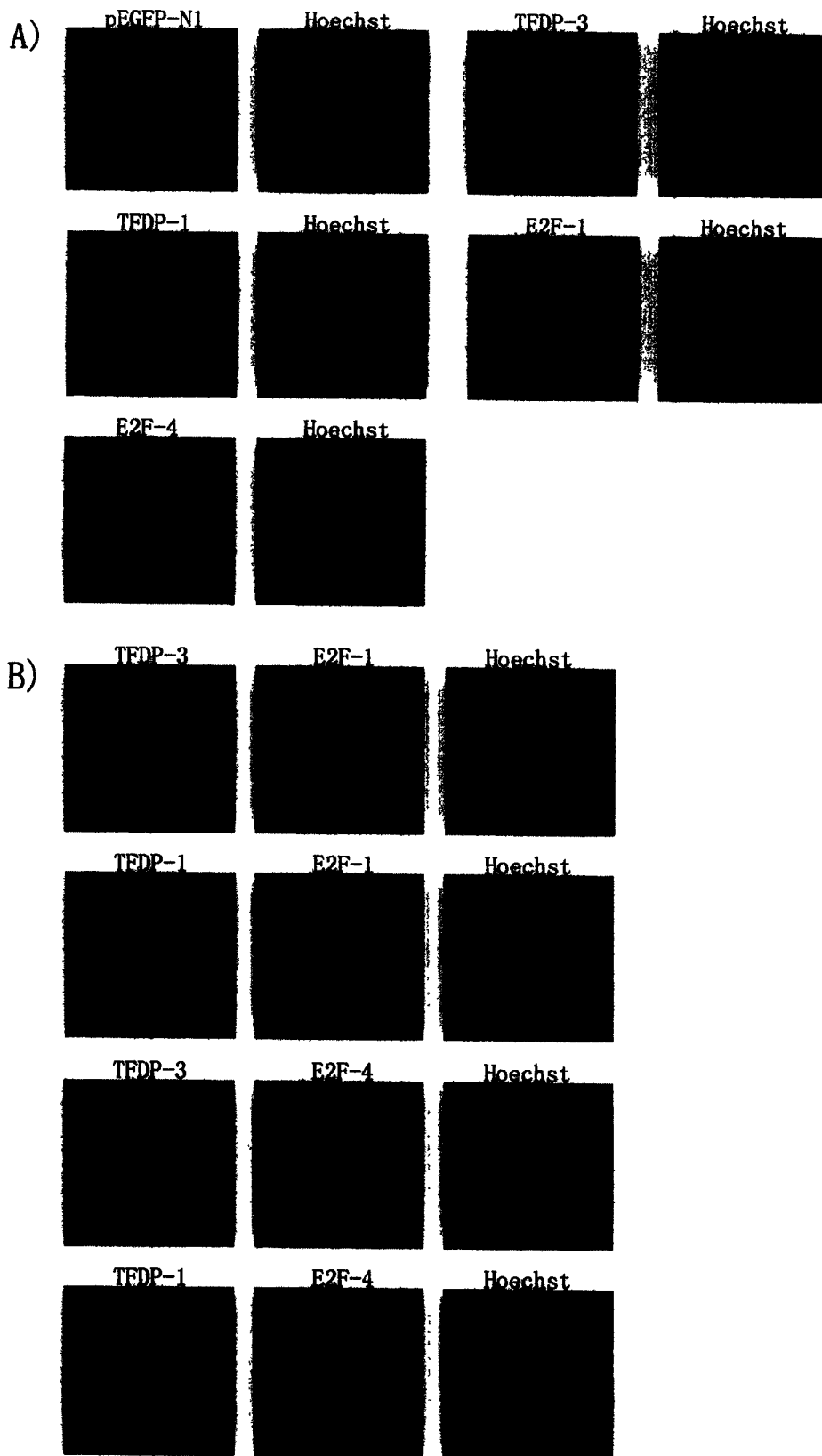


图 9

