

猪内源性反转录病毒囊膜蛋白基因的克隆及其蛋白二级结构和 B 细胞表位预测

吴健敏^{1,2}, 孙建华², 吕茂民¹, 陈忠伟², 阳玉彪¹, 赵武², 陈凤莲², 黄红梅², 章金刚^{1*}

(1. 军事医学科学院野战输血研究所, 北京 100850; 2. 广西兽医研究所, 南宁 530001)

摘要: 猪内源性反转录病毒(PERV)是与猪-人异种移植病原安全性密切相关的一类病毒。*env* 基因编码病毒的囊膜蛋白,它与病毒的亚型分类、宿主感染范围、细胞的嗜性以及对宿主细胞的感染机制、诱导宿主产生中和抗体等密切相关。本研究利用 RT-PCR 的方法,从五指山小型猪外周血淋巴细胞中扩增 PERV 的囊膜蛋白基因并进行测序,随后用生物信息学相关软件和方法,对 PERV-Env 蛋白二级结构及 B 细胞表位进行预测。经综合分析评价,结果发现 PERV-Env 蛋白有 18 个可能的 B 细胞优势抗原表位区域,7 个可能的糖基化位点。该分析预测结果不但有利于 PERV 疫苗的设计、单抗及诊断试剂研制,而且将有助于分析 Env 蛋白的功能及 PERV 对来源细胞的感染机制。

关键词: PERV; Env 蛋白; B 细胞表位; 二级结构; 预测

中图分类号: S852.65

文献标识码: A

文章编号: 0366-6964(2007)04-0412-05

Clone and Prediction of the Secondary Structure and B Cell Epitopes for the Env Protein of Porcine Endogenous Retrovirus

WU Jian-min^{1,2}, SUN Jian-hua², LÜ Mao-min¹, CHEN Zhong-wei², YANG Yu-biao¹, ZHAO Wu², CHEN Feng-lian², HUANG Hong-mei², ZHANG Jin-gang^{1*}

(1. Institute of Transfusion Medicine, AMMS, Beijing 100850, China ;

2. Guangxi Veterinary Institute, Nanning 530001, China)

Abstract: Porcine endogenous retrovirus(PERV) is a kind of virus correlatively to the safety of pathogens during pig xenografts. The *env* gene encodes the Env protein, which plays an important role in virus subtyping, host ranging, cells tropism and mechanism of virus infection, as well as induction of neutralizing antibody secretion. In this research, the envelope gene of PERV was amplified by RT-PCR and sequenced from the peripheral blood lymphocyte of WZS minipigs. The secondary structure and B cell epitopes of the Env protein of PERV were analysed and predicted subsequently by the Bioinformatics software. The results showed that the PERV-Env protein was supposed to exist 18 potential antigen epitopes and 7 potential glycosylated sites. These results indicated that it would be helpful not only to the PERV vaccine designing and the preparation of both of monoclonal antibody and diagnostic reagent, but also to research of Env protein function and mechanism of PERV infecting human-derived cells.

Key words: porcine endogenous retrovirus; Env protein; B cell epitope; the secondary structure; prediction

收稿日期: 2006-05-22

基金项目: 国家“十五”重点攻关项目(2004BA717B-02-04); 国家自然科学基金(30270989); 广西壮族自治区应用基础研究专项基金(桂科基 0448041)

