

甲型流感病毒 PB1-PA 蛋白作用位点的发现与分析*

何 森¹, 林强乾¹, 赵艳超¹, 张淑波¹, 陈维田²

(1. 中山大学生命科学学院, 广东 广州 510275;

2. 香港理工大学屋宇设备工程系, 香港 九龙)

摘 要: 基于共进化理论, 探究了甲型流感病毒 PB1 蛋白与 PA 蛋白上具有共同进化可能性的保守九聚片段 (C9MP)。结构信息显示 PB1 蛋白的第 1-15 位氨基酸与 PA 蛋白的第 239-716 位氨基酸具有相互作用域; 对该区域变异分布的分析发现, PA 蛋白第 670 位氨基酸 Q 所在的 C9MP 与 PB1 蛋白的第 9 位氨基酸 F、第 12 位氨基酸 V 和第 13 位氨基酸 P 所在的 C9MP 在 PB1^{MPI} 相互作用面上具有最低的共进化值。结合 DSSP 程序的分析表明, 由 PA 蛋白第 670 位氨基酸 Q 与 PB1 蛋白的第 9 位氨基酸 F、第 12 位氨基酸 V 与第 13 位氨基酸 P 构成的区域可能成为潜在的相互作用位点。

关键词: 共进化理论; 甲型流感病毒; PB1 蛋白; PA 蛋白; 相互作用位点

中图分类号: Q354 **文献标志码:** A **文章编号:** 0529-6579 (2011) 02-0081-06

Discovery and Analysis of Protein-Protein Interaction Sites of PB1-PA Nuclease Complex of Influenza A Virus

HE Miao¹, LIN Qiangqian¹, ZHAO Yanchao¹, ZHANG Shubo¹, CHAN Daniel Wai Tin²

(1. School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China;

2. Department of Building Services Engineering, The Hong Kong Polytechnic University, Kowloon, China)

Abstract: Based on co-evolution theory, this research tries to locate the protein interaction sites in conserved 9-mer peptides (C9MP) between PB1 protein and PA protein in the nuclease complex of influenza A virus. There exist interaction regions between sites 1-15 amino acids of PB1 protein and sites 239-716 amino acids of PA protein from structure message. It is revealed by variation analysis to this region that, on PB1^{MPI} contact surface, there is a minimum co-evolution value between the conserved 9-mer peptide where Q670 of PA protein is located and the conserved 9-mer peptide where F9, V12 and P13 of PB1 protein are located. Combined with DSSP analysis results, the regions formed by four amino acids, Q670 of PA protein, F9, V12 and P13 of PB1 protein, are potential interaction sites.

Key words: co-evolution theory; influenza A virus; PB1; PA; interaction site

甲型流感病毒包含 8 个独一无二的单链 RNA 片段, RNA 松散地包埋在许多核蛋白 (Nucleoprotein, NP) 分子中, 其中, 包含 3 个病毒 RNA 聚合酶 (PB1、PB2 及 PA) 的复合体位于病毒颗粒的内核^[1]。1990 年, Aya 等^[2]用梯度氯化铯离心法从甲型流感病毒中分离出 RNA 聚合酶, 随后分离出了 PB1 蛋白、PB2 蛋白和 PA 蛋白。

PB1 蛋白, 又称为聚合酶碱性蛋白 1 (Polymerase Basic Protein 1), 由 RNA 片段 2 编码。它在 RNA 聚合酶复合体中的作用是负责病毒 RNA 分子的延伸, 同时也在模版 RNA 合成的延伸过程中起作用。PB1 蛋白酶定位在感染细胞的细胞核中^[1]。研究表明, PB1 蛋白的第 506-659 位氨基酸的保守性不显著, 但是, 仍分布有许多零星的保守片

* 收稿日期: 2010-03-06

基金项目: 国家自然科学基金重点资助项目 (60736028); 广东省科技计划项目重大专项资助项目 (2003A3080503)

