

DOI: 10.3724/SP.J.1005.2010.00701

不同宿主 H1N1 病毒血凝素蛋白(HA)受体结合位点的变异特征

傅天韵¹, 娄维义¹, 石铁流²

1. 华东师范大学第二附属中学, 上海 201203;
2. 华东师范大学生命科学学院, 上海 200062

摘要: 2009 年全球性爆发的 H1N1 病毒已经导致 213 个国家和地区受到感染, 有 16 226 人死亡。病毒与宿主细胞表面受体的结合是病毒感染不可缺少的第一步, 从而导致病毒膜与宿主细胞膜的融合。血凝素(Hemagglutinin, HA)就是介导这种受体结合与膜融合的病毒蛋白, 受体结合位点(Receptor binding sites, RBSs)位于 HA 蛋白三聚体中每个单体的球形头部, 主要由 190 位螺旋(190~198aa)、130 位环(135~138aa)和 220 位环(221~228)3 个二级结构域组成。文章收集了 1918~2009 年间 1 221 株 H1N1 病毒株的 HA1 序列(长度为 327 个氨基酸残基), 通过序列比对、各位点氨基酸残基的熵值以及 3D 结构模拟等生物信息学研究。结果显示不同宿主的病毒 RBSs 具有不同的熵值, 而且不同宿主的病毒 HA1 其 RBSs 具有不同的优势序列。3D 结构模拟也显示了 H1N1 不同 HA1 之间在 190 位螺旋构象上的细微差异。该研究揭示了不同 HA1 上 RBSs 的一些新的特征, 为进一步探讨病毒感染的机理提供了新的信息。

关键词: H1N1 流感病毒; 受体结合位点; 变异特征; 生物信息学分析

Variation feature of receptor binding sites of H1N1 influenza hemagglutinin in different hosts

FU Tian-Yun¹, LOU Wei-Yi¹, SHI Tie-Liu²

1. No. 2 High School of East China Normal University, Shanghai 201203, China;
2. School of Life Sciences, East China Normal University, Shanghai 200062, China

Abstract: Recent outbreak of H1N1 virus worldwide has caused 16 226 deaths in over 213 countries and districts. Binding between the virus and the receptor on the host cell surface is the key initial event for the infection, which results in the fusion of viral host cell membrane. Hemagglutinin (HA) is the viral protein that mediates the receptor binding and membrane fusion. The receptor binding sites (RBSs) are located at the membrane-distal part of each subunit of the HA trimer and are formed by three secondary structure elements, 190 helix (residues 190 to 198), 130 loop (residues 135 to 138), and 220 loop (residues 221 to 228). HA1 with 327 amino acid sequences in length was collected from 1221 H1N1 viruses between 1918 and 2009, and bioinformatic studies were carried out through sequence comparison, entropy calculation for each amino acid residue, and 3D structure modeling. The results showed that the RBSs of different viruses with different hosts have different entropies, and the RBSs in HA1 with different hosts have different favorite amino acid sequences. The 3D modeling indicates the subtly conformation changes in the 190 helix region between different HA1s in H1N1. This study explores new characters of the RBS structure in different HA1s, and provides new information for the further investigation

收稿日期: 2009-10-10; 修回日期: 2010-02-02

作者简介: 傅天韵(1992-), 女, 高二。E-mail: futianyun@hsefz.com

通讯作者: 石铁流(1964-), 男, 教授, 博士, 博士生导师, 研究方向: 生物信息学和计算生物学。E-mail: tishi@sibs.ac.cn

of the infection mechanism.

Keywords: H1N1 influenza; receptor binding site; variation feature; bioinformatics analysis

2009 年春至今, 爆发了甲型 H1N1 全球性“流感大流行”。截止至 2010 年 2 月 26 日, WHO 宣布流感病毒扩散到世界 213 个国家和地区, 全球确诊死亡病例为 16 226 人, 而且目前仍在持续增加。

当流感病毒进入宿主体内后, 病毒感染宿主细胞不可缺少的第一步就是病毒要与细胞表面的受体结合, 从而导致两者膜融合。血凝素(Hemagglutinin, HA)病毒三聚体球形头部的受体结合位点(Receptor binding site, RBS)就是病毒识别并结合宿主细胞表面受体的部位, 主要由 190 位螺旋、130 位环和 220 位环 3 个结构域组成, 呈浅口袋状凹陷^[1, 2]。而宿主细胞膜上的受体是含唾液酸(Sialic acid, SA)的糖蛋白或糖脂, 该唾液酸通过其第二位碳原子以 α -2,3 (SA α 2,3Gal)或 α -2,6(SA α 2,6Gal)糖苷键与糖链末端的半乳糖相连。不同病毒的 RBS 对这些唾液酸的亲和力是不同的, 这就决定了病毒感染宿主的范围^[3, 4]。人流感病毒趋向于结合人上呼吸道细胞表面的 SA α 2,6Gal 唾液酸; 禽流感病毒则主要结合禽类肠道细胞表面的 SA α 2,3Gal 唾液酸, 而对人细胞受体的亲和力较低; 但在猪的呼吸道上皮细胞膜受体上这两种唾液酸都有, 因此猪可能是人、禽、猪等多种宿主感染病毒交叉基因重组的中间宿主, 被称为流感病毒的“混合器”。

