

• 论 著 •

IL-6 基因结构和功能生物信息学预测*

于 欣¹, 杨 震^{2△}, 楚元奎¹, 张怡清¹, 于美玲³

(宁夏医科大学:1. 临床学院检验系;2. 临床学院心脏中心;3. 药剂科,银川 750004)

摘 要:目的 以白细胞介素(IL)-6 为研究对象,对其基因结构和功能进行生物信息学分析。方法 采用生物信息学软件对基因编码的蛋白质结构、理化性质、信号肽、跨膜结构、亚细胞定位、二级结构及高级结构进行生物信息学分析。结果 IL-6 为稳定亲水性跨膜蛋白,定位于细胞外,二级结构以 α -螺旋和无规则卷曲为主,有 2 个 Ser 和 2 个 Thr 可成为蛋白激酶磷酸化位点。结论 分析预测 IL-6,在研究其参与高血压、炎症反应等方面有重要意义。

关键词:白细胞介素-6; 生物信息学; 基因结构

DOI:10. 3969/j. issn. 1673-4130. 2016. 21. 004

文献标识码:A

文章编号:1673-4130(2016)21-2959-03

Bioinformatics prediction of IL-6 gene structure and function*

YU Xin¹, YANG Zhen^{2△}, CHU Yuankui¹, ZHANG Yiqing¹, YU Meiling³

(1. Department of Clinical Laboratory; 2. Heart Center; 3. Department of Pharmacy, Ningxia Medical University, Yinchuan, Ningxia 750004, China)

Abstract: **Objective** To analysis the bioinformatics of IL-6 gene structure and function. **Methods** By means of bioinformatics software, the structure, physical and chemical properties, signal peptide, transmembranes structure, subcellar localization, secondary structure and advanced structure of the protein were analyzed. **Results** IL-6 protein was stable hydrophilic transmembrane protein, which were located outside the cells. The secondary structure was primarily composed of α -helix and randmo coil. And it had two serines and two threonines, which might become phosphorylation sites of the proteinkinase. **Conclusion** Analysis and prediction of IL-6 might be significant in research of clinical diseases such as hypertension, inflammation and so on.

Key words: IL-6; bioinformatics; gene structure

近年来,人类基因组计划和其他物种基因组计划使人类在生命科学领域,尤其是核酸和蛋白质等生物大分子的序列结构与功能等领域迅速积累了大量数据,由生命科学和信息技术等学科相结合形成的新兴交叉学科——生物信息学^[1]。伴随分子遗传学在检验医学中的进步,某些临床遗传学研究已从简单的单基因疾病转向复杂的多基因疾病,如心血管疾病、高血压、各种肿瘤及炎症反应等^[2]。白细胞介素(IL)-6 是具有多种生物活性的细胞因子,也是与内分泌、炎症反应关系最为密切的炎症因子^[3],其以自分泌、旁分泌等多种形式对不同器官产生不同生物学效应。笔者采用生物信息学软件分析 IL-6 不同特点,为研究其与高血压等临床常见疾病发生发展相关关系提供思路。现报道如下。

1 资料与方法

1.1 一般资料 以登陆序列号(NP_000591.1)在 Genbank 数据库检索得到人蛋白序列。

1.2 方法 采用在线获取生物信息学软件 protparam 分析 IL-6 蛋白的物理和化学特性;采用软件 TargetP(由丹麦 TeKniske 大学生物学序列分析中心提供)和 PSORT II 综合分析其亚细胞定位;采用软件 SignalP4.0 分析信号肽;采用 TM-HMM2.0 Sever 对 IL-6 跨膜结构进行预测;采用 Netphos2.0 Server 和 NetOGlyc4.0 Server 对其磷酸化位点和糖基化位点进行前瞻性分析;采用软件 Smart 在线预测 IL-6 保守结构域;采用 DNASar 在线分析其二级结构及抗原系数。所有在线工具网址见文献^[1,4]。

