抗原肽与 MHC 分子相互作用 QM/MM 分子动力学模拟研究

冯石磊 胡墅 刘兵 刘伟*

(大连理工大学高科技研究院 大连 116023)

摘要 在 MHC I 类(major histocompatibility complex class I)分子抗原加工提呈过程中抗原蛋白在抗原提呈细胞(antigen-presenting cells, APC)的胞浆中被蛋白酶体(proteasome)裂解成短肽 peptide, 由转运相关蛋白(transporter associated with antigen processing, TAP)将蛋白酶体裂解产生的短肽片段从胞浆转运至内质网腔. 短肽 peptide 在内质网中与新生成的 MHC I 类分子结合, 形成 peptide-MHC 复合体被提呈到 APC 细胞表面, 与 T 细胞表面抗原受体(T cell receptor, TCR)特异性识别结合, 使得 CTL 细胞开始活化、增殖、分化, 进而对肿瘤细胞进行特异性杀伤. 目前对 CTL 细胞如何识别抗原肽-MHC 复合物分子及抗原短肽 peptide 如何与主要组织相容性复合体 MHC 分子的相互作用识别结合的机理还不是很清楚. 传统的预测 CTL 细胞表位的方法没有考虑受体与配体结合过程中电子结构的变化, 电子结构的变化需要用量子力学方法来处理. 本文采用 QM/MM 多尺度生物大分子的分子动力学模拟方法, 以天然抗原肽 TAX (LLFGYPVYVYU)与 HLA-A*0201 分子结合的晶体结构为模板, 替换抗原肽“锚点”氨基酸, 将口袋氨基酸残基的原子极化电荷在空间形成的静电势用电多极矩分量表示. 用箱线图分析每个口袋氨基酸分子静电势变化和功能, 确定 Pocket B 的 Glu63 和 Lys66 的功能是精细识别氨基酸和一级结合氨基酸, Pocket F 的 Asp77, Tyr84 的功能是精细识别氨基酸, 而 Asp77, Lys146 是一级结合氨基酸, 表明 QM/MM 方法在提取抗原肽与 MHC I 类分子识别结合特异性信息是可行的, 这对了解免疫识别机理和指导肿瘤疫苗的开发都具有指导意义.

关键词 QM/MM 分子动力学模拟; 电多极矩; T 细胞表位; 外源性抗原; MHC I 类分子

QM/MM Molecular Dynamics Simulation for the MHC class I Molecule Interacting with Antigen Peptide

Feng, Shilei Hu, Shu Liu, Bing Liu, Wei*

(College of Advanced Science and Technology, Dalian University of Technology, Dalian 116023)

Abstract In the process of antigen presenting of major histocompatibility complex (MHC) class I molecule, antigen protein, which is in the cytoplasm of antigen-presenting cells (APC), is cleaved into short peptide fragments by proteasome, and the short peptide fragments will be transferred from the cytoplasm to the endoplasmic reticulum lumen by transporter associated with antigen processing (TAP). The peptide will combine with the MHC class I molecule, and the forming pMHC complex will be presented to the surface of APC cells, and it will be recognized by T cell receptor (TCR). And then, the CTL cells will be activated, and it will begin to proliferate and differentiate, and it will kill the tumor cells. The mechanisms how does the CTL cell recognize the peptide-MHC complex and how does the MHC molecule interact with the peptide are not clear. In traditional method, electronic structure changes is not considered during the peptide binding to the MHC molecules. The electronic structure changes need to use the method of Quantum Mechanics to deal with. In this paper, we use QM/MM multi-scale molecular dynamics simulation of biological macromolecules, and take the crystal structure of TAX-HLA-A*0201 molecules complex as template. We would replace the anchor amino acid of the peptide, and use the electric multipole moment component to present the electrostatic potential formed by the atomic polarization charge of pocket amino acid residues. We analyze the electrostatic potential changes and the functions of each pocket amino acid by box plots, we would consider that Glu63 and Lys66 of Pocket B are primary anchoring amino acids and fine recognition amino acid, Asp77 and Tyr84 of Pocket F are the fine recognition amino acid, Asp77 and Lys146 are the primary anchoring amino acids. It shows that QM/MM method is very effective in extracting the specific information of antigen peptide binding to MHC class I molecules. And it possesses some guiding significance to understand the mechanism of immune recognition and the development of tumor vaccine.

Keywords QM/MM molecular dynamic simulation; electric multipole moment; T cell epitope; exogenous antigen; MHC class I molecules

* E-mail: jchjys@dlut.edu.cn; Tel.: 0411-84707872

Received April 11, 2013; published June 27, 2013.