作者简介: 吴健敏(1963-), 女, 博士, 主要从事生物制品的病毒安全性研究

* 通讯作者: 章金刚(1963-), 男, 研究员, 博导, E-mail: zhangjg@nic.bmi.ac.cn

猪内源性反转录病毒(Porcine endogenous retrovirus, PERV),是与猪-人异种移植病原安全性密切相关的一类病毒。自 1997 年发现其在体外能够感染多种人源细胞以来,引起了人们对异种移植安全性的广泛关注^[1~3]。PERV 是以前病毒 DNA 形式整合进宿主细胞基因组中,并随细胞染色体复制而复制的反转录病毒。在分类学上属于反转录病毒科哺乳类 C 型反转录病毒属,其基因组为单股、正链、线性 RNA(ssRNA)的二聚体,单体长约 7~11kb。PERV 的基因组结构包括两端非编码区(5′、3′LTR)及 *gag*、*pol*、*env* 等 3 个编码基因。其中 *env* 基因编码病毒的囊膜蛋白(Env),由二条肽键组成,较小的那条键贯穿病毒的囊膜,称为穿膜蛋白(Transmembrane, TM),较大的那条键通过二硫键和氢键与 TM 相连,暴露于囊膜蛋白,称为表面蛋白(Surface glycoprotein, SU)。二者均来自 *env* 基因编码前体的蛋白质,经水解后产生。SU 蛋白分子上有许多糖基化位点,它不但具有受体结合的功能,还具有诱导宿主产生中和抗体的作用,同时还非常容易变异。而 TM 蛋白参与感染过程中病毒与宿主细胞膜的融合,因此 Env 蛋白一直是 PERV 研究的重点。

蛋白质分子的 B 细胞表位预测,最初只用于对病原微生物的表位研究,以寻找疾病特异性相关表位于疫苗的设计、单抗及诊断试剂研制。但之后发现,除抗体主要与蛋白质表面的高电荷及高极性区域相结合以外,其它类型的蛋白质与蛋白质间相互作用也与此相类似,因而基于亲水性参数的预测方案也可作为研究蛋白质与蛋白质间相互作用位点提供有价值的参考信息。为此本研究以来源于五指山猪的内源性反转录病毒为对象,在分析 Env 蛋白二级结构特征的基础上,运用计算机软件分析 Env 的亲水性和可及性,从而对 Env 的 B 细胞表位进行初步探索,旨在为进一步研究 Env 蛋白的功能及 PERV 的感染机制提供科学依据。

1 材料与方法

1.1 材料

PERV-WZS 来自中国农业科学院畜牧研究所五指山小型猪。

1.2 工具酶和试剂

Pyrobest *Taq* DNA 聚合酶、Oligo(dT) 15、AMV 反转录酶、pGEM-T 载体、DNA Purification

Kit、dNTPs 购自 Promega 公司,TRIzol 试剂为 Gibco 公司产品,*E. coli* DH5 α 为本室保存。常规化学试剂由军事医学科学院器材处提供。

1.3 总 RNA 的制备

无菌自小型猪前腔静脉采集 EDTA 抗凝血,0.83% NH₄Cl 低渗破坏红细胞,2 000 r/min 离心 5 min, PBS 洗涤 2 次,即为外周血白细胞(PBL),利用 TRIzol 试剂制备 PBL 总 RNA。

1.4 PERV *env* 基因的克隆和测序

根据 GenBank 已发表的 PERV 核苷酸序列,设计并合成了一对引物,用于扩增 PERV 囊膜全基因。引物序列如下, Pf: 5′-ATGCATCCCACGT-TAAACCG-3′; Pr: 5′-CCTTTCATTCCCCACT-TCTTC-3′。具体操作参照以往报道的方法^[4]进行。最后将鉴定为阳性的菌株送上海华诺生物科技有限公司测序。测序结果登陆 GenBank。

1.5 PERV-Env 蛋白二级结构预测

采用 2 种不同原理的预测方案,一为 DNASTar 分析软件中的 Gamier-Robson 方法预测蛋白质的二级结构,另一方案为 Rost 建立的基于多重序列比较的 PHDsec 二级结构预测方案,在 EMBL 提供的 PHD 服务器(hppt://www. embl-heidelberg. de/predictprotein/predictprotein. html)上进行计算^[5]。

1.6 PERV-Env 蛋白 B 细胞表位预测

选取 3 个主要参数——亲水性参数(Hydrophilicity)^[6]、抗原指数(Antigenicity index)^[7]及可及性参数(Accessibility)^[8]分别进行单参数预测。其中亲水性参数采用 Kyle-Doolittle 的氨基酸亲水性标准,预测其亲水性;抗原性指数按 Jameson wolf 方案预测;蛋白质的表面可能性按 Eminni 方案预测。各参数的意义见相应参考文献。对各个参数的预测结果进行比较,最后进行综合分析以确定 PERV-Env 蛋白的 B 细胞表位。

