

结核分枝杆菌 clpP2 抗原表位预测与分析*

刘思静, 蒋明娟, 蒲启康, 任晨艳, 苏琳, 张祥, 汪川[△]

四川大学华西公共卫生学院 卫生检验与检疫系(成都 610041)

【摘要】 目的 预测结核分枝杆菌酪蛋白酶蛋白水解亚基单位 2(clpP2)的抗原表位,为结核病疫苗、药物研制等提供新的靶点。**方法** 以 clpP2 蛋白的氨基酸序列为基础,采用生物信息软件 DNA Star 预测二级结构;运用 SWISS-MODEL 在线软件进行同源建模;运用 VaxiPred 软件综合预测 T 细胞抗原表位,用 ABCpred、COBEPro 和 BepiPredPred 软件综合预测其 B 细胞抗原表位;采用 DNA Star Protean、SignalP、TMHMM 在线软件和 NCBI 数据库分析其免疫学相关信息。**结果** clpP2 蛋白结构丰富,含有较多潜在细胞毒性 T 细胞和辅助 T 细胞表位;也含有较多的 B 细胞线性和构象优势表位,这些区域亲水性、表面可能性和柔韧性指数都较高。**结论** clpP2 蛋白具有诱导免疫应答的潜能,可以作为结核监测、治疗、预防的新靶点。

【关键词】 结核分枝杆菌 clpP2 抗原表位

Prediction and Analysis of Epitopes in clpP2 of *Mycobacterium tuberculosis* LIU Si-jing, JIANG Ming-juan, PU Qi-kang, REN Chen-yan, SU Lin, ZHANG Xiang, WANG Chuan[△]. Department of Public Health Laboratory Sciences, West China School of Public Health, Sichuan University, Chengdu 610041, China

[△] Corresponding author, E-mail: wangchuan@scu.edu.cn

【Abstract】 Objective To predict and analyze the antigenic epitopes in *Mycobacterium tuberculosis* protein caseinolytic protease P2 (clpP2), and explore its possibility to be applied as a new tuberculosis (TB) vaccine and drug development target. **Methods** Secondary structure of clpP2 based on nucleic sequence was predicted by DNA Star software. The homologous sequence conformation were analyzed by Swiss-Model online software. T cells antigenic epitopes were predicted through VaxiPred, and B cell epitopes were predicted by combining use of several different prediction programs, such as ABCpred, COBEPro and BepiPredPred. The immune characteristics of clpP2 were analyzed by DNA Star, SignalP, TMHMM online software and were searched through NCBI database. **Results** clpP2 protein was diverse in structure, composing with a great deal of CTL and Th cell epitopes. clpP2 was also predicted to comprise rich potential liner and discontinuous B-cell epitopes. These epitopes were accessible on the protein surface, located in flexible and hydrophilic regions. **Conclusion** clpP2 is prompted to induce immune responses and develops a novel target in surveillance, treatment and vaccine.

【Key words】 *Mycobacterium tuberculosis* clpP2 Epitope

结核病(tuberculosis, TB)是一个重要的全球性公共卫生问题,2014 年全球有 600 万结核新发病例,150 万死亡病例^[1]。多重耐药结核、人免疫缺陷病毒(HIV)/TB 共感染、潜伏性结核感染(latent tuberculosis infection, LTBI)等一系列问题,使结核发病率仍居高不下。据世界卫生组织(WHO)推算,全球约 1/3 人口感染结核分枝杆菌(*Mycobacterium tuberculosis*, MTB),其中大部分属于无临床症状的 LTBI,约 10% 的 LTBI 者会发展为活动性结核病^[2]。因此,研究 MTB 潜伏性感染相关蛋白的结构和免疫原性等特征,对于结核的