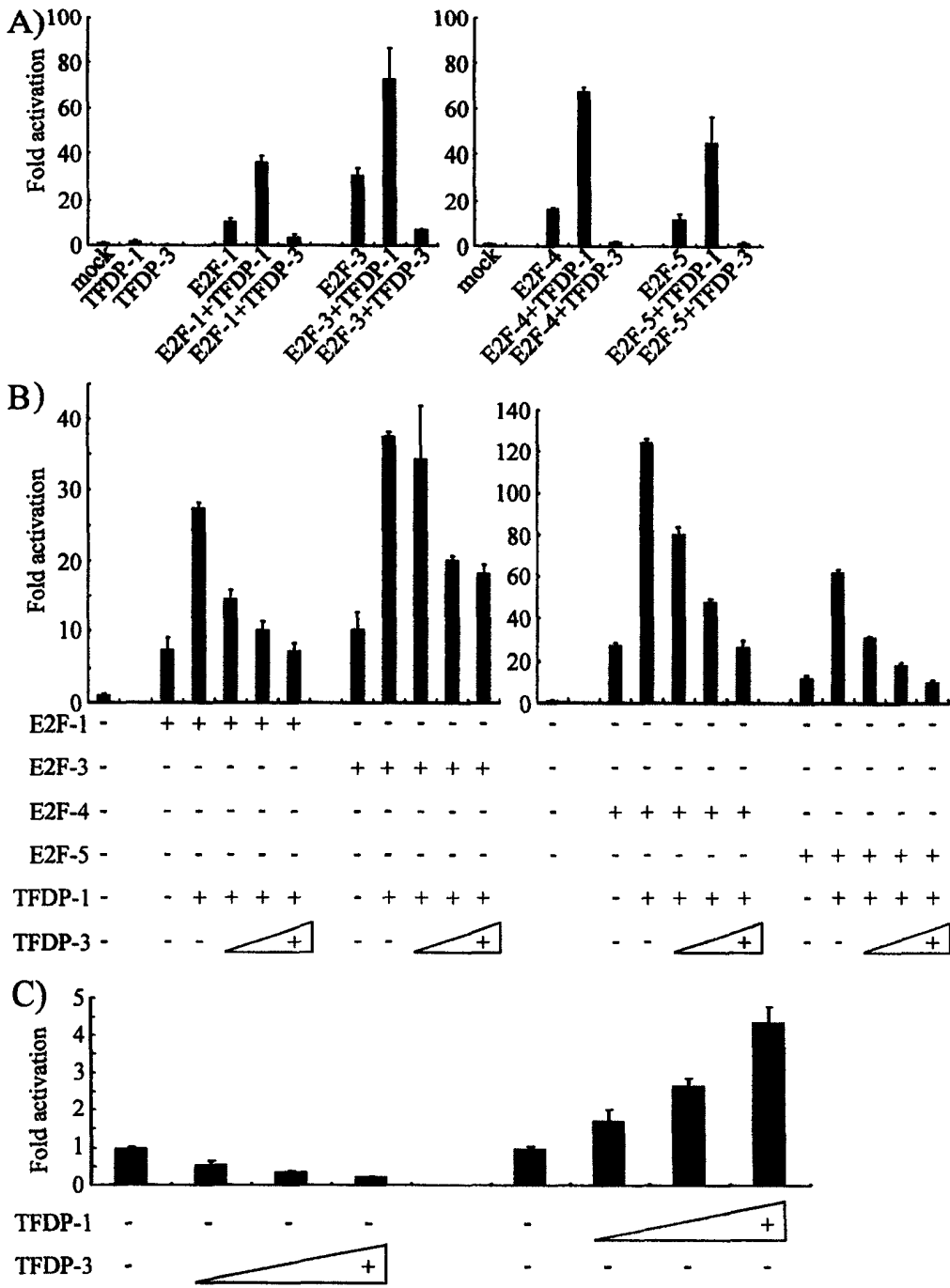


图 10



