

基础研究

DOI:10.13406/j.cnki.cyxh.001911

HIV-1 CRF01_AE 毒株 gp41 蛋白生物信息学预测

邵继平, 谢学立, 刘莹, 郭燕子, 符健

(海南医学院海南省药物临床前药理毒理学研究重点实验室, 海口 571199)

【摘要】目的:对 HIV-1 gp41 蛋白的膜外区和跨膜区进行预测,并分析该蛋白的结构和功能。**方法:**从 NCBI GenBank 数据库获取 gp41 基因及其蛋白编码序列,利用 ExPASy ProtParam tool、ProtScale、SignalP 4.0 Server、TMHMM Server v.2.0、SWISS-MODEL、PredictProtein 等工具对 gp41 蛋白的生物学特性进行预测。**结果:**HIV-1 gp41 基因 ORF 长度为 666 bp,编码 222 个氨基酸,蛋白分子式为 $C_{1164}H_{1815}N_{317}O_{325}S_6$,等电点(isoelectric point, pI)=8.60,属于稳定、疏水性蛋白。该蛋白具有跨膜区,信号肽切割位点位于第 20 和 21 位氨基酸处。其二级结构由无规则卷曲、 β 折叠和 α 螺旋组成,属于全螺旋型蛋白。而且该蛋白具有 N-端信号肽,能引导蛋白分泌到胞外。此外 gp41 蛋白具有 4 个蛋白质-蛋白质和 1 个蛋白质-核酸结合位点,这些位点能调控 gp41 蛋白的多种生物学功能。**结论:**对 gp41 蛋白的结构和功能进行了预测,为深入研究其参与疾病发生发展的分子机制奠定了基础。

【关键词】HIV-1;包膜糖蛋白 gp41;CRF01_AE 毒株;生物信息学预测

【中图分类号】R373

【文献标志码】A

【收稿日期】2018-06-06

Bioinformatic prediction of gp41 protein in HIV-1 CRF01_AE strain

Shao Jiping, Xie Xueli, Liu Ying, Guo Yanzi, Fu Jian

(Drug Preclinical Pharmacological Toxicology Key Laboratory, Hainan Medical University)

【Abstract】Objective: To predict the extracellular domain and transmembrane domain of HIV-1 gp41 protein, and to investigate its structure and function. **Methods:** The gp41 gene and its protein-encoding sequence were obtained from the NCBI GenBank database. ExPASy ProtParam tool, ProtScale, SignalP 4.0 Server, TMHMM Server v.2.0, SWISS-MODEL and Predict Protein software were used to predict the biological properties of gp41 protein. **Results:** The HIV-1 gp41 gene had an ORF length of 666 bp and encoded 222 amino acids. Its molecular formula was $C_{1164}H_{1815}N_{317}O_{325}S_6$, and the isoelectric point was 8.60. It was a stable and hydrophobic protein. This protein had a transmembrane region and the signal peptide cleavage site was located between the amino acids 20 and 21. Its secondary structure mostly consisted of random coil, β -fold, and α -helix, and thus it was an all-helix protein. It also had an N-terminal signal peptide and guided the extracellular secretion of protein. In addition, gp41 protein had 4 protein-protein binding sites and 1 protein-DNA binding site, which could regulate various biological functions of gp41 protein. **Conclusion:** This study predicts the structure and function of gp41 protein and lays a foundation for research on the molecular mechanism of gp41 protein in disease development and progression.

【Key words】 HIV-1; envelope glycoprotein gp41; CRF01_AE strain; bioinformatics prediction

HIV-1 病毒(human immunodeficiency virus type 1, HIV-1)是艾滋病的致病媒介,目前在全球造成 3 500 万人感染。HIV-1 能攻击人体免疫系统 CD4 T 淋巴细胞,使免疫功能丧失,引起各种机会感染,病死率高。HIV 感染者要经过长达 10 年或更长的潜伏期后才发展成艾滋病,后期死于恶性肿瘤或全身衰竭。虽然科学家们研制出了近百种治疗性药物,