防治具有重要意义。

酪蛋白酶(caseinolytic protease, clp)是高度保守的丝氨酸蛋白酶,广泛存在于原核生物和真核生物中^[3],主要发挥蛋白水解酶作用,降解异常蛋白;维持内环境稳态,在细胞应激反应方面起着重要作用^[4]。MTB 的 clpP2 蛋白是 clpP 蛋白酶蛋白水解亚基单位 2,在处于持留状态的细菌中维持表达,而且在环境条件恢复正常后,表达上调^[3]。由此可见,该蛋白可能成为指示 MTB 潜伏感染以及结核复燃的标记,可望作为新型结核疫苗及药物的靶点。但是目前关于 clpP2 在指示 MTB 潜伏感染或结核复燃中的应用,以及作为结核预防和治疗的候选蛋白等的相关研究还很少,有关该蛋白的结构及免疫学信息的全面分析资料也不足。因此,本研究应用生

* 国家自然科学基金面上项目(No. 31570924)和四川省科技厅国际合作项目(No. 2017HH0080)资助

[△] 通信作者, E-mail: wangchuan@scu.edu.cn

物信息方法,对结核 clpP2 蛋白的结构、抗原表位、免疫学相关信息等进行初步预测,可望为该蛋白作为结核疫苗、药物、诊断试剂的新靶点提供依据。

1 材料与方法

1.1 蛋白质氨基酸序列检索与分析

从 NCBI 网站上获取 clpP2 的氨基酸序列, GenBank 登录号为 NP_216976.1, 含有 214 个氨基酸。利用 ExPaSy 网上在线软件 (<http://www.expasy.org/>) 对 clpP2 的相对分子质量、等电点、氨基酸等生物学信息进行分析。

1.2 结构预测

利用 DNA Star Protean 软件分析 clpP2 蛋白的二级结构, 利用 SWISS-MODEL 网上在线软件^[5-7] (<http://www.expasy.org/swissmod/SWISS-MODEL.html>) 进行同源建模, 获得其三维结构。

1.3 抗原表位预测

利用 VaxiPred 网上在线软件分别预测 clpP2 蛋白的细胞毒性 T 细胞 (CTL) 抗原表位 (<http://www.imtech.res.in/raghava/ctlpred>) 和辅助 T 细胞 (Th) 抗原表位 (<http://www.imtech.res.in/raghava/propred>), 分别选择 4 种计算方法进行预测: 基于支持向量机法 (only Support Vector Machine based, SVM)、基于人工神经网络法 (only Artificial Neural Network based, ANN)、一致法 (Consensus)、结合法 (Combined)。利用 ABCpred (<http://www.imtech.res.in/raghava/abcpred>)、COBEPro (<http://scratch.proteomics.ics.uci.edu>) 和 BepiPredPred (<http://www.cbs.dtu.dk/services>) 网上在线软件分别预测 clpP2 蛋白的 B 细胞线性表位; 利用 Ellipro Antibody Epitope (<http://tools.immuneepitope.org/tools/Ellipro>) 在线软件预测 clpP2 蛋白的 B 细胞构象表位。

1.4 免疫学相关信息分析

1.4.1 clpP2 蛋白的亲水性、可塑性、表面可及性 利用 VaxiPred 网上在线软件 (<http://www.imtech.res.in/raghava/bcpred/>) 和 DNA Star Protean 软件分析 clpP2 蛋白的亲水性、可塑性、表面可及性。

1.4.2 clpP2 的信号肽结构和跨膜结构的预测 利用 SignalP 在线软件 (<http://www.cbs.dtu.dk/services/SignalP>) 对 clpP2 的信号肽结构进行预测, 利用 TMHMM 在线软件 (<http://www.cbs.ndtu.dk/services/TMHMM>) 对 clpP2 的跨膜结构进行

预测^[8]。

1.4.3 BLAST 分析 利用 NCBI Homo sapiens (human) Nucleotide BLAST (<http://blast.ncbi.nlm.nih.gov/>) 分析 clpP2 蛋白氨基酸序列与人类蛋白质的同源性。

