

# 分子对接预测 H5 亚型禽流感病毒的广谱中和表位

颜渊清, 李少伟\*, 杨春燕, 罗文新, 王明桥, 陈毅歆, 罗海峰, 吴婷, 张军, 夏宁邵

国家传染病诊断试剂与疫苗工程技术研究中心(厦门大学), 厦门大学生命科学院, 厦门大学细胞生物学与肿瘤细胞工程教育部重点实验室, 厦门 361005

\* 联系人, E-mail: shaowei@xmu.edu.cn

2007-10-15 收稿, 2007-12-26 接受

国家自然科学基金项目(批准号: 30500092, 30640017, 30600106)、国家“十五”攻关项目(批准号: 2004BA519A73)、国家科技支撑计划(批准号: 2006BA101B06)、福建省重大科技专项(批准号: 2004YZ01-1)和福建省重点科技项目(批准号: 2005Y020)资助

**摘要** H5N1 禽流感病毒已经在亚洲、欧洲和非洲广泛传播, 造成了巨大损失. 最近我们鉴定出一株对多种来源的 H5N1 代表株均有良好中和活性的、识别 H5 亚型血凝素(hemagglutinin, HA)的广谱单克隆抗体 8H5, 它对寻找克服禽流感高变性的广谱治疗性抗体、疫苗和药物具有重要价值. 本研究应用分子模拟技术, 采用“典范结构”方法对 8H5 抗体 Fab 片段进行结构建模, 并通过能量分析、SAS 值分析、“拉曼强传图”检验、profile-3D 分析等理论验证, 获得较为合理的 8H5Fab 的三维空间结构. 8H5Fab 与 3 种 HA 蛋白晶体结构的分子对接结果表明, 8H5 抗体与 HA 蛋白的作用模式与 HA 宿主来源无关, 但与 HA 结构的亚型相似性相关. 综合抗原同源比对结果, 推测 8H5 抗体识别的广谱中和表位是由 HA 上的 Asp<sup>68</sup>, Asn<sup>72</sup>, Glu<sup>112</sup>, Lys<sup>113</sup>, Ile<sup>114</sup>, Pro<sup>118</sup>, Ser<sup>120</sup>, Tyr<sup>137</sup>, Tyr<sup>252</sup> 不连续氨基酸残基组成的构象表位. 这一模型将为 H5 亚型禽流感病毒广谱疫苗和治疗性药物的分子设计提供依据.

**关键词**

H5N1 禽流感病毒  
分子对接  
血凝素蛋白  
中和表位

2003 年底以来, 高致病性禽流感病毒 H5N1 在全球持续蔓延, 导致了多起大规模家禽死亡, 甚至引起了极高致死率的人类感染. 血凝素(hemagglutinin, HA)是病毒感染宿主细胞的受体结合部位, 是禽流感病毒表面最主要的结构蛋白之一, 也是机体保护性中和抗体的最主要识别抗原. 禽流感难以控制的重要原因之一在于其 HA 蛋白的极高变异率. 最近, 我们构建了包含 388 株具有血凝抑制(hemagglutinin inhibition, HI)活性的 H5 特异性单克隆抗体库. 用 46 株来自不同时间、不同地点、不同宿主分离到的 H5N1 病毒代表株对这些单抗的 HI 活性和中和活性进行测定, 发现其中 1 株单抗(8H5)对全部病毒株都有较强中和活性, 意味着 H5N1 病毒高保守性中和表位的存在<sup>[1]</sup>. 研究这些广谱中和单抗的结构对寻找克服 H5N1 高变异的突破口具有重要意义.

抗体分子具有典型的三维空间结构, 根据已知的抗体晶体结构可对已知可变区序列的抗体进行结

构建模<sup>[2,3]</sup>. 位于抗体 Fab 片段的 6 个抗原互补决定区(complementary-determining region, CDR)(轻链 3 个, 重链 3 个)的结构决定了抗体的识别特异性以及亲和力, 其氨基酸序列高度可变; 除了 CDR-H3 外, 其余的 CDR 可以组成含有特征结构的“典范结构”(canonical structure)的集合<sup>[3,4]</sup>. 基于这种“典范结构”的分子模拟是一种较为精确的抗体分子结构模拟方法<sup>[5]</sup>. 本研究运用该方法对 8H5 抗体的 Fab 片段分子结构进行模拟, 并将其与已报道的 3 种 H5 亚型 HA 蛋白晶体结构进行分子对接, 预测 8H5 单抗对应的中和表位组成, 为禽流感病毒广谱疫苗、药物的分子设计和广谱治疗性抗体研究提供重要依据.

## 1 材料与方法

( ) 抗体序列和抗原结构优化. 单克隆抗体 8H5 细胞株由本实验室制备并保存<sup>[1]</sup>. 从该细胞株中克隆可变区基因, 由上海博亚公司完成测序工作, 获得编码可变区的氨基酸序列为: 轻链可变区, EIVLT-

QSPAIMSASLGEKVTMSCRASSSVNFVYWYQQRSDASPKLLIYSSNLPAGVPPRFSGSGSNGNSYSLTISGLEGEDAATYYCQHFTSSPYTFGGGKLEIKRLE;重链可变区, QVQLQQSGAELMKPGASVKISCKATGYTFSNYWIEWIKRPGHGLEWIGEILPGSDRTNYNGKFKGKATFTADTSSNTAHMQLSSLTSEDSAVYYCANRYDGYFGLDYWGQGTSVTVSS.

目前已公开的 H5 亚型禽流感病毒 HA 蛋白的晶体结构有 3 个, 其 PDB 编码为 1j5m (病毒株: A/DUCK/SINGAPORE/3/97; 来源: 禽), 2fk0 (病毒株: A/VIETNAM/1203/2004; 来源: 人) 和 2ibx (病毒株: VN1194; 来源: 人)<sup>[6-8]</sup>. 在进行抗原-抗体分子对接前, 抗原 HA 蛋白均以单体形式的结构在 Insight 软件包中进行结构优化.

( ) 抗体的一级结构序列编号和“典范结构”分析. 根据 Chothia 的抗体序列编码规则, 抗体链中氨基酸残基的缺失或者插入经常发生在 L31, L95, L106, H31, H82 和 H100 位点<sup>[2]</sup>. 在轻链的可变区(V<sub>L</sub>)中, 定位于 CDR-L1 和 CDR-L3 前一位的两个保守的半胱氨酸(Cys)残基总是在抗体的第 23 和 88 位. CDR-L1 从第 24 位开始, 到第 34 位结束; 而 CDR-L3 从第 89 位开始, 到第 97 位结束. CDR-L1 和 CDR-L3 结束后的氨基酸残基经常是 Trp 和 Phe-Gly. 而 CDR-L2 经常在 CDR-L1 结束后的第 16 个氨基酸残基开始, 其前面的氨基酸残基一般是 Ile-Tyr, 序列定位在第 50~56 位之间. 在重链的可变区中, 第一个保守的 Cys 在第 22 位上, 刚好在 CDR-H1 开始前的第 4 个氨基酸残基; 第二个保守的 Cys 在第 92 位上, 在 CDR-H3 开始前的第 3 个氨基酸残基. CDR-H1 从第 26 位开始到 32 位结束, 结束后的第一个氨基酸残基经常是 Trp. CDR-H3 从第 95 位开始到 102 位结束, 结束后的氨基酸残基经常是 Trp-Gly. CDR-H2 定位在第 52 位到第 56 位间. 即使这些 CDR 是高可变性的, 其中的 5 个(除了 CDR-H3)可以组成一个叫做“典范结构”的集合, 每一个 CDR 都可以在这个集合里找到它的主链构象. 每个 CDR 属于哪种“典范结构”是由这个 CDR 的长度、特异性位点的氨基酸残基类型决定的<sup>[5]</sup>.