但至今尚无根治艾滋病的特效药^[1-2]。

目前科学界尝试利用免疫疗法来治疗 HIV 感染。HIV-1 包膜刺突(envelope spike, Env)是病毒粒子表面的唯一抗原,是中和抗体的唯一靶点。作为疫苗免疫原,Env 是该领域的研究热点^[3-5]。HIV-1 Env 基因全长约 2.6 kb,是变异性最强的基因。Env 前体 gp160 被高甘露糖型聚糖广泛修饰,蛋白酶解后部分前体变成复杂的碳水化合物。成熟的 Env 是由 3 个表面糖蛋白(exterior glycoproteins)gp120 和 3 个跨膜糖蛋白(transmembrane glycoproteins)gp41 组成的三聚体,由于 gp41 比 gp120 的序列更为保守且糖基化位点少,基于 gp41 的疫苗免疫原可能具有

作者介绍:邵继平, Email: jpsaosl@163.com,

研究方向:抗病毒药物。

通信作者:符健, Email: fujian@163.com。

基金项目:海南省科技厅社会发展资助项目(编号:ZDYF2016126)。

优先出版: <http://kns.cnki.net/kcms/detail/50.1046.R.20181105.0905.018.html>

(2018-11-06)

更广谱的保护作用。

近年来,中国新发 HIV-1 感染病例中,CRF01_AE 毒株的比例逐渐上升,因此本研究综合应用生物信息学方法预测了 HIV-1 CRF01_AE gp41 跨膜糖蛋白的一般理化性质、亲疏水性、跨膜区和信号肽、二级结构及结构域等方面的特点,提高对 gp41 蛋白结构和功能的整体认识,为深入研究 gp41 蛋白参与艾滋病发生发展的分子机制奠定了理论基础^[6-8]。

1 材料与方法

1.1 材料

HIV-1 gp41 基因组信息及其编码蛋白质的氨基酸序列由 NCBI GenBank (www.ncbi.nlm.nih.gov/genbank/) 提供,其 GenBank 登录号为 KY580724.1。

1.2 方法

1.2.1 一般理化性质分析 利用 ExPASy ProtParam tool (<http://expasy.org/tools/portparam.html>) 软件对 gp41 蛋白的一级结构和基本理化性质进行预测,包括蛋白的分子量、氨基酸组成、等电点 (isoelectric point, pI) 及稳定性等。

1.2.2 亲疏水性分析 利用 ExPASy ProtParam (<http://www.expasy.org/tools/protscale.html>)、ProtScale (<https://web.expasy.org/protscale/>) 等在线工具对 gp41 蛋白的亲疏水性进行预测。

1.2.3 信号肽和跨膜区预测 在起始密码子后面有一段编码疏水性氨基酸的序列,被称为信号肽。信号肽把蛋白质引到亚细胞器内,对于蛋白质的转运具有重要意义^[9-10]。本文利用 SignalP 4.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) 和 TMHMM Server v.2.0 工具 (<http://www.cbs.dtu.dk/services/TMHMM/>) 预测了 gp41 蛋白的信号肽和跨膜区特征。

1.2.4 二级结构及空间结构预测 二级结构 (secondary structure) 是指多肽链本身的折叠和盘绕方式,其基本单位主要有 α 螺旋、 β 折叠、 β 转角和无规则转曲 4 种。根据 gp41 蛋白的氨基酸序列,本文采用 PredictProtein (<https://www.predictprotein.org/>) 在线工具对 HIV-1 gp41 蛋白的二级结构进行预测,并应用 SWISS-MODEL 软件 (<http://swissmodel.expasy.org/interactive>) 进行空间结构的同源建模^[11-14]。

2 结果

2.1 一般理化性质分析

在本实验室前期研究的基础上^[15],选择 HIV-1 CRF01_AE 毒株 gp41 基因上的 666 bp 作为 ORF,共 222 个氨基酸。利用 ExPASy ProtParam tool 软件对其进行分析,结果如下:该蛋白的分子式为 $C_{1164}H_{1815}N_{317}O_{325}S_6$,包含 C、H、N、O、S 5 种元素,由 3 627 个原子构成,原子构成比为 C:H:N:O:S=1 164:1 815:317:325:6,分子量为 25 642.50。该蛋白含 20 种氨基酸,其中亮氨酸 (Leu) 25 个,占 11.3%;异亮氨酸 (Ile) 21 个,占 9.5%;谷氨酰胺 (Gln) 20 个,占 9.0%;其余氨基酸所占比例均在 9% 以下,如图 1 所示。氨基酸的负电荷总数 (Asp+Glu) 为 19,