2 结果

2.1 clpP2 的基本信息

clpP2 蛋白由 214 个氨基酸组成, 其中丙氨酸 (Ala) 含量最多。蛋白相对分子质量 2 307.74, 等电点为 4.99。

2.2 clpP2 的结构预测

2.2.1 clpP2 的二级结构预测 应用 Gamier-Robson 法预测 clpP2 的二级结构, 结果见图 1, clpP2 蛋白 α 螺旋较多, 位于 21~22、35~44、49~62、66、68~70、89~92、106~119、128~131、150~171、175、182~198、205~214 位氨基酸残基; β 折叠也较多, 位于 1、7~20、28、45~48、63~65、71~77、82~88、93~105、123~127、132~135、141~146、176~181、199~204、214 位氨基酸残基; 转角结构较少, 长度较短, 多为 1~2 个氨基酸; 在 3、6、23~24、27、30~31、34、67、78~79、81~82、120~121、137~140、147~149、172 位氨基酸处有无规则弯曲。

2.2.2 clpP2 蛋白的三级结构预测 采用同源建模 (homology modeling) 法, 利用 SWISS-MODEL 网上在线软件预测 clpP2 蛋白的三级结构 (图 2), 与 Gamier-Robson 法预测的二级结构结果一致, clpP2 蛋白二级结构以稳定牢固的 α 螺旋和 β 折叠为主, β 转角和无规则卷曲较少, 提示该蛋白上可能的 B 细胞抗原表位相对较少。

2.3 clpP2 抗原表位分析

2.3.1 clpP2 CTL 表位预测 结果表明基于 SVM 法进行预测获得 48 个潜在 CTL 表位, 基于 ANN 法预测获得 30 个潜在 CTL 表位, 基于 Consensus 和 Combined 计算方法分别获得 6 个和 72 个潜在 CTL 表位。综合 4 种计算方法预测结果见图 3, 在 clpP2 蛋白的第 12~20、26~34、109~117、124~133、155~163 位氨基酸区段可能存在 CTL 表位。

使用在线程序对 clpP2 的 Th 表位进行预测, 部分结果见表 1, 其中有两段结合肽为两类结合肽等位基因所共有, 即 205~213 位的 YRKLSAQTA 和 159~167 位的 MRTLMETTL。