1 引言

T 细胞是重要的免疫细胞之一, 它的主要作用是介导细胞免疫, 调节机体的免疫功能. 外周成熟的 T 细胞实际上是由异质性群体构成的, 按功能不同可分为辅助性 T 细胞(helper T cell, Th)、细胞毒性 T 细胞(cytotoxic T lymphocyte, CTL)和抑制性 T 细胞(suppressor T cell, Ts). 现已证实 CTL 细胞能够对肿瘤靶细胞直接进行特异性杀伤^[1], 其过程是肿瘤抗原蛋白在抗原提呈细胞(antigen-presenting cells, APC)的胞浆中被蛋白酶体(proteasome)裂解成短肽 peptide(短肽长度一般为 8~11), 由转运相关蛋白(Transporter associated with antigen processing, TAP)将蛋白酶体裂解产生的短肽片段从胞浆转运至内质网腔, 短肽在内质网中与新合成的主要组织相容性复合体(major histocompatibility complex class I, MHC I 类)结合, 形成 peptide-MHC 复合体被提呈到 APC 细胞表面, 供 T 细胞表面抗原受体 TCR 特异性识别结合, 使得 CTL 细胞开始活化、增殖、分化, 从而对肿瘤细胞进行特异性杀伤. 通常将能够启动细胞免疫应答的抗原短肽称为 CTL 细胞表位.

抗原肽-MHC I 类分子复合体对 CD8⁺T 细胞的激活起着十分重要的作用. MHC I 类分子晶体结构是由 α 重链和 β_2 微球蛋白轻链以非共价键组成的异源二聚体. α 重链由 $\alpha 1$, $\alpha 2$ 和 $\alpha 3$ 结构域构成, $\alpha 1$ 和 $\alpha 2$ 结构域各含有 1 个 α 螺旋和 4 个 β 片层, 它们形成了抗原肽结合槽(peptide-binding groove), α 螺旋为结合槽壁, β 片层为结合槽底, 结合槽是中部宽两端窄, 槽底部有 6 个深浅不同的“口袋”(Pocket A, Pocket B, Pocket D, Pocket C, Pocket E 和 Pocket F), 这些 Pockets 富含多态性氨基酸, 确定某一等位基因编码的 MHC I 类分子只能与具有某一特定结构的一组抗原肽结合^[2]. 长度为 8~13 mer 抗原肽借助特定“锚点”残基(anchor residue)驻扎在 MHC I 类分子相应不同“口袋”(Pocket)中. 在抗原肽氨基酸序列 $P_1P_2P_3P_4P_5P_6P_7P_8P_9$ 中, 驻扎在 Pocket A 中的锚定残基位置被命名为 Position 1 (P_1), 驻扎在 Pocket B 中的锚定残基位置被命名为 Position 2 (P_2), 驻扎在 Pocket D 中的锚定残基位置被命名为 Position 3 (P_3), 驻扎在 Pocket C 中的锚定残基位置被命名为 Position 6 (P_6), 驻扎在 Pocket E 中的锚定残基位置被命名为 Position 7 (P_7), 驻扎在 Pocket F 中的锚定残基位置被命名为 Position 9 (P_9). 抗原肽氨基酸序列及其结构功能的不同可直接影响 T 细胞效应. 对于一个已知氨基酸序列抗原蛋白, 实验上选取 T 细胞表位是采用先合成大量的交叠肽, 再通过抗原肽结合实验或 CTL 杀伤效应检测进行选取. 这种方法不仅费时费力, 费用大, 而且预测 CTL 细胞表位的效率也低, 因而有必要采用理论方法预测 T 细胞表位. 通过理论模型预测候选 CTL 表位, 可以减少实验的成本投入和提高 CTL 表位鉴定的效率. 从理论

方法上预测抗原表位主要分为 4 大类: (1)结合基序法^[3](binding motif), (2)定量矩阵法^[4] (quantitative matrix), (3)机器学习法(machine learning)(即人工神经网络、隐马尔可夫、支持向量机等^[5]), (4)分子建模(molecular modeling)和分子动力学(molecular dynamics)法^[6,7]. 文献^[8,9]报道已有大量的经合成和试验的 CTL 细胞表位, 但对 CTL 细胞如何识别抗原肽-MHC 复合物分子, 及抗原短肽 peptide 如何与主要组织相容性复合体 MHC 分子的相互作用识别结合的机理还不是很清楚. 目前预测 CTL 细胞表位的方法没有考虑受体与配体结合过程中电子结构的变化^[10~12], 电子结构的变化需要用量子力学方法 (Quantum Mechanics, QM) 来处理. 分子力学 (Molecular Mechanics, MM)方法是用简单的势函数来描述分子体系中原子之间的相互作用. 本文试图将量子力学(QM)计算精确性与分子力学(MM)计算高效性结合起来, 采用 QM/MM 多尺度生物大分子的分子动力学模拟方法, 以天然抗原肽 LLFGYPVYV 为模板, 替换抗原肽“锚点”氨基酸, 通过比较替换前后 MHC I 类分子对应“口袋”(Pocket)的分子静电势变化, 提取 MHC I 类分子与抗原肽的识别结合特异性信息, 这对了解免疫识别机理和指导肿瘤疫苗的开发都具有重大意义.