( ) 轻链模型(L)和重链模型(H)的构建. 在 Insight 软件包的同源建模程序块(homology)中, 轻链和重链的氨基酸序列通过 FASTA 方法与蛋白质数据库的序列进行比对<sup>[9,10]</sup>. 同源性高于 50%且 X 射线衍射解析度好的结构作为模板. 通过与模板序列的比对, 保守区域的坐标位点首先被赋予目标序列. 对

于 CDR 区域(除了 CDR-H3), 通过搜索蛋白质结构数据库而找到相同类型的“典范结构”坐标, 将这个坐标赋予 CDR 区域的序列. 对于 CDR-H3, 由于其极高的可变性, 在蛋白质结构数据库的搜索结果中, 将 RMSD 值最小的结构的坐标值赋予 CDR-H3. 得到的模型通过 Refine 程序对 N 端、C 端和结合间的区域进行修补. 在进入下一步程序前控制其优化前后的 RMSD 值小于 0.05 nm 的情况下, 逐步释放模型. 在 Discover 程序块中, 固定保守框主链, 在 cvff 力场下采用最陡下降法、共轭梯度法和分子动力学对保守框侧链及其 CDR 进行能量优化.

( ) L 链模型和 H 链模型的配对. 在 L 模型和 H 模型间界面的氨基酸残基是高度保守的. 这些保守的氨基酸残基定位在 L33~L39, L43~L47, L84~L90, L98~L104, H34~H40, H44~H48, H88~H94 及 H103~H109<sup>[3,11]</sup>. 通过最小平方法, 在保证 RMSD 在小于 0.05 nm 的情况下, L 模型和 H 模型相互配对在一起. 这个配对好的模型命名为 8H5Fab. 通过 BUMP 程序, 找出相互间有可能碰撞的氨基酸残基, 有可能严重碰撞的氨基酸残基通过改变二面角优化模型. 随后, 在 Discover 程序块中, 固定 8H5Fab 的骨架, 在 cvff 力场下先后通过最陡下降法和共轭梯度法进行 1000 步的能量最小化. 得到最低能量的 8H5Fab 模型进行残基相容性、氢键组成和相互作用能量等测试.

( ) 8H5Fab 和 HA 抗原的分子对接. 在 InsightII 软件包的 Zdockpro 模块中, 8H5Fab 和优化后的抗原通过 Zdock 进行分子对接<sup>[12]</sup>. 其中 8H5Fab 作为受体, 而抗原作为配体. 对接过程中为了节省计算量, 抗原结构中的 HA2 区域和 8H5Fab 中的恒定区域被阻止参与对接. 从 Zdock 的输出文件中, 前 30 个复合物被筛选出来进行进一步的 Rdock 分析<sup>[13]</sup>. 综合 Zdock 分值、Rdock 分值和 CDR 结合区域, 抗原和抗体复合物被选定进行下一步分析.

( ) 复合物溶剂可及表面分析和表位确定. 以水分子半径 0.14 nm 作为半径标准, 计算 HA 蛋白单体和 HA-8H5Fab 复合物的氨基酸残基溶剂可及表面(solution accessible surface, SAS)面积, 进而比较氨基酸残基在单体和复合物上所得 SAS 值的差异, SAS 减少值大于 0.07 nm<sup>2</sup>的氨基酸残基被认为是 8H5 抗体作用的潜在表位位点<sup>[14-16]</sup>. 计算 HA-8H5Fab 复合物相互作用界面(interface)上重原子间的距离, 满足形成氢键、盐键或者范德华力的原子所在的氨基酸残基

被认为是潜在表位位点。同时满足上述两个条件,并综合3种HA蛋白结构同源比对的结果,HA蛋白上的氨基酸残基定义为计算机模拟下的HA蛋白的8H5表位。

## 2 结果与分析

### 2.1 8H5 抗体 Fab 片段的分子模建

( ) 8H5Fab 结构模板的获得和一级结构分析。将8H5的V<sub>L</sub>序列和V<sub>H</sub>序列与PDB中的蛋白序列进行FASTA比对,其比对结果如图1所示,根据模板的结构域的分布,我们可以得到待模建的8H5抗体可变区中的保守氨基酸、保守框架和抗原互补区域。

根据 Chothia 的抗体命名法则,将8H5抗体可变区序列进行标准化命名(表1)。结果显示,CDR-L1的31位氨基酸残基发生缺失;而在CDR-H2的52位插入1个氨基酸残基P52a,CDR-H3的100位插入3个氨基酸残基F100a,G100b和L100c。根据CDR的长度及其保守氨基酸残基的分布,可对8H5抗体进行“典范结构”分类,即CDR-L1,CDR-L3和CDR-H1属于分类中的第1类,而其余的两个区域为第2类<sup>[5]</sup>。

( ) 8H5Fab 片段的分子模建。分子模建的结果如图2所示,L链和H链都具有典型的抗体结构。L链有211个氨基酸残基,其中前109个属于可变区域,剩下的102个残基为恒定区的部分。H链共有219个

