

· 研究原著 ·

文章编号 1000-2790(2007)14-1260-04

# 人宫颈癌基因蛋白 B 细胞表位及其 HLA 限制性细胞毒性 T 细胞表位预测分析

刘安定<sup>1</sup> 杨燕<sup>1</sup> 李方和<sup>1</sup> 陆蒙吉<sup>2</sup> 龚非力<sup>3</sup> 杨东亮<sup>1,4</sup> (华中科技大学同济医学院附属同济医院:<sup>1</sup> 实验医学研究中心,<sup>2</sup> 病原生物学系,<sup>3</sup> 免疫学系,<sup>4</sup> 临床免疫研究室 湖北 武汉 430030)

## Prediction of B cell epitopes and HLA restricted CTL epitopes of human cervical cancer oncogene protein

LIU An-Ding<sup>1</sup>, YANG Yan<sup>1</sup>, LI Fang-He<sup>1</sup>, LU Meng-Ji<sup>2</sup>, GONG Fei-Li<sup>3</sup>, YANG Dong-Liang<sup>1,4</sup>

<sup>1</sup>Experimental Medical Center, <sup>4</sup>Institute of Clinical Immunology Tongji Hospital, <sup>2</sup>Department of Pathogenic Biology, <sup>3</sup>Department of Immunology, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

**【Abstract】** AIM: To predict the secondary structure, the B cell epitopes and the HLA-A, B restricted T cell epitopes of human cervical cancer oncogene (HCCR) protein. **METHODS:** The secondary structure was predicted by the methods of Proteus and SOPMA. The hydrophilicity, surface probability, flexibility and antigenic index were predicted by the methods of Kyte-Doolittle, Emini, Karplus-Schultz and Jameson-wolf, respectively. According to the above methods, the B cell epitopes for HCCR protein were predicted. HLA-A \* 0201-restricted T cell epitopes were predicted by BIMAS, SYFPEITHI and NetCTL. And the restricted T cell epitopes of HLA-B and the other alleles of HLA-A were predicted by NetCTL. **RESULTS:** The secondary structure of HCCR1 protein was mainly composed of  $\alpha$ -helix. The B cell epitopes were probably located at or adjacent to the N-terminal No. 41 - 53, 216 - 228, 310 - 325 and 355 - 360 regions. And the five predominant HLA-A \* 0201-restricted T cell epitopes were YLVFLLMYL<sub>(152-160)</sub>, YLFPRQLL<sub>(159-167)</sub>, LLLHNVVLL<sub>(343-351)</sub>, CLFLGIISI<sub>(138-146)</sub> and SIPPFANYL<sub>(145-153)</sub>. The HLA-A, B restricted T cell epitopes of HCCR protein were probably located at extracellular domain of protein. **CONCLUSION:** Prediction of the epitopes of HCCR protein can provide a basis for production of the monoclonal antibody and development of some promising antigen peptides for tumor vaccines.

收稿日期 2007-01-26; 接受日期 2007-02-07

基金项目: 国家自然科学基金(30571646); 国家重大基础研究项目(973)(2005CB522901)

