

DOI:10.3724/SP.J.1008.2010.00822

肿瘤相关蛋白 EIF4G1 亚型的 B 细胞表位预测

曹廷虎¹, 常文军², 胡锦涛¹, 朱 军¹, 曹广文^{2*}

1. 上海市宝山区中心医院外科, 上海 201900
2. 第二军医大学基础部流行病学教研室, 上海 200433

[摘要] **目的** 预测肿瘤相关蛋白 EIF4G1 亚型的 B 细胞表位。**方法** 下载 EIF4G1 蛋白各种亚型的氨基酸序列, 以单参数(亲水性、可及性、柔韧性、抗原性)预测为基础, 结合 ABCpred 和二级结构预测筛选等方法综合筛选 EIF4G1 蛋白质亚型特异性的 B 细胞表位。**结果** 目前已知 EIF4G1 蛋白的亚型有 5 种, 亚型间的氨基酸变化局限在一个由 300 个氨基酸形成的序列区域, 亚型特异性的 B 细胞表位可能位于该区域的 14~19、21~27、52~61、106~112、113~139、183~189、201~216 和 217~224 位氨基酸残基区域内或附近。**结论** EIF4G1 蛋白质亚型间存在特异性 B 细胞表位, 这些表位对应用合成肽抗原检测蛋白质亚型和进行肿瘤患者的早期诊断具有重要指导意义。

[关键词] 真核细胞起始因子 4G; 蛋白质亚型; B 淋巴细胞表位; 肿瘤

[中图分类号] R 730.4 **[文献标志码]** A **[文章编号]** 0258-879X(2010)08-0822-04

Prediction of B cell epitopes of tumor-associated protein EIF4G1 subtypes

CAO Ting-hu¹, CHANG Wen-jun², HU Jin-xin¹, ZHU Jun¹, CAO Guang-wen^{2*}

1. Department of Urology, Shanghai Baoshan Central Hospital, Shanghai 201900, China
2. Department of Epidemiology, College of Basic Medical Sciences, Second Military Medical University, Shanghai 200433, China

[Abstract] **Objective** To predict the B cell epitopes of tumor-associated protein EIF4G1 subtypes. **Methods** The sequences of all the protein subtypes of EIF4G1 were retrieved from NCBI protein database. Based on single parameter evaluation, including hydrophilicity, flexibility, antigenicity, the B cell epitopes of the EIF4G1 protein subtypes were predicted using NPS@structure software and ABCpred software. **Results** EIF4G1 protein had five subtypes. The variation of the five different EIF4G1 subtypes was limited within a 300aa region. We identified eight epitopes locating in or near 14-19, 21-27, 52-61, 106-112, 113-139, 183-189, 201-216, and 217-224, which can be used to identify specific B cell epitopes of different protein subtypes. **Conclusion** B cell epitopes of EIF4G1 protein subtypes do exist, and they may be used for the protein subtypes evaluation and early diagnosis of tumor patients using artificially-produced matched peptides.

[Key words] eukaryotic initiation factor-4G; protein isoforms; B lymphocytes epitopes; neoplasms

[Acad J Sec Mil Med Univ, 2010, 31(8):822-825]

EIF4G1 蛋白已被发现在炎症性乳腺癌^[1]、鼻咽癌^[2]、肺癌^[3]、前列腺癌^[4]中高表达, 利用噬菌体肽库技术表达 EIF4G1 蛋白的一段氨基酸序列能够有效鉴别前列腺癌患者血清中 EIF4G1 蛋白的自身抗体, 具有明显的诊断价值^[4]。肿瘤患者在疾病早期往往针对肿瘤特异性抗原产生自身抗体, 这种自身抗体可出现在临床可见肿瘤形成之前, 检测血清中自身抗体来早期筛查恶性肿瘤是一种早期、非侵入性和廉价的检测手段^[5], 具有广阔的应用前景。