综上所述, 受体结合位点的变异会导致病毒感染宿主范围和致病力的变化, 也与流感的流行规模密切相关。因此, 对血凝素 HA 上的受体结合位点及其变异特征和规律的研究极为重要的。有关 H1N1 甲型流感病毒血凝素变异的研究已经有不少报道, 但大多为某一段时间的短期研究, 而长期、系统的研究较少。本文通过收集 NCBI 流感中心病毒库的数据, 获得了 1918~2009 年 92 年间所有 H1N1 甲型流感病毒血凝素 HA 的氨基酸序列, 并通过生物信息学软件重点分析了不同宿主 H1N1 亚型流感病毒的 HA1 受体结合位点, 拟阐明其演变的特征和规律, 从而为防控流感大流行以及常规流感的监测提供科学依据。

1 材料和方法

1.1 病毒序列及其来源

从 NCBI(National Center for Biotechnology Information)数据库(NCBI Influenza Virus Sequence Database)(<http://www.ncbi.nlm.nih.gov>)中, 收集流感病毒血凝素 HA 蛋白的氨基酸序列。选择的参数设置为:

病毒种类(Virus Species): A 型流感病毒(Influenza virus A); 宿主(Host): 任意(any); 国家/地区(Country/Region): 任意(any); 片段/蛋白质(Segment/Protein): HA 片段; 亚型(Subtype): H1N1。只选择全长序列(Full-length sequences only)并去除相同的序列(Remove identical sequences), 其余条件系统默认。序列统一采用 FASTA 格式保存。

1.2 HA1 氨基酸序列比对和分析

氨基酸编号按照国际规定以 H3 的 HA 序列为比对标准^[1, 2]。

多序列比对采用 Clustalx1.83 软件, 初步比对后对结果进行初步分析, 对于个别影响比对结果的氨基酸插入序列进行人工编辑, 保留序列比对时产生的空位。然后利用 BioEdit 软件对序列进行编辑, 去除 HA 近 N 端的信号肽序列以及 C 端的 HA2 序列部分。然后再次利用 Clustalx1.83 进行比对和编辑, 直到序列全部对齐^[4]。序列对齐有利于氨基酸残基序列位数一致和统计结果的准确性。

1.3 HA1 受体结合位点熵值计算和分析

流感病毒受体结合位点(Receptor binding sites, RBS)的选择是以 Gamblin 等^[1]为标准, 即位于 HA 蛋白三聚体中每个单体的头部, 主要由 190-螺旋(190~198)、130-环(135~138)和 220-环(221~228)3 个结构域以及周围 3 个保守的氨基酸残基: Tyr(Y)⁹⁸、Trp(W)¹⁵³ 和 His(H)¹⁸³ 组成。

采用 BioEdit version7.0.9.0 对这些病毒序列进行分析, 计算熵值(Entropy), 熵值的范围在 0~4.392

之间(只出现一种氨基酸残基的位点其熵值为 0; 某一个位点出现所有 20 种氨基酸残基且概率均等, 包含因氨基酸残基缺失而插入的空位, 则该位点的熵值为 4.392)。关于位点熵值的保守性定义参考文献^[5]: 氨基酸残基缺失位点的熵值为: 1.000; 当氨基酸残基位点的熵值 1.000, 代表高突变位点; 熵值 0.600, 代表易突变位点; 熵值 0.600, 则代表保守位点。

为了避免在熵值计算时存在由目前数据库收录的样本量的差异而可能带来的影响, 本文采用 Bootstrap 方法^[6, 7], 对人源 H1N1 和猪源 H1N1 模拟 100 组等样本量(72 条序列)的熵值, 并进行单侧 Wilconx 秩和检验^[8, 9]。

同时对每个位点进行了氨基酸数量的统计和百分率计算, 用于分析优势序列。

1.4 空间结构模拟分析

通过 SWISS-MODEL Workspace(<http://swissmodel.expasy.org>), 直接向服务器 web 界面提交拟建模的氨基酸序列以及指定模板“template”(来源于 PDB 数据库: <http://www.pdb.org>), 对 H1N1 病毒的 HA1 进行同源建模, 获其 3D 结构。并利用 Swiss-PdbViewer 4.0.1 软件进行空间构象的模拟分析^[10-12]。