通讯作者: 杨东亮. Tel: (027) 83662894 Email: dlyang@tjh.tjmu.edu.cn

作者简介: 刘安定. 硕士. Tel: (027) 83663659 Email: liuandingtjmu@163.com

**【Keywords】** cervix neoplasms; oncogene proteins; epitopes; B-Lymphocyte; CTL epitopes

**【摘要】** 目的: 预测人宫颈癌基因(human cervical cancer oncogene, HCCR)蛋白的二级结构, B 细胞表位及其 HLA-A, B 限制性细胞毒性 T 细胞表位. 方法: 综合分析二级结构、亲水性、柔韧性、表面可及性与抗原性指数, 预测 HCCR 蛋白的 B 细胞抗原表位, 利用 BIMAS, SYFPEITHI 和 NetCTL 方法预测分析其 HLA-A \* 0201 限制性 CTL 表位, 运用 NetCTL 方法对 HLA-A 的其他等位基因和 HLA-B 限制性 CTL 表位进行预测分析. 结果: HCCR 蛋白的二级结构主要由  $\alpha$ -螺旋结构组成, B 细胞优势表位位于 N 端第 41 ~ 53, 216 ~ 228, 310 ~ 325 和 355 ~ 360 区段; 预测得到 5 个 HLA-A \* 0201 限制性 CTL 优势表位分别为 YLVFLLMYL<sub>(152-160)</sub>, YLFPRQLL<sub>(159-167)</sub>, LLLHNVVLL<sub>(343-351)</sub>, CLFLGIISI<sub>(138-146)</sub> 和 SIPPFANYL<sub>(145-153)</sub>, HCCR 蛋白 HLA-A, B 限制 CTL 表位主要位于胞外区. 结论: 应用多参数预测 HCCR 蛋白 B 细胞表位及其 HLA-A, B 限制性细胞毒性 T 细胞表位, 为进一步实验鉴定其表位进而制备单克隆抗体和基于 HCCR 抗原的肿瘤免疫学治疗奠定了基础.

**【关键词】** 宫颈肿瘤 癌基因蛋白质类 表位 B 淋巴细胞 CTL 表位  
【中图分类号】R392.1 【文献标识码】A

## 0 引言

人宫颈癌基因(human cervical cancer oncogene, HCCR)是近年来发现的与多种肿瘤相关的基因, HCCR 可能作为抑癌基因 p53 的负调控因子, 引起肿瘤的发生<sup>[1]</sup>. 其编码的蛋白可以作为一种重要的肿瘤标志物, 在监测和诊断肝细胞癌、乳腺癌、结肠直肠癌等肿瘤方面具有重要的意义, 是一种极为灵敏的指标<sup>[2-4]</sup>. HCCR 蛋白在肿瘤治疗中可以作为一个重要的靶位点<sup>[5]</sup>. 但到目前为止, 人们对其蛋白特性及其表位生物学了解甚少. 我们根据人 HCCR 蛋白质的氨基酸序列, 采用免疫信息学, 在方法学比较的基础上, 联合运用多种方法对 HCCR 蛋白的二级结构、B 细胞表位及其 HLA-A, B 限制性细胞毒性 T 细胞表位进行预测分析. 旨在为进一步鉴定其表位, 研制单克隆抗体和基于 HCCR 抗原的肿瘤免疫学治疗奠定基础.

# 1 材料和方法

1.1 材料 HCCR 蛋白的氨基酸全长序列从 NCBI 网站上获取 (GenBank accession no. AAK34885), 含有 360 个氨基酸 相对分子质量为 41 803.

## 1.2 方法

1.2.1 HCCR 蛋白二级结构预测 采用 Proteus<sup>[6]</sup>, SOPMA<sup>[7]</sup> 蛋白质结构预测服务器分别预测 HCCR 蛋白的二级结构.

1.2.2 HCCR 蛋白 B 细胞抗原表位的多参数预测 按 Kyte-Doolittle 方案预测氨基酸亲水性; Emini 方案预测蛋白质的表面可及性; Karplus-Schultz 方案预测蛋白质的柔韧性; Jameson-Wolf 方案预测蛋白质的抗原指数.

1.2.3 B 细胞表位的综合分析 综合比较分析上述方法 辅以对 HCCR 蛋白的二级结构的分析, 排除二级结构位于  $\alpha$ -螺旋和  $\beta$ -折叠内不易形成表位的序列, 预测 HCCR 蛋白的 B 细胞表位.

1.2.4 HCCR 抗原 HLA-A \* 0201 限制性细胞毒性 T 细胞表位的预测 运用 BIMAS, SYFPEITHI 和 NetCTL<sup>[8]</sup> 预测服务器, 分别对 HCCR 蛋白 HLA-A \* 0201 限制性细胞毒性 T 细胞表位进行预测, 然后按文献<sup>[9]</sup>的方法, 同时结合 NetCTL 预测结果, 综合分析 HCCR 蛋白 HLA-A \* 0201 限制性细胞毒性 T 细胞

表位.