作者简介: 何森 (1963 年生), 男, 副教授; E-mail: lsshem@mail.sysu.edu.cn

断, 这些保守片段是否代表蛋白表面氨基酸或相互作用位点仍需检验。

PA 蛋白, 又称为聚合酶酸性蛋白 (Polymerase Acid Protein), 由 RNA 片段 3 编码。是 RNA 聚合酶复合体中的最后一个蛋白, 它对病毒 RNA 聚合过程的作用未知。有证据表明, 它可能起着蛋白激酶或螺旋解构酶的作用。PA 聚合酶也定位在感染细胞的细胞核中^[1]。Amelia 等通过重组的猿 40 病毒, 用蛋白片断缺失法分析了 PA 蛋白的区域, 研究显示, 区域 I (124 - 139 位氨基酸) 与区域 II (186 - 247 位氨基酸) 具有重要功能; 两个核定位信号均存在于区域 I 与区域 II 中, 结论表明, PA 蛋白包含一个核定位信号区域, 是病毒执行繁殖复制的重要区域^[3]。

大部分蛋白质需要相互作用才能发挥出其生物学作用。通过构成复合体才能发挥生物学作用的蛋白质, 在相互作用面上的结构域或序列, 具有比一般功能的结构域更高的保守度。Caffrey 发现相互作用表面氨基酸比其它暴露在表面的氨基酸具有更高的保守度^[4]; Minsteris 和 Wang 发现相互作用的表面氨基酸具有与蛋白伴侣共进化的属性^[5]。

研究表明诸如 NLS 序列、聚合酶功能区序列、PB2 蛋白与 PA 蛋白结合区序列完全没有发生重叠的现象^[6]。PB2 蛋白需要与 PB1 蛋白结合才能发挥出转录酶的功能。PB1 蛋白的 N 端 78 个氨基酸和 PA 蛋白 C 端的 3/4 个蛋白片断结合^[7]; PB1 蛋白的第 506 - 659 位氨基酸与 PB2 蛋白的 124 个氨基酸相互作用^[8]。研究显示, PB1、PB2 与 PA 蛋白在病毒体外亦能自动形成复合体^[2]。

本文基于共进化理论, 从 PB1 蛋白的 N 端结构域与 PA 蛋白的 C 端结构域入手, 研究了甲型流感病毒的 PB1 蛋白与 PA 蛋白间具有潜在共进化关系的肽段, 创新性推测出复合体中的相互作用位点, 并在三维结构上标记出相互作用位点。

1 数据和方法

1.1 数据来源

甲型流感病毒序列全部来自于 NCBI 的流感病毒专门数据库 Influenza Virus Resource (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>)。本文提取了 1428 株病毒的 PB1 蛋白序列和 1580 株病毒的 PA 蛋白序列。序列下载后, 以 FASTA 格式

保存。

PB1 与 PA 的结构信息来自于 Protein Data Base (PDB)。在 PDB 上检索文件名为 2ZNL.pdb 的蛋白结构数据, 作为蛋白结构信息的来源。

1.2 计算方法

本文自主设计了 CD search (Conserved Domain Search) 软件, 主要用于“保守九聚氨基酸肽片段 (Conserved 9-mer Peptide, 简记为 C9MP)”的筛选, 以 C9MP 为单位的序列比对和共进化参数计算结果为筛选依据; 所谓 C9MP 是指九个在一级结构上连续排列的氨基酸所构成的短肽片段。利用软件进行序列比对, 筛选出所有序列中 90% 以上完全匹配的 9 个连续氨基酸, 即保守九聚氨基酸肽片段; 计算一个蛋白所有保守片段与另一个蛋白的保守片段相关的 η 、Z、 n_a 、 n_b 、 N_A 、 N_B 、 M_A 、 M_B 等 8 个共进化参数。