正电荷总数 (Arg+Lys) 为 21,两者的比例为 19:21。理论 pI 为 8.60。蛋白的不稳定指数为 39.85,提示该蛋白能稳定存在于溶液中。

HIV-1 gp41 蛋白的氨基酸序列如下:

```

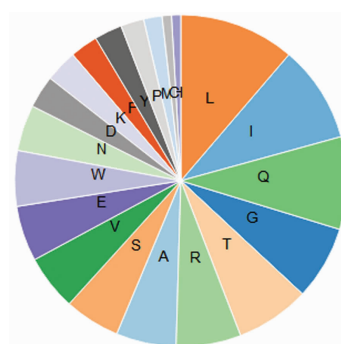
10      20      30      40      50      60
GIGAMIFGFL GAAGSTMGAA SITLTVQARQ LLTGIVQQQS NLLRAIEAQQ HLLQLTVWGI

70      80      90      100     110     120
KQLQARVLAV ERYLRDQRFL GLWGCSGKTI CTTSVPWNST WSNKTYKEIW DNMTWIEWER

130     140     150     160     170     180
EIGNYTNQIY TLLTESQNQQ DKNEKDLLAL DQWASLWNWF DITRWLWYIK IFIMVVGGLI

190     200     210     220
GLRIHVAVLV IVNRVRQCYS PLSFQTPHIQ QREPDRPERIEE

```



注 :L, Leu; I, Ile; Q, Gln; G, Gly; T, Thr; R, Arg; A, Ala; S, Ser; V, Val; E, Glu; W, Trp; N, Asn; D, ASP; K, Lys; F, Phe; Y, Tyr; P, Pro; M, Met; C, Cys; H, His

图 1 gp41 蛋白的氨基酸组成

2.2 亲疏水性分析

氨基酸亲疏水性是分析蛋白质二级结构的前提,在蛋白质折叠、预测蛋白质作用位点和抗原位点、分析蛋白质跨膜区等方面起重要作用。本研究通过 ExPASy ProtParam 和 ProtScale 在线工具对 HIV-1 gp41 蛋白的亲疏水性进行分析,结果如图 2 所示:该蛋白的脂肪系数 (aliphatic index) 为 102.34,单个氨基酸的极性最小值 (MIN) 为 4.9,最大值 (MAX) 为 13,总平均亲水性 (grand average of hydropathicity, GRAVY) 为 -0.132,属于疏水性蛋白。

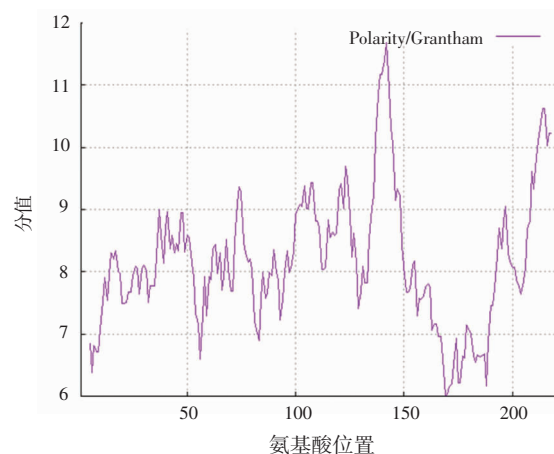


图 2 gp41 蛋白的亲疏水性分析

2.3 跨膜区和信号肽分析

SignalP 4.0 Server 是一个信号肽预测服务器,本文采用 SignalP 4.0 server(<http://www.cbs.dtu.dk/services/SignalP/>)在线工具预测 gp41 蛋白有无信号肽,采用神经网络法(neural network algorithm, NN)计算 gp41 氨基酸序列中是否存在潜在的信号肽剪切位点。信号肽剪切位点的特征是:C 值最大, S 值陡峭, Y 值具有最高峰。NN 法预测主要涉及 C、S 和 Y 3 个数值。此外,还有 S-mean 和 D 2 个判定值。C 为剪切位点值, D 值主要用来区分蛋白是否为分泌蛋白。图 3 右上角有不同颜色标识的 C 值、S 值和 Y 值曲线,其变化趋势显示蛋白是否含有信号肽以及潜在的剪切位点。gp41 蛋白的 C-max 在第 21 位氨基酸处,为 0.179,根据 C、S 和 Y 算出潜在的信号肽剪切位点位于第 20 和 21 位氨基酸之间, D 值为 0.519 (>0.450),提示该蛋白是分泌蛋白。