2 结果

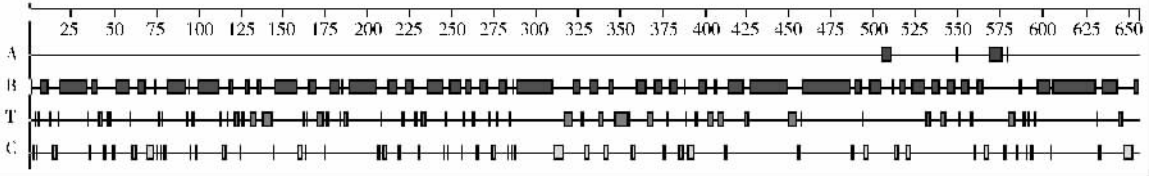
2.1 五指山小型猪 PERV *env* 基因的克隆、测序

经 RT-PCR 扩增得到大小约 2 000 bp 的基因片段,与预期结果一致。将该片段克隆到 T-载体后,进行 PCR 和 *Spe* I 和 *Sph* I 双酶切鉴定,结果与预期相符,表明正确插入。对鉴定正确的重组质粒进行测序,结果来源于 PERV-WZS 的 *env* 基因序列长度为 1 968 bp,编码 656 aa, GenBank 注册号为 AY850198。

2.2 PERV-Env 蛋白二级结构预测

一级结构分析表明 PERV-Env 蛋白由 656aa 组成,其中 1-456 为 SU 蛋白,457-656 为 TM 蛋白。应用 Gamier-Robson 方法分析 Env 蛋白二级结构,第 503-516 区段、第 550 和第 570-576 区段为 α 螺旋区域, β 折叠区域有 38 个区段,不规则地分布于整个 Env 蛋白,与相应的转角区域基本一致(见图 1);以 PHDsec 二级结构预测方案分析计算,在受

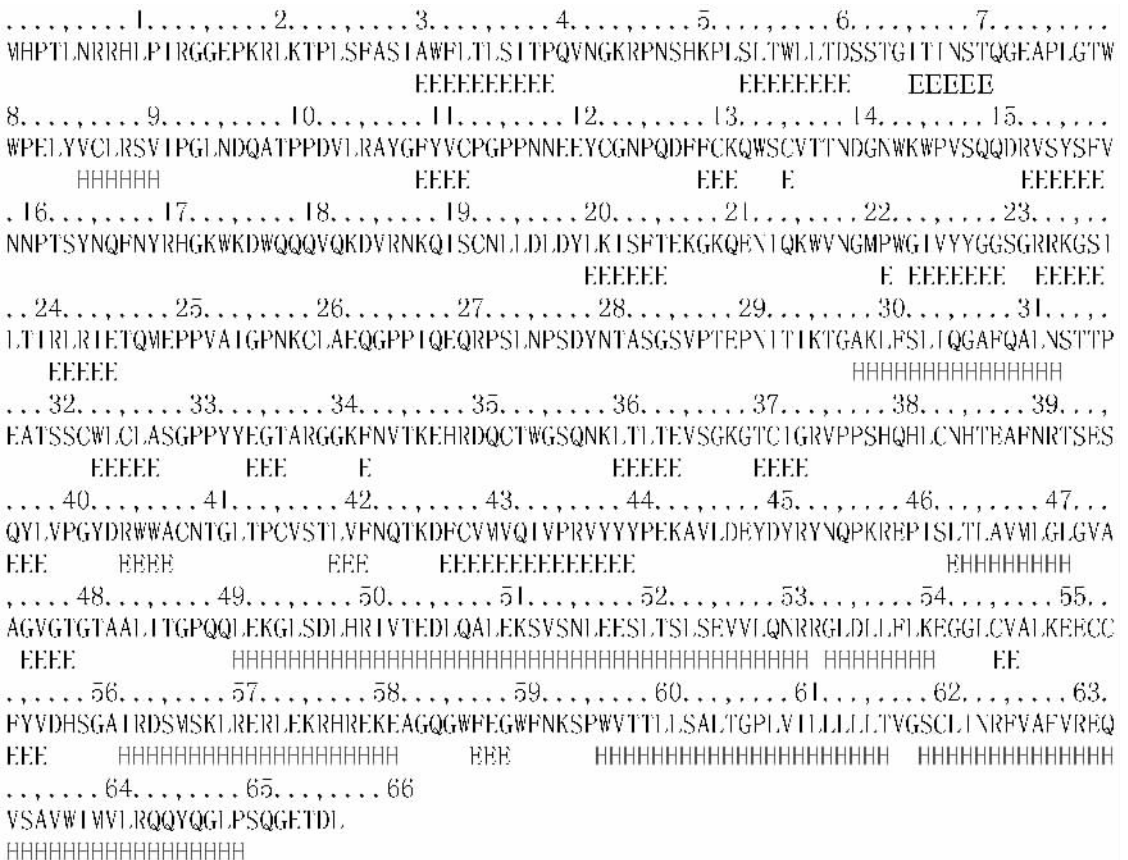
体结合区及产生中和抗体的抗原区 SU 蛋白中,有 7 个可能的糖基化位点,位于 gp70 蛋白(Asn-X-Thr/Ser)中,此外还发现了 Env 蛋白其它的功能区及其作用位点(详见表 1)。用此方案预测 Env 蛋白 α 螺旋区有 8 个区段,其中 491-531、562-581 区段与 Gamier-Robson 方法分析的结果相一致。 β 折叠区有 22 个区段,与 Gamier-Robson 方法分析的结果基本吻合(图 2)。