利用 CLUSTALX2.0 进行第 1 次序列比对, 选择 BLOSUM 矩阵, 输出得到 PB1.fasta 和 PA.fasta。获得的.fasta 文件中序列总数为 M 值, 其中, M_A 值表示 PB1.fasta 中包含的序列总数, M_B 表示 PA.fasta 中包含的序列总数。

对于每一个保守多肽片段来说, 相同的片段占了 90% 以上, 不相同的片段总数 N 值被精确统计出来。在 PB1 蛋白中不相同的片段数为 N_A ; 在 PA 蛋白中, 对应的值为 N_B 。

提取两个病毒蛋白.fasta 文件中的 GI 号及病毒名, 构建两个仅包含 GI 号及病毒名的数据文件 PB1_gz.txt 与 PA_gz.txt。提取 PB1.fasta 与 PA.fasta 所包含的共有病毒, 统计共有的病毒株数, 得到 η 值。 $n_{a(0-8)}$ 值代表 PB1 蛋白的氨基酸位置 0 - 8 在 η 条件下的变异数目, $n_{b(0-8)}$ 值代表 PA 蛋白的氨基酸位置 0 - 8 在 η 条件下的变异数目。假设 PB1 蛋白含有 X 个 C9MP, 在 PA 蛋白含有 Y 个 C9MP, 则总共有 $Z = X \times Y$ 个潜在相互作用位点。

假设在甲型流感病毒毒株中, 我们考虑 PB1 蛋白的第一个显著保守片段 (记为 A1) 和 PA 蛋白的第一个显著保守片段 (记为 B1), A1 片段在 M_A 中有 N_A 个突变序列, B1 片段在 M_B 中有 N_B 个突变序列, 那么, 在没有进化压力的情况下, A1 和 B1 在同一个病毒株中出现变异的概率是:

$$P(Z | N_A, N_B, M_A, M_B) = \frac{(\text{在 } Z \text{ 中分布的变异方式数目}) \times (\text{不在 } Z \text{ 中分布的变异方式数目})}{\text{可选择的变异方式总数}}$$

如果 Z 值在 η 值范围内。在 η 范围能够找到包含两个变异 C9MP 的方法数为:

$$w_Z = \binom{\eta}{Z}$$

假设在 η 值范围内, A1 片段的变异数目是 n_a , B1 片段的变异数是 n_b , 那么在 η 范围 (但在 Z 值内) 变异数目为:

$$\overline{w_Z} = \frac{(\eta - Z)!}{(\eta + Z - n_a - n_b)!(n_a - Z)!(n_b - Z)!}$$

由于分布在所有甲型流感病毒毒株变异方式的数目受到 η 以外变异数目的约束, 定义 w_A 与 w_B 作为在 η 以外 PB1 蛋白和 PA 蛋白的变异方式数目:

$$w_A = \frac{(M_A - \eta)!}{(N_A - n_a)!(M_A - \eta - (N_A - n_a))!}$$

$$w_B = \frac{(M_B - \eta)!}{(N_B - n_b)!(M_B - \eta - (N_B - n_b))!}$$

计算在 Z 中可能发生重叠的变异方法总数, 需要将 n_a 和 n_b 的所有可能值累加求和:

$$W_Z = w_Z \sum_{n_a=Z} \sum_{n_b=Z} \overline{w_Z} w_A w_B$$

最后, P 值表示为下式:

$$P(Z | N_A, N_B, M_A, M_B) = \frac{W_Z}{W} = \frac{w_Z \sum_{n_a=Z} \sum_{n_b=Z} \overline{w_Z} w_A w_B}{\binom{M_A}{N_A} \binom{M_B}{N_B}}$$