1.2.5 HCCR 抗原 HLA-A 的其他等位基因和 HLA-B 限制性 CTL 表位的预测 运用 NetCTL 在线预测服务器对 HCCR 抗原 HLA-A 的其他等位基因和 HLA-B 限制性 CTL 表位进行预测和分析.

# 2 结果

2.1 HCCR 蛋白的二级结构 Proteus 分析结果显示其二级结构以  $\alpha$ -螺旋为主, 占 73%, 且多集中在中间区域, C 末端次之. 无规卷曲占 27%, 主要分布在 N 端的第 14 ~ 21, 30 ~ 50, 115 ~ 121, 162 ~ 172, 221 ~ 228, 239 ~ 250, 263 ~ 270, 311 ~ 318 和 355 ~ 360 区域; 未预测到有  $\beta$ -折叠结构. 而 SOPMA 方法预测结果为 HCCR 蛋白 N 端第 13 ~ 52, 66 ~ 85, 111 ~ 124, 139 ~ 148, 160 ~ 175, 221 ~ 229, 263 ~ 272, 310 ~ 319, 349 ~ 360 区段主要由构成蛋白质柔性区域的无规卷曲结构、 $\beta$ -转角结构和伸展结构所组成, 其他区域主要由  $\alpha$ -螺旋结构组成, 两种方法所预测的结果基本一致. 综合以上两种方法的预测结果, 认为 HCCR 蛋白以  $\alpha$ -螺旋结构出现的区域最多, 其柔性区域即功能区主要位于 N 端第 14 ~ 21, 30 ~ 50, 111 ~ 123, 162 ~ 172, 221 ~ 228, 263 ~ 272, 310 ~ 319 和 349 ~ 360 区段 (图 1).



C 无规卷曲区域;  $\beta$   $\beta$ -折叠区域; h  $\alpha$ -螺旋区域; } 转角区域. 图 1 根据 SOPMA 方法预测的 HCCR 蛋白二级结构

2.2 多参数预测 HCCR 蛋白表位 HCCR 蛋白亲水性、表面可及性、柔韧性和抗原性指数单参数预测结果如图 2 ~ 5, 其中高于阈值的肽段即为预测的抗原表位. 将各种参数预测的可能性表位见表 1.

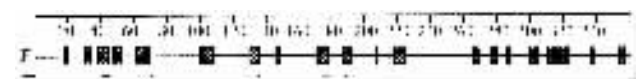


图 4 HCCR 蛋白骨架的柔韧性分析

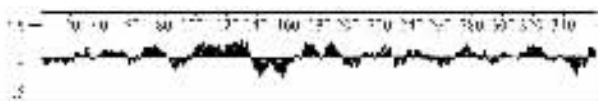


图 2 HCCR 蛋白的亲水性分析



图 5 HCCR 蛋白的抗原指数性分析

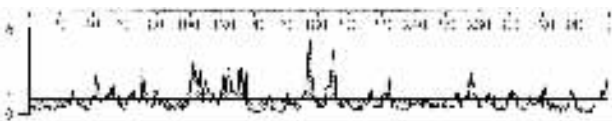


图 3 HCCR 蛋白的表面可及性分析

2.3 HCCR 蛋白 B 细胞抗原表位的综合预测 通过联合预测亲水性、表面可及性、柔韧性和抗原性指数的方法分析 HCCR 蛋白潜在的抗原决定簇, 结果显示在 N 端第 41 ~ 53, 63 ~ 83, 99 ~ 112, 121 ~ 139,