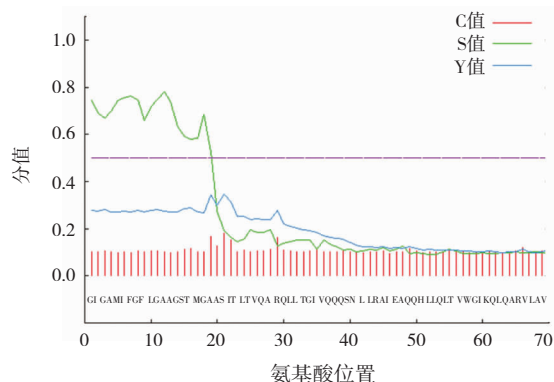


图 3 gp41 蛋白的信号肽分析

α 螺旋跨膜区由 20~30 个疏水性氨基酸组成,本文采用 TMHMM Server v.2.0 软件预测 gp41 蛋白有无跨膜区。图 4 的纵坐标是概率,横坐标是氨基酸序列。结合图 4 和表 1,不难判断 gp41 的跨膜区位于第 165~187 位氨基酸处,是具有 N-端信号肽的跨膜蛋白。

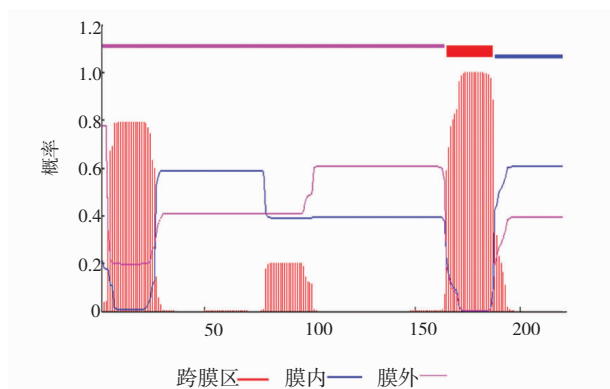


图 4 gp41 蛋白的跨膜区分析

2.4 二级结构及空间结构分析

蛋白质二级结构分型标准为:H、E、T 和 C 四型,其中 H 表示 α 螺旋, E 表示 β 折叠, L 表示无规则卷曲, C 表示线圈。当 $H\% > 45\%$ 且 $E\% < 5\%$ 时,为全螺旋型;当 $H\% < 5\%$ 且 $E\% > 45\%$ 时,为全折叠型;当 $H\% > 30\%$ 且 $E\% > 20\%$ 时,为螺旋-折

叠型;其他情况则为混合型。本文采用 PredictProtein 在线工具对 HIV-1 gp41 蛋白的二级结构进行预测,结果如图 5 所示:红色表示螺旋区域,蓝色为片层结构,绿色为无规则卷曲。gp41 蛋白有 168 个 α 螺旋结构,占 75.68%;8 个 β 折叠结构,占 3.60%;46 个无规则卷曲结构,占 20.72%,因此 gp41 为全螺旋型蛋白。

表 1 gp41 蛋白的跨膜区参数分析

信号肽和跨膜区预测	参数
长度	222
跨膜区数目	1
跨膜螺旋的氨基酸数目	44.690 41
期望值	17.573 17
总概率	0.220 19
可能的信号肽	N-端
膜外区	1 164
跨膜螺旋	165 187
膜内区	188 222

此外,本研究利用 SWISS-MODEL 在线工具预测 gp41 蛋白三级结构,并进行同源模型。从众多的模板中选择 1env. 1.A 模板序列,该序列与 gp41 蛋白的氨基酸序列同源性高达 74%(图 6)。该图来自 <https://swissmodel.expasy.org/interactive/uNDEXA/models/> 网站。

2.5 蛋白质结合位点预测

研究蛋白质-蛋白质和蛋白质-核酸的相互作用是理解蛋白质生物学功能的基础,是后基因组时代的研究核心。本研究通过 PredictProtein 在线工具预测了 gp41 蛋白的结合位点,如图 7 所示:gp41 具有 1 个蛋白质-核酸结合位点,黄色圆圈表示;4 个蛋白质-蛋白质结合位点,红色方块表示。通过预测 gp41 蛋白结合位点增加了对其表达规律和生物功能的了解。