其中, W 为没有任何限制条件下, 筛选自所有 PB1 蛋白和 PA 蛋白变异的总方式数。

这里 P 值被 Wang 等定义为共进化值, 即文中计算的 P_{cov} 值^[9]。

共进化相关变量具体计算和图形输出均利用 Matlab7.0 完成。

DSSP 是用于计算多肽在模拟过程中二级结构变化的专用软件。将 2znl.pdb 输入 DSSP 软件, 分析蛋白质氨基酸的溶剂可及性 (ACC), 程序自动列表会产生 ACC 数据。文中分别计算了 PB1 蛋白与 PA 蛋白的 ACC 值。根据 Rost 的工作, 暴露于水环境中的表面积大于其整体表面积 25% 的氨基酸为表面氨基酸。当两个表面氨基酸相互作用, 其亲水表面积减少超过 $1A^2$ 时, 定义为相互作用表面氨基酸^[10]。文中分别计算了氨基酸在形成复合体前单体状态下的 ACC 值和形成复合体后的 ACC 值。

本文采用 ZDOCK2.3 软件对蛋白质相互作用的

三维结构进行了模拟。

2 结果与分析

2.1 CD Search 结果与分析

根据 CD search 分析的结果, PB1 蛋白具有 447 个 C9MP, 对位置重叠片段进行整合, 得到 26 个域。这里的“域”指的是由数个 C9MP 连结在一起, 按照前后顺序构成的一个更大片段。如 386-396 的 KKIEKIRPLLI, 实际上是由 3 个 C9MP, 386-394 (KKIEKIRPL) 与 387-395 (KKIEKIRPLL) 与 388-396 (IEKIRPLLI), 重叠构成。PA 蛋白具有 373 个 C9MP, 对重叠片段整合, 可以得到 29 个域。

2.2 共进化数值计算与分析

本文计算了 PB1 蛋白的 447 个 C9MP 与 PA 蛋白的 373 个 C9MP 所有可能的相互作用, 即 447×373 个相互作用, 包含了 166 731 个数值点。利用 Matlab 的可视化工具输出图形 (见图 1), 图中的小色块单位是 $-\log(P_{cov})$ 值。为了更直观找出相互作用关系, 忽略渐变的可能性, 直接将 $-\log(P_{cov})$ 大于 5 (即 $P_{cov} > 10^{-5}$) 的区域用显著的红色标注出来, Wang 等^[9]认为这些红色区域是“毫无疑问地”在共进化上具有高度关联的区域。蓝色区域意味着低的共进化值, 蓝色区域在复合体结构和功能的稳定性方面可以被认为做出较少贡献。

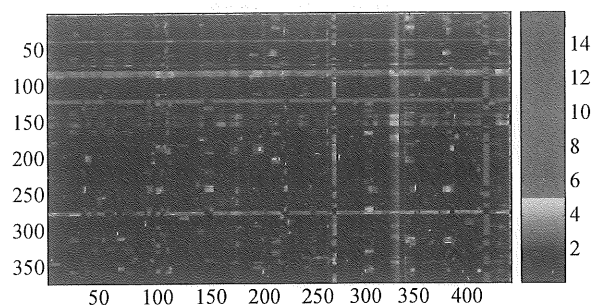


图 1 在假定极端保守情况下 PB1-PA 完整的 C9MP 共进化关系 (图中纵轴表示 PA 蛋白的 C9MP, 横轴表示 PB1 蛋白的 C9MP)

Fig. 1 The complete C9MP co-evolution between PB1 and PA in the condition of presumptive ultra-conservative. (y-coordinate indicates C9MP of PA protein, x-coordinate indicates C9MP of PB1 protein)

2.3 PB1 蛋白与 PA 蛋白相互作用位点预测与分析

Carol 等发现 PB1 蛋白、PB2 蛋白与 PA 蛋白都能在宿主受感染细胞中大量表达。这三个蛋白在宿主细胞中相互作用形成一个核酸酶复合体, 其中, PB1 蛋白可以与 PB2 蛋白相互作用而不需要依赖于其它因素; PA 蛋白可能需要经过修饰后才能够与 PB1 蛋白相互作用^[11]。Yasushi 研究小组找到了 PB1 蛋白与 PA 蛋白的相互作用域, 分别位于 PB1 蛋白的 1-140 处与 PA 蛋白的 201-716^[12]。