2 结 果

2.1 IL-6 蛋白理化性质的预测分析 采用软件 protparam 对 IL-6 蛋白的理化性质进行预测分析,发现 IL-6 由 212 个氨基酸组成,分子式为 $C_{1049}H_{1685}N_{283}O_{321}S_{10}$,相对分子质量为 $23\,718.2 \times 10^3$,理论等电点 pI 为 6.17。在构成 IL-6 蛋白的 20 种氨基酸中,Leu 所占成分比例最高,占 13.2%;而 His 和 Trp 比例较少,分别占 0.9%和 0.5%;未发现 Pyl 和 Sec。带负电荷氨基酸残基数为 24,带正电荷氨基酸残基数为 23;不稳定系数为 57.7。依据 Guruprasad 方法,分析结果提示 IL-6 蛋白不稳定性;脂肪系数为 87.5,属于脂溶性蛋白;总平均亲水性为(-0.271 0),见图 1。

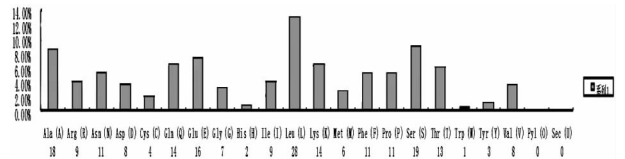


图 1 IL-6 氨基酸组成

2.2 IL-6 蛋白信号肽分析 IL-6 分泌蛋白 N 端 15~35 位氨基酸的疏水性短肽区域可引导蛋白质的肽链通过内质网进入腔内,此疏水短肽即为信号肽。本研究采用 SignalP4.0 对 IL-6 蛋白信号肽进行预测。首先,输入前 70 个氨基酸,采用神经网络模型 NN 和隐马尔可夫模型 HMM 进行预测,随后给出

* 基金项目:宁夏回族自治区自然科学基金资助项目(NZ12175)。

作者简介:于欣,女,副教授,主要从事心血管疾病临床血液免疫分子生物学实验室方面的研究。△ 通讯作者,E-mail: yangzhen080@

Y、C、S-score 计测结果,通过 Y 最大值判断分析信号肽剪切位点;通过 S 平均值判断分泌蛋白(若 S 平均值高于 0.500,则为分泌蛋白,且存在信号肽)。预测表明,C 结果值为 0.257,Y 结果值为 0.477,S 结果值为 0.963,其剪切位点位于 27~28 位氨基酸,且 S 平均值 0.887,说明 IL-6 是具有信号肽的分泌蛋白。

2.3 IL-6 蛋白跨膜结构预测 膜蛋白常有胞外区、跨膜区和胞内区等。跨膜区为蛋白的细胞膜内部分,可 1 个或多个。笔者采用生物 TMHMM2.0 Sever 对 IL-6 蛋白跨膜结构进行分析,结果表明,IL-6 蛋白可能是由跨膜区(5~27 位氨基酸)、胞外区(1~405 位氨基酸处于细胞膜表面)及胞内区(28~101 位氨基酸)构成的膜蛋白。

2.4 IL-6 蛋白的疏水性预测分析 采用软件 protscale 分析 IL-6 蛋白疏水性,结果显示该蛋白多肽链第 18 号分值最高(2.756,疏水性最强);第 83 号分值最低(-2.422,亲水性最强)。比较亲水性区域和疏水性区域,发现前者区域显著大于后者疏水性区域,提示 IL-6 蛋白是 1 种亲水性蛋白。见图 2。

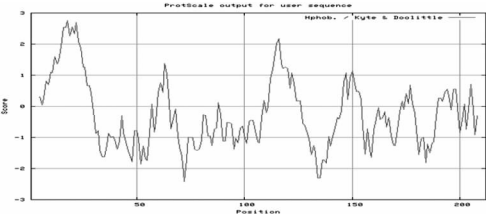


图 2 IL-6 蛋白的疏水性预测分析

2.5 IL-6 蛋白的亚细胞定位分析 把 IL-6 蛋清蛋白序列输入软件 TargetP,序列来源项选择 noplant,其他选项用默认值。结果表明,该蛋白大部分集中在细胞其他位置。进一步采用软件 PSORT II 的 the k-NN Prediction 分析,基本确认 IL-6 蛋白分子属于分泌蛋白,分别在胞核(占 43.5%)、细胞质(占 39.1%)、细胞骨架(占 8.7%)、线粒体(占 8.7%)中发挥生物学作用。