3 讨论

免疫疗法通过激活自身的免疫系统来治疗疾病,其难点在于如何提高免疫应答。研究表明,通过将多个优势表位组合,引起强烈的免疫反应能解决这一难题。HIV-1 gp41 为跨膜蛋白,是艾滋病病毒表面重要的糖蛋白,参与病毒进入靶细胞的融合过程。gp41 分为膜外区(aa512~683)、跨膜区(aa684~705)和胞内区(aa705~856)三部分。膜外区直接参与 HIV 包膜和靶细胞膜的融合,跨膜区介导多肽链跨越脂双层,富含色氨酸的 MPER 区域具有高度保守的抗原表位,是免疫治疗的理想靶点。Narasimhulu 等^[16]采用定点突变的方法研究了 HIV-1 gp41 MPER 区域与病毒感染力的机制。结果表明当把 MPER 第 666 位的色氨酸 W 突变为丙氨酸 A 和第 675 位的

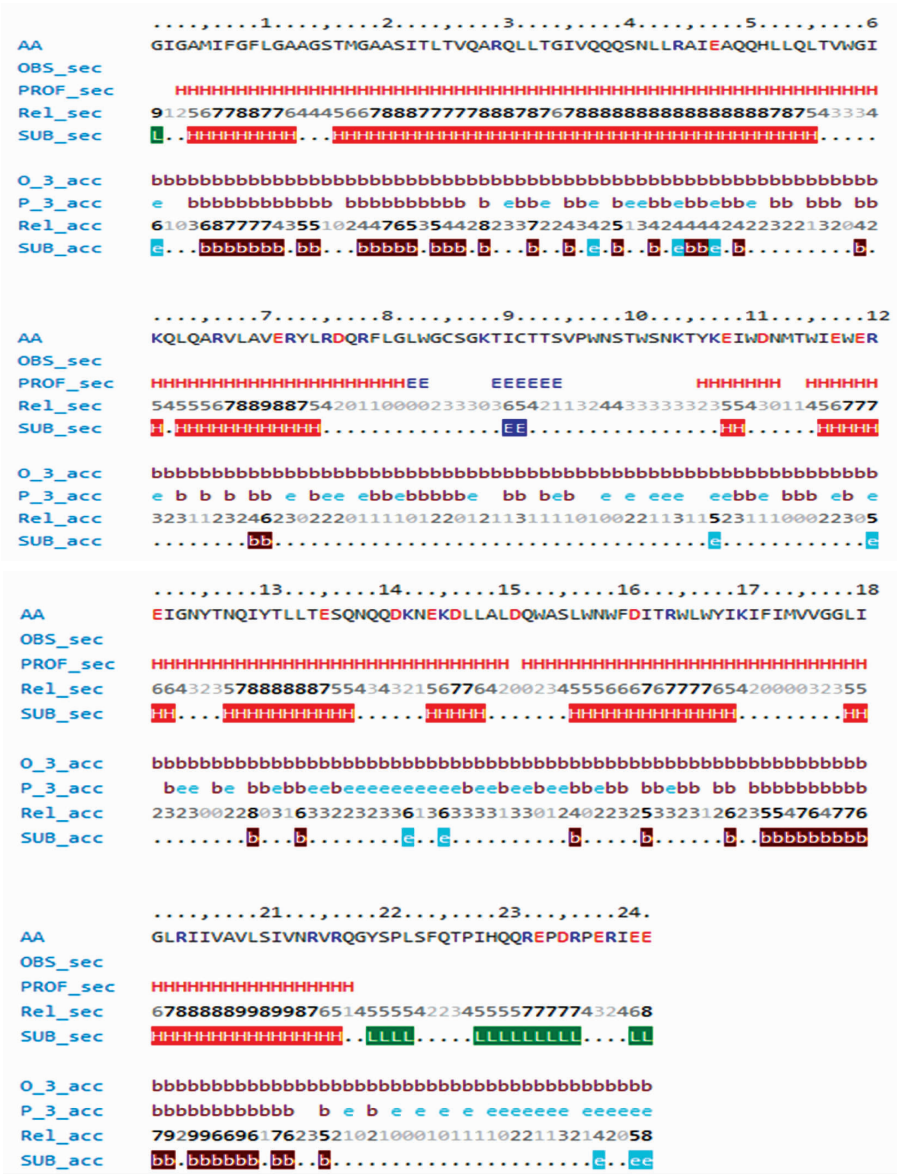


图 5 gp41 蛋白的二级结构分析

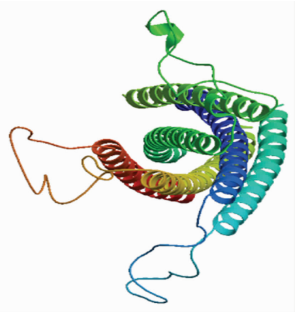


图 6 gp41 蛋白的三级结构分析

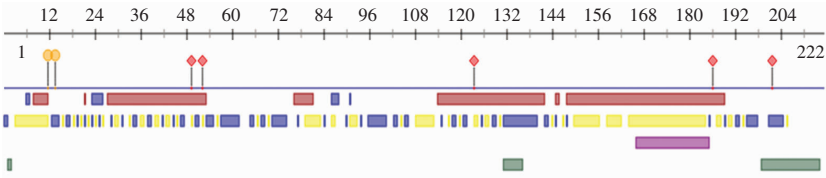


图 7 gp41 蛋白的蛋白质 - 蛋白质和蛋白质 - 核酸结合位点

异亮氨酸 I 突变为 A 时, HIV-1 突变子的感染力下降。Madani 等^[17]设计了拟 CD4 小分子化合物, 它与 HIV-1 Env 结合后能产生抗体, 增加 HIV-1 Env 疫苗的保护性免疫响应, 这种保护作用至少能持

续 6 个月。Keller 等^[18]研究了 gp41 膜外区 E560K 和 Q577R HR1 突变子对六股螺旋 (six-helix bundle, 6HB) 的影响, 发现这些突变子能减少构象变化, 发动 6HB 形成并增加病毒对 mCD4 类似物的敏感性。