依据 CD Search 结果, 对于 PB1-PA 蛋白复合体来说, PB1 蛋白的氨基酸位点 1-140 包含 87 个 C9MP, PA 蛋白的氨基酸位点 201-716 包含 245 个 C9MP。

图 1 表明, PB1-PA 的结构域有许多在进化上具有相关联的域, 其中, 不少具有极高的 Pcov 值。但是, 由于目前 PDB 数据库中只有一个 PB1 蛋白与 PA 蛋白相互作用的结构图, 为了便于验证与对照, 本文选择了 PDB 数据库中的 PB1^{MP1}, 探讨其与 PA 蛋白的相互作用, 探究相互作用位点, 该方法可以推广至其它区域相互作用位点的预测分析。

表 1 PB1^{MP1} 的基本信息

Table 1 Basic information of PB1^{MP1}

编号	保守九聚片段编号	实际氨基酸	氨基酸的位置	包含氨基酸数
1	1-5	MDVNPTLLFLKVP	1-13	13

计算发现 PB1^{MP1} 与 PA 蛋白上的 C9MP 仅在 1、10、18-22、27-33、40-46、58-59、144-145、151-154、220-221、231-234 位置上 (见表 2) 同时具有 PB1^{MP1} 与 PA 蛋白残基 201-716 区域内最低的共进化值。

通过蛋白质的三维结构模拟, 可以显著地发现 PA 蛋白上与 PB1^{MP1} 可能存在共进化关系的各个 C9MP 与 PB1^{MP1} 在结构上距离较远。但是, 在这 12 个 C9MP 中, 惟一例外的是编号 10 的 C9MP, 模拟显示其与 PB1^{MP1} 具有显著的接触。为了便于下文引用, 参照 Wang 的命名规则, 即依据氨基酸片段的首尾字母和起始位置编码, 命名 10 号 C9MP 为 PA^{QC670}。

对 PA^{QC670} 片段变异统计分析发现 (见表 3), 同为接触面的氨基酸, Q670 比 R673 更为保守。统计发现 R673 位置上共有 13 个 R 变异为 K。由

于相互作用位点在复合体中的功能十分重要, 检索分析表明, PA^{QC670} 与 PB1^{MP1} 的变异实际上没有存在于同一个病毒中, 与 R673 对位的 PB1^{MP1} 的氨基酸没有发生协同突变, 因此, 可以排除 R673 作为潜在在相互作用位点的可能。在对这些变异检索过程没有发现 Q670 突变。在此, 定义共同序列为剔除变异后的剩余序列。序列比对发现, Q670 在接近 1580 个病毒株中没有发生突变, 因此, Q670 属于高度保守位点, 也是可能潜在的相互作用位点。

表 2 PA 蛋白与 PB1 蛋白可能存在共进化关系的 C9MP 位置

Table 2 The co-evolution positions of C9MP between PA protein and PB protein

编号	C9MP 位置	实际氨基酸	氨基酸的位置	含氨基酸数
1	1	RLADQSLPP	213-221	9
2	10	YVDGFEPNG	232-240	9
3	18-22	RSKFLMLDALKLS	279-291	13
4	27-28	LKLSIEDPSH	288-297	10
5	29-33	IEDPSHEGEGIPL	292-304	13
6	40-46	KNMKKTSQLLGENMA	358-375	15
7	58-59	IASMRRNYFT	438-447	10
8	144-145	EFSLTDPRL	524-533	10
9	151-154	EPHKWEKYCVLE	533-544	12
10	220-221	QALRDNLEPG	670-679	10
11	231-234	GLYEAIEECLIN	685-696	12
12	236	IEECLINDP	690-698	9

