

# HCV NS3/4A 蛋白酶与 Faldaprevir 类似物的分子识别机制及其复合物运动模式

梁立<sup>1</sup>, 胡建平<sup>1,2</sup>, 杜文义<sup>1</sup>, 左柯<sup>1</sup>, 刘崑<sup>1</sup>, 苟小军<sup>1</sup>

1 成都大学 四川抗菌素工业研究所 药食同源植物资源开发四川省高校重点实验室, 四川 成都 610106

2 乐山师范学院 化学学院, 四川 乐山 614004

梁立, 胡建平, 杜文义, 等. HCV NS3/4A 蛋白酶与 Faldaprevir 类似物的分子识别机制及其复合物运动模式. 生物工程学报, 2016, 32(5): 669–682.

Liang L, Hu JP, Du WY, et al. Molecular recognition mechanism and motion of HCV NS3/4A protease with Faldaprevir analogue. Chin J Biotech, 2016, 32(5): 669–682.

**摘要:** Faldaprevir 类似物 (Faldaprevir analogue molecule, FAM) 能有效抑制 HCV NS3/4A 蛋白酶的催化活性, 是一种潜在抗 HCV 先导化合物。通过生物信息学统计分析了已报道的 HCV NS3/4A 蛋白酶晶体结构, 得到了 FAM-HCV NS3/4A 蛋白酶晶体结构。对 FAM-HCV NS3/4A 蛋白酶复合物进行了 20.4 ns 的分子动力学模拟, 重点从氢键和结合自由能两个角度分析了二者分子识别中的关键残基及结合驱动力。氢键和范德华力是促使 FAM 特异性结合到蛋白 V132–S139、F154–D168、D79–D81 和 V55 的活性口袋中的主要驱动力, 这与实验数据较为吻合。耐药性突变实验分析了 R155K、D168E/V 和 V170T 定点突变对 FAM 分子识别的影响, 为可能存在的 FAM 耐药性提供了分子依据。最后, 用自由能曲面和构象聚类两个方法探讨了体系的构象变化, 给出体系的 4 种优势构象, 为后续的基于 HCV NS3/4A 蛋白酶结构的 Faldaprevir 类似物抑制剂分子设计提供一定的理论帮助。

**关键词:** HCV NS3/4A 蛋白酶, Faldaprevir 类似物分子, 分子动力学模拟, 结合自由能, 构象变化

**Received:** August 31, 2015; **Accepted:** March 16, 2016

**Supported by:** National Natural Science Foundation of China (Nos. 11247018, 11147175), Research Key Project of Sichuan Province Education Department (No. 12ZA066), Leshan Science and Technology Project (No. 14SZD018).

**Corresponding authors:** Jianping Hu. Tel: +86-833-2152297; E-mail: hujianping@emails.bjut.edu.cn  
Xiaojun Gou. Tel: +86-28-84616301; E-mail: gxjmeprd@163.com

国家自然科学基金 (Nos. 11247018, 11147175), 四川省教育厅科研重点项目 (No. 12ZA066), 乐山市科技计划项目 (No. 14SZD018) 资助。

# Molecular recognition mechanism and motion of HCV NS3/4A protease with Faldaprevir analogue

Li Liang<sup>1</sup>, Jianping Hu<sup>1,2</sup>, Wenyi Du<sup>1</sup>, Ke Zuo<sup>1</sup>, Wei Liu<sup>1</sup>, and Xiaojun Gou<sup>1</sup>

*1 Key Laboratory of Medicinal and Edible Plants Resources Development of Sichuan Education Department, Sichuan Industrial Institute of Antibiotics, Chengdu University, Chengdu 610106, Sichuan, China*

*2 College of Chemistry, Leshan Normal University, Leshan 614004, Sichuan, China*

**Abstract:** Faldaprevir analogue molecule (FAM) has been reported to effectively inhibit the catalytic activity of HCV NS3/4A protease, making it a potential lead compound against HCV. A series of HCV NS3/4A protease crystal structures were analyzed by bioinformatics methods, and the FAM-HCV NS3/4A protease crystal structure was chosen for this study. A 20.4 ns molecular dynamics simulation of the complex consists of HCV NS3/4A protease and FAM was conducted. The key amino acid residues for interaction and the binding driving force for the molecular recognition between the protease and FAM were identified from the hydrogen bonds and binding free energy analyses. With the driving force of hydrogen bonds and van der Waals, FAM specifically bind to the active pocket of HCV NS3/4A protease, including V130-S137, F152-D166, D77-D79 and V55, which agreed with the experimental data. The effect of R155K, D168E/V and V170T site-directed mutagenesis on FAM molecular recognition was analyzed for their effect on drug resistance, which provided the possible molecular explanation of FAM resistance. Finally, the system conformational change was explored by using free energy landscape and conformational cluster. The result showed four kinds of dominant conformation, which provides theoretical basis for subsequent design of Faldaprevir analogue inhibitors based on the structure of HCV NS3/4A protease.