172 ~ 193 ,221 ~ 228 ,274 ~ 280 ,310 ~ 325 和 355 ~ 360 区段内或者附近存在着抗原表位 ,根据蛋白质的二级结构分析结果显示 ,在第 101 ~ 109 ,130 ~ 136 ,172 ~ 193 和 275 ~ 280 区段主要是作为骨架起稳定作用的  $\alpha$ -螺旋结构所组成 ,形成表位的可能性比较

小。综合比较分析 ,N 端第 41 ~ 53 ,216 ~ 228 ,310 ~ 325 和 355 ~ 360 区段满足亲水性、可及性、柔韧性和抗原性 ,在二级结构上主要是由无规卷曲结构组成 ,最可能为 HCCR 蛋白的优势 B 细胞表位。

表 1 HCCR 蛋白优势 B 细胞表位的综合比较

预测方案	预测表位
亲水性	23 ~ 32 ,39 ~ 57 ,65 ~ 83 ,98 ~ 138 ,171 ~ 195 ,210 ~ 229 ,239 ~ 248 ,271 ~ 282 ,314 ~ 328 ,355 ~ 360
表面可及性	41 ~ 53 ,63 ~ 82 ,100 ~ 136 ,172 ~ 193 ,221 ~ 229 ,275 ~ 282 ,315 ~ 325 ,355 ~ 360
柔韧性	31 ~ 53 ,62 ~ 70 ,100 ~ 109 ,130 ~ 136 ,171 ~ 178 ,187 ~ 193 ,219 ~ 226 ,266 ~ 281 ,311 ~ 323 ,355 ~ 360
抗原性指数	26 ~ 55 ,64 ~ 82 ,99 ~ 110 ,121 ~ 140 ,172 ~ 195 ,213 ~ 229 ,269 ~ 281 ,287 ~ 293 ,300 ~ 328 ,337 ~ 343 ,355 ~ 360
二级结构	14 ~ 21 ,30 ~ 50 ,111 ~ 123 ,162 ~ 172 ,221 ~ 228 ,263 ~ 272 ,310 ~ 319 ,349 ~ 360
综合分析	41 ~ 53 ,216 ~ 228 ,310 ~ 325 ,355 ~ 360

**2.4 HCCR 抗原 HLA-A \* 0201 限制性细胞毒性 T 细胞表位的预测** 将 BIMAS 和 SYFPEITHI 的综合预测结果结合 NetCTL( 阈值设为 0.80 )预测结果进行综合分析 ,获取评分显著高与其他的 5 条 HCCR 抗原 HLA-A \* 0201 限制性细胞毒性 T 细胞候选表位( 表 2 ) ,分别为 YLVFLLMYL<sub>(152~160)</sub> , YLFPRQLLI<sub>(159~167)</sub> , LLLHNVVLL<sub>(343~351)</sub> , CLFLGIISI<sub>(138~146)</sub> 和 SIPPFA-NYL<sub>(145~153)</sub>。 **2.5 HCCR 抗原 HLA-A 的其他等位基因和 HLA-B 限制性 CTL 表位的预测** 利用远程预测 ,登陆 NetCTL ,对 A1 ,A3 ,A24 ,A26 ,B7 ,B8 ,B2 ,B39 ,B44 ,B58 和 B62 限制性 CTL 表位的预测 ,选取综合评分前 5 名的序列作为优势候选表位( 表 3 ) ,分析发现 HCCR 蛋白 HLA-A ,B 限制 CTL 表位主要位于蛋

白的胞外区。

表 2 BIMAS 结合 SYFPEITHI 和 NetCTL 分别预测 HLA-A \* 0201 限制性 CTL 表位

位置	序列	SYFP-EITHI	BIMAS	总分	NetCTL 综合评分
152	YLVFLLMYL	26	1 243	32 320	1.4202
343	LLLHNVVLL	30	309	9 270	1.1598
159	YLFPRQLLI	24	178	4 272	1.4448
138	CLFLGIISI	27	89	2 403	0.9455
145	SIPPFANYL	24	61	1 464	0.8134