表 3 中的黑斜体字母表明某个位置氨基酸的变异情况。在我们所关心的氨基酸 Q670 及 R673 中, Q670 的保守度要远远高于 R673 的保守度。E677, P687, G679 等位点高度保守, 如果两者存在相互作用, 需要确认与 Q670 对位的 PB1 氨基酸具有稳定性的特征。与 Q670 对位的 PB1 氨基酸分别是 F9、V12、P13 和 A14。在这 4 个氨基酸中, A14 由于其在所有病毒株的 PB1 蛋白变异率高于 10%, 因而被放弃。在表 4 中, 黑体字母同样代表该位置氨基酸的变异情况, 计算表明, 除了 6、8 位置的氨基酸外, 其它氨基酸发生变异的程度都较低; 可以看出, F9、V12 和 P13 均具有很高的保守度, 表明这些位点均有参与相互作用的可能。

利用三维模拟的方式, 本文重构了 PA^{QC670} 和 PB1^{MP1} 的相互接触情况 (见图 2)。局部放大了关键相互作用位点 (图 3)。图中的网格表示了整个蛋白的空间范围, Q670 上下方的镂空区是氨基酸填充区; 从图 3 中可以看出, Q670 与 PB1^{MP1} 的 F9、

表 3 PA^{QC670} 变异情况

Table 3 The mutations of PA^{QC670}

PA 670 - 679	670	671	672	673	674	675	676
共同序列	Q	A	L	R	D	N	L
> gi 110733271	Q	A	V	R	D	N	L
> gi 113496094	Q	A	L	R	D	S	L
> gi 115278258	Q	A	L	R	D	K	L
> gi 115279288	Q	A	L	K	D	N	L
> gi 149930197	Q	A	F	R	D	N	L
> gi 149930216	Q	A	F	R	D	N	L
> gi 149930235	Q	A	F	R	D	N	L
> gi 157283209	Q	A	L	K	D	N	L
> gi 159150248	Q	A	L	R	D	N	M
> gi 172052937	Q	A	L	K	D	N	L
> gi 190352428	Q	A	L	R	D	Y	L
> gi 190682210	Q	A	L	R	D	Y	L
> gi 193805126	Q	A	L	K	D	N	L
> gi 194304644	Q	A	L	K	D	N	L
> gi 194304682	Q	A	L	K	D	N	L
> gi 194304720	Q	A	L	K	D	N	L
> gi 194304739	Q	A	L	K	D	N	L
> gi 194304777	Q	A	L	K	D	N	L
> gi 194310544	Q	A	L	R	D	Y	L
> gi 194352124	Q	A	L	K	D	N	L
> gi 211997904	Q	A	L	K	D	N	L
> gi 211997923	Q	A	L	K	D	N	L
> gi 211997942	Q	A	L	K	D	N	L
> gi 212506480	Q	A	F	R	D	N	L
> gi 215538978	Q	A	L	R	D	K	L
> gi 215539019	Q	A	F	R	D	N	L
> gi 215539172	Q	A	F	R	D	N	L
> gi 71568555	Q	A	L	R	D	S	L
> gi 78097188	Q	A	L	R	D	S	L
> gi 83744365	Q	A	F	R	D	N	L

表 4 PB1^{MPI} 变异情况

Table 4 The mutations of PB1^{MPI}

PB1 1 - 13	1 - 8	9	10 - 13	PB1 14 - 15	14	15
共同序列	...	F	...	共同序列	.	A
> gi 115279364	...	F	...	> gi 110733273	.	A
> gi 115289089	...	F	...	> gi 113497118	.	V
> gi 138392702	...	F	...	> gi 134044411	.	A
> gi 145278824	...	F	...	> gi 158454194	.	A
> gi 158957652	...	F	...	> gi 189230912	.	A
> gi 172052953	...	F	...	> gi 209978260	.	A
> gi 188504421	...	F	...	> gi 211998317	.	V
> gi 218663901	...	F	...	> gi 218874837	.	A
> gi 218664015	...	L	...	> gi 218874898	.	A
> gi 76411281	...	F	...	> gi 73665893	.	A
> gi 77746868	...	F	...	> gi 76411281	.	A
> gi 191177921	...	F	...	> gi 77543657	.	A