A. α 螺旋区段; B. β 折叠区段; T. 转角区域; C. 不规则卷曲区域
A. α -helix region; B. β -sheet region; T. Turn region; C. Random coil region

图 1 PERV-Env 蛋白二级结构预测结果 (Gamier-Robson 方法)

Fig. 1 Secondary structure of PERV-Env protein predicted by the method of Gamier-Robson



E. Extended/ β -sheet; H. Heleix/ α -helix

图 2 PERV-Env 蛋白二级结构预测结果 (PHDsec 方案)

Fig. 2 Secondary structure of PERV-Env protein predicted by the method of PHDsec

表 1 PERV-Env 蛋白功能区及作用位点

Table 1 The functional domain and activity site of PERV-Env protein

名称 Name	氨基酸序列及结合位点 Amino acid sequence and binding sites
1. 糖基化位点 N-glycosylation site	NSTQ(68), NITI(291), NSTT(312), NVTK(343), NHTE(384), NRTS(390), NQTK(422)
2. 糖胺聚糖附着位点 Glycosaminoglycan attachment site	SGKG(366)
3. 依赖于 cAMP 和 cGMP 的蛋白磷酸化激酶位点 cAMP- and cGMP-dependent protein kinase phosphorylation site	RKGS(234)
4. 蛋白激酶 C 磷酸化位点 Protein kinase C phosphorylation site	SHK(48), TEK(205), SGR(230), TIR(239), TIK(293), TAR(336), SGK(366)
5. 酪蛋白激酶 II 磷酸化位点 Casein kinase II phosphorylation site	TQGE(70), TPPD(99), TTND(137), SQQD(148), SFTE(203), TQME(246), TTPE(314), TLTE(361), SNLE(514), SLSE(522), SQGE(650)
6. 酪氨酸激酶磷酸化位点 Tyrosine kinase phosphorylation site	RTSESQY(391), KAVLDEYDY(443)
7. 十四(烷)酰化位点 N-myristoylation site	GITINS(64), GMPWGI(219), GSILTI(235), GSVPT(284), GAFQAL(306), GTARGG(335), GTCIGR(369), GLTPCV(411), GLGVAA(470), GVG TGT(476), GLCVAL(543)
8. 酰胺位点 Amidation site	NGKR(42), SGRR(230)
9. 原核(生物)膜脂蛋白类脂(脂质)附着位点 Prokaryotic membrane lipoprotein lipid attachment site	ILLLLLTVGSC(610)