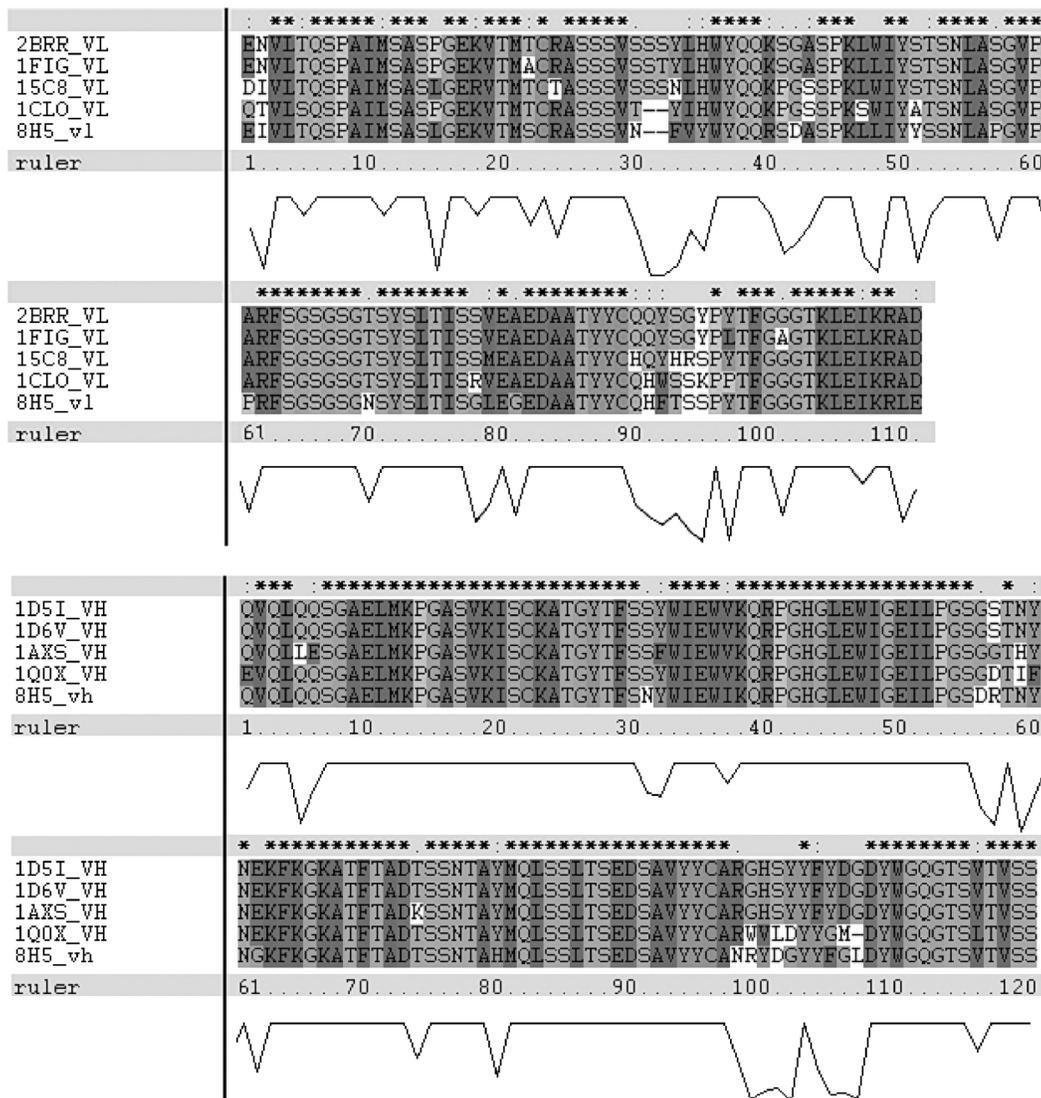


图1 8H5 抗体 V<sub>L</sub>, V<sub>H</sub> 序列与已知结构模板的多重比对结果(Clustal X)

表 1 8H5 抗体序列编号和“典范结构”类型分类

	氨基酸残基	编号	Chothia 编号	典范结构类型
CDR-L1	RASSSVNFVY	24~33	24 25 26 27 28 29 30 32 33 34	1
CDR-L2	YSSNLAP	49~55	50 51 52 53 54 55 56	2
CDR-L3	QHFTSSPYT	88~96	89 90 91 92 93 94 95 96 97	1
CDR-H1	GYTFSNY	26~32	26 27 28 29 30 31 32	1
CDR-H2	LPGSDR	52~57	52 52a 53 54 55 56	2
CDR-H3	RYDGYFFGLDY	99~109	95-100 100a 100b 100c 101 102	
其他	LSSLT	83~87	82 82a 82b 82c 83	

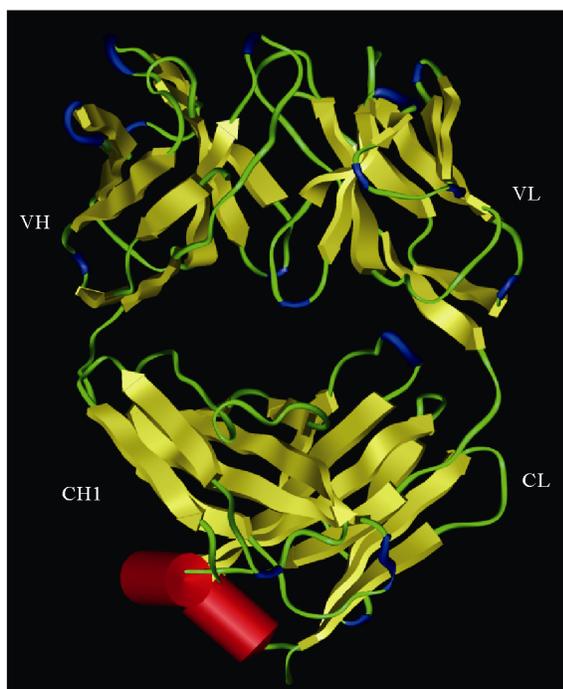


图 2 同源模建的 8H5Fab 三维空间结构

氨基酸残基, 其中 120 个属于可变区域, 剩下的 99 个残基为恒定区部分. 组成 L 链和 H 链的 4 个结构域都是由  $\beta$ -sheet 面对面包裹形成的, 每个结构域都有一对保守的二硫键:  $V_L$  的 C23 和 C88 间、 $V_H$  C22 和 C92 间以及  $C_{H1}$  和  $C_L$  各一对. 和其他典型的抗体结构一样, 二硫键键长约 0.2 nm.

( ) 8H5Fab 模建结构的评估. (1) 能量分析. 如表 2, 能量优化后的 L 链和 H 链的总能量分别为 3637.96289 和 4551.27780 kcal (1 cal = 4.1868 J). 而配对后的 8H5Fab 的总能量为 7875.70700 kcal, 比单独的 H 链和 L 链总能量之和少 313.53369 kcal. H 链和 L 链形成完整的 8H5Fab 结构时能量的下降表明, 完整的 8H5Fab 结构比单独暴露的 H 链和 L 链三维结构更为稳定. 如表 3, 8H5Fab 形成后减少的能量与 H

表 2 L 链、H 链和 8H5Fab 能量分析

	L 链/kcal	H 链/kcal	8H5Fab/kcal
范德华能量	1903.96936	1830.84920	3506.66400
排斥能量	9630.30371	9506.77250	19408.84000
色散能量	-7726.33447	-7675.92300	-15902.18000
库仑能量	1.61702	789.50665	705.74580
键能	536.83484	457.26556	994.10040
Theta 能量	874.28912	1071.52870	1945.81800
Phi 能量	251.47975	270.37045	521.85020
Out of plane 能量	7.66596	15.04769	22.71366
键-键能量	15.56962	33.08727	48.65688
键-theta 能量	44.38714	81.77821	126.16540
Theta-theta 能量	21.44454	31.37887	52.82341
Theta-theta-phi 能量	-19.29458	-29.53335	-48.82793
键-键(1-3)能量	0	0	0
Op-op 能量	0	0	0
Phi-phi 能量	0	0	0
氢键能量	0	0	0
总能量	3637.96289	4551.2778	7875.707