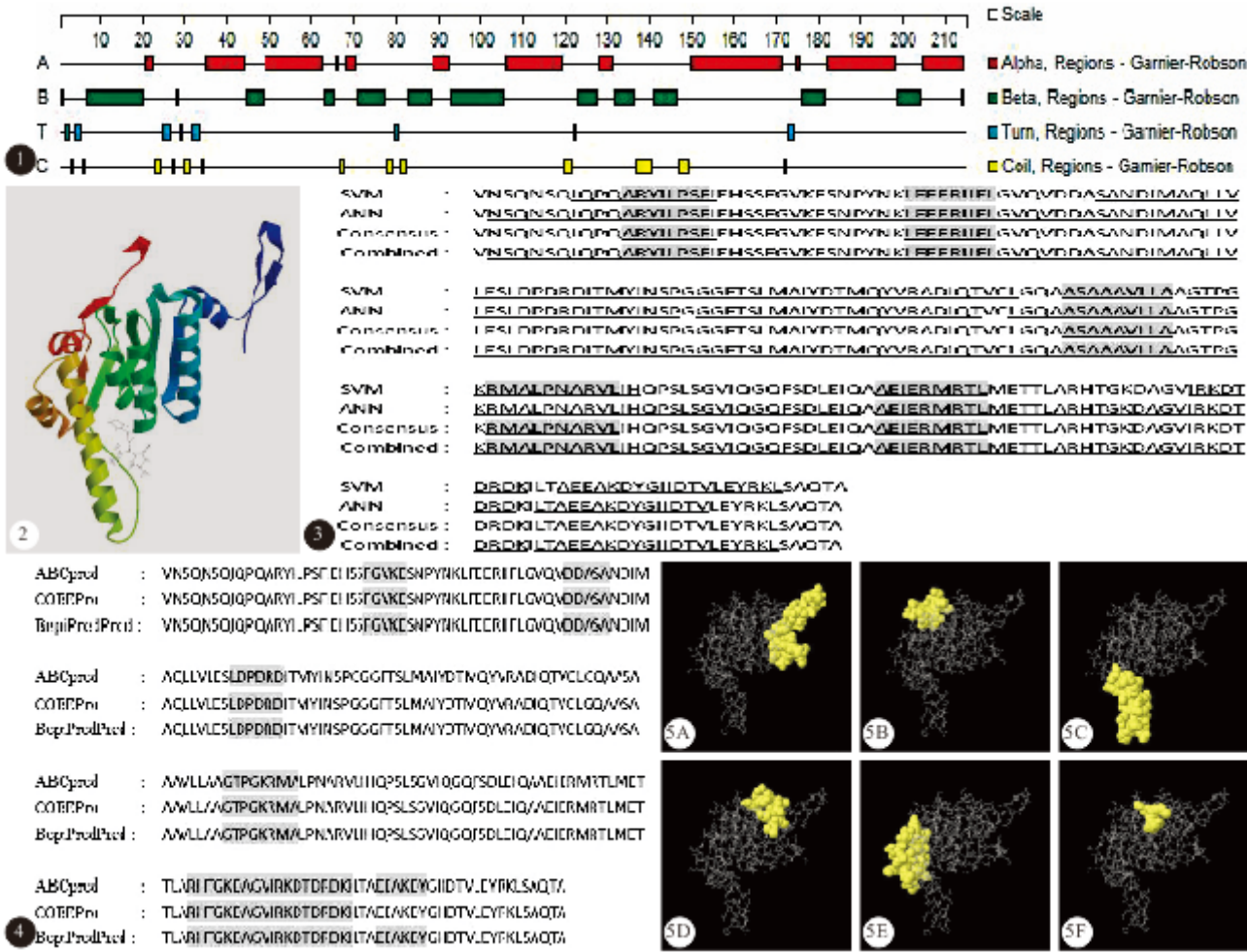


图 1 结核分枝杆菌 clpP2 二级结构预测 图 2 结核分枝杆菌 clpP2 的三级结构预测 图 3 SVM、ANN、Consensus、Combined 4 种方法对 clpP2 蛋白 CTL 表位预测结果 图 4 ABCpred、COBEPro 和 BepiPred 3 种方法对 clpP2 蛋白 B 细胞线性表位预测结果 图 5 clpP2 蛋白 B 细胞构象表位预测结果

Fig 1 Prediction of secondary structure of *M. tuberculosis* clpP2 protein Fig 2 Conformation prediction of *M. tuberculosis* clpP2 protein Fig 3 Potential CTL epitopes in clpP2 predicted by only Support Vector Machine based (SVM), only Artificial Neural Network based (ANN), Consensus and Combined algorithms Fig 4 Potential liner B-cell epitopes in clpP2 predicted by ABCpred, COBEPro and BepiPredPred Fig 5 Three-dimensional representation of predicted discontinuous B-cell epitopes

Fig 1: Alpha: Alpha-spiral region; Beta: Beta-collapse region; Tum: Region of the corner; Coil: Random-coli region; Fig 3: Sequences with underline represent CTL epitopes detected by each prediction algorithm. The 4 epitopes detected by all 4 algorithms are in the shaded areas; Fig 4: Sequences with underline represent B-cell epitopes detected by each B-cell prediction algorithm. The 6 epitopes detected by all 3 algorithms are in the shaded areas; Fig 5: Yellow areas highlight discontinuous epitopes in representing amino acids 1-31 (A), 61-71 (B), 95-99 (C), 138-162 (D), 173-176 (E), 202-214 (F)

表 1 VaxiPred 对 clpP2 蛋白 Th 表位预测结果			
Table 1 Potential Th-cell epitopes in clpP2 predicted by VaxiPred			
Allele	Sequence	Start position	
HLA-DRB1 * 0101	YRKLSAQTA	205	
	YVRADIQTV	94	
	MRTLMMETTL	159	
HLA-DRB1 * 0102	YRKLSAQTA	205	
	MRTLMMETTL	159	
	VLIHQPSLS	131	
	LLVLESLDP	59	
	VLLAAGTPG	113	
	IQPQARYZL	7	