2 材料与方法

2.1 QM/MM 体系建立

本文从 TCR-TAX-HLA-A*0201 蛋白质晶体结构(PDB code:1ao7)上载取 TAX-HLA-A*0201 复合物分子晶体结构作为模板(Template), 见图 1(VMD^[13]生成). HLA-A*0201 分子与抗原肽 LLFGYPVYV 非键接触的口袋氨基酸见表 1, 定义“非键接触”的两个氨基酸中至少有 2 个原子之间的距离小于 4 Å^[14].

图 1 TAX-HLA-A*0201 复合物分子晶体结构

Figure 1 Crystal structure of TAX-HLA-A*0201 complex molecules

本文选取 QM 区域为 抗原肽“锚点”氨基酸残基和与其非键接触的 HLA 分子口袋氨基酸残基, MM 区域为剩余部分, 将 QM 与 MM 区域边界处的化学键断开添加连接原子, 连接原子位于断开的成键原子之间, 且三者共线. 连接原子是一个虚拟原子 L(通常为氢原子), 其质量和带电量均为零, 将其归到 QM 区域.

表1 与抗原肽各锚点“非键接触”的HLA-A*0201分子中的氨基酸^[15]Table 1 Residues on HLA molecule nonbonding contact with each anchor amino acids of peptide^[15]

Anchor amino acid	Pocket	Pocket amino acid (Chain A)
Leu	Pocket A	Tyr7, Tyr59, Glu63, Tyr159, Thr163, Trp167, Tyr171
Leu	Pocket B	Tyr7, Phe9, Met45, Glu63, Lys66, Val67, Tyr99, Tyr159
Phe	Pocket D	Lys66, His70, Tyr99, Gln155, Leu156, Tyr159
Pro	Pocket C	His70, Thr73
Val	Pocket E	Arg97, Trp147, Val152, Gln155
Val	Pocket F	Asp77, Thr80, Leu81, Tyr84, Tyr116, Thr143, Lys146, Trp147

则有 QM/MM 体系总能量为^[16]

$$E_{\text{total}} = E_{\text{QM}}(\text{QM}+\text{L}) + E_{\text{MM}}(\text{total}) - E_{\text{MM}}(\text{QM}+\text{L}) \quad (1)$$

Eq. 1 中的 E_{QM} ^[17]和 E_{MM} ^[18]分别为

$$E_{\text{QM}} = \frac{1}{2} \sum_{\mu, \nu} P_{\mu, \nu} (H_{\mu, \nu} + F_{\mu, \nu}) + \sum_{\alpha} \sum_{\beta > \alpha} \frac{Z_{\alpha} Z_{\beta}}{R_{\alpha \beta}} \quad (2)$$

$P_{\mu, \nu}$ 为密度矩阵元, $H_{\mu, \nu}$ 为单电子哈密顿量相对原子轨道的矩阵元, $F_{\mu, \nu}$ 为 Fock 矩阵元, $\sum_{\alpha} \sum_{\beta > \alpha} \frac{Z_{\alpha} Z_{\beta}}{R_{\alpha \beta}}$ 为核排斥能.

$$E_{\text{MM}} = \sum_{\text{bonds}} \frac{k}{2} (l - l_0)^2 + \sum_{\text{angles}} \frac{k}{2} (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] + \sum_{i < j} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \quad (3)$$

用 MOPAC 软件^[19]的 AM1 半经验量化方法^[20]计算 QM 区域量子化能量 $E_{\text{QM}}(\text{QM}+\text{L})$, 选用 STO-3G 基组; 用 GROMACS 分子动力学软件^[21], 选用 opls 全原子力场计算 QM 区域能量 $E_{\text{MM}}(\text{QM}+\text{L})$ 及 QM 和 MM 两区域能量 $E_{\text{MM}}(\text{total})$.

2.2 QM/MM 体系分子动力学模拟参数设置

以 TAX-HLA-A*0201 复合物分子晶体结构模板 (Template) 为例:

(i) 选择 opls 全原子力场, 将氢原子添加到 TAX-HLA-A*0201 复合物分子晶体结构 PDB 文件, 每个原子赋予部分电荷值, 生成力场参数拓扑文件.