表 3 L 和 H 链的相互作用能量分析

能量分析模型: L 和 H	
范德华能量	= -225.68803 kcal
排斥能量	= 271.72363 kcal
色散能量	= -495.94510 kcal
库仑能量	= -85.37601 kcal
总能量	= -311.06405 kcal

链和 L 链间的相互作用能量相当, 说明两条链的配对是自由能驱使的, 符合热力学第二定律, 因此, 8H5Fab 的结构是合理的. (2) SAS 值分析. 以水分子半径 0.14 nm 作为半径标准, L 链的 SAS 值为 115.8041 nm<sup>2</sup>, H 链的 SAS 值为 118.5229 nm<sup>2</sup>. 而 8H5Fab 的 SAS 值为 197.3709 nm<sup>2</sup>, 比单独 L 链和 H 链的 SAS 值总和少了约 37 nm<sup>2</sup>. Chothia 等人<sup>[11]</sup>在研究  $V_L$  和  $V_H$  间的 SAS 后, 得到配对后的两个结构域交界面积大约为 18 nm<sup>2</sup> 左右. 由于 8H5Fab 由  $V_L$ - $V_H$  和  $C_L$ - $C_{H1}$  两对对等的结构域组成, 其配对后丢失的 SAS 值进一步说明了 8H5Fab 模建是合理的. (3) 氢键分析. 如表 4, 8H5Fab 在 L 链和 H 链之间的交界

表4 L链和H链间的氢键分析

H 供体	H 受体	键长/nm	键角(°)
L:38:HE22	H:39:OE1	0.205	155.28
L:1:HE2	H:46:OE1	0.183	156.42
L:94:HG	H:50:OE2	0.218	153.78
H:60:HD22	L:1:OE1	0.241	145.53
H:100B:HN	L:34:OH	0.163	152.72
L:36:HH	H:100C:O	0.193	172.15
L:43:HG	H:104:O	0.193	156.76
L:174:HG	H:164:ND1	0.243	148.73
L:174:HG	H:164:NE2	0.230	151.36
H:167:N	L:162:OG	0.290	NA <sup>a)</sup>
L:137:HD22	H:180:OG	0.234	165.49
L:122:HN	H:212:O	0.198	160.82
L:122:HN	H:212:OXT	0.238	130.38

a) NA 表示软件未检测到数值

面形成 13 对氢键, 其中 7 对位于可变区, 6 对位于恒定区。氢键是生物分子用于维系相互作用结构的重要作用力之一, 大量的氢键有利于 8H5Fab 的结构稳定。(4) 二面角分析. 如图 3, 对 8H5Fab 模型进行二面角分析, 有 80% 的氨基酸残基位于拉曼强传(Ramachandran)图的最优区域, 说明模建的 8H5Fab 结构是合理的. 在 8H5Fab 和 X 射线衍射解析的抗体结构进行叠加后, 模型 8H5Fab 和实验解析的抗体结构除了在 CDR 不同外其余的区域都有很好的重合性。(5) profile-3D 分析. 如图 4 对 8H5Fab 进行 profile-3D 测试, 其全局自相容数值(overall selfcompatibility)

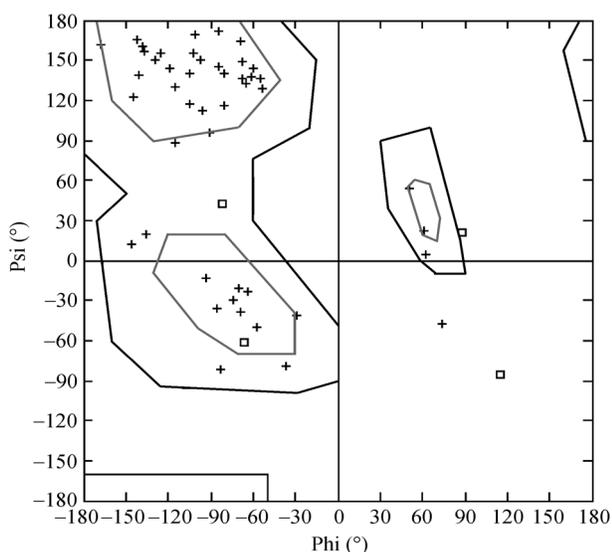


图3 抗原互补区域的“拉曼强传图”

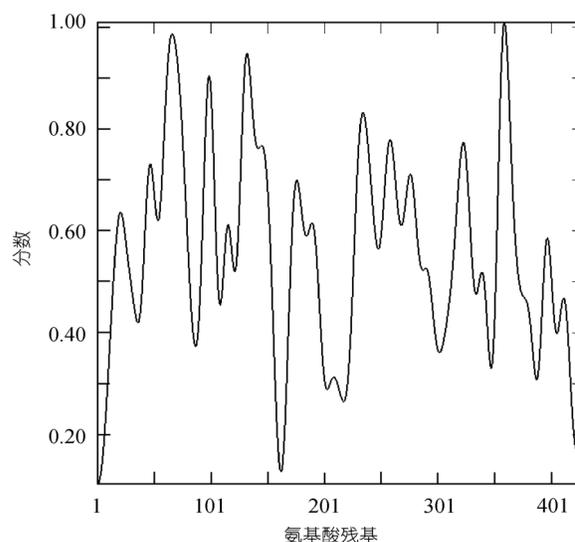


图4 8H5Fab模型的 profile-3D 程序测试

为 216.91. 根据软件的测试说明, 对于 8H5Fab 来说, 当全局自相容数值大于等于 196.23 时表示这个模型是合理的, 数值低于 88.30 时, 表示模型不正确. 此外, 通过 profile-3D 程序对 8H5Fab 的每个氨基酸残基进行测试, 其数值都大于 0, 表明模建的 8H5Fab 结构是处在较为合理的构象.

## 2.2 8H5Fab 和 HA 抗原的分子对接及其生物学功能

将上述模建的 8H5Fab 分子与 3 种 H5 亚型的 HA 蛋白进行分子对接, 如图 5 所示, 8H5Fab 的 CDR 区与 HA 抗原的结合位点位于抗原的受体结合亚结构域(receptor subdomain) aa110~260<sup>[8]</sup>, 提示 8H5 抗体作为 H5 亚型禽流感病毒中和抗体的生物学功能, 8H5 抗体可能竞争 HA 抗原与受体的结合, 从而行使中和作用.