总分 = ProPred-I 和 SYFPEITHI 各自评分之乘积。

表 3 NetCTL 预测 HCCR 蛋白 HLA-A ,B 限制性 CTL 表位

等位基因	位置	序列	位置	序列	位置	序列	位置	序列	位置	序列
A1	189	QSHPEIISY	63	KTKAINGKY	176	QTDFLDIYH	300	AQEVKSACY	175	QQTDFLDIY
A3	57	MSYVVTKTK	166	LIRHFWTPK	55	NLMSYVVTK	63	KTKAINGKY	251	ALHVKALSR
A24	178	DFLDIYHAF	81	RFYILYTIF	324	TWLGEWLQI	328	EWLQISCSL	196	SYLEKVIPL
A26	52	DVKNLMSYV	193	EISYLEKV	347	NVLLSTNY	114	NIKFHQLPY	63	KTKAINGKY
B7	39	APRSSKLHL	5	RVCWARSAV	259	RAMLLTSYL	120	LPYREMEHL	102	KARRIKTNM
B8	253	HVKALSRAM	275	RLKTHTTVI	282	VIHQDKAL	27	RLQLGRSGL	59	YVVTKTKAI
B27	104	RRIKTNMWK	129	RQFRQDVTK	258	SRAMLLTSY	274	HRLKTHTTV	284	HQLDKALAK
B39	77	RHFPRFYIL	278	THTTVIHQL	225	THPAIHDL	131	FRQDVTKCL	241	NHPLGMNQL
B44	236	RECFSNHPL	301	QEVKSACYL	318	GEDRCRTWL	337	KEAELSLLL	123	REMEHLRQF
B58	204	LISDAGLRW	15	GSAVTPGHF	144	ISIPPFANY	89	FMKGLQMLW	29	QLGRSGLAW
B62	189	QSHPEIISY	300	AQEVKSACY	144	ISIPPFANY	63	KTKAINGKY	249	LQALHVKAL

### 3 讨论

HCCR 是近年发现的与人类多种肿瘤相关的基因,在宫颈癌、肝细胞癌、乳腺癌、白血病等肿瘤中呈过量表达<sup>[1]</sup>。HCCR 蛋白在肿瘤免疫治疗中可以作为一个重要的靶位点<sup>[5]</sup>,预测其抗原表位,可为制备 HCCR 特异性单克隆抗体、HCCR 诊断试剂的开发和针对 HCCR 抗原的肿瘤免疫预防和免疫学治疗奠定基础。

根据 Kyte-Doolittle 预测氨基酸亲水性方案、Emini 预测蛋白质的表面可及性方案、Jameson-Wolf 预测蛋白质的抗原指数方案分别对 HCCR 蛋白可能的 B 细胞表位进行了预测,综合各单参数预测的结果,发现 N 端第 41~53、63~83、99~112、121~139、172~193、221~228、274~280、310~325 和 355~360 区段具有较好的亲水性、可及性、柔韧性和抗原性指数。蛋白质的二级结构也是预测抗原表位的重要信息, $\alpha$ -螺旋、 $\beta$ -折叠等二级结构的化学键键能比较高,不易形成抗原表位序列,而蛋白质二级结构中  $\beta$ -转角及无规卷曲结构,往往突出在蛋白的表面,从而构成蛋白质的功能区,因此该区域内常含有 B 细胞的优势抗原表位。因此得到单参数的预测结果后,还要结合二级结构的预测结果来排除不易形成表位的序列。综合亲水性、可及性、柔韧性和抗原指数和二级结构预测的结果,最终确定了 HCCR 蛋白的潜在 B 细胞优势表位。