**Keywords:** HCV NS3/4A protease, Faldaprevir analogue molecule, molecular dynamics simulation, binding free energy, conformational change

丙型肝炎病毒 (Hepatitis C virus, HCV) 感染是造成慢性肝病和肝移植的主要原因之一<sup>[1]</sup>。据估计, 全球范围内有 1.6–1.8 亿人长期感染 HCV<sup>[2-3]</sup>, 其中一部分会发展为进行性肝纤维化, 最终导致肝硬化, 晚期肝病及肝癌<sup>[4]</sup>。目前的丙型肝炎标准诊疗方案 (Standard of care, SOC) 是联合使用长效干扰素 (PEG-IFN $\alpha$ ) 和广谱抗病毒药物利巴韦林 (Ribavirin)<sup>[5]</sup>。但这种治疗方案的持续病毒学应答 (Sustained viral responses, SVR) 受治疗周期、HCV 基因型等因素的影响, 在 40%–80% 大范围内波动<sup>[6]</sup>, 治疗效果往往不够理想。Pawlotsky 课题组的研究结果表明<sup>[7]</sup>, 在 SOC 中加入直接作用性抗病毒

(Directly acting antiviral, DAA) 药物后 SVR 显著提高, 疗效令人满意。

HCV NS3/4A 蛋白酶 (Protease, PR) 是由 NS3 丝氨酸蛋白与辅助因子 NS4A 蛋白以非共价键结合而成的二聚体。NS3 丝氨酸 PR 是 HCV 非结构蛋白加工成熟过程的关键酶, 与三磷酸核苷酸酶、RNA 解螺旋酶结合, 构成 HCV 复制所必需的非结构蛋白 NS3。NS3 属于胰 PR 超家族, 拥有 PR 和 RNA 解螺旋酶双重活性<sup>[8]</sup>, 位于由 HCV RNA 编码的前体多聚蛋白的开放阅读框 (Open reading frame, ORF) 中的非结构蛋白区中。另外, NS3/4A PR 能水解 TRIF 和 MAVS 两种人类免疫蛋白, 从而干扰人体对

HCV 的先天性免疫应答<sup>[9-10]</sup>。NS3/4A PR 还可通过调控 IFN-3 的表达水平,破坏细胞内免疫通路,诱导宿主产生免疫逃逸<sup>[11]</sup>。

HCV NS3/4A PR 抑制剂是首类进入临床试验的 DAA 药物。Faldaprevir 类似物分子 (Faldaprevir analogue molecule, FAM)<sup>[12]</sup> 作为一种 HCV NS3/4A PR 抑制剂,能有效抑制 NS3/4A PR 活性。阻断 HCV 的复制、翻译和翻译后多聚蛋白的加工成熟,还可使 IFN 更好地发挥抗病毒疗效。

构象变化和分子识别是药物设计的基础。目前有关 HCV NS3/4A PR 与 FAM 之间识别机制和复合物体系在水中构象变化等问题尚未见报道。本文用分子动力学 (Molecular dynamics, MD) 模拟方法对这些问题进行了研究,为后续的结构改造、新药研发提供结构及理论支撑。图 1 给出了 FAM 的化学结构,图中数字表示原子序号,括号里的数字用来表示极性氢,这些极性氢有可能和蛋白质部分形成氢键<sup>[12]</sup>。

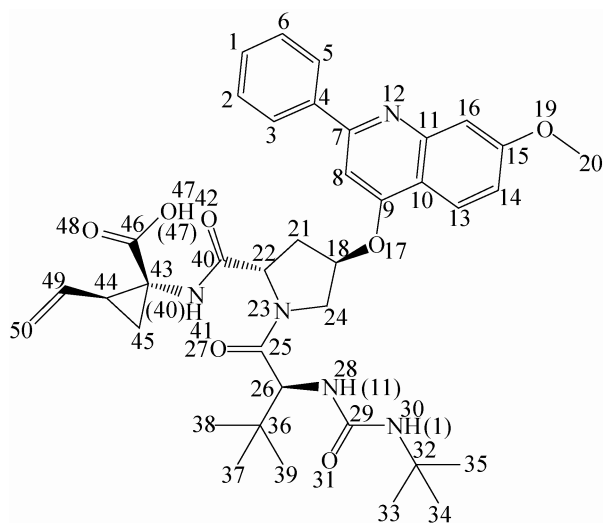


图 1 FAM 的分子结构<sup>[12]</sup>

Fig. 1 Molecular structure of FAM<sup>[12]</sup>.