2 结果与分析

2.1 H1N1 血凝素蛋白(HA)受体结合位点的分析

2.1.1 病毒 HA 序列及其来源

截止至 2009 年 8 月 15 日, 从 GenBank 数据库中收集符合条件的序列共 1 221 条, 时间跨度从 1918~2009 年, 宿主范围包括人、猪、禽、雪貂和食蚊兽等:

(1) 人 H1N1 病毒株的 HA1 序列: 1918~2009 年共 953 条;

(2) 猪 H1N1 病毒的 HA1 序列: 1930~2009 年共 194 条;

(3) 禽 H1N1 病毒的 HA1 序列: 1976~2009 年共 72 条;

(4) 雪貂 H1N1 病毒的 HA1 序列: 仅 2008 年 5 月 1 条;

(5) 食蚊兽 H1N1 病毒的 HA1 序列: 仅 2007 年 1 条。

2.1.2 HA1 受体结合位点的变异情况

H1N1 病毒完整的血凝素蛋白(HA0)一般由 562~566 个氨基酸组成, 酶解后为两部分: HA1 和 HA2。对 NCBI 数据库中获得 1 221 株 H1N1 病毒株的 HA1 共 327 个氨基酸序列进行编辑和比对, 并计算各位点氨基酸残基的熵值, 结果见表 1。

从表 1 可以看出, HA1 蛋白的 327 个氨基酸残基位点中, 涉及到的受体结合位点共 25 个, 其中高突变位点有 4 个(16%)、易突变位点 7 个(28%)、保守位点 14 个(56%), 包括 3 个是没有突变的位点。在非受体结合位点共 302 个氨基酸残基中, 高突变位点有 23 个(7.6%)、易突变位点 45 个(14.9%)、保守位点 234 个(77.5%), 包括 50 个是没有突变的位点。

所以, 受体结合位点的熵值相比非受体结合位点的熵值差异要大, 其氨基酸残基位点突变的比率也明显增大。说明感染不同宿主的 H1N1 流感病毒其受体结合位点各有其一定的特征, 以适合不同宿主细胞表面的受体结构。

2.1.3 HA1 受体结合位点在不同宿主中的变异情况

表 2~表 4 显示了不同宿主的 H1N1 病毒 RBS 各位点的氨基酸残基熵值及变化。从表 2~表 4 可以看到, 感染不同宿主的 H1N1 流感病毒其受体结合位点的熵值是不同的, 除了感染宿主雪貂、食蚊兽的 H1N1 病毒只有 1 例外, 流感病毒 RBS 的情况分别为:

(1) 感染人类的 H1N1 病毒 RBS 熵值变异最大, 尤其在 190 位环区域的 10 个氨基酸位点中, 存在 2 个高突变位点: 192 位(熵值为 1.245)、193 位(熵值为 1.092), 以及 3 个易突变位点: 190 位(熵值为 0.967)、196 位(熵值为 0.944)和 197 位(熵值为 0.826)。而在 130 位环中, 有 2 个易突变位点氨基酸位点 133 位、136 位(熵值分别为 0.865、0.647), 其他位点以及 220 位环的所有氨基酸残基位点都属于保守性位点, 其中熵值为 0 的位点共 4 个。

(2) 感染猪的 H1N1 病毒 RBS 熵值变异中等, 在 130 位环中 133 位、135 位、137 位为易突变位点(熵值分别为 0.843、0.850、0.800); 在 220 位环中 225 位为易突变位点(熵值为 0.860)其他位点都属于保守性位点, 其中熵值为 0 的位点共 3 个; 而在 190 位环中, 存在 193 位(熵值为 0.682)、198 位(熵值为 0.716)2 个易突变位点。