Duchemin 等^[19]使用 gp41 Env 特异性 IgA 进行实验,结果证实 2F5-IgA 能诱导携带 HIV-1 Clade A 和 B 毒株的靶细胞,而且 2F5-IgA 和 2F5-IgG 能协同增加 ADCC 介导的靶细胞裂解。这些研究为理解 gp41 蛋白构象变化和 design 新颖的进入抑制剂及稳定的疫苗免疫原提供了新思路。

本研究结果表明, gp41 的不稳定指数为 39.85, 是稳定蛋白。这一特征有利于 gp41 蛋白的提取和纯化,维持天然的构象,不易变性。蛋白质二级结构中 β 转角和无规卷曲区域较易形成抗原表位, B 细胞表位多位于表面可及性好、亲水性高、柔性好的区域,转角区域成为优势抗原表位的可能性较大。这些区域暴露于细胞膜外,易于结合抗体。从 gp41 蛋白的二级结构来看, β 折叠结构 8 个,占 3.60%; 无规则卷曲结构 46 个,占 20.72%,基本满足上述条件。此外, gp41 具有 4 个蛋白质-蛋白质和 1 个蛋白质-核酸结合位点,从而有利于该蛋白质与机体相互作用,参与细胞调控。

生物信息学用数学和计算机知识处理生物数据。通过统计学方法预测其功能,确定数据所含的生物学意义,具有目的性强、信息量大和选择范围多等优点,已被科学界广泛采用。但由于预测方法种类繁多,且各种方法均存在一定的误差。因此,本文在设计基于 gp41 蛋白的药物及免疫原时,采用多种生物信息学软件对 gp41 基因及其编码的蛋白质进行综合分析,并对所有预测结果进行比较,选择不同预测结果中重叠率较高的区域,后期再结合体内外实验加以验证^[20]。本研究在构建 HIV-1 免疫治疗疫苗及多肽药物开发方面具有重要意义,为深入研究 gp41 参与疾病发生发展的分子机制奠定了基础。