(ii) 再将蛋白质分子置入立方体盒子中心, 距盒子边界的最小距离为 1.0 nm. 添加约 4×10^4 个溶剂分子 (TIP4P 显式水分子模型) 和若干个 Na^+ 或 Cl^- 离子, 使要模拟的蛋白质分子体系外显电中性. 用最速下降法对溶液中蛋白质分子进行结构优化, 位移步长设为 0.01 nm, 经过约 1418 步后, 达到能量梯度小于 1000 kJ/mol/nm 的收敛条件.

(iii) 平衡溶剂分子, 约束溶质蛋白质分子用 LINCS 方法^[22], 在 NVT 条件下用 V-rescale 温度调节算法^[23], 耦合时间为 0.1 ps, 经过 50000 步的 100 ps 模拟时间使

溶剂分子达到 300 K 温度平衡. 温度平衡后, 在 NPT 条件下用 Parrinello-Rahman 压力调节算法^[24,25], 耦合时间为 2.0 ps, 控制水的模拟的等温可压缩比为 $4.5 \times 10^{-5} \text{ bar}^{-1}$, 经过 50000 步的 100 ps 模拟时间使溶剂分子达到 1 bar 压强平衡. 最终压强平均值为 1 bar 左右, 水的密度平均值为 1017 g/cm³ 左右.

(iv) 体系温度和压强平衡后, 在常温 300 K, 常压 1 atm 下, 对水溶液中的蛋白质分子进行 100 ps, 50000 步, 分子动力学模拟, 实现蛋白质分子和水分子构成的溶液体系结构充分松弛和稳定, 以防止在 QM/MM 模拟过程中出现 QM 区域的计算不收敛的情况, 同时也可以有效地减少计算时间. 处理长程静电作用采用周期性边界条件和 Particle Mesh Ewald (PME) 方法^[26,27], 其中的栅格宽度为 1.6 Å. 设置勒拿德-琼斯相互作用的截断半径为 10 Å.

(v) 模拟体系达到稳定后, 在蛋白质分子结构中选择 QM 和 MM 区域. 进行 QM/MM 体系分子动力学模拟经过 100000 步的 10 ps 模拟时间使体系达到稳定. 在 QM/MM 分子动力学模拟计算中, GROMACS 负责分子动力学计算, 将 QM 区域的原子坐标等信息传递给 MOPAC, MOPAC 用 AM1 方法计算 QM 区域的能量, 并且计算出每个原子处的能量梯度和每个原子的极化电荷, 然后将这些数据传递回 GROMACS. MM 区域的原子通过其力场参数计算出其能量梯度和能量, GROMACS 将 QM 区域和 MM 区域的原子一起进行动力学运算, 调整整个分子体系的结构, 完成一个时间步长的计算.

(vi) 为了采样分析, 继续进行 QM/MM 体系分子动力学模拟, 步长 0.0001 ps, 经过 10000 步的 1 ps 模拟时间, 每 10 步取样一次, 得到 1000 个不同时间点的结果, 用于后期数据处理.

2.3 氨基酸残基电多极矩和静电势的计算

为了考察 MHC I 类分子每个口袋对抗原肽不同氨基酸锚点的识别功能, 将口袋氨基酸残基的原子极化电荷在空间形成的静电势用电多极矩分量表示^[28], 口袋氨基酸残基的原子极化电荷是在 QM/MM 分子动力学模拟中通过 Mulliken 方法^[29]计算获得.

每一个残基的电单极矩 q_{aa} 、电偶极矩矢量模 p_{aa} 和

电四极矩张量范数 D_{aa} 分别按照下式(Eq. 4)计算:

$$\begin{aligned} q_{aa} &= \sum_{k=1}^N q_k \\ p_{aa} &= \sqrt{d_{aa} d_{aa}} \\ d_{aa} &= \sum_{k=1}^N q_k r_k \\ D_{aa} &= \sqrt{\sum_{i=1, j=1}^3 D_{ij}^2} \\ D_{ij} &= \sum_{k=1}^N q_k (3x_i x_j - r_k^2 \delta_{ij}) \\ r_k &= \sqrt{(x_k - x_{Ca})^2 + (y_k - y_{Ca})^2 + (z_k - z_{Ca})^2} \end{aligned} \quad (4)$$

其中 r_k 是指口袋中第 aa 号残基第 k 原子距离同一 aa 残基的 C- α 原子的距离, N 是指第 aa 号残基的侧链上的原子总数目. q_k 为原子极化电荷, d_{aa} 电偶极矩, D_{ij} 为电四极矩分量, x_i, y_j 为残基原子笛卡尔坐标.