2.6 IL-6 蛋白翻译后修饰位点分析 真核生物中,蛋白质翻译后修饰过程常见有糖基化和磷酸化 2 种。糖基化通常修饰天冬酰胺上 N 端,其氨基酸特征序列为 Asn-X-Ser-Thr(X 可表示为任一类氨基酸)。机体某些蛋白需要糖分子/糖链等与其他分子相互作用发挥生物学作用。磷酸化多具有功能开关作用,当机体蛋白被磷酸化后才具备调控生物体功能代谢的作用。笔者采用 Netphos2.0 Server 和 NetOGlyc4.0 Server 对 IL-6 蛋白进行预测,预测结果表明其有 6 个 Ser、2 个 Thr、1 个 Tyr 可能成为蛋白激酶磷酸化位点,其糖基化位点分别为 Thr2、Ser3、Ser16、Ser26、Ser29、Thr30、Thr39、Ser41、Thr42、Ser57、Ser64、Thr71。

2.7 IL-6 蛋白的二级结构分析 多肽链的主链借助氢键形成有规则卷曲折叠的一维构象称为蛋白质二级结构。本研究采用 DNASTar 预测 IL-6 蛋白的二级结构和抗原表位,见图 3。综合 Gamien-Robson 和 Chou-Famsman2 方法,预测发现 IL-6 的二级结构主要为螺旋结构(75.9%和 51.9%)、折叠(4.72%和 1.46%)、转角(5.19%和 2.45%)和无规则卷曲(1.42%)。根据 Jameson woff 预测 IL-6 氨基酸,抗原指数较高区域主要为 29~56、65~86、90~91、94~111、132~141、154~162、165~172、180~181、197~199、202~210 区段,其他部位的可能较小或表现为负值。综合分析发现,除第 30~56 外,其他区域亲水性和氨基酸表面指数偏低,提示该区域可能是优势抗原表位。

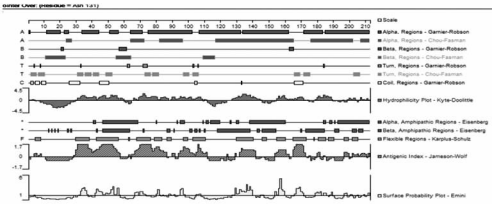


图 3 IL-6 蛋白的二级结构

2.8 IL-6 蛋白的保守功能域分析 保守结构域指生物进化或者 1 个蛋白家族中不变或相同的结构域,具有重要功能。采用 Smart(网址为 <http://smart.embl-heidelberg.de>)推测,IL-6 蛋白的低复杂度区域(LCR)为 11~33 氨基酸;第 57~210 氨基酸为 IL-6 结构域。

3 讨论

有学者于 1986 年首次证实单核细胞可产生 1 种中性粒细胞趋化因子(即 IL),且可分为多种细胞 IL。IL 最初指白细胞产生,于白细胞间发挥功能的细胞因子,后发现 IL 还可由其他细胞分泌,亦可作用其他细胞。其中,IL-6 又称前炎性细胞因子[5],主要被单核-巨噬细胞、成纤维细胞、T 淋巴细胞[6]、内皮细胞、肥大细胞产生,可刺激 B 细胞产生抗体、T 细胞增殖、细胞毒性 T 淋巴细胞(CTL)分化、造血、肝细胞增殖和分化等。作为重要参与物质在某些疾病炎症过程中发生作用(如慢性阻塞性肺疾病)[7]。分析表明,IL-6 是 1 种脂溶性信号肽分泌蛋白,亦为亲水性蛋白,主要由细胞核(占 43.5%)、细胞质(占 39.1%)发挥生物学作用。预测表明其有 6 个 Ser、2 个 Thr、1 个 Tyr 可成为蛋白激酶磷酸化位点;主要为螺旋结构;其第 30~56 可能为优势抗原表位,结构域为第 57~210 氨基酸。本研究通过生物信息学方法对 IL-6 蛋白结构和功能进行了生物信息学分析,为研究其参与高血压、肺动脉高压或乳腺癌等疾病炎症损伤的研究结构和功能提供思路[8-10],并为其参与某些临床疾病炎症反应时的作用提供理论依据。

参考文献

[1] 刘贤慧,杨章平,吴正常,等. 奶牛 IL-8 基因结构和功能的生物信息学分析[J]. 扬州大学学报:农业与生命科学版, 2013,24(1):18-22.