肿瘤患者自身抗体的产生往往是针对于肿瘤相

关抗原的特定氨基酸序列(即 B 细胞抗原表位)而产生。B 细胞抗原表位是一段具有较强抗原性的氨基酸序列, 往往位于蛋白质结构的表面。近年来的研究发现, 许多蛋白都存在蛋白质亚型, 例如 P53 目前已知 6 个蛋白质亚型^[6], VEGF 目前已知 8 个蛋白质亚型^[7], 这些蛋白质亚型在不同肿瘤以及肿瘤发生发展的不同阶段具有表达水平和功能上的差异^[6-7]。蛋白质亚型往往表现为氨基酸序列或(和)空间构象上的差异, 而这种差异可能导致蛋白质亚型间抗原表位的不同, 从而刺激机体产生蛋白质亚

[收稿日期] 2010-05-10 **[接受日期]** 2010-06-12

[基金项目] 国家自然科学基金(30872562), 上海市教育委员会科研创新项目(08ZZ39). Supported by the National Natural Science Foundation of China (30872562) and Innovation Program of Shanghai Municipal Education Commission(08ZZ39).

[作者简介] 曹廷虎, 硕士, 主治医师. E-mail: cgr70@sina.com

* 通讯作者(Corresponding author). Tel: 021-81871060, E-mail: gcgao@smmu.edu.cn

型特异性的自身抗体。因此,对蛋白质亚型特异性抗原表位进行预测,对将来合成抗原表位短肽并用于肿瘤相关抗原亚型的发现及恶性肿瘤的早期诊断具有重要意义。本研究拟以 EIF4G1 蛋白为例,对蛋白质亚型特异性抗原表位预测中的相关问题进行探讨。

1 材料和方法

1.1 EIF4G1 蛋白质亚型氨基酸序列及其比对 EIF4G1 蛋白质亚型序列从 NCBI 网站中进行搜索和下载,利用 MEGA4.0 软件^[8]比对,明确各蛋白质亚型氨基酸序列变化的重点区域;同时从文献中^[4]下载 EIF4G1 蛋白的一个已知噬菌体肽表位,以便比较分析。

1.2 EIF4G1 蛋白质亚型的二级结构预测 采用 NPS@structure server 服务器^[9]在线对各种蛋白质亚型的二级结构进行预测,该预测软件整合了多种蛋白质二级结构预测方法并对各种预测结果进行相互校正,最后获得更加可靠的预测结果。为了达到 70%~80% 的精确度,本研究中 EIF4G1 蛋白质亚型的二级结构预测选用了 DSC、HNN、PHD、PRED-ATOR 和 MLRC 等 5 种预测方法,预测的结构状态包括 α -螺旋、 β -折叠、 β -转角和无规则卷曲 4 种。

1.3 EIF4G1 蛋白质亚型的亲水性、可及性、柔韧性和抗原性的预测 根据 MEGA 4.0 软件所确定蛋白质亚型氨基酸序列变化区域,采用 DNASTAR 软件中的相应菜单命令对该区域的氨基酸序列进行亲水性、可及性、柔韧性和抗原性的预测,并比较 4 个参数的计算结果。

1.4 ABCpred 服务器对 EIF4G1 蛋白质亚型 B 细胞抗原表位的预测 根据 MEGA 4.0 软件所确定蛋白质亚型氨基酸序列变化区域,采用 ABCpred (<http://www.imtech.res.in/raghava/abcpred/>) 服务器^[10]预测蛋白质亚型的 B 细胞表位。

1.5 综合预测蛋白质亚型特异性的 B 细胞抗原表位 综合蛋白质二级结构和氨基酸序列单参数预测的结果,确定蛋白质亚型相关变化区域的可能 B 细胞表位;通过对蛋白质亚型特异性氨基酸序列的比较来确定亚型特异性的 B 细胞表位。

2 结果

2.1 EIF4G1 蛋白质亚型的概况 EIF4G1 蛋白目前已知有 5 个蛋白质亚型,其名称依次为 A、B、C、D 和 E。根据 MEGA 4.0 软件的比对结果,EIF4G1 蛋白的 A 亚型有 1 599 个氨基酸,其他亚型依次在羧基端减少 40、87、164 和 196 个氨基酸。从文献中

下载的噬菌体表位序列^[3]位于 EIF4G1 蛋白 A 亚型羧基端的第 182~299 个氨基酸,提示噬菌体肽中所含的 B 细胞表位可能存在于 EIF4G1 蛋白的所有已知 5 个亚型中。