每个氨基酸残基的电四极矩分量可以按下式标量化:

$$\mu'_{aa} = \frac{\mu_{aa} - \mu_{\min}}{\mu_{\max} - \mu_{\min}}$$

以空口袋为参照, 分别用 20 个氨基酸为锚点依次驻扎在口袋 Pocket A, Pocket B, Pocket D, Pocket C, Pocket E, Pocket F, 每个口袋上第 aa 个氨基酸的分子静电势变化用变量 S_{aa} 表示:

$$S_{aa} = [(q'_{aao} - q'_{aac})^2 + (p'_{aao} - p'_{aac})^2 + (D'_{aao} - D'_{aac})^2]^{1/2} \quad (5)$$

下标 aao 表示有锚点驻扎的口袋氨基酸, 下标 aac 表示无锚点驻扎的口袋氨基酸.

$\bar{S}_{aa}$ 表示每个口袋氨基酸对 20 个锚点氨基酸残基静电势变化 S_{aa} 的平均值.

用变量 Tot_{Dif} 表示一个口袋(例如 Pocket A)所有氨基酸分子静电势变化总和:

$$\text{Tot}_{\text{Dif}} = \sum_{aa} S_{aa} \quad (6)$$

以模板抗原肽 LLFGYPVYVYU 为参照, 分别用不同氨基酸依次替换模板原有锚点氨基酸, 整个口袋的分子静电势变化用变量 Ref_{Dif} 表示:

$$\begin{aligned} \text{Ref}_{\text{Dif}} &= \sum_{aa} [(q_{aaR} - q_{\text{ref}})^2 + (p_{aaR} - p_{\text{ref}})^2 + \\ &\quad (D_{aaR} - D_{\text{ref}})^2]^{1/2} \end{aligned} \quad (7)$$

下标 ref 表示模板原有锚点氨基酸, 下标 aaR 表示替换后的锚点氨基酸.

2.4 箱线图分析方法

用箱线图(图 2)分析判断数据集异常值的主要步骤是将一组数据由大到小排列, 分别计算出 5 个分位点: 第 1 分位点 Q_1 , 使四分之一数据小于 Q_1 ; 第 2 分位点为中值, 使数据集分为相等的两部分; 第 3 分位点 Q_3 , 使四分之一数据大于 Q_3 ; 第 4 分位点下限 $Q_1 - 1.5\text{IQR}$; 第 5 分位点上限 $Q_3 + 1.5\text{IQR}$; 其中 $\text{IQR} = Q_3 - Q_1$, 上限和下限统称为内限, 处于内限之外的点称为异常值.

图 2 箱线图示例

Figure 2 The example of box plot

3 结果与讨论

3.1 每个口袋氨基酸 S_{aa} 值的箱线图

按 Eq. 5 计算, 以空口袋为参照, 分别用 20 个氨基酸为锚点依次驻扎在每个空口袋中, 则每个口袋上第 aa 个氨基酸有 20 个不同的分子静电势变化 S_{aa} 值, 每个口袋氨基酸分子静电势变化和功能作用可以采用箱线图方法分析, 如图 3 所示. 图 3 中的虚线为每一个口袋氨基酸的 S_{aa} 平均值 $\bar{S}_{aa}$. 例如, Pocket A 上的氨基酸 Glu63 的箱线图上的 S_{aa} 平均值 $\bar{S}_{aa}$ 为 0.33064, IQR 值为 0.18218.

3.2 IQR 值的箱线图

由图 3 可知, 一个口袋氨基酸 S_{aa} 值分布范围大, 即 IQR 值较大, 说明该口袋氨基酸电子结构对不同锚点氨基酸残基的极化作用比较敏感, 具有对锚点氨基酸的识别功能, 称该口袋氨基酸为识别氨基酸(recognition amino acid)^[28]. 为了更好地鉴别不同口袋氨基酸对锚点氨基酸的识别能力, 我们又给出 6 个口袋中每个氨基酸对 20 个锚点氨基酸残基静电势变化 IQR 值的箱线图, 如图 4 所示. 将 IQR 值超上限的口袋氨基酸定义为精细识别氨基酸, 而 IQR 值在上限与第三分位点之间的口袋氨基酸定义为次级识别氨基酸.