表 1 H1N1 病毒株 HA1 各位点氨基酸残基的熵值

氨基酸残基序号				HA1 各位点氨基酸残基的熵值						
1				0.042	0.042	0.613*	0.035	0.351	0.042	0.047
11	0.042	0.035	0.035	0.035	0.042	0.035	0.029	0.007	0.007	0.013
21	0.000	0.250	0.040	0.013	0.012	0.000	0.012	0.033	0.000	0.000
31	0.000	0.007	0.027	0.007	0.055	0.000	0.007	0.673*	1.012**	0.000
41	0.022	0.007	0.022	0.000	0.000	1.210**	0.020	1.015**	0.007	0.804*
51	0.108	0.000	0.012	0.595	0.000	0.013	0.686*	0.000	0.677*	0.646*
61	0.024	0.007	0.024	0.519	0.000	0.013	0.000	0.007	0.080	0.012
71	0.429	0.667*	0.019	1.241**	0.786*	0.991*	1.177**	0.007	0.000	0.000
81	0.000	0.046	0.635*	0.007	0.734*	0.656*	0.490	0.709*	0.750*	0.020
91	0.000	0.195	0.000	0.000	0.000	0.000	1.369**	0.007	0.755*	0.354
101	0.007	0.039	0.013	0.007	0.084	0.007	0.057	0.000	0.029	0.115
111	0.066	0.019	0.012	0.052	0.031	0.238	0.000	0.007	0.046	0.000
121		0.012	0.031	1.122**	0.022	0.000	0.013	0.364	0.012	1.163**
131	1.047**	0.653*	1.089**	0.000	0.245	0.698*	0.254	0.666*	0.000	0.583
141	0.471	1.148**	0.035	1.308**	1.017**	0.032	0.012	0.007	0.562	0.007
151	0.020	0.584	0.007	0.345	0.712*	1.109**	0.082	0.831*	0.885*	0.718*
161	0.000	0.000	0.818*	0.073	0.273	0.353	0.019	0.007	1.260**	0.017
171	0.574	0.043	0.762*	0.038	0.029	0.007	0.000	0.033	0.228	0.000
181	0.012	0.579	0.000	0.046	0.019	0.620*	0.893*	1.106**	1.726**	1.011**
191	0.007	1.254**	1.161**	0.354	0.007	0.998*	0.907*	0.959*	0.779*	0.180
201	0.033	0.038	0.571	0.013	0.757*	0.276	0.013	1.106**	0.007	0.716*
211	0.573	0.709*	0.013	0.713*	0.013	0.000	0.066	0.196	1.201**	0.000
221	0.007	0.056	0.056	0.076	0.801*	0.095	0.666*	0.000	0.012	0.667*
231	0.062	0.007	0.044	0.000	0.013	0.029	0.716*	0.431	0.345	0.066
241	0.034	0.447	0.013	0.692*	0.012	0.019	0.019	0.714*	0.000	0.000
251	0.000	0.584	0.569	0.000	0.618*	0.450	0.013	0.000	0.047	0.593
261	1.063**	0.364	0.493	1.045**	0.441	0.013	0.000	0.013	0.183	1.257**
271	0.007	0.660*	0.548	0.562	0.715*	1.027**	1.329**	0.012	0.738*	0.701*
281	0.720*	0.007	0.000	0.007	0.007	0.997*	0.000	0.443	0.349	0.119
291	0.596	0.189	0.012	0.000	0.039	0.076	0.000	0.605*	0.012	0.012
301	0.641*	0.040	0.000	0.007	0.429	0.000	0.000	0.019	0.007	0.042
311	0.655*	0.000	0.728*	0.244	0.007	0.000	0.435	0.665*	0.013	0.000
321	0.007	0.013	0.007	0.580	0.007	0.007	0.060	0.020	0.055	0.007

注: (1) 黑体字表示位于 HA1 上的受体结合位点; (2)**: 代表高突变位点(氨基酸残基位点的熵值 > 1.000); *: 代表易突变位点(1.000 > 熵值 > 0.600); 没有标记: 代表保守位点(熵值 < 0.600)。

(3) 感染禽类的 H1N1 病毒 RBS 熵值变异最小, 具体表现在: 130 位环、220 位环的各氨基酸残基位点的熵值全部小于 0.4, 都属于保守性位点, 尤其是这两个区域共 15 个氨基酸位点中有 11 个位点是相同的(熵值为 0); 而在 190 位环区域的 10 个氨基酸位点中, 只有 198 位为高突变位点(熵值为 1.040)、193 位为易突变位点(熵值为 0.654)外, 其他 8 个位点都

属于保守性位点。

应用 Wilcoxon 秩和检验, 在 0.01 的置信水平上, 人流感病毒、猪流感病毒与禽流感病毒在 190 位、192 位、192 位、197 位、130 位上的变异依次减少。从这些数据可以发现, 禽流感病毒在 RBS 的氨基酸位点变异较小, 猪流感病毒次之, 而人流感病毒在这些位点变化最大。