2.2 EIF4G1 蛋白质亚型的二级结构预测 相对于蛋白质二级结构中 α -螺旋和 β -折叠、 β -转角和无规则卷曲区域等结构比较松散,容易扭曲盘旋到蛋白质表面,因此, β -转角和无规则卷曲区域常存在 B 细胞抗原表位。如图 1 所示,将 5 个蛋白质亚型的二级结构相似区域对齐,发现各亚型的二级结构差异主要局限在 A 亚型的前 196 个氨基酸,前 400 个氨基酸对应的二级结构区域存在更多的无规则卷曲,噬菌体展示技术发现的已知 EIF4G1 蛋白 B 细胞表位序列(182~299 氨基酸)也存在丰富的卷曲区域。为了便于研究蛋白质亚型特异性的抗原表位,我们选择前 300 个氨基酸序列作为研究对象。以 EIF4G1 蛋白 A 亚型为例,1~61 位、69~171 位、180~189 位和 199~289 位等氨基酸序列中均发现无规则卷曲二级结构,考虑到 1~40 氨基酸为 A 亚型特有氨基酸序列,提示该区域可能存在 A 亚型的特异性 B 细胞抗原表位。

2.3 EIF4G1 蛋白质亚型的亲水性、可及性、柔韧性及抗原性 根据氨基酸序列各个参数与 B 细胞抗原表位的关系,目前认为氨基酸序列的亲水性、可及性、柔韧性和抗原性参数数值越大,该区域存在 B 细胞抗原表位的可能性也越大。如图 2 所示,EIF4G1 蛋白质亚型特异性或者相对特异性的氨基酸区域(如 1~40、41~87、88~164、165~196 和 197~300 位)均存在亲水性、可及性、柔韧性和抗原性等参数均较大的氨基酸序列,提示可能存在蛋白质亚型特异性的抗原表位。文献中用噬菌体展示肽技术所发现的 EIF4G1 抗原表位序列(182~299)也存在较多的亲水性、可及性、柔韧性和抗原性等参数均较大的氨基酸序列,提示单参数预测结果和实验结果存在较好的一致性。

2.4 ABCpred 对 EIF4G1 蛋白质亚型相关重点变化区域的表位预测 ABCpred 是 Saha 和 Raghava 以人工神经网络(recurrent neural network)的运算方法为基础建立的预测 B 细胞表位的网络工具,其结果按回归神经网络所得评分排列,评分最高的肽段成为表位的可能性最大。当设置表位肽段长度为 16 个氨基酸和界点值为 0.51 时,测得 EIF4G1 蛋白质亚型相关重点变化区域的氨基酸残基序列及评分如表 1,其中最高分 0.95,最低分 0.66,EIF4G1 蛋白质亚型特异性或者相对特异性的氨基酸区域均存在 ABCpred 预测的抗原表位。

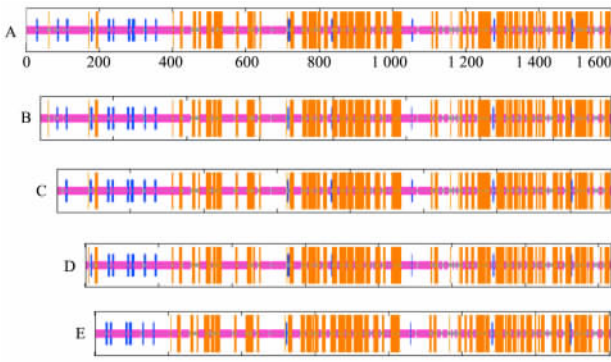


图 1 EIF4G1 蛋白质亚型的二级结构预测

Fig 1 Predicted secondary structure of

EIF4G1 protein subtypes by using NPS@structure server

α -helix, blue; β -sheet, red; coil, magenta; unclassified, gray

2.5 综合预测蛋白质亚型特异性的 B 细胞抗原表位 综合比较单参数(亲水性、可及性、柔韧性及抗原性等)抗原表位预测、蛋白质二级结构抗原表位预测和ABCpred软件数学模型预测的结果,我们选择各预测方法都合格的氨基酸线性区域作为可能的表位,结果发现 EIF4G1 蛋白质亚型相关重点变化区域的 14~19、21~27、52~61、106~112、113~139、183~189、201~216 和 217~224 位均存在符合综合预测法抗原表位标准的区域。对于 A 亚型,由于 1~40 位氨基酸是其特有序列,因此,14~19 和 21~27 区域的抗原表位区域是 A 亚型特有的抗原表位;对于 B、C、D、E 亚型,本研究没有发现单一的具有特异性的亚型特异性抗原表位。