[2] 王晶,李敏,杜晓杰,等. 2 型猪链球菌表面蛋白 Sao 的生物信息学分析及基因工程疫苗的设计[J]. 中国病原生物学杂志,2014,9(3):211-215.

[3] 丛楠. 重组靶向毒素 IL-6(T23)-PE38KDEL 可溶性表达优化和纯化方法的研究[Z]. 2013.

[4] 李宗吉,楚元奎,于欣,等. 细粒棘球蚴 GST 抗原表位的生物信息学预测[J]. 国际检验医学杂志,2015,36(13): 1818-1820.

[5] 鹿秋玉,李才锐,孙曙光. IL-6 与糖尿病性视网膜病变的相关研究进展[J]. 国际眼科杂志,2015,15(1):52-54.

[6] Jones LL, Allli R, Li B, et al. Differential T cell cytokine receptivity and not signal quality distinguishes IL-6 and IL-10 signaling during Th17 differentiation[J]. J Immunol, 2016,196(7):2973-2985.

[7] Comer DM, Kidney JC, Ennis M, et al. Airway epithelial cell apoptosis and inflammation in COPD, smokers and nonsmokers[J]. Eur Respir J, 2013,41(5):1058-1067.

[8] Li J, Song J, Jiang MH, et al. Interleukin-(下转第 2963 页)

T₄ 期组间差异有统计学意义 ($P < 0.05$)。Zhang 等^[11]也证实 EBV-DNA 拷贝数与 T 分期有显著正相关关系。

鼻咽癌患者在初诊时有 65.00%~80.00% 患者发现癌细胞转移。本研究中鼻咽癌患者血浆 EBV-DNA 拷贝数与 N 分期有显著相关性 ($\chi^2 = 61.77, P = 0.000$)，笔者认为随 N 分期的递增，患者血浆 EBV-DNA 拷贝数增高，N0 分期 EBV-DNA 拷贝数最低，N3 分期 EBV-DNA 拷贝数最高。有学者报道血浆 EBV-DNA 拷贝数与 N 分期呈正相关关系，从另一个角度说明血浆 EBV-DNA 拷贝数越高，淋巴转移越广泛。本研究发现，鼻咽癌患者血浆 EBV-DNA 拷贝数与 M 分期有显著相关性 ($\chi^2 = 71.82, P = 0.000$)，研究结果显示 M1 分期 EBV-DNA 拷贝数高于 M0 分期 ($P < 0.05$)，其机制可能与血浆 EBV-DNA 拷贝数与转移的肿瘤负荷有关，EBV-DNA 可能由肿瘤细胞释放进入外周血，随着病情的进展，血浆中 EBV-DNA 拷贝数会相应增加，Zhang 等^[11]报道鼻咽癌远处转移患者血浆 EBV-DNA 拷贝数显著高于无远处转移者。

本研究评价荧光定量 PCR 和 ELISA 2 种方法对鼻咽癌的诊断价值。荧光定量 PCR 具有常规 PCR 的高敏感性，能克服常规 PCR 技术不能定量及易污染问题；此外，采用荧光探针，可直接检测到 PCR 扩增中荧光信号变化，实现实时荧光定量检测。研究中发现，荧光定量 PCR 方法测定 EBV 感染阳性率显著高于 ELISA 血清学方法，且阳性率随着鼻咽癌进展而增加。近年来，有学者认为荧光定量 PCR 法对鼻咽癌分期诊断有较高的敏感性^[12]，可作为鼻咽癌预后评价指标之一，而 ELISA 方法对鼻咽癌分期诊断敏感性较低。笔者研究显示，ELISA 方法检测 EBV 感染阳性率显著低于荧光定量 PCR 方法且对分期并不敏感。ELISA 方法作为医院免疫室常规检测方法，与荧光定量 PCR 方法比较价格相对低廉、无需特殊设备，但操作较繁琐、主观性强，只适用于大规模人群筛查。

综上所述，血浆 EBV-DNA 拷贝数与鼻咽癌 TNM 的临床分期呈正相关，鼻咽癌 TNM 分期越高，患者血浆 EBV-DNA 拷贝数越高。因此，血浆中 EBV-DNA 拷贝数可在分子水平上反映鼻咽癌患者的病情程度，对鼻咽癌的预后评价具有较大的临床意义^[13]。