V12 和 P13 在复合体中是紧密结合的，此处可能是 PA^{QC670} 与 PB1^{MPI} 的相互作用位点之一。

利用 DSSP 程序，如果将 PA 蛋白与 PB1 蛋白分别独立计算 ACC 值，PA 蛋白的 Q670，PB1 蛋白的 F9、V12、P13 的 ACC 值分别为 82、117、43 和 103 A²；如果考虑将 PA 蛋白与 PB1 蛋白结合形成复合体后，计算 ACC 值，则 PA 蛋白的 Q670、PB1 蛋白的 F9、V12、P13 的 ACC 值分别下降至 13、56、14 和 79 A²。计算结果见表 5。

表 5 PA^{QC670} 的 Q670 与 PB1^{MPI} 复合体形成前后参数变化

Table 5 The changes of parameters after the Q670 of PA^{QC670} and PB1^{MPI} forming a complexes

氨基酸位点	Q670	F9	V12	P13				
指标	ACC/A ²	ARR/%	ACC/A ²	ARR/%	ACC/A ²	ARR/%	ACC/A ²	ARR/%
复合体前	82	26.8	117	36.3	43	16.7	103	43.1
复合体	13	4.2	56	17.4	14	5.4	79	33.1

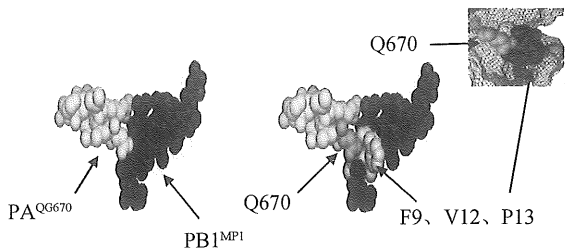


图 2 PA^{QC670} 的 Q670 与 PB1^{MPI} 的 F9、V12 和 P13 相互作用三维模拟效果

Fig. 2 The 3 - D simulation results of interactions between Q670 of PA^{QC670} to F9, V12 and P13 of PB1^{MPI}

在成为复合体前，上述 4 个氨基酸暴露在水环境中的面积与整体面积比 (ARR) 是 26.8%、36.3%、16.7% 和 43.1%；成为复合体后，上述 4 个氨基酸的 ARR 值下降到 4.2%、17.4%、5.4% 和 33.1%。

当 PB1 蛋白与 PA 蛋白尚未相互作用时，依据生物学定义，PA 的 Q670、PB1 的 F9 和 P13 都是表面氨基酸。特别 PB1 的 P13 (ARR_{max} = 43.1%) 高度暴露于水环境中。当 PB1 蛋白与 PA 蛋白相互作用时，PA 的 Q670、PB1 的 F9 和 P13

暴露在水环境中的表面积均迅速下降, 降幅 (比例) 分别达到 69 \AA^2 (22.6%)、 61 \AA^2 (18.9%) 和 136 \AA^2 (10.0%), 均显著超过了 Rost (1994)^[10] 给出的临界值 1 \AA^2 ; 因此, 它们均可成为相互作用的表面氨基酸。

3 讨论

本文的主要结论是 PA^{QC670} 的 Q670 与 PB1 的 F9、P13 构成了一个潜在的相互作用域, 该相互作用域由这 3 个氨基酸构成。

研究表明, 越是功能重要的区域, 如果出现变异, 蛋白 (或蛋白复合体) 失活的可能性越高^[13]。类似核酸酶复合体这种关乎病毒复制、繁殖的精密蛋白复合体, 出现变异的大部分结果是导致病毒无法繁殖子代病毒^[14]。相互作用表面氨基酸的保守性要明显高于其它表面氨基酸^[15]。