3.3 $\bar{S}_{aa}$ 值的箱线图

由图 3 可知, 口袋氨基酸有 IQR 值小而 $\bar{S}_{aa}$ 值大的现象, 该口袋氨基酸对锚点氨基酸起结合结合作用, 称为结合氨基酸(Anchoring amino acids)^[28]. 为了更好地鉴别不同口袋氨基酸对锚点氨基酸的结合能力, 我们又

图3 口袋氨基酸静电势变化箱线图

Figure 3 The boxplot of the electrostatic field change of the pocket amino acid

图4 口袋氨基酸对20锚点残基静电势变化IQR值的箱线图

Figure 4 Boxplot of the IQR values of each pocket amino acid about the 20 anchor amino acids

给出6个口袋中每个氨基酸对20个锚点氨基酸残基静电势变化平均值 $\bar{S}_{aa}$ 的箱线图,如图5所示.将 $\bar{S}_{aa}$ 值超上限的口袋氨基酸定义为一级结合氨基酸,而 $\bar{S}_{aa}$ 值在上限与第三分位点之间的口袋氨基酸定义为次级结合氨基酸.

3.4 口袋氨基酸 Tot_{Dif} 值计算结果

以空口袋为参照,分别用20个氨基酸为锚点依次驻扎在每个空口袋中,按Eq. 6计算MHC I类分子结合槽底部每个口袋所有氨基酸分子静电势变化总和 Tot_{Dif},如图6所示, Pocket B, Pocket A, Pocket F 的 Tot_{Dif} 值远大于 Pocket C, Pocket D, Pocket E 的 Tot_{Dif} 值. 而实验的结果表明,对于MHC I类分子,口袋B包含第2锚点残

图5 口袋氨基酸对20锚点残基静电势变化平均值 $\bar{S}_{aa}$ 的箱线图
Figure 5 The boxplot of the average $\bar{S}_{aa}$ of each pocket amino acid about the 20 anchor amino acids

基和口袋F包含第9锚点残基大体上决定了MHC分子对抗原肽的结合能力,被称为锚点残基,口袋A包含的第1锚点残基和口袋D包含的第3锚点残基,被称为次级锚点残基.我们模拟的结果与实验结果基本一致.

图6 20个锚点氨基酸残基对应每个口袋所有氨基酸分子静电势变化
Figure 6 The total change of the electrostatic field of each pocket about the 20 anchor amino acids

3.5 口袋氨基酸 Ref_{Dif} 值计算结果

以天然抗原肽序列 LLFGYPVYV 为参照,分别用不同氨基酸依次替换天然抗原肽 LLFGYPVYV 驻扎在每个对应口袋中的原始锚点氨基酸,按 Eq. 7 计算,每个口袋氨基酸分子静电势变化 Ref_{Dif} ,如图7所示, Ref_{Dif} 的值愈小,表明替换氨基酸对口袋分子静电势的变化影响小,它具有与天然锚点氨基酸相似的结合特异性.

3.6 结合特异性分析

MHC I 类分子结合槽底部的6个口袋(Pocket A, Pocket B, Pocket D, Pocket C, Pocket E, Pocket F)分别是由多态性的氨基酸序列组成的,具有对抗原肽识别和结合的功能.本文以天然抗原肽 LLFGYPVYV 为模板,替换天然锚点氨基酸,用 QM/MM 分子动力学模拟抗原肽-HLA A*0201 复合物生物大分子的电子结构变化,比较替换前后口袋的分子静电势变化,将 QM/MM 计算结

果与文献免疫学实验统计结果^[30](图8)比较,图8中字母的高度代表某种氨基酸在某位置上出现的频率大小,对每个口袋氨基酸功能分析如下:

Pocket A 是由氨基酸 Tyr7, Tyr59, Glu63, Tyr159, Thr163, Trp167 和 Tyr171 组成的,天然抗原肽 P_1 锚点位置是氨基酸 Leu (L)(表1). Saper 等^[31]通过实验证实:HLA*A0201 分子中,5个Tyr的羟基(Tyr7, Tyr59, Tyr99, Tyr159 和 Tyr171)及 Glu63 的羧基和 Lys66 氨基酸组成了 Pocket A 的极性表面,抗原肽 P_1 位置氨基酸侧链会进入到 Pocket A 中.由图4和图5可以确定 Pocket A 的 Glu63 为精细识别氨基酸, Tyr171 为次级识别氨基酸,而 Tyr159 和 Glu63 是2个次级结合氨基酸,在结合绑定锚点氨基酸起重要作用.由图7可以确定: Ref_{Dif} 值小的替换氨基酸 I, M, A, V, W, F, K, R 具有与原始锚点氨基酸 Leu (L)相似的结合特异性,本文计算结果与文献免疫学实验统计结果(图8)是一致的.