表 2 H1N1 受体结合位点的熵值结果——130 位环

宿主	氨基酸位点 及出现频率	153 位	133 位	130 位环			
				135 位	136 位	137 位	138 位
人类	氨基酸组成	W(100)	~(66.74) K(26.13) R(5.35) N(1.26) I(0.31) G(0.10) X(0.10)	V(99.27) I(0.31) A(0.31)	S(66.21) T(33.68) X(0.10)	A(98.43) T(0.73) V(0.52) X(0.21) S(0.10)	S(77.02) A(22.98)
	熵值	0.000	0.865*	0.051	0.647*	0.099	0.539
猪	氨基酸组成	W(99.48) M(0.52)	R(67.01) K(27.84) ~(2.06) I(1.55) G(1.03) X(0.52)	V(74.74) T(14.43) A(6.7) I(1.55) S(2.06) L(0.52)	T(97.94) S(2.06)	A(78.87) V(11.34) S(3.61) I(3.09) T(1.55) D(1.03) K(0.52)	A(92.78) S(7.22)
	熵值	0.032	0.843*	0.850*	0.100	0.800*	0.259
禽类	氨基酸组成	W(100)	K(93.06) R(5.56) ~(1.39)	V(100)	T(100)	A(100)	A(97.22) S(2.78)
	熵值	0.000	0.287	0.000	0.000	0.000	0.127
雪貂	氨基酸组成	W	K	V	T	A	A
食蚊兽	氨基酸组成	W	T	V	S	A	S
综合	氨基酸组成	W(99.92) M(0.08)	~(52.50) K(30.38) R(15.15) N(0.98) I(0.49) G(0.25) X(0.16) T(0.08)	V(95.41) T(2.29) A(1.31) I(0.49) S(0.33) G(0.08) L(0.08)	S(52.09) T(47.83) X(0.08)	A(95.41) V(2.21) T(0.82) S(0.66) I(0.49) D(0.16) X(0.16) K(0.08)	S(61.51) A(38.49)
	熵值	0.007	1.089**	0.245	0.698*	0.254	0.666*

注: **: 代表高突变位点(氨基酸残基位点的熵值 1.000); *: 代表易突变位点(1.000 熵值 0.600); 没有标记: 代表保守位点(熵值 0.600)。

2.1.4 HA1 受体结合位点在不同宿主中具有优势氨基酸序列

表 5 是感染不同宿主的 H1N1 受体结合位点出现频率最高的氨基酸残基, 即优势氨基酸。RBS 除了周围的保守氨基酸位点: Tyr(Y)⁹⁴、Trp(W)¹⁵³ 和 His(H)¹⁸³ 外^[1], 在 130 位环内有: Val(V)¹³⁵、Ala(A)¹³⁷ 为保守氨基酸位点; 190 位螺旋内有: Gln(Q)¹⁹¹、Leu(L)¹⁹⁴ 和 Tyr(Y)¹⁹⁵ 为保守氨基酸位点; 220 位环内有: Pro(P)²²¹、Lys(K)²²²、Val(V)²²³、Arg(R)²²⁴、Gln(Q)²²⁶ 和 Gly(G)²²⁸ 为保守氨基酸位点。

据报道, 226、228 位点位于受体结合位点“Pocket”(袋状凹陷)的底部, 在与宿主细胞表面受体的唾液酸结合过程中发挥重要作用, 而不同宿主的 H1N1 病毒中该位点氨基酸残基属于保守位点, 表明该位点可能结合不同宿主受体的共同部位。

另外, 从表 5 可以发现, H1N1 受体结合位点中的有一些位点在不同宿主中明显不同, 如: 133 位, 在人的 H1N1 中大多为缺失, 在猪中为 Arg(R)¹³³, 禽类中为 Lys(K)¹³³; 198 位, 在人类中 Glu(E)¹⁹⁸, 猪中为 Ala(A)¹⁹⁸, 禽类中为 Thr(T)¹⁹⁸。而在 136 位、138 位、196 位、197 位、225 位、227 位, 人 H1N1 明显与其他种类的不同。提示感染不同宿主的 H1N1 受体结合位点具有其相对特征性的氨基酸序列, 说明 H1N1 病毒具有较强的宿主特异性, 相对比较保守。

2.2 H1N1 血凝素蛋白(HA)受体结合位点的空间结构模拟

在表 5 的基础上, 采用空间结构模拟的方法, 探索感染不同宿主的 H1N1 病毒 RBS 其相同、不同氨基酸残基位点对于空间结构的影响。

为了便于同时对比人类、猪、禽类 HA1 尤其是 RBS 在立体结构上的差别, 本文选择了 PDB 数据库中 1934 年人 H1 血凝素 1ru7K 链的空间结构为模板, 分别进行同源建模。获得相应病毒 HA 的 3D 结构后, 利用 Swiss-PdbViewer4.0.1 软件对人类、猪、禽类三类宿主的病毒血凝素 HA1 氨基酸序列进行空间构象进行重叠、优化并标识, 结果见图 1。