2.3 PERV-Env 蛋白 B 淋巴细胞表位预测

首先进行 Env 蛋白亲水性预测,应用 Kyle-Doolittle 的方法进行,PERV-Env 蛋白的亲水性较高的区域主要有 19 个区段。再按 Jameson wolf 方案预测 Env 蛋白的抗原指数较高的部位,结果发现 Env 蛋白的大多数区域的抗原指数都比较高,包括两端。最后按 Eminni 方案预测 Env 蛋白的表面可能性区域,结果见图 3。图 3 显示,Env 蛋白的多数区段出现在蛋白质表面的可能性比较大,而且这些区段都位于 Env 蛋白的亲水区域。应用不同的预测方法,其预测的抗原表位的个数和抗原表位可能出现的肽段有所不同,但在 2-23、68-80、93-107、115-128、138-145、158-191、206-220、229-235、247-250、255-295、311-322、327-332、338-360、374-396、423-426、487-492、589-596、651-654 区段则显示为一致性较高的区段,这些区段有较好的亲水性和可及性,在二级结构上含有易形成抗原表位的无规则卷曲和 β -转角结构^[9]。根据以上多种预测结果,提示 Env 蛋白 B 淋巴细胞表位可能在这些片段或它们附近。

3 讨论

PERV-Env 蛋白不仅能刺激宿主产生中和抗

体,而且还参与感染过程中病毒与宿主细胞膜的融合,是 PERV 最重要的结构蛋白之一。因此对其进行蛋白质二级结构及抗原表位的预测,对于 PERV 多肽疫苗的合成、诊断试剂的研制、单克隆抗体的筛选以及 Env 蛋白的功能研究等至关重要。目前已有多种单参数预测模型和相应的分析软件用于蛋白质抗原决定簇预测。如传统的二级结构预测方案、Gamier-Robson 方案等。新近发展的基于多重序列比较的 PHDsec 蛋白质二级结构预测方案,如 PHD,是根据氨基酸的残基替换通常发生在蛋白质表面回折区域,蛋白质主链结构,特别是疏水的核心结构,如 α 螺旋和 β 折叠,受序列变异的影响很小。因此,通过蛋白质的多重序列比较,获得保守序列,然后通过神经网络算法(Neural network algorithm)来预测蛋白质二级结构^[10]。

通常蛋白质二级结构分析与表位分布也有较大关系,螺旋区段及折叠区段因其结构规则,有氢键维持结构,不易形变,主要起稳定结构作用,而转角区域和无规则卷曲区域多位于球蛋白分子表面,在此改变多肽链方法的阻力较小,故易于形变,有利于与抗体结合,从而成为抗原表位的可能性较大。本研究采用 2 种方法和多参数进行综合预测,可提高预

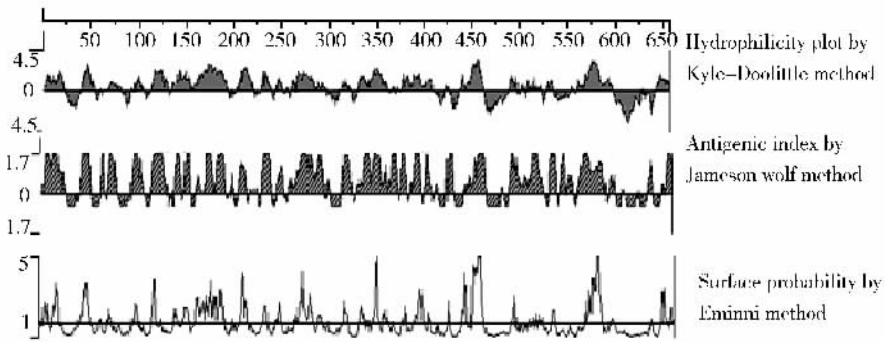


图 3 PERV-Env 蛋白的亲水性、抗原指数及其表面可能性分析

Fig. 3 Analysis of hydrophilicity plot, antigenic index and surface probability of the PERV-Env protein