Pocket B: 是由氨基酸 Tyr7, Phe9, Met45, Glu63, Lys66, Val67, Tyr99 和 Tyr159 组成的,天然抗原肽 P_2 锚点位置是氨基酸 Leu (L)(表1). Phe9, Met45 和 Val67 构成 Pocket B 的疏水性内壁^[31].由图4和图5可以确定 Pocket B 的 Glu63 和 Lys66 为2个精细识别氨基酸,也是2个一级结合氨基酸.可见 Pocket B 在 MHC 分子结合抗原短肽的过程中有着重要作用.由图7可以确定: Ref_{Dif} 值小的替换氨基酸 I, M, V, T, Q, A 具有与原始锚点氨基酸 Leu (L)相似的结合特异性,本文计算结果与文献免疫学实验统计结果(图8)是一致的.

Pocket D 是由氨基酸 Lys66, His70, Tyr99, Gln155, Leu156 和 Tyr159 组成的,天然抗原肽 P_3 锚点位置是氨基酸 Phe (F)(表1). His114, Leu156, Tyr159, Tyr99 和 Leu160 组成了疏水性 Pocket D, 侧链带有苯环的氨基酸适合进入到 Pocket D 中^[31].由图4和图5可以确定 Pocket D 的 Gln155 为次级识别氨基酸, Gln155 为次级结合氨基酸.由图7可知相对原始锚点氨基酸 Phe (F)替换氨基酸的 Ref_{Dif} 最大值约 0.25, 替换氨基酸 Tyr (Y), Trp (W)具有与原始锚点氨基酸 Leu (L)相似的结合特异性,本文计算结果与文献免疫学实验统计结果(图8)是一致的.

Pocket C 是由氨基酸 His70 和 Thr73 组成的,天然抗原肽 P_6 锚点位置是氨基酸 Pro (P)(表1). Thr73, His70, His74 和 Arg97 组成极性的 Pocket C^[31].由于 Pocket C 的氨基酸选择较少,由图7可知相对原始锚点氨基酸 Phe (P)替换氨基酸的 Ref_{Dif} 最大值约 0.06, 所以替换氨基酸对 Pocket C 的分子静电势变化影响很小,本文计算结果与文献免疫学实验统计结果(图8)是一致的.

Pocket E 是由氨基酸 Arg97, Trp147, Val152 和 Gln155 组成的,天然抗原肽 P_7 锚点位置是氨基酸 Val(V)(表1). Pocket E 的 2/3 表面是疏水性的, 1/3 表面是极性的^[31], 它既能结合疏水性氨基酸残基,也能结合

图7 相对于原始锚点氨基酸的静电场变化分析

Figure 7 The electrostatic field changes compared with the original anchor amino acid

亲水性氨基酸残基. 由图4和图5可以确定 Pocket E 的 Arg97 为次级识别氨基酸, Arg97 为一级结合氨基酸. 由图7可知相对原始锚点氨基酸 Val (V) 替换氨基酸的 Ref_{Dif} 最大值约 0.35, 与口袋 E 结合较好的包含亲水性的 H, K, S 也包含疏水性的 W, Y, A, F, I, L, 所以替换氨基酸既可是疏水性氨基酸也可是亲水性氨基酸.

Pocket F: 是由氨基酸 Asp77, Thr80, Leu81, Tyr84, Tyr116, Thr143, Lys146 和 Trp147 组成的, 天然抗原肽 P9 锚点位置是氨基酸 Val (V)(表1). Pocket F 的底部被 Tyr116 侧链占据, 只能使短小侧链的疏水性氨基酸易进入, 如 Val^[31]. 由图4和图5可以确定 Pocket F 有2个精细识别氨基酸 Asp77, Tyr84, 2个一级结合氨基酸

Lys146, Asp77. 由图7可以确定: Ref_{Dif} 值小的替换氨基酸 L, I, A, M, 具有与原始锚点氨基酸 Val (V) 相似的结合特异性, 本文计算结果与文献免疫学实验统计结果(图8)是一致的.