保持 8H5Fab 的空间位置不变, 将 3 个对接结果重叠, 结果显示(图 5), 虽然抗原与 8H5Fab 的结合部位相当, 但结合角度却有明显的差异, 其中 8H5Fab 与 1jism (红色)、2ibx (绿色)的结合角度几乎相同, 而与 2fk0 (蓝色)的结合角度差别明显, 参与结合的 8H5 氨基酸侧链也有很大的不同(图 5(b)~(d)). 1jism HA 抗原分离自禽类动物, 而 2ibx 分离自人类, 提示 8H5 抗体与 H5 亚型的 HA 抗原的结合与病毒宿主无关; 与 1jism 及 2ibx 相比, 2fk0 与 H1 亚型 HA 结构更为接近<sup>[6]</sup>, 提示其可能是 H5 亚型向 H1 亚型突变的过渡结构.

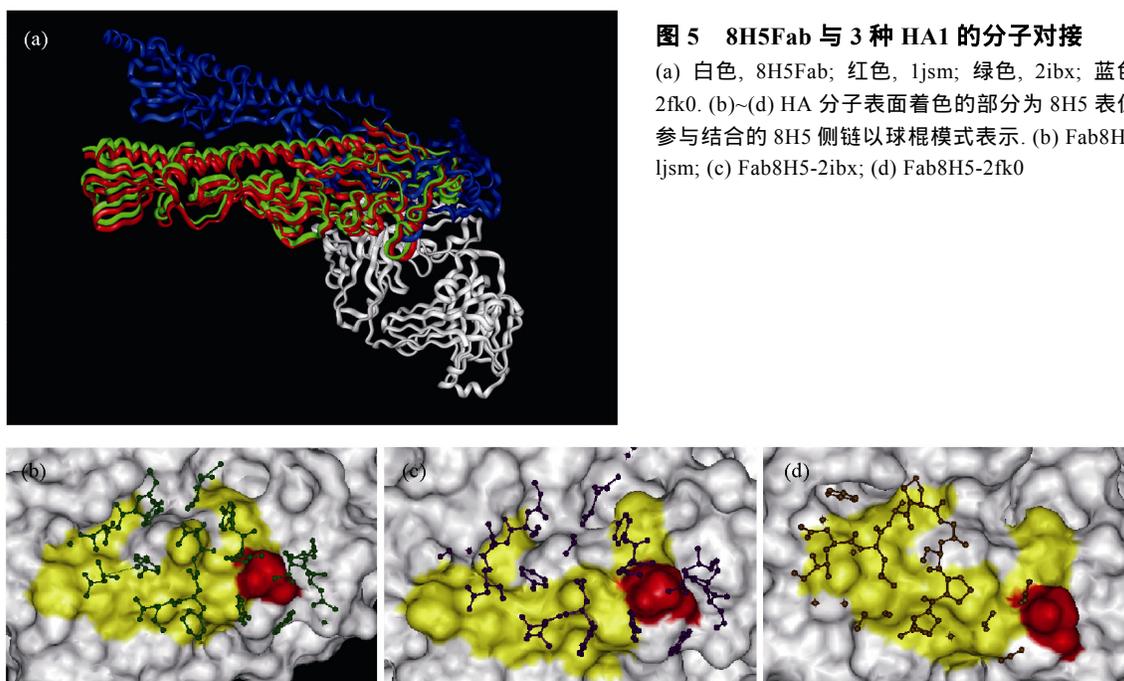


图5 8H5Fab与3种HA1的分子对接

(a) 白色, 8H5Fab; 红色, 1j5m; 绿色, 2ibx; 蓝色, 2fk0. (b)-(d) HA 分子表面着色的部分为 8H5 表位, 参与结合的 8H5 侧链以球棍模式表示. (b) Fab8H5-1j5m; (c) Fab8H5-2ibx; (d) Fab8H5-2fk0

表5 溶剂可及表面积分析 HA 蛋白的 8H5 表位( $\times 0.01 \text{ nm}^2$ )<sup>a)</sup>

2fk0			1j5m			2ibx		
位置	氨基酸	$\Delta$ SAS	位置	氨基酸	$\Delta$ SAS	位置	氨基酸	$\Delta$ SAS
A57	Lys	78.361	<b>A68</b>	<b>Asp</b>	27.716	<b>A72</b>	<b>Asp</b>	18.550
A60	Ile	23.262	A71	Leu	37.573	A75	Ile	66.375
A62	Arg	70.997	<b>A72</b>	<b>Asn</b>	86.130	<b>A76</b>	<b>Asn</b>	65.577
A75	Met	8.945	<b>A112</b>	<b>Glu</b>	28.131	<b>A116</b>	<b>Glu</b>	18.239
<b>A77</b>	<b>Asp</b>	40.472	<b>A113</b>	<b>Lys</b>	30.212	<b>A117</b>	<b>Lys</b>	28.056
A78	Glu	26.316	<b>A114</b>	<b>Ile</b>	34.115	<b>A118</b>	<b>Ile</b>	34.576
A79	Phe	143.975	A115	Arg	133.757	A119	Gln	84.019
<b>A81</b>	<b>Asn</b>	126.082	A116	Ile	7.654	<b>A122</b>	<b>Pro</b>	35.802
A82	Val	51.191	<b>A118</b>	<b>Pro</b>	48.655	A123	Lys	52.910
A82	Pro	69.096	A119	Arg	28.444	<b>A124</b>	<b>Ser</b>	104.211
A83	Glu	10.926	<b>A120</b>	<b>Ser</b>	109.740	A125	Ser	41.273
A117	His	18.864	A121	Ser	30.395	A126	Trp	9.254
A118	Phe	7.108	A123	Ser	68.298	A127	Ser	69.711
<b>A119</b>	<b>Glu</b>	56.822	A124	Asn	23.362	A130	Glu	17.771
<b>A120</b>	<b>Lys</b>	49.100	A126	Asp	8.974	A135	Val	11.537
<b>A121</b>	<b>Ile</b>	31.501	A127	Ala	17.753	<b>A141</b>	<b>Tyr</b>	54.689
A122	Gln	73.071	A131	Val	8.408	A142	Gln	28.543
A123	Ile	8.250	<b>A137</b>	<b>Tyr</b>	47.138	A144	Lys	43.149
<b>A125</b>	<b>Pro</b>	12.846	A140	Arg	37.735	A146	Ser	13.358
<b>A125</b>	<b>Ser</b>	12.562	A142	Ser	15.934	A149	Arg	60.859
<b>A141</b>	<b>Tyr</b>	33.140	A145	Arg	66.956	A157	Lys	69.367
A142	Gln	45.850	A153	Lys	50.086	A158	Asn	10.379
<b>A256</b>	<b>Tyr</b>	37.194	A154	Asn	64.563	A255	Glu	11.505
A273	Glu	73.261	A164	Tyr	9.150	<b>A256</b>	<b>Tyr</b>	65.716
A274	Tyr	56.550	A251	Glu	7.993			
A276	Asn	77.474	<b>A252</b>	<b>Tyr</b>	52.789			

a) 黑体+下画线示 3 种 8H5-HA 对接结果共有的 9 个氨基酸残基

### 2.3 HA 蛋白 8H5 表位的预测

( )  $\Delta$ SAS 值分析 8H5 表位. 根据对接结果计算 HA-8H5Fab 复合物中 HA 蛋白上的氨基酸残基的  $\Delta$ SAS 值,  $\Delta$ SAS 值大于 0 说明氨基酸残基位于复合物的相互作用界面, 值的大小表示氨基酸埋藏于界面中溶剂不可及的程度, 其中有些残基(如 2fk0 中的 Phe79, 1jism 中的 Ser120 和 2ibx 中的 Ser124)高达 90% 的区域是被 8H5Fab 包裹着. 以  $C\alpha$  相互距离为 0.4 nm 作为基准, 将 HA 蛋白上的  $\Delta$ SAS 值大于 0.07 nm<sup>2</sup> 的氨基酸残基列于表 5, 获得 HA 针对 8H5 抗体的表位氨基酸.