HCCR 蛋白在多种肿瘤中过度表达,可以作为免疫治疗的一个重要靶点,研究 HCCR 蛋白序列的 CTL 表位可为肿瘤免疫治疗开辟新的途径。寻找和鉴定 CTL 识别的肿瘤特异性抗原肽是肿瘤免疫治疗的关键,而利用免疫信息学进行抗原表位的预测是重要的手段之一。BIMAS 和 SYFPEITHI 是最常用的两大 CTL 表位预测程序。由于数据库和算法的差异,将导致预测结构存在差异。将两种预测程序结合起来,可以提高表位预测的准确率<sup>[9]</sup>。由于 BIMAS 和 SYFPEITHI 都是基于抗原肽与 MHC I 类结合特性而建立起来的方法,未考虑针对抗原加工处理过程的预测。而蛋白酶体的裂解和 TAP 转运对 CTL 表位的形成至关重要。NetCTL 是近年来 Larsen MV 等建立起来 CTL 表位的整合预测方法,包括对抗原肽与 MHC I 类分子结合特性、TAP 转运效率、蛋白酶体裂解基序的综合预测,其综合性能要比目前公认的两大预测程序 BIMAS、SYFPEITHI 要高<sup>[8]</sup>。我们将 BIMAS、SYFPEITHI 和 NetCT 预测的结果进行综合分析,结果发

现联合 BIMAS 和 SYFPEITHI 所得到的高分序列,同样在 NetCTL 预测中具有很高的评分。综合分析以上分析,获取 5 条 HCCR 抗原 HLA-A\*0201 限制性细胞毒性 T 细胞优势候选表位。同时,通过 NetCTL 对 HLA-A 的其他等位基因和 HLA-B 限制性 CTL 表位进行了预测和分析,可为构建 HLA 不同等位基因限制的、基于混合 T 细胞表位的肿瘤治疗性疫苗提供理论依据。

本研究通过生物信息学方法对 HCCR 蛋白二级结构 B 细胞表位及其 HLA-A、B 限制性细胞毒性 T 细胞表位进行了预测分析,为进一步研究该蛋白单克隆抗体的制备奠定了理论基础,也为以 HCCR 蛋白作为靶分子的肿瘤免疫治疗提供了新的途径,其预测的结果有待进一步的实验加以验证。

### 【参考文献】

- [1] Ko J, Lee YH, Hwang SY, et al. Identification and differential expression of novel human cervical cancer oncogene HCCR-2 in human cancers and its involvement in p53 stabilization [J]. *Oncogene*, 2003, 22(30):4679-4689.
- [2] Yoon SK, Lim NK, Ha SA, et al. The human cervical cancer oncogene protein is a biomarker for human hepatocellular carcinoma [J]. *Cancer Res*, 2004, 64(15):5434-5441.
- [3] Jung SS, Park HS, Lee IJ, et al. The HCCR oncoprotein as a biomarker for human breast cancer [J]. *Clin Cancer Res*, 2005, 11(21):7700-7708.
- [4] Shin SM, Chung YJ, Oh ST, et al. HCCR-1-interacting molecule "deleted in polyposis 1" plays a tumor-suppressor role in colon carcinogenesis [J]. *Gastroenterology*, 2006, 130(7):2074-2086.
- [5] Chung YJ, Kim JW. Novel oncogene HCCR: Its diagnostic and therapeutic implications for cancer [J]. *Histol Histopathol*, 2005, 20(3):999-1003.
- [6] Montgomerie S, Sundararaj S, Gallin WJ, et al. Improving the accuracy of protein secondary structure prediction using structural alignment [J]. *BMC Bioinformatics*, 2006, 7:301.
- [7] Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments [J]. *Comput Appl Biosci*, 1995, 11(6):681-684.
- [8] Larsen MV, Lundegaard C, Kasper L, et al. An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions [J]. *Eur J Immunol*, 2005, 35(8):2295-2303.
- [9] Panagiotopoulos C, Qin H, Tan R, et al. Identification of a  $\beta$ -cell-specific HLA class I Restricted epitope in type 1 diabetes [J]. *Diabetes*, 2003, 52(11):2647-2651.