4 结论

在 HLA-A*0201 分子与抗原肽相互作用识别结合中, Pocket B 的 Glu63 和 Lys66 功能是精细识别氨基酸和一级结合氨基酸, Pocket F 的 Asp77, Tyr84 功能是精细识别氨基酸, 而 Asp77, Lys146 是一级结合氨基酸. Pocket B, Pocket A, Pocket F 的 Tot_{Dif} 值远大于 Pocket C, Pocket D, Pocket E 的 Tot_{Dif} 值, 本文计算结果与实验结

图8 20个氨基酸在HLA-A*0201分子结合抗原肽的不同位点出现频率统计图^[32,33]

Figure 8 The frequency statistical figure of 20 amino acids appeared on different site of the peptide binding to HLA-A*0201 molecule^[32,33]

果基本一致, 表明QM/MM分子动力学模拟的方法在提取抗原肽与MHC I类分子识别结合特异性信息是可行的。

References

- [1] Van der Bruggen, P.; Zhang, Y.; Chaux, P.; Stroobant, V.; Panichelli, C.; Schultz, E. S.; Chapiro, J.; vall der Eynde, B. J.; Brasseur, F.; Boon, T. *Immunol. Rev.* **2002**, *188*, 51.
- [2] Garboczi, D. N.; Ghosh, P.; Utz, U.; Fan, Q. R.; Biddison, W. E.; Wiley, D. C. *Nature* **1996**, *384*, 134.
- [3] Rammensee, H.; Bachmann, J.; Emmerich, N. P. *Immunogenetics* **1999**, *50*, 213.
- [4] Nielsen, M.; Lundegaard, C.; Lund, O. *BMC Bioinformatics* **2007**, *8*, 238.
- [5] Honeyman, M. C.; Brusica, V.; Stone, N. L. *Nat. Biotechnol.* **1998**, *16*, 966.
- [6] Hattotuwa, C. K.; Guan, P.; Docychnova, I. A. *J. Mol. Graph. Modell.* **2004**, *22*, 195.
- [7] Dong, S. M.; Song, Z.; Liu, T.; Zhu, M. H.; Liu, W. *Acta Chim. Sinica* **2010**, *68*, 1821. (董素梅, 宋哲, 刘涛, 朱鸣华, 刘伟, 化学学报, **2010**, *68*, 1821.)
- [8] Kessler, J. H.; Beekman, N. J.; Bres-Vloemans, S. A.; Berdijk, P.; van Veelen, P. A.; Kloosterman-Joosten, A. M.; Vissers, D. C.; ten Bosch, G. J.; Kester, M. G.; Sijts, A.; Wouter Drijfhout, J.; Ossendorp, F.; Offringa, R. *Exp. Med.* **2001**, *193*, 73.
- [9] Trautwein, N.; Stevanović, S. *Methods Mol. Biol.* **2013**, *960*, 159.
- [10] Liao, W. W.; Arthur, J. W. *PLoS One* **2011**, *6*, e25055.
- [11] Lee, Y.; Ferrari, G.; Lee, S. C. *Biomed. Microdevices* **2010**, *12*, 207.
- [12] Binkowski, T. A.; Marino, S. R.; Joachimiak, A. *PLoS One* **2012**, *7*, e41710.
- [13] Humprey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.
- [14] Yael, A.; Hanah, M. *Methods* **2004**, *34*, 454.
- [15] Song, Z.; Liu, T.; Liu, W.; Zhu, M. H.; Wang, X. G. *Acta Phys.-Chim. Sin.* **2007**, *23*, 198. (宋哲, 刘涛, 刘伟, 朱鸣华, 王晓刚, 物理化学学报, **2007**, *23*, 198.)
- [16] Morokuma, K.; Svensson, M.; Humbel, S. J. *Phys. Chem.* **1996**, *100*, 19357.
- [17] Xu, G. X.; Li, L. M.; Wang, D. M. *Quantum Chemistry (Middle)*, Science Press, Beijing, **2009**. (徐光宪, 黎乐民, 王德民, 量子化学(中册), 科学出版社, 北京, **2009**.)
- [18] Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- [19] <http://openmopac.net/>.
- [20] Dewar, M. J. S. *J. Mol. Struct.* **1983**, *100*, 41.
- [21] Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701.
- [22] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463.
- [23] Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- [24] Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.
- [25] Nosé, S.; Klein, M. L. *Mol. Phys.* **1983**, *50*, 1055.
- [26] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- [27] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
- [28] Agudelo, W. A.; Galindo, J. F.; Patarroyo, M. E. *Biochem. Biophys. Res. Commun.* **2011**, *410*, 410.
- [29] Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
- [30] Nicolas, R.; Ilka, H.; Ole, L.; Morten, N. *Immunogenetics* **2008**, *60*, 759.
- [31] Saper, M. A.; Bjorkman, P. J.; Wiley, D. C. *J. Mol. Biol.* **1991**, *219*, 277.
- [32] Lund, O.; Nielsen, M.; Kesmir, C.; Petersen, A. G.; Lundegaard, C. *Immunogenetics* **2004**, *55*, 797.
- [33] http://www.cbs.dtu.dk/researchgroups/immunology/HLA/A_cont_files/Media/A0201/A0201.jpg?disposition=download.

(Zhao, C.)