( ) HA 抗原序列同源比对分析 8H5 表位. 实验上, 8H5 抗体几乎能够与目前分离到的所有 H5 亚型禽流感病毒发生中和反应<sup>[1]</sup>, 提示 8H5 识别 H5 亚型 HA 蛋白的一个广谱中和表位, 因此, 通过综合上述  $\Delta$ SAS 值分析的 8H5 表位氨基酸(图 6 中灰背景示意)在 3 种 HA 蛋白的同源比对位置(图 6), 可以得到 8H5 表位的氨基酸组成:(以 1jism 氨基酸编号计)Asp<sup>68</sup>, Asn<sup>72</sup>, Glu<sup>112</sup>, Lys<sup>113</sup>, Ile<sup>114</sup>, Pro<sup>118</sup>, Ser<sup>120</sup>, Tyr<sup>137</sup>, Tyr<sup>252</sup>, 即 HA 蛋白的 8H5 表位是由 9 个不连续氨基酸残基

组成的构象表位(表 5, 下画线+黑体所示).

( ) 自然突变率分析 8H5 表位的保守性. 通过 Blast 分析 1000 个不同来源的 H5 亚型 HA 蛋白序列(统计于表 6), 8H5 表位氨基酸在序列的进化上高度保守的, 除了 Ser<sup>120</sup>, Tyr<sup>252</sup> 保守性为 95.9%和 87.2% 以外, 其他氨基酸保守性都在 99%以上, 提示这些中和表位氨基酸对于 H5 亚型病毒的生存是重要的, 同时说明本方法预测的 8H5 广谱中和表位的位置是较为可信的. 氢键相互作用分析 8H5 表位区: 分析 HA-8H5Fab 复合物中分子间的氢键相互作用现象, 如表 7, 相互作用界面上存在多对符合氢键形成条件的氨基酸残基侧链供受体对(属于中和表位的氨基酸分子以下画线所示). Lys<sup>113</sup>(对应在 2ibx 是 Lys<sup>117</sup>, 2fk0 是 Lys<sup>120</sup>, 表 7 中以下画线黑斜体所示)在 3 种 HA 分子中均可与 8H5Fab 形成氢键, 提示 Lys<sup>113</sup> 是 HA 蛋白上 8H5 中和表位的关键氨基酸. 然而, 3 个 HA 分子的 Lys<sup>113</sup> 参与形成氢键的原子不尽相同, 提示 1jism 和 2ibx 结构类似(O 作为 H 受体), 而与 2fk0 (HN 作为 H 供体)不同, 这符合上述的三者与 8H5Fab 作用角度的区别.

2ibx	.DQICIGYHA	NNSTEQVDTI	MEKNVTVTHA	QDILEKTHNG	KLCDLDGVKP	LILRDCSVAG
2fk0	GDQICIGYHA	NNSTEQVDTI	MEKNVTVTHA	QDILEKTHNG	KLCDLDGVKP	LILRDCSVAG
1jism	.DQICIGYHA	NNSTEQVDTI	MEKNVTVTHA	QDILEKTHNG	KLCDLNGVKP	LILRDCSVAG
2ibx	WLLGNPMCDE	FINVPEWSYI	VEKANPVNDL	CYPGDFNDYE	ELKHLLSRIN	HFEKIQIIPK
2fk0	WLLGNPMCDE	FINVPEWSYI	VEKANPVNDL	CYPGDFNDYE	ELKHLLSRIN	HFEKIQIIPK
1jism	WLLGNPMCDE	FLNVPEWSYI	VEKDNPVNGL	CYPENFNDYE	ELKHLLSSTN	HFEKIRIIPR
2ibx	SSWS <sup>68</sup> SHEASL	GVSSAC <sup>72</sup> PYQG	KSSFFR <sup>112</sup> NVW	LIKK <sup>113</sup> NSTYPT	IKRSYNN <sup>114</sup> TNQ	EDLLV <sup>118</sup> LWGIH
2fk0	SSWS <sup>68</sup> SHEASL	GVSSAC <sup>72</sup> PYQG	KSSFFR <sup>112</sup> NVW	LIKK <sup>113</sup> NSTYPT	IKRSYNN <sup>114</sup> TNQ	EDLLV <sup>118</sup> LWGIH
1jism	SSWS <sup>68</sup> NHDASS	GVSSAC <sup>72</sup> PYNG	RSSFFR <sup>112</sup> NVW	LIKK <sup>113</sup> NAYPT	IKRSYNN <sup>114</sup> TNQ	EDLL <sup>118</sup> LWGIH
2ibx	HPNDA <sup>137</sup> A <sup>137</sup> AEQTK	LYQNPTTYIS	WGTSTLNQRL	VPRIATRSKV	NGQSGRMEFF	WTILKPNDAI
2fk0	HPNDA <sup>137</sup> A <sup>137</sup> AEQTK	LYQNPTTYIS	WGTSTLNQRL	VPRIATRSKV	NGQSGRMEFF	WTILKPNDAI
1jism	HPNDA <sup>137</sup> A <sup>137</sup> AEQTK	LYQNPTTYVS	WGTSTLNQRS	VPEIATRPKV	NGQSGRMEFF	WTILKPNDAI
2ibx	NFESNGNFIA	PEYAYKIVKK	GDSTIMKSEL	EYGCNTRCQ	TPMGAINSSM	PFHNIHPLTI
2fk0	NFESNGNFIA	PEYAYKIVKK	GDSTIMKSEL	EYGCNTRCQ	TPMGAINSSM	PFHNIHPLTI
1jism	NFESNGNFIA	PEYAYKIVKK	GGSAIMKSG	EYGCNTRCQ	TPMGAINSSM	PFHNIHPLTI
2ibx	GECPKYVKS	RLV <sup>252</sup> LATGLRN	SP			
2fk0	GECPKYVKS	RLV <sup>252</sup> LATGLRN	SP			
1jism	GECPKYVKS	RLV <sup>252</sup> LATGLRN	VP			