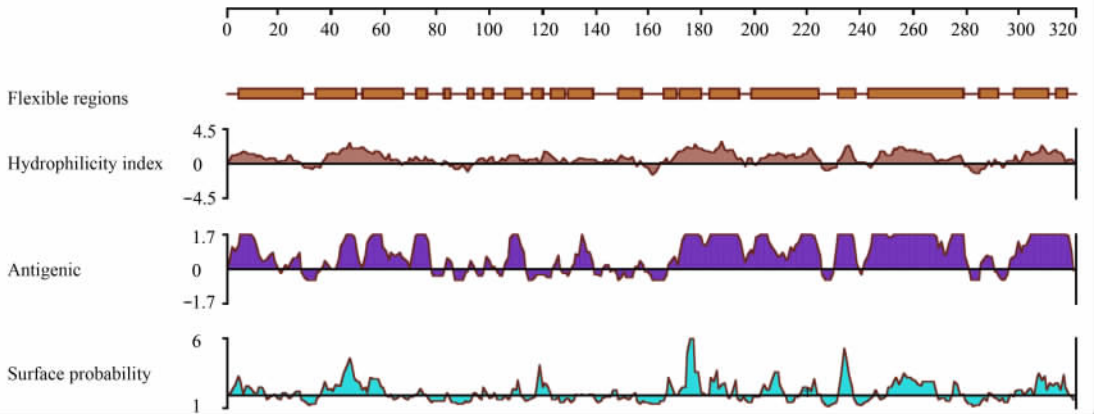


图 2 EIF4G1 蛋白质亚型相关重点变化区域的亲水性、可及性、柔韧性和抗原性分析

Fig 2 Analysis of hydrophilicity, surface, flexible, and antigenic index of variable region of EIF4G1 protein subtypes

表 1 ABCpred 预测的抗原表位情况

Tab 1 EIF4G1 protein epitopes predicted by ABCpred sever

| Rank | Sequence | Start position | Score | Subtypes specific position | Rank | Sequence | Start position | Score | Subtypes specific position |
|------|-------------------|----------------|-------|----------------------------|------|------------------|----------------|-------|----------------------------|
| 6 | PSPGLPQAFPPGQTA | 14 | 0.89 | 1-40 | 17 | QPPQIAPKRERKTIRI | 167 | 0.76 | 165-196 |
| 12 | PPGQTAPVVFSTPQAT | 24 | 0.81 | 1-40 | 1 | TIRIRDPNQGGKDITE | 179 | 0.95 | 165-196 |
| 17 | SQPRQHFYPSRAQPPS | 45 | 0.76 | 41-87 | 8 | ARTASTPTPPQTGGGL | 200 | 0.86 | 197-199 |
| 15 | PSRAQPSSAASRVQS | 53 | 0.78 | 41-87 | 7 | PQTGGGLEPQANGETP | 209 | 0.87 | 197-199 |
| 4 | QSAAPARPGPAAHVYP | 67 | 0.91 | 41-87 | 19 | PQANGETPQVAVIVRP | 217 | 0.66 | 197-199 |
| 5 | PSQISYPASQGAYYIP | 91 | 0.90 | 88-164 | 9 | VAVIVRPDDRSQGAII | 226 | 0.85 | 197-199 |
| 3 | PASQGAYYIPGQGRST | 97 | 0.92 | 88-164 | 2 | AIIADRPGLPGPEHSP | 239 | 0.94 | 197-199 |
| 11 | YIPGQGRSTYVVPTQQ | 104 | 0.83 | 88-164 | 10 | PGPEHSPSESQPSSPS | 248 | 0.84 | 197-199 |
| 3 | YPVQPGAPGFYPGASP | 120 | 0.92 | 88-164 | 6 | PSSPSPSPSPVLEP | 259 | 0.89 | 197-199 |
| 10 | FYPGASPTFEFGTYAGA | 129 | 0.84 | 88-164 | 9 | TPSPSPVLEPGSEPNL | 265 | 0.85 | 197-199 |
| 8 | FGTYAGAYYPAQGVQQ | 138 | 0.86 | 88-164 | 16 | GSEPNLAVLSIPGDTM | 275 | 0.77 | 197-199 |
| 13 | YYPAQGVQQFPTGVAP | 145 | 0.80 | 88-164 | 18 | AVLSIPGDTMTTIQMS | 281 | 0.71 | 197-199 |