参考文献

- [1] Liu LT, Tang LQ, Chen QY, et al. The prognostic value of plasma Epstein-Barr viral DNA and tumor response to neoadjuvant chemotherapy in Advanced-Stage nasopharyngeal carcinoma[J]. Int J Radiat Oncol Biol Phys, 2015, 93(4):862-869.
- [2] Wille CK, Nawandar DM, Panfil AR, et al. Viral genome methylation differentially affects of BZLF1 to activate Epstein-Barr virus lytic gene expression and viral replication[J]. J Virol, 2013, 87(2):935-950.
- [3] 陈文学, 邹学森, 李金高, 等. 血浆 EB 病毒 DNA 检测对鼻咽癌诊断的价值研究[J]. 实用癌症杂志, 2014, 29(12):1526-1528.
- [4] 罗耀凌, 陈浩, 彭颂国, 等. 联合检测 EB 病毒不同抗体及 EB 病毒 DNA 在鼻咽癌血清学诊断中的价值[J]. 中华医学杂志, 2013, 93(44):3516-3519.
- [5] Liu Z, Ji MF, Huang QH, et al. Two Epstein-Barr virus-related serologic antibody tests in nasopharyngeal carcinoma screening: results from the initial phase of a cluster randomized controlled trial in southern China[J]. Am J Epidemiol, 2013, 177(3):242-250.
- [6] Liu WL, Guo XZ, Zhang LJ, et al. Prognostic relevance of Bmi-1 expression and autoantibodies in esophageal squamous cell carcinoma[J]. BMC Cancer, 2010, 10:467.
- [7] Hui EP, Ma BB, Chan KC, et al. Clinical utility of plasma Epstein-Barr virus DNA and ERCC1 single nucleotide polymorphism in nasopharyngeal carcinoma[J]. Cancer, 2015, 121(16):2720-2729.
- [8] Zheng XH, Lu LX, Li XZ, et al. Quantification of Epstein-Barr virus DNA load in nasopharyngeal brushing samples in the diagnosis of nasopharyngeal carcinoma in southern China[J]. Cancer Sci, 2015, 106(9):1196-1201.
- [9] Shao JY, Li YH, Gao HY, et al. Comparison of plasma Epstein-Barr virus (EBV) DNA levels and serum EBV immunoglobulin A/virus capsid antigen antibody titers in patients with nasopharyngeal carcinoma [J]. Cancer, 2004, 100(6):1162-1170.
- [10] Sun PL, Chen C, Cheng YK, et al. Serologic biomarkers of Epstein-Barr virus correlate with TNM classification according to the seventh edition of the UICC/AJCC staging system for nasopharyngeal carcinoma[J]. Eur Arch Otorhinolaryngol, 2014, 271(9):2545-2554.
- [11] Zhang L, Tang LQ, Chen QY, et al. Plasma Epstein-Barr viral DNA complements TNM classification of nasopharyngeal carcinoma in the era of intensity-modulated radiotherapy[J]. Oncotarget, 2016, 7(5):6221-6230.
- [12] Zhao FP, Liu X, Chen XM, et al. Levels of plasma Epstein-Barr virus DNA prior and subsequent to treatment predicts the prognosis of nasopharyngeal carcinoma[J]. Oncol Lett, 2015, 10(5):2888-2894.
- [13] Chan KA. Plasma Epstein-Barr virus DNA as a biomarker for nasopharyngeal carcinoma[J]. Chin J Cancer, 2014, 33(12):598-603.
- [10] Hsu EC, Kulp SK, Huang HL, et al. Integrin-linked kinase as a novel molecular switch of the IL-6-NF-B signaling loop in breast cancer[J]. Carcinogenesis, 2016, 37:430-442.

(收稿日期:2016-02-01 修回日期:2016-04-21)

(收稿日期:2016-02-05 修回日期:2016-04-25)

(上接第 2960 页)

- 6 promoter polymorphisms and susceptibility to atrial fibrillation in elderly Han Chinese patients with essential hypertension[J]. J Int Cytok Res, 2012, 32(11):542-547.
- [9] Heresi GA, Aytekin M, Hammel JP, et al. Plasma interleukin-6 adds prognostic information in pulmonary arterial hypertension[J]. Eur Respir J, 2014, 43(3):912-914.