图 6 3 种 HA 蛋白 8H5 结合区的氨基酸序列同源比对

灰背景字表示溶剂可及表面分析得到的 3 种 HA 的 8H5 表位氨基酸残基, 方框内示意已报道的受体结合区域

表 6 8H5 表位氨基酸的自然突变频率

	8H5 表位氨基酸残基								
	Asp <sup>68</sup>	Asn <sup>72</sup>	Glu <sup>112</sup>	Lys <sup>113</sup>	Ile <sup>114</sup>	Pro <sup>118</sup>	Ser <sup>120</sup>	Tyr <sup>137</sup>	Tyr <sup>252</sup>
Cys	-	-	-	-	-	-	0.1%	-	-
Asp	<b>99.9%</b>	-	0.1%	-	-	-	0.6%	-	-
Glu	-	-	<b>99.6%</b>	-	-	-	-	-	-
Phe	-	-	-	-	-	-	-	0.1%	-
Gly	0.1%	-	0.2%	-	-	-	0.2%	-	-
His	-	0.1%	-	-	-	-	-	0.7%	0.2%
Ile	-	-	-	-	<b>99.3%</b>	-	-	-	-
Lys	-	0.3%	0.1%	<b>99.9%</b>	-	-	-	-	-
Leu	-	-	-	-	0.5%	-	-	-	-
Met	-	-	-	-	-	-	-	-	-
Asn	-	<b>99.5%</b>	-	-	-	-	3.2%	-	12.6%
Pro	-	-	-	-	-	<b>99.8%</b>	-	-	-
Gln	-	-	-	-	-	-	-	-	-
Arg	-	-	-	0.1%	-	-	-	-	-
Ser	-	-	-	-	-	0.2%	<b>95.9%</b>	0.1%	-
Thr	-	0.1%	-	-	-	-	-	-	-
Val	-	-	-	-	0.3%	-	-	-	-
Tyr	-	-	-	-	-	-	-	<b>99.1%</b>	<b>87.2%</b>

表 7 8H5Fab 抗体与血凝素蛋白间的氢键

H 供体	H 受体	键长/nm	键角(°)
8H5Fab:H25:HG1	<u>1j5m:A68:OD2</u>	0.236	139.11
<u>1j5m:A72:HD2<sup>a)</sup></u>	8H5Fab:H76:OD1	0.228	156.55
<u>1j5m:A72:HD22</u>	8H5Fab:H27:O	0.178	129.68
8H5Fab:H31:HD22	1j5m:A112:OE2	0.193	139.49
8H5Fab:H28:HG1	<b><u>1j5m:A113:O<sup>b)</sup></u></b>	0.178	130.03
1j5m:A115:HN	8H5Fab:H32:OH	0.181	124.02
1j5m:A115:HN	8H5Fab:H97:OD1	0.206	152.29
1j5m:A115:HH22	8H5Fab:H94:OD1	0.170	153.24
8H5Fab:L55:HN	1j5m:A125:O	0.187	138.59
8H5Fab:L55:HN	2ibx:A129:O	0.233	120.20
2ibx:A157:HZ1	8H5Fab:L55:O	0.181	126.66
2ibx:A157:HZ1	8H5Fab:L56:N	0.245	126.52
<u>2ibx:A76:HD22</u>	8H5Fab:H28:N	0.241	135.04
8H5Fab:H28:HG1	<b><u>2ibx:A117:O</u></b>	0.181	123.53
8H5Fab:H31:HD22	<u>2ibx:A116:OE2</u>	0.191	142.38
8H5Fab:H97:HN	2ibx:A119:OE1	0.211	150.73
2ibx:A119:HN	8H5Fab:H97:OD1	0.189	136.25
2fk0:A117:NE2	8H5Fab:L53:OD1	0.257	NA <sup>c)</sup>
<b><u>2fk0:A120:HN</u></b>	8H5Fab:L54:O	0.237	136.82
2fk0:A256:HH	8H5Fab:L57:O	0.194	147.24
2fk0:A62:HH21	8H5Fab:H26:O	0.168	127.18
8H5Fab:H28:HG1	2fk0:A273:OE1	0.225	143.37
8H5Fab:H28:HG1	2fk0:A273:OE2	0.189	124.43
2fk0:A57:HZ2	8H5Fab:H98:O	0.176	125.63
2fk0:A276:HD21	8H5Fab:H94:OH	0.203	138.56
2fk0:A79:HN	8H5Fab:H102:OH	0.206	151.32

a) 下画线示参与形成氢键的 HA 的 8H5 表位氨基酸残基的原子; b) 下画线+黑斜体示 3 种 HA 共有的参与形成氢键的 8H5 表位氨基酸残基的原子; c) NA 表示软件未检测到数值

### 3 讨论

HA 作为单独的多肽蛋白(HA0)定位在病毒的膜蛋白上. 如果要使病毒具有感染性, 那么 HA0 必须通过宿主的类胰岛素蛋白酶分解成两个肽段: HA1 和 HA2. 通过和细胞表面的唾液酸受体结合, HA 蛋白能够调节病毒与宿主细胞的吸附, 进而进入宿主细胞; 同时, HA 也是机体抗流感病毒体液免疫的重要靶抗原, 其抗体大多具有中和活性<sup>[17]</sup>. 因此, HA 是针对 H5N1 药物和疫苗开发的重要靶标. 不仅如此, HA 还能够用于监测现有的疫苗和诊断方法对新病毒株的有效性及其潜在免疫交叉反应. 流感疫苗的一大缺陷在于这种疫苗只能对特异的病毒株有免疫保护作用, 因此每年都要研制针对每一个不同病毒株的流感疫苗. 如果能够找到流感病毒株上的广谱表位, 针对这个表位所研制出来的流感疫苗和诊断试剂将可能克服目前疫苗或者诊断的缺陷, 大大增强人们对流感疫情的控制能力<sup>[18]</sup>.

Damien 实验室报道了两个基于 X 射线衍射的 HA 抗原表位<sup>[14]</sup>, 其中 HA 抗原来自引起 1968 年流感爆发的 H3 亚型病毒. 自 1997 年开始, H5N1 已逐渐成为目前最具破坏性的流感病毒. 对于 H5 亚型的中和表位的研究, 尚无成功获得结构解析的报道. 在前期获得的禽流感病毒广谱中和单抗 8H5 基础上, 本研究通过分子建模和分子对接分析其识别的 H5 广谱中和表位.