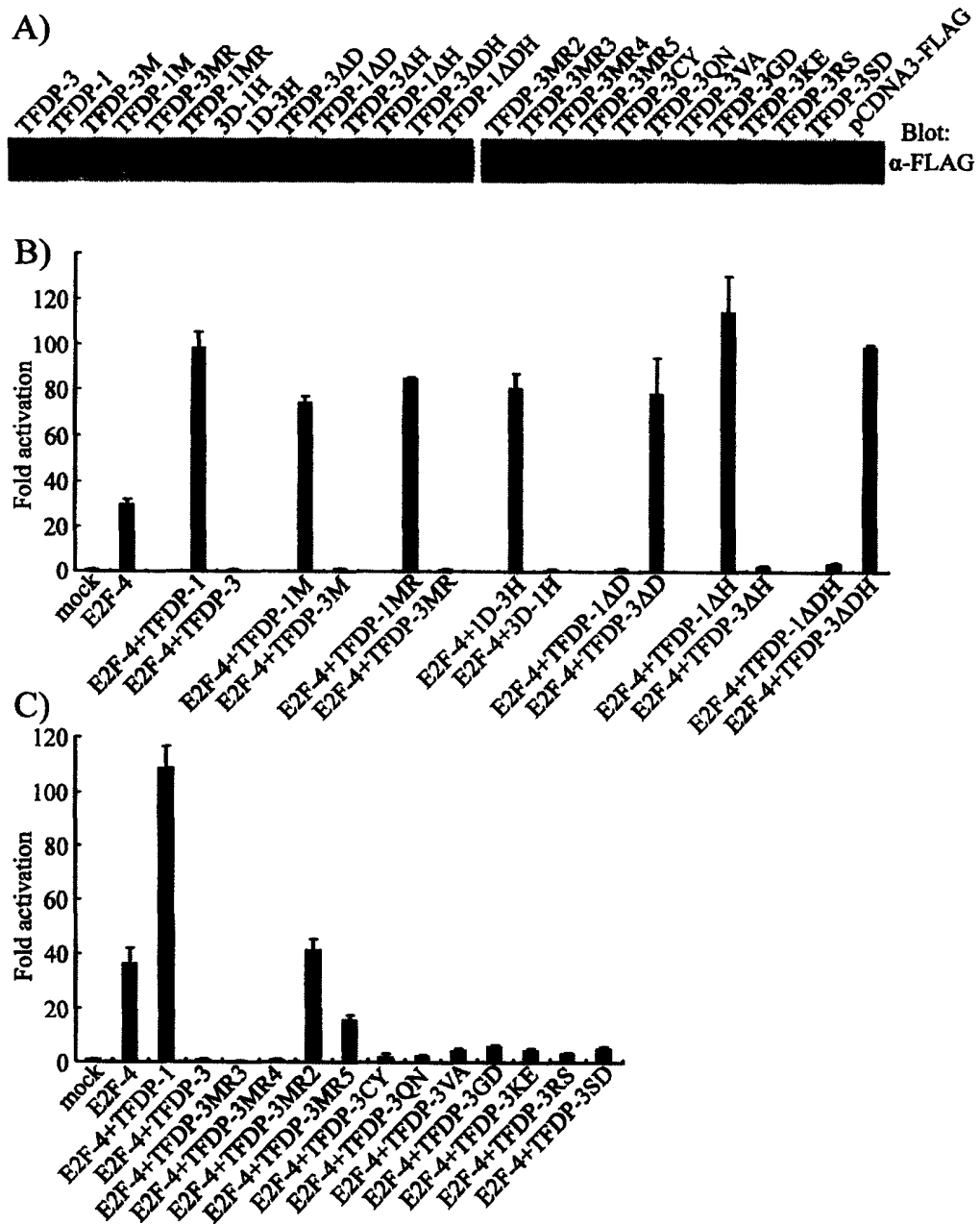


图 12