感染人类的 H1N1 病毒 RBS 区域中, 190 位环氨基酸位点变异最大, 尤其是在 192 位和 193 位为高突变位点。仔细观察其氨基酸序列组成, 主要有 3 种组合: KT 最多见, 其次为 RA, 第三为 QS。通过对比对包含这 3 种组合的 HA1 序列空间结构(图 2), 可以看到它们之间有一些细微的差别。但 192 位和 193 位氨基酸位点并不直接与宿主受体 SA α 2,6Gal 唾液酸发生结合, 因此这些变化可能不影响 RBS 与受体结合的亲和力。

3 讨论

流感病毒血凝素受体结合位点与宿主细胞表面受体的结构是决定病毒感染宿主特异性的主要因素。甲型 H1N1 流感病毒血凝素 HA 由流感基因片段 4 编码, 一般包含 1742~1778 个核苷酸, 所编码的相应 HA 蛋白(HA0)一般包含 562~566 个氨基酸。1~17 位氨基酸为信号肽, 18~344 位为 HA1, 345~566 位为 HA2。HA1 与 HA2 这两个亚单位间由链间二硫键相连, 并形成 HA 三聚体, 由球形头部(HA1)和茎部(部分 HA1 和 HA2)组成。HA1 具有与宿主细胞受体结合的特性, HA2 则能够与宿主细胞膜融合。当 HA0 与细胞表面受体结合后可活化蛋白酶, 将 HA0 酶切为 HA1 和 HA2, HA2 通过其 N 末端的疏水序列与细胞膜融合, 使病毒进入细胞内。因此 HA 蛋白具有 3 个功能: 即与细胞结合、介导病毒外膜与细胞