抗体与抗原的活性主要是通过轻链可变区和重链可变区来实现的, 为了便于后续分子对接结果的筛选, 本研究通过模拟 8H5 抗体的 Fab 段而不是通过模拟具有同样抗体活性的单链抗体模型来进行<sup>[19,20]</sup>. 通过分析 8H5 单抗的一级结构, 找到了可变区的保守框架和 CDR(表 1). 抗体保守框架具有较高的保守性, 模拟所得结构准确性也较高. 而 CDR 区可以通过“典范结构”进行分类, 类别分类基于 CDR 区的长度, 抗体关键位点氨基酸的类别, 在“典范结构”中, 相同的类别构象相似, 而不同的类别构象差异大. 采用“典范结构”方法模拟得到的抗体结构, 在能量测试, SAS 分析, “Ramachandran 图”检验, profile-3D 分

析及其与 X 射线衍射得到的晶体结构叠加后均验证这个模型的正确性. 分子对接采用的是 InsightII 软件中的 Zdockpro 模块, 为了验证其相对准确性, 我们将上面提及的 H3 亚型的 HA 及抗体复合物中分离出单独的 HA 和抗体的结构, 然后, 在综合预设对接参数的情况下使用 Zdock 和 Rdock 对接, 获得与晶体结构几乎重叠的对接结果(结果略), 这些对接参数应用于 8H5Fab 与 3 种 HA 蛋白的分子对接中.

比较 8H5Fab 与 3 种 HA 蛋白的对接结果, 虽然 1j5m 和 2ibx 在宿主上有所不同, 但是抗体与抗原结合的角度几乎相同, 说明 8H5 表位与 HA 抗原的宿主类型无关, 提示 8H5 作为广谱抗体可用于 H5 亚型的不同 HA 抗原的检测. 而 1j5m 和 2ibx 的抗体作用模式与 2fk0 有很大的差异. 当对接结果叠加时, 2ibx 与 1j5m 的 RMSD 值小于 0.01 nm, 2fk0 与 1j5m 的 RMSD 值却在 0.15 nm 左右. 结构上, 2fk0 HA 的结构更接近于 H1 亚型; 实验上, 8H5 对 H1 亚型无中和作用, 提示 8H5 抗体与 HA 结合的亚型特异性, 为研究 H1 和 H5 亚型的转变提供了线索, 相互界面上氢键的分析也印证了这一结构差异. 运用  $\Delta$ SAS 值分析 HA-8H5Fab 的相互作用界面, 综合 HA 蛋白的同源比对结果, 我们获得了 H5 亚型 HA 蛋白上 8H5 表位的组成氨基酸(以 1j5m 氨基酸编号计): Asp<sup>68</sup>, Asn<sup>72</sup>, Glu<sup>112</sup>, Lys<sup>113</sup>, Ile<sup>114</sup>, Pro<sup>118</sup>, Ser<sup>120</sup>, Tyr<sup>137</sup>, Tyr<sup>252</sup>, 即 HA 蛋白的 8H5 广谱中和表位是由 9 个不连续氨基酸残基组成的构象表位. 该表位分布于 HA 受体结合亚结构域的 3 个区域, 除了有 4 个氨基酸残基在  $\beta$  折叠外, 其余的均位于血凝素蛋白的无规则卷曲结构中. 对接结果中在 1j5m 抗原中的 Glu<sup>112</sup> 到 Ser<sup>121</sup> 和 Tyr<sup>137</sup> 到 Lys<sup>140</sup> 的区域落在了 Stevens 等人<sup>[6]</sup>2006 年报道的 H5N1 抗原表位区域. 另外, 自然突变率的统计也显示本研究预测的中和表位的高度保守性. 氢键在蛋白质分子的相互作用中行使着重要的作用, 氢键分析结果显示 8H5 表位中的 Lys<sup>113</sup> 可能对于抗原抗体的相互作用有着重要的贡献, 为针对该表位的药物设计提供重要的信息. 当然, 分子对接的准确性有待于基因突变表达、表位分析<sup>[21]</sup>和结构解析等实验的最终验证.

致谢 感谢美国 Scripps 研究所的林天伟博士、新加坡国立大学的孔勇博士和韩晋华博士对本工作的指导和帮助.

## 参考文献

- 1 罗海峰, 陈毅歆, 陈自敏, 等. 一株抗 H5 亚型禽流感病毒血凝素蛋白单克隆抗体的广谱中和活性. 病毒学报, 2007, 23: 85—90
- 2 Brucoleri R E, Haber E, Novotny J. Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature*, 1988, 335: 564—568
- 3 Chothia C, Lesk A M, Tramontano A, et al. Conformations of immunoglobulin hypervariable regions. *Nature*, 1989, 342: 877—883
- 4 Chothia C, Lesk A M. Canonical structures for the hypervariable loops of immunoglobulins. *J Mol Biol*, 1987, 196: 901
- 5 Veronica M, Arthur M L, Anna T. Antibody modeling: Implications for engineering and design. *Methods*, 2000, 20: 267—279
- 6 Stevens J, Blixt O, Tumpey T M, et al. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science*, 2006, 312: 404—410
- 7 Yamada S, Suzuki Y, Suzuki T, et al. Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to Human-type receptors. *Nature*, 2006, 444: 378—382
- 8 Ha Y, Stevens D J, Skehel J J, et al. H5 avian and H9 swine influenza virus haemagglutinin structures: Possible origin of influenza subtype. *EMBO J*, 2002, 21: 865—875
- 9 Lipman D J, Pearson W R. Rapid and sensitive protein similarity searches. *Science*, 1985, 227: 1435—1441
- 10 Pearson W R. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol*, 1990, 183: 63—98
- 11 Chothia C, Novotny J, Brucoleri R, et al. Domain association in immunoglobulin molecules. The packing of variable domains. *J Mol Biol*, 1985, 186: 651—663
- 12 Chen R, Weng Z. A novel shape complementarity scoring function for protein-protein docking. *Proteins*, 2003, 51: 397—408
- 13 Li L, Chen R, Weng Z. RDOCK: Refinement of rigid-body protein docking predictions. *Proteins*, 2003, 53: 693—707
- 14 Fleury D, Daniels R S, Skehel J J, et al. Structural Evidence for recognition of a single epitope by two distinct antibodies. *Proteins*, 2000, 40: 572—578
- 15 Lee B, Richards F M. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol*, 1971, 55: 379—400
- 16 Connolly M L. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 1983, 221: 709—713
- 17 Thomas J K, Noppenberger J. Avian influenza: A review. *Am J Health Syst Pharm*, 2007, 64: 149—165
- 18 Bui H H, Peters B, Assarsson E, et al. Ab and T cell epitopes of influenza A virus, knowledge and opportunities. *Proc Natl Acad Sci USA*, 2007, 104: 246—251
- 19 Morioka H, Miura H, Kobayashi H, et al. Antibodies specific for (6-4) DNA photoproducts: Cloning, antibody modeling and construction of a single-chain Fv derivative. *Biochim Biophys Acta*, 1998, 1385: 17—32
- 20 Guo C Z, Wu J H, Wang Y X, et al. Molecular simulation of a single-chain antibody against AChE to explore molecular basis of inhibitory effect of 3F3 McAb to enzyme activity. *Acta Pharmacol Sin*, 2003, 24: 460—466
- 21 Li G, Tao S, Wang X J. Sequence and epitope analysis of surface proteins of avian influenza H5N1 viruses from Asian patients. *Chin Sci Bull*, 2006, 51: 2472—2481