3 讨论

EIF4G1 蛋白的高表达与鼻咽癌、前列腺癌和乳腺癌的发生密切相关,但该蛋白目前已知 5 个蛋白质亚型的存在,具体哪个亚型与肿瘤的关系更密切以及

不同亚型其临床意义是否不同尚不清楚。

以往对蛋白质亚型的分析主要是从信使 RNA 剪切模式入手,但这种方法不能在蛋白质水平反映各种亚型的表达水平,更无法阐述各蛋白质亚型的生物学功能。研究发现许多肿瘤抗原能刺激机体产生自身抗

体,通过检测血清自身抗体水平则能反映相应肿瘤抗原的表达状况。从免疫学角度来看,抗体是针对于蛋白质表面的抗原决定簇(即抗原表位)而产生的,而抗原表位是肿瘤抗原蛋白质中能激发机体产生抗体并与抗体分子结合的一段序列。利用蛋白质亚型特异性的抗原表位来合成短肽,理论上能通过检测相应自身抗体来反映机体内肿瘤相关抗原蛋白质亚型的表达水平。在前列腺癌的研究中,已发现 EIF4G1 基因的一段 mRNA 序列能在噬菌体载体中正确表达 EIF4G1 的一段氨基酸序列,且该噬菌体肽能用于检测肿瘤患者体内的自身抗体并用于前列腺癌早期诊断。因此,利用 EIF4G1 蛋白质亚型特异性抗原表位来检测亚型特异性抗体能用于反映体内 EIF4G1 蛋白质亚型的表达情况,从而为阐述 EIF4G1 蛋白质亚型在肿瘤中的作用及在临床诊断中的价值奠定基础。

随着计算机技术的不断发展和生物学理论方法的逐步丰富,B 细胞表位的预测技术也在不断发展。目前的 B 细胞表位预测技术常综合蛋白质二级结构预测、氨基酸序列参数(亲水性、可及性、柔韧性和抗原性等)以及其他预测技术来预测可能的 B 细胞表位^[11-12]。本研究首先下载 EIF4G1 蛋白质亚型的氨基酸序列,然后通过序列比对和二级结构分析确定与亚型序列变化密切相关的氨基酸序列区域,并通过与已知的噬菌体肽序列比对来判断预测结果的可信程度。结果表明 EIF4G1 蛋白的 5 个亚型的变化区域主要集中在 A 亚型的前 200 个氨基酸,考虑到预测结果与已知 EIF4G1 蛋白的噬菌体肽序列比对,本研究选择前 300 个氨基酸进行重点分析。

二级结构分析发现所研究的 300 个氨基酸存在大量的无规则卷曲区域,提示可能蕴含着抗原表位,且在各 EIF4G1 蛋白质亚型特异性或者相对特异性区域均有抗原表位存在的可能性。氨基酸序列的亲水性、可及性、柔韧性和抗原性等参数也提示在各蛋白质亚型特异性或者相对特异性区域均有抗原表位存在的可能性。综合二级结构、氨基酸序列参数和 ABCpred 预测结果,我们预测在 300 个氨基酸序列的 14~19、21~27、52~61、106~112、113~139、183~189、201~216 和 217~224 位可能存在 B 细胞抗原表位,合成这些多肽序列可用于 EIF4G1 蛋白质亚型自身抗体的检测并用于临床诊断和肿瘤疫苗设计。