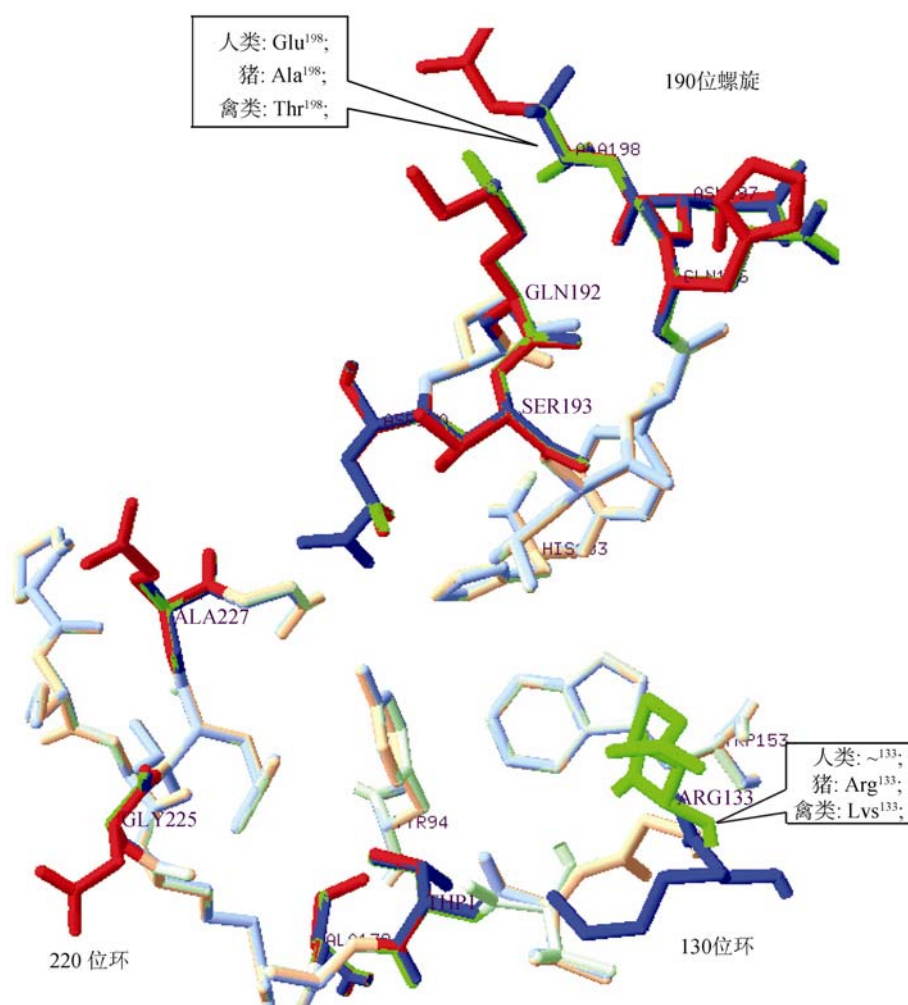


图 1 人类、猪、禽类 H1N1-RBS 的空间差异

人类 H1N1-RBS 以红色标识, 猪 H1N1-RBS 以绿色标识, 禽类 H1N1-RBS 以蓝色标识, 存在差异的位点以相应的深色标识。

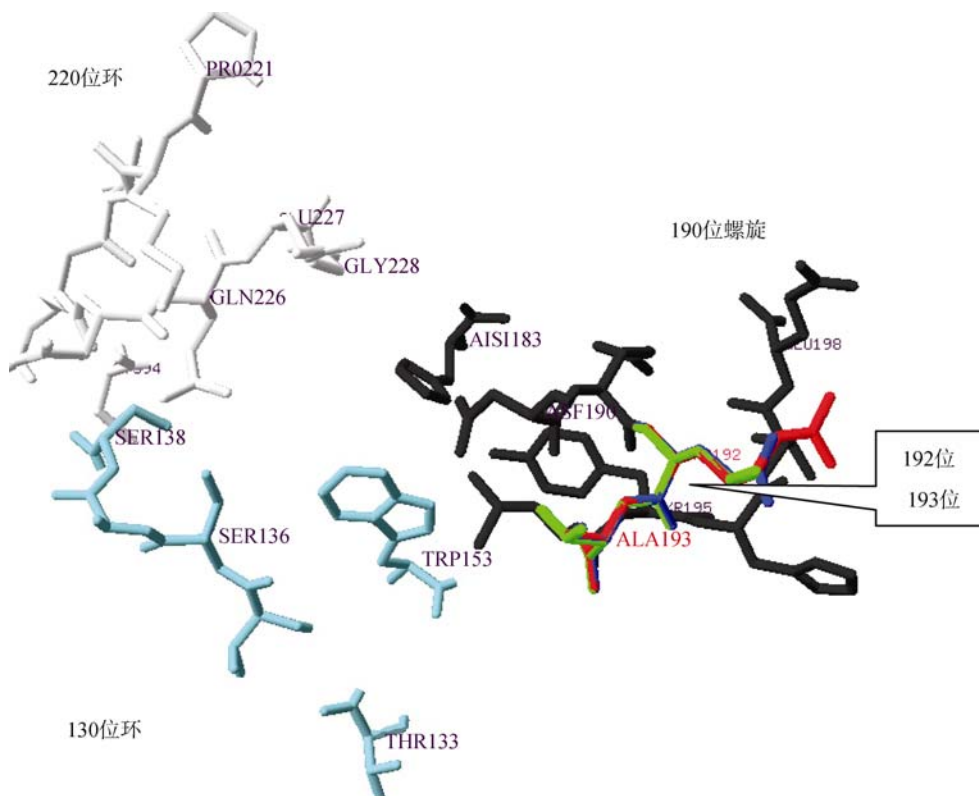


图 2 人类 H1N1-RBS 的 190 位螺旋中 192-193 位 3 种优势氨基酸组合的空间差异

人类 H1N1-RBS 的 190 位螺旋中 192-193 位存在 3 种优势氨基酸组合, KT 组合以绿色标识, RA 组合以红色标识, QS 组合以蓝色标识。

膜融合使病毒进入细胞,同时诱导宿主免疫系统产生对病毒的中和抗体。

甲型流感病毒上的 RBS 呈浅口袋状凹陷,其独特形状决定了结合宿主受体的特异性,也就决定了病毒感染宿主的种类。但对于这些构成这一袋状凹陷的氨基酸位点来说,在决定与宿主受体结合的亲和力时所发挥的作用是不同的。Connor 等^[13]认为在 H2 和 H3 流感病毒中,具有 Leu(L)²²⁶ 和 Ser(S)²²⁸ 位点的 HA(为人流感病毒)优先识别 SAa2,6Gal, 而 Gln(Q)²²⁶ 和 Gly(G)²²⁸(为禽流感病毒)则优先识别 SAa2,3Gal。Vines 等^[14]通过氨基酸定点替换实验证明了流感病毒 RBS 第 226 位和 228 位氨基酸的变化决定了病毒感染宿主的特异性。当 H3 病毒的 HA 同时发生 Leu226Gln、Ser228Gly 突变时,则可使感染人的 H3 流感病毒能够感染鸭,并在鸭肠道细胞中增殖,同时其凝血活性也发生了变化。陈君彦等^[15]对 2001 年中国华南及东北地区的 15 株 H1N1 猪流感病毒分离株的血凝素基因进行了序列测定和分析,发现这些病毒与所有参考毒株的 HA 第 226 位都是谷氨酰胺(Q),第 228 位都为甘氨酸(G),由此认为这