测的准确性和表位的特异性。

从本试验结果可以看出依据不同的预测方法,其预测的抗原表位的数目和所处的肽段位置有所不同,但在氨基酸序列片段 2-23、68-80、93-107、115-128、138-145、158-191、206-220、229-235、247-250、255-295、311-322、327-332、338-360、374-396、423-426、487-492、589-596、651-654 等几种预测方法的结果基本一致,而且这些肽段恰好分别位于 β -折叠区域的交界区,可能形成 β 转角或不规则卷曲,是较柔性的区域。因此 B 细胞表位可能在这些区域或它们附近。此外根据 Rost 建立的 PHDsec 二级结构预测方案进行分析,PERV-Env 蛋白中可能存在的 7 个糖基化位点(N-glycosylation site)的氨基酸肽段正好与我们分析预测的 B 细胞表位的区域(68-80、255-295、311-322、338-360、374-396、423-426)相一致。而 Uwe 等^[11]通过表位图分析发现 PERV-TM 蛋白上能够产生中和抗体的 2 个表位 E1(GPQQLEK)、E2(FEGWFN),分别位于 Env 蛋白 N 端的 487-493 和 587-592 区段,也与我们的 B 细胞表位预测结果(487-492、589-596)相同。

本试验基于 PERV-Env 蛋白的氨基酸序列,用多参数预测 PERV-Env 蛋白的二级结构和 B 细胞表位,为寻找特异性相关表位用于疫苗的设计、单抗及诊断试剂研制提供理论依据。同时也为 Env 蛋白功能研究提供了线索。

参考文献:

[1] Ptience C, Takeuchi Y, Weiss R A. Infection of human cells by an endogenous retrovirus of pigs [J]. Nat Med, 1997, 3(3): 282~286.
[2] Kuddus R H, Gandhi C R, Rehman K K, *et al.* Some morphological, growth, and genomic properties of

human cells chronically infected with porcine endogenous retrovirus (PERV) [J]. Genome, 2003, 46(5): 858~869.

- [3] Wood J C, Quinn G, Suling K M, *et al.* Identification of exogenous forms of human-tropic porcine endogenous retrovirus in miniature swine [J]. J Virol, 2004, 78(5): 2 493~2 500.
[4] 吕茂民,贾莉纬,赵宝全,等.中国实验小型猪内源性反转录病毒囊膜糖蛋白基因的克隆与序列分析[J].中国预防兽医学报,2002,24(增刊):143~146.
[5] Rost B. PHD: Predicting one-dimensional protein structure by profile-based neural networks [J]. Methods Enzymol, 1996, 266:525.
[6] Hopp T P. 12 years of antigenic determinant predictions and more [J]. Pept Res, 1995, 6(4):183~190.
[7] Penton N, Musacchio A, Rivera J M, *et al.* Antigenicity of a recombinant NS3 protein representative of At-pase/helicase domain from hepatitis C virus [J]. Clin Biochem, 2003, 36(1):41~49.
[8] West M, Wilsom V G. Hydrophobic residue contributions to sequence-specific DNA binding by the bovine papilloma virus helicase E1 [J]. Virology, 2002, 296(1):52~61.
[9] Krchna K V, Mach O, Mal Y A. Computer prediction of B-cell determinants from protein amino acid sequences base incidence of β -turns [J]. Methods in Enzymology, 1989, 178: 586~ 611.
[10] 邹强,谢佩蓉,李华,等.人衰变加速因子的二级结构与 B 细胞表位预测 [J]. 中国免疫学杂志, 2001, 17(1):388~392.
[11] Uwe F, Oliver S, Reinhard K, *et al.* Neutralizing antibodies against conserved domains of p15E of porcine endogenous retroviruses; basis for a vaccine for xenotransplantation [J]. Virology, 2003, 307:406~413.