2.3.2 clpP2 B 细胞表位预测 采用 ABCpred(窗口长度为 16 mer,阈值为 0.8)对 clpP2 进行评分预测,获得 10 个可能的 B 细胞表位(表 2);采用 COBEPro 进行筛选,获得 8 个可能的 B 细胞表位;采用 BepiPredPred(阈值为 0.35,此时灵敏度和准确性最合适)进行筛选,获得 9 个可能的 B 细胞表位;综合 3 种方法预测结果如图 4,在 clpP2 蛋白的第 25~29、49~53、66~71、119~126、170~188、192~197 位氨基酸区段可能存在 B 细胞线性表位。

表 2 ABCpred 对 clpP2 蛋白 B 细胞表位预测结果

Table 2 Potential liner B-cell epitopes in clpP2 predicted by ABCpred

Number	Sequence	Start position	Score
1	AGVIRKDTDRDKILTA	176	0.96
2	YGIIDTVLEYRKLSAQ	197	0.95
3	GVQVDDASANDIMAQL	45	0.90
4	PSFIEHSSFGVKESNP	17	0.89
5	LLAAGTPGKRMALPNA	115	0.87
6	RDITMYINSPGGGFTS	70	0.85
7	QLLVLESLDAPRDITM	59	0.84
8	RVLIHQPSLSGVIQGG	131	0.84
9	ILTAEAAKDYGIIDTV	188	0.82
10	RHTGKDAGVIRKDTDR	170	0.81

利用在线软件预测 clpP2 蛋白的 B 细胞构象表位,结果见图 5,预测出 6 个可能的 B 细胞构象表位,分别是位于 1~31 位的 VNSQNS QIQPQAR YILPSFIEHSSFGVKESN (图 5A)、61~71 位的 LVLESLDAPDRD (图 5B)、95~99 位的 YVRAD (图 5C)、138~162 位的 SLSGVIQGGQFSDLEIQA AEIERMRT (图 5D)、173~176 位的 GKDAGVIRKDTDRDKILTAEAAKD (图 5E)、202~214 位的 TVLEYRKLSAQTA (图 5F)。

2.4 免疫学相关信息分析

2.4.1 clpP2 蛋白的亲水性、可塑性和表面可及性
利用 DNASTar 分析 clpP2 蛋白的亲水性、可塑性和表面可及性,与 2.3 中预测的抗原表位的位置比较发现,B 细胞潜在表位多位于亲水性强、可塑性强、表面可及性指数较高的区域,提示该区域暴露于表面的概率较大,且易与抗体结合,作为 B 细胞抗原表位的可能性很大。

2.4.2 clpP2 的信号肽结构和跨膜结构的预测
利用 SignalP 和 TMHMM 软件对 clpP2 蛋白的预测结果表明,clpP 蛋白中不存在信号肽结构,也不存在疏水性较强的跨膜结构,因此将 clpP2 蛋白作为亚单位疫苗和检测抗原的候选蛋白时,能够获得较高的表达量。

2.4.3 BLAST 分析
通过 BLAST 分析,发现 clpP2 蛋白的氨基酸序列与人类线粒体前体——ATP 依赖的 clp 蛋白酶亚基(ATP-dependent Clp protease proteolytic subunit, mitochondrial precursor)的同源性较高,达到 44%(190 个氨基酸中有 84 个),同源性主要位于 15~204 位氨基酸之间,同源性高达 67%。

3 讨论

结核卡介苗(bacille calmette-guerin, BCG)是目前唯一批准使用的结核疫苗,但近年众多研究发现,BCG 虽然可以有效预防儿童结核病^[9],但对青少年及成人肺结核的保护效力争议很大,各项研究显示不同地区及人群保护效力为 0%~80% 不等^[10],而且接种 BCG 后结核菌素试验呈阳性,增加了临床诊断的难度;另外大量的 LTBI 患者无临床症状,难以被发现,成为了活动的传染源。这些都给结核的防治工作带来阻碍。因此,研发新的检测方法、抗结核药物和有效的疫苗成为当今防治结核病的重点。