参 考 文 献

- [1] Garber DA, Silvestri G, Feinberg MB. Prospects for an AIDS vaccine: three big questions, no easy answers[J]. *Lancet Infect Dis*, 2004, 4(7): 397-413.
- [2] 邵继平, 姜世勃, 刘叔文. 基于 gp41 的 HIV 亚单位疫苗研究进展[J]. *细胞与分子免疫学杂志*, 2011, 27(9): 1047-1050.
- [3] Mathys L, Balzarini J. The role of N-glycans of HIV-1 gp41 in virus infectivity and susceptibility to the suppressive effects of carbohydrate-binding agents[J]. *Retrovirology*, 2014, 11: 107.
- [4] Derdeyn CA, Decker JM, Bibollet-Ruche F, et al. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission[J]. *Science*, 2004, 303(5666): 2019-2022.
- [5] Bunnik EM, Euler Z, Welkers MR, et al. Adaptation of HIV-1 envelope gp120 to humoral immunity at a population level [J]. *Nat Med*, 2010, 16(9): 995-997.
- [6] Hoxie JA. Toward an antibody-based HIV-1 vaccine[J]. *Annu Rev Med*, 2010, 61: 135-152.
- [7] Wei X, Decker JM, Wang S, et al. Antibody neutralization and escape by HIV-1[J]. *Nature*, 2003, 422(6929): 307-312.
- [8] Shi Y, Feng Y, Ding X, et al. Short communication: the distribution of potential N-linked glycosylation sites in Gp120 differs among major hiv-1 subtypes circulating in China[J]. *AIDS Res Hum Retroviruses*, 2016, 32(1): 101-108.
- [9] Haruyama Y, Kataoka H. Glypican-3 is a prognostic factor and an immunotherapeutic target in hepatocellular carcinoma[J]. *World J Gastroenterol*, 2016, 22(1): 275-283.
- [10] Massahi A, Callk P. In-silico determination of Pichia pastoris signal peptides for extracellular recombinant protein production[J]. *J Theor Biol*, 2015, 364(7): 179-188.
- [11] Hornbeck PV, Kornhauser JM, Tkachev S, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse[J]. *Nucleic Acids Res*, 2012, 40: D261-270.
- [12] Huang HD, Lee TY, Tzeng SW, et al. KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites[J]. *Nucleic Acids Res*, 2005, 33: W226-229.
- [13] Eisenhaber B, Eisenhaber F. Prediction of posttranslational modification of proteins from their amino acid sequence[J]. *Methods Mol Biol*, 2010, 609: 365-384.
- [14] Gao J, Thelen JJ, Dunker AK, et al. Musite, a tool for global prediction of general and kinase-specific phosphorylation sites [J]. *Mol Cell Proteomics*, 2010, 9(12): 2586-2600.
- [15] 邵继平, 姜世勃, 刘叔文. 基于中国 HIV-1 CRF01_AE 重组亚型 gp41 NHR 结构域亚单位疫苗的研究[J]. *细胞与分子免疫学杂志*, 2012, 28(12): 1250-1257.
- [16] Narasimhulu VGS, Bellamy-McIntyre AK, Laumaea AE, et al. Distinct functions for the membrane-proximal ectodomain region (MPER) of HIV-1 gp41 in cell-free and cell-cell viral transmission and cell-cell fusion[J]. *J Biol Chem*, 2018, 293(16): 6099-6120.
- [17] Madani N, Princiotto AM, Mach L, et al. A CD4-mimetic compound enhances vaccine efficacy against stringent immunodeficiency virus challenge[J]. *Nat Commun*, 2018, 9(1): 2363.
- [18] Keller PW, Morrison O, Vassell R, et al. HIV-1 gp41 residues modulate CD4-induced conformational changes in the envelope glycoprotein and evolution of a relaxed conformation of gp120[J]. *J Virol*, 2018, 92(16): 583.
- [19] Duchemin M, Khamassi M, Xu L, et al. IgA targeting human immunodeficiency virus-1 envelope gp41 triggers antibody-dependent cellular cytotoxicity cross-clade and cooperates with gp41-specific IgG to increase cell lysis[J]. *Front Immunol*, 2018, 9: 244.
- [20] 邵继平, 姜世勃, 刘叔文. HIV-1 CRF07_BC 毒株 gp41 NHR 结构域 N51 的表达及结构分析[J]. *南方医科大学学报*, 2012, 32(12): 1737-1740.

(责任编辑: 冉明会)