BLAST 和 C9MP 搜索结果显示, 甲型流感病毒的 PB1 蛋白与 PA 蛋白比 HIV 病毒的 RT 蛋白和 IN 蛋白保守许多^[9]。本文的研究过程与 Wang et al 的显著不同, 主要差异在于本文研究的 PA^{QC670} 与 PB1^{MP1} 的变异没有同时存在于同一个病毒中。

本研究也初步发现, 共进化理论不仅可用于寻找相互作用位点, 而且, 也有可能用于寻找存在共进化可能性的区域。本研究结果对于临床治疗的靶标发现, 疫苗和药物的研发具有重要指导意义。

参考文献:

- [1] WEBSTER R G, BEAN W J, GORMAN O T, et al. Evolution and ecology of influenza A viruses[J]. *Microbiol Rev*, 1992, 56(1): 152 - 179.
- [2] HONDA A, MUKAIGAWA J, YOKOIYAMA A, et al. Purification and molecular structure of RNA polymerase from influenza virus A/PR8 [J]. *J Biochem*, 1990, 107(4): 624 - 628.
- [3] NIETO A, DE LA LUNA S, BARCENA J, et al. Complex structure of the nuclear translocation signal of influenza virus polymerase PA subunit [J]. *J Gen Virol*, 1994, 75 (Pt 1): 29 - 36.
- [4] CAFFREY D R, SOMAROO S, HUGHES J D, et al. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? [J]. *Protein Sci*, 2004, 13(1): 190 - 202.
- [5] MINTSERIS J, WENG Z. Structure, function, and evolution of transient and obligate protein-protein interactions [J]. *Proc Natl Acad Sci U S A*, 2005, 102(31): 10930 - 10935.
- [6] GONZALEZ S, ZURCHER T, ORTIN J. Identification of two separate domains in the influenza virus PB1 protein involved in the interaction with the PB2 and PA subunits: a model for the viral RNA polymerase structure [J]. *Nucleic Acids Research*, 1996, 24(22): 4456.
- [7] ZURCHER T, de la LUNA S, SANZ-EZQUERRO J J, et al. Mutational analysis of the influenza virus A/Victoria/3/75 PA protein: studies of interaction with PB1 protein and identification of a dominant negative mutant [J]. *J Gen Virol*, 1996, 77 (Pt 8): 1745 - 1749.
- [8] PÉRALES B, de la LUNA S, PALACIOS I, et al. Mutational analysis identifies functional domains in the influenza A virus PB2 polymerase subunit [J]. *J Virol*, 1996, 70(3): 1678 - 1686.
- [9] WANG Y E, DELISI C. Inferring protein-protein interactions in viral proteins by co-evolution of conserved side chains [J]. *Genome Inform*, 2006, 17(1): 23 - 35.
- [10] ROST B, SANDEr C. Conservation and prediction of solvent accessibility in protein families [J]. *Proteins*, 1994, 20(3): 216 - 226.
- [11] ST A C, SMITH G E, SUMMERS M D, et al. Two of the three influenza viral polymerase proteins expressed by using baculovirus vectors form a complex in insect cells [J]. *J Virol*, 1987, 61(2): 361 - 365.
- [12] OHTSU Y, HONDA Y, SAKATA Y, et al. Fine mapping of the subunit binding sites of influenza virus RNA polymerase [J]. *Microbiol Immunol*, 2002, 46(3): 167 - 175.
- [13] DARLIX J L, LAPADAT-TAPOLSKY M, de ROCQUIGNY H, et al. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses [J]. *J Mol Biol*, 1995, 254(4): 523 - 537.
- [14] STEINHAEUER D A, HOLLAND J J. Rapid evolution of RNA viruses [J]. *Annu Rev Microbiol*, 1987, 41: 409 - 433.
- [15] OFRAN Y, ROST B. Predicted protein-protein interaction sites from local sequence information [J]. *FEBS Lett*, 2003, 544(1/3): 236 - 239.