生物信息学已被广泛应用于疫苗设计、新基因的鉴定、新病毒的发现等方面^[11-12],利用生物信息学方法分析 MTB 蛋白的结构和潜在抗原表位有助于发现重要的诊断、疫苗蛋白靶点和药物作用靶点。本研究对 MTB 的 clpP2 蛋白结构、B 细胞线性表位和构象表位、CTL 表位和 Th 表位、蛋白亲水性、可塑性等进行分析表明,clpP2 存在多个潜在的抗原表位。由于目前的预测方法均存在一定程度的误差,因此本研究选用多个在线软件进行预测后,综合分析选出可能性较大的抗原表位,结果表明,clpP2 蛋白预测出 6 个 B 细胞线性表位、6 个 B 细胞构象表位、5 个 CTL 表位以及多个 Th 表位。由此表明,clpP2 蛋白不仅能够被 CD4⁺ T 细胞和 CD8⁺ T 细胞识别,还能被 B 细胞识别,具有诱导细胞免疫和体液免疫的潜能;而且 clpP2 蛋白这些区域亲水性较高,无疏水性较高的跨膜结构,易表达;另外,通过 BLAST 分析,发现 clpP2 与人 ATP 依赖的 clp 蛋白酶亚基同源性较高,因此,在抗原表位设计时应避开这些同源性区域。上述发现有助于我们进一步深入研究 clpP2 蛋白的免疫功能,为结核诊断、治疗、预防的候选蛋白提供理论依据。

参 考 文 献

[1] World Health Organisation. Global tuberculosis report 2015 [2015-10-22]. http://www.who.int/tb/publications/Global_report/en/.

[2] World Health Organisation. Guidelines on the Management of Latent Tuberculosis Infection [2015-10-22]. <http://www.who.int/tb/publications/latent-tuberculosis-infection/en>.

[3] PERSONNE Y, BROWN AC, SCHUESSLER DL, et al. *Mycobacterium tuberculosis* ClpP proteases are cotranscribed but exhibit different substrate specificities. *PloS One*, 2013, 8 (4): e60228 [2015-01-19]. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3613350/>. doi: 10.1371/journal.pone.0060228.

(下转第 281 页)