些 H1N1 亚型都是优先结合 SAa2,3Gal,可能具有感染禽的潜力。但从本文的分析结果来看,无论是感染人类、猪、禽类,还是感染雪貂、食蚊兽的 H1N1 病毒,其 226 位除了在人类 H1N1 病毒中有极少数的变异外(小于 3%),其他都为 Q;而 228 位是非常保守的,都为 G。因此,在 H1N1 病毒的 RBS 中,第 226 位 Q 和第 228 位 G 在结合 SAa2,3Gal 或者 SAa2,6Gal 时所起的作用还有待于研究。

另外,H1N1 流感病毒 HA1 蛋白上的 RBS 共 25 个氨基酸残基,其中 14 个氨基酸位点都是高度保守的,除了上述 Gln(Q)²²⁶、Gly(G)²²⁸ 位点外,还有 Trp(W)¹⁵³、Val(V)¹³⁵、Ala(A)¹³⁷(130 位环),His(H)¹⁸³、Gln(Q)¹⁹¹、Leu(L)¹⁹⁴、Tyr(Y)¹⁹⁵(190 位螺旋),以及 Tyr(Y)⁹⁴、Pro(P)²²¹、Lys(K)²²²、Val(V)²²³、Arg(R)²²⁴(220 位环)。这些位点在人类、猪、禽类,甚至雪貂、食蚊兽中几乎都是相同的,因而在宿主结合特异性方面应该不会发挥重要作用。这就提示我们:RBS 的其他 11 个氨基酸位点或者联合另外某些氨基酸位点可能在病毒感染宿主的特异性方面起着主导作用,这还有待于进一步研究。

总之, 对于 H1N1 流感病毒 HA 蛋白氨基酸序列中特殊位点的突变对其受体结合特性、感染范围, 甚至毒力等都有显著影响, 这些特殊位点的识别及其突变规律的揭示对流感跨宿主传播、致病机理, 以及抗病毒药物的筛选等研究都具有重要意义。

参考文献(References):

- [1] Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A, Wiley DC, Skehel JJ. The structure and receptor binding properties of the 1918 influenza Hemagglutinin. *Science*, 2004, 303(5665): 1838–1842.
- [2] Lin T, Wang G, Li A, Zhang Q, Wu C, Zhang R, Cai Q, Song W, Yuen KY. The hemagglutinin structure of an avian H1N1 influenza A virus. *Virology*, 2009, 392(1): 73–81.
- [3] Matrosovich MN, Gambaryan AS, Teneberg S, Piskarev VE, Yamnikova SS, Lvov DK, Robertson JS and Karlsson KA. Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. *Virology*, 1997, 233(1): 224–234.
- [4] Suzuki Y, Ito T, Suzuki T, Holland RE, Chambers TM, Kiso M, Ishida H, Kawaoka Y. Sialic acid species as a determinant of the host range of influenza A viruses. *J Virol*, 2000, 74(24): 11825–11831.
- [5] 徐进, 冯宝龙, 王清艳, 梁瑾, 王靖飞. 氨基酸序列集熵值计算工具实现及应用. *生命科学*, 2008, 20(3): 415–420.
- [6] Bickel PJ, Freedman DA. Some asymptotic theory for the bootstrap. *Ann Stat*, 1981, 9: 1196–1217.
- [7] Singh K. On the asymptotic accuracy of Efron's bootstrap. *Ann Stat*, 1981, 9: 1187–1195.
- [8] Hollander M, Wolfe DA. *Nonparametric Statistical Methods*. New York: John Wiley & Sons. 1973, 27–33 (one-sample), 68–75 (two-sample). Or second edition (1999).
- [9] Bauer DF. Constructing confidence sets using rank statistics. *J Am Stat Assoc*, 1972, 67: 687–690.
- [10] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics*, 2006, 22(2): 195–201.
- [11] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*, 2003, 31(13): 3381–3385.
- [12] Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, 1997, 18(15): 2714–2723.
- [13] Connor RJ, Kawaoka Y, Webster RG, Paulson JC. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology*, 1994, 205(1): 17–23.
- [14] Vines A, Wells K, Matrosovich M, Castrucci MR, Ito T, Kawaoka Y. The role of influenza A virus hemagglutinin residues 226 and 228 in receptor specificity and host range restriction. *J Virol*, 1998, 72(9): 7672–7631.
- [15] 陈君彦, 李海燕, 申之义, 陈化兰, 于康震, 毕英佐. H1N1 亚型猪流感病毒中国分离株血凝素基因分子演化的研究. *中国预防兽医学报*, 2005, 27(1): 14–17.