针对于 EIF4G1 蛋白质亚型集中变化的 300 个氨基酸序列,其表位序列中仅 14~19 和 21~27 蛋白多肽能特异性检测 EIF4G1 蛋白质 A 亚型;而对于其他亚型的检测则需通过不同表位肽的联合应用来评价,例如,在检测自身抗体时,当 14~19 或 21~27 表位肽阴性且 52~61 表位肽阳性时,才能准

确判断 B 亚型的存在。因此,利用表位肽抗体检测来判断 EIF4G1 蛋白质亚型的表达虽然可行,但却具有一定局限性。这种局限性也许可以通过发现新的蛋白质亚型特异性空间构象表位来解决^[13],但目前非连续表位的预测存在一定难度。

本研究针对于蛋白质亚型和肿瘤相关抗原的自身抗体现象提出了利用蛋白质亚型特异性抗原表位来检测患者血清中蛋白质亚型特异性自身抗体的设想,并通过生物信息学分析预测了 EIF4G1 蛋白质亚型特异性的抗原表位,为今后利用抗原表位研究 EIF4G1 蛋白质亚型在恶性肿瘤中的诊断和疫苗开发奠定了基础。

[参考文献]

- [1] Silvera D, Arju R, Darvishian F, Levine P H, Zolfaghari L, Goldberg J, et al. Essential role for eIF4G1 overexpression in the pathogenesis of inflammatory breast cancer[J]. *Nat Cell Biol*, 2009, 11: 903-908.
- [2] Tu L X, Liu Z, He X F, He Y, Yang H L, Jiang Q P, et al. Over-expression of eukaryotic translation initiation factor 4 gamma 1 correlates with tumor progression and poor prognosis in nasopharyngeal carcinoma[J]. *Mol Cancer*, 2010, 9: 78.
- [3] Comtesse N, Keller A, Diesinger I, Bauer C, Kayser K, Huwer H, et al. Frequent overexpression of the genes FXR1, CLAPM1 and EIF4G located on amplicon 3q26-27 in squamous cell carcinoma of the lung[J]. *Int J Cancer*, 2007, 120: 2538-2544.
- [4] Wang X, Yu J, Sreekumar A, Varambally S, Shen R, Giacherio D, et al. Autoantibody signatures in prostate cancer[J]. *N Engl J Med*, 2005, 353: 1224-1235.
- [5] Lu H L, Goodell V, Disis M L. Humoral immunity directed against tumor-associated antigens as potential biomarkers for the early diagnosis of cancer[J]. *J Proteome Res*, 2008, 7: 1388-1394.
- [6] Bourdon J C. p53 and its isoforms in cancer[J]. *British J Cancer*, 2007, 97: 277-282.
- [7] Qiu Y, Hoareau-Aveilla C, Olteas S, Harper S J, Bates D O. The anti-angiogenic isoforms of VEGF in health and disease [J]. *Biochem Soc Trans*, 2009, 37: 1207-1213.
- [8] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4. 0 [J]. *Mol Biol Evol*, 2007, 24: 1596-1599.
- [9] Combet C, Blanchet C, Geourjon C, Deleage G. NPS@: network protein sequence analysis[J]. *Trends Biochem Sci*, 2000, 25: 147-150.
- [10] Saha S, Raghava G P S. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network [J]. *Proteins*, 2006, 65: 40-48.
- [11] 刘争春, 王颖, 王树军, 张惠珍, 荣婷婷, 王庆飞, 等. 人卵巢癌相关抗原 OVA66 的 B 细胞表位预测 [J]. *实用医学杂志*, 2009, 25: 513-516.
- [12] 孟露露, 荫俊, 王慧, 史晶, 侯晓军. B 型肉毒毒素重链抗原表位的预测及表位合成肽抗原性的检测 [J]. *生物技术通讯*, 2006, 17: 183-185.
- [13] Moreau V, Granier C, Villard S, Laune D, Molina F. Discontinuous epitope prediction based on mimotope analysis [J]. *Bioinformatics*, 2006, 22: 1088-1095.