- Rev,2014,40(6):730-738.
- [4] ZHAU HE, LI CL, CHUNG LW. Establishment of human prostate carcinoma skeletal metastasis models. *Cancer*,2000, 88(12 Suppl):2995-3001.
- [5] ROSOL TJ, TANNEHILL-GREGG SH, LEROY BE, *et al.* Animal models of bone metastasis. *Cancer*, 2003, 97 (3 Suppl):748-757.
- [6] VIGNANI F, BERTAGLIA V, BUTTIGLIERO C, *et al.* Skeletal metastases and impact of anticancer and bone-targeted agents in patients with castration-resistant prostate cancer. *Cancer Treat Rev*,2016 (44): 61-73 [2015-06-13]. <http://www.sciencedirect.com/science/article/pii/S0305737216000177>. doi: 10.1016/j.ctrv.2016.02.002.
- [7] KUPERWASSER C, DESSAIN S, BIERBAUM BE, *et al.* A mouse model of human breast cancer metastasis to human bone. *Cancer Res*,2005,65(14):6130-6138.
- [8] ANGELUCCI A, GRAVINA GL, RUCCI N, *et al.* Evaluation of metastatic potential in prostate carcinoma: an *in vivo* model. *Int J Oncol*,2004,25(6):1713-1720.
- [9] FISHER JL, SCHMITT JF, HOWARD ML, *et al.* An *in vivo* model of prostate carcinoma growth and invasion in bone. *Cell Tissue Res*,2002,307(3):337-345.
- [10] BALAJI KC, HURYK RF, VERHULST S, *et al.* Growth of heterotopic LNCaP prostate cancer tumor in nude mice is not affected by dietary calcium. *Prostate*,2001,48(4):265-273.
- [11] HRICAK H, CHOYKE PL, EBERHARDT SC, *et al.* Imaging prostate cancer: a multidisciplinary perspective. *Radiology*,2007,243(1):28-53.
- [12] LECOUVET FE, LHOMMEL R, PASOGLOU V, *et al.* Novel imaging techniques reshape the landscape in high-risk prostate cancers. *Curr Opin Urol*,2013,23(4):323-330.
- [13] FRADET A, SOREL H, DEPALLE B, *et al.* A new murine model of osteoblastic/osteolytic lesions from human androgen-resistant prostate cancer. *PLoS One*,2013,8(9): e75092[2016-02-02]. <http://dx.doi.org/10.1371/journal.pone.0075092>.
- [14] OYEN WJ, WITJES JA, CORSTENS FH. Nuclear medicine techniques for the diagnosis and therapy of prostate carcinoma. *Eur Urol*,2001,40(3):294-299.
- [15] BLAKE GM, PARK-HOLOHAN SJ, COOK GJ, *et al.* Quantitative studies of bone with the use of ^{18}F -fluoride and $^{99\text{m}}\text{Tc}$ -methylene diphosphonate. *Semin Nucl Med*,2001,31(1):28-49.
- [16] LANGSTEGER W, HEINISCH M, FOGELMAN I. The role of fluorodeoxyglucose, ^{18}F -dihydroxyphenylalanine, ^{18}F -choline, and ^{18}F -fluoride in bone imaging with emphasis on prostate and breast. *Semin Nucl Med*,2006,36(1):73-92.
- [17] GRECCHI E, O' DOHERTY J, VERONESE M, *et al.* Multimodal partial-volume correction: application to ^{18}F -fluoride PET/CT bone metastases studies. *J Nucl Med*, 2015,56(9):1408-1414.

(2016-08-17 收稿,2016-12-25 修回)

编辑 吕 熙

(上接第 247 页)

- [4] 陈 林, 蓝林华, 何海栋, 等. ClpP 蛋白酶研究进展: 从细菌到人线粒体. *中国细胞生物学学报*,2014,36(6):717-725.
- [5] BIASINI M, BIENERT S, WATERHOUSE A, *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*, 2014,42 (Web Server issue): W252-W258 [2015-01-19]. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4086089/>. doi: 10.1093/nar/gku340.
- [6] ARNOLD K, BORDOLI L, KOPP J, *et al.* The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*,2006,22(2): 195-201.
- [7] BENKERT P, BIASINI M, SCHWEDE T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*,2011,27(3):343-350.
- [8] 扈鸿霞, 方维焕, 刘建华, 等. PRRSV GP5 基因的原核表达及其免疫性分析. *动物医学进展*,2008,29(5):32-35.
- [9] BISHAI W, SULLIVAN Z, BLOOM BR, *et al.* Bettering BCG: a tough task for a TB vaccine? *Nat Med*,2013,19(4): 410-411.
- [10] NARAYANAN PR. Influence of sex, age & nontuberculous infection at intake on the efficacy of BCG: reanalysis of 15 year data from a double-blind randomized control trial in south India. *Indian J Med Res*,2006,123(2):119-124.
- [11] PETRIZZO A, TORNESELLO M, BUONAGURO FM, *et al.* Immunogenomics approaches for vaccine evaluation. *J Immunotoxicol*,2012,9(3):236-240.
- [12] LIU EB, WADFORD DA, SETO J, *et al.* Computational and serologic analysis of novel and known viruses in species human adenovirus D in which serology and genomics do not correlate. *PloS One*,2012,7(3):e33212[2015-01-19]. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3302849/>. doi: 10.1371/journal.pone.0033212.

(2016-07-26 收稿,2016-10-15 修回)

编辑 余 琳