## 1 模拟方法

### 1.1 分子动力学模拟

在 HCV NS3/4A PR 和 FAM 的晶体结构<sup>[12]</sup> (PDB 代码:4K8B) 中,包含有辅助结晶的离子和结晶水,考虑到这些水分子等并未与 FAM 结合,也未参与活性中心的构成。故删去 4 条链的非蛋白部分,包括 A 链的 ZN201、SO4203 和 SO4204, B 链的 ZN201、SO4203, 以及 D 链的 SO4301 及所有的结晶水分子。仅保留体系中 PR 的氨基酸与 FAM 部分,并以此作为 MD 模拟的初始结构。MD 模拟采用 AMBER<sup>[13]</sup> 程序以及 AMBER 力场,其中生物大分子的力场参数基于实验值拟合<sup>[14]</sup>。模拟温度为 26.85 °C,溶剂使用 TIP3P 水模型<sup>[15]</sup>。

首先,将复合物置于去头八面体水盒子中,使复合物边缘距盒子壁 0.8 nm,此时盒子里共加入 9 820 个水分子,体系总原子数为 15 614。随后进行两次能量优化,第一次是约束溶质的 10 000 步能量优化,约束力常数为  $2.091 \times 10^5$  kJ/(mol·nm<sup>2</sup>),由 5 000 步最陡下降法优化和 5 000 步共轭梯度能量优化组成。第二次是去约束的 5 000 步最陡下降和 5 000 步共轭梯度能量优化,收敛条件为能量梯度小于  $4.182 \times 10^{-4}$  kJ/(mol·nm<sup>2</sup>)。能量优化完成后,开始进行 MD 模拟,整个模拟过程分为两步分进行:首先进行约束动力学模拟,约束力常数为  $4.182 \times 10^3$  kJ/(mol·nm<sup>2</sup>),期间温度使用 Langevin dynamics 升温机制,温度偶联因子设为 1.0,温度从 -273.15 °C 逐步升高到 26.85 °C,模拟 400 ps;接着,再进行 20 ns 的无约束恒温 MD 模拟。用 SHAKE 算法<sup>[16]</sup>约束键长,非键半径为 1 nm;模拟积分步长为 2 fs,每隔 1 ps 采集一次构象,共采集 20 400 个构象。并使用 VMD 软件对体

系模拟全过程跟踪监测。

## 1.2 能量分解

在本文中,用 MM/GBSA (Molecular mechanics/generalized Born surface area) 方法<sup>[17-18]</sup>进行能量分解分析。其基本思想是把每个残基的能量贡献近似分为力学 (Molecule mechanics, MM) 方法计算的真空分子内能、用广义波恩 (Generalized Born, GB) 模型<sup>[19-20]</sup>计算的极性溶剂化能,以及用 LCPO 算法<sup>[21-22]</sup>计算的非极性溶剂化能,并且把能量分解到残基的主链原子核侧链原子上。这里,真空分子内能可分为极性的静电相互作用和非极性的范德华相互作用两部分。LCPO 算法认为非极性溶剂化能与溶剂可接近表面积 (Solvent accessible surface area, SASA) 呈正相关。通过 MM/GBSA 能量分解,可以观察到 HCV NS3/4A PR 中的主要残基对 FAM 结合所做出的贡献。

## 1.3 自由能曲面

自由能曲面 (Free energy landscape, FEL) 是通过研究自由能曲面图的最小值 (即体系最稳定的状态) 和分界点 (即体系变化过程中的一个短暂的状态) 来表示生物学过程中发生的动力学变化的一种分析方法<sup>[23-26]</sup>。一些主成分 (Principal component, PC) 可以用以表示重要的动力学过程,这些 PC 也可以来绘制自由能曲面图<sup>[27]</sup>,具体根据公式:

$$\Delta G(X) = -k_B T \ln P(X)$$

式中,坐标  $X$  表示 PC,  $k_B$  表示玻尔兹曼常数,  $T$  是绝对温度,  $P(X)$  是体系 PC 的概率。

## 1.4 聚类分析

使用 MMTSB 工具<sup>[28]</sup>,对 MD 模拟得到的 20 400 个体系构象进行聚类分析 (Cluster analysis)。分析过程依据 Daura 等<sup>[29]</sup>提出的观

点:逐一计算各构象之间的 RMSD 值,以此为基础建立 RMSD 矩阵 ( $N \times N$ ,  $N$  为构象数)。人为设定一个 RMSD 阈值,在任意两构象 RMSD 值之间进行比较,若其差值小于此阈值则归为一类。以每一类中能量最低的构象作为该类的代表构象。本文中,以体系整体 RMSD 值建立矩阵,对体系整体构象的聚类情况进行了讨论。

## 2 结果与讨论

### 2.1 HCV NS3/4A PR 的生物信息学分析

截止目前,在蛋白质数据库 (www.rcsb.org) 中检索 HCV NS3/4A PR,得到相关结构 104 个。其中仅来源于丙型肝炎病毒的结构有 81 个,而其中结合了 NS4A 蛋白的成熟复合物结构仅有 17 个。表 1 给出了蛋白质数据库中收录的成熟 HCV NS3/4A PR 结构生物学信息。表中 PDB 表示解析结构的 PDB 代码,Res.表示解析的分辨率,Sequence 为 HCV NS3/4A PR 的氨基酸序列,ligand 则表示结构中的小分子结合信息。

从表 1 可以看出,该 17 个结构分辨率相差较大,但均在可接受范围内。除 4U01 外均为二聚体结构,其中 3LOX、3LON、3OYP、2XCN、2OBO、2A4Q、2F9U 和 1RTL 的 NS3A、NS4A 蛋白单体氨基酸序列之间不一致,表中仅给出了相对较全的单体序列。考虑到后续对二聚体结构整体的模拟工作,避免二聚体结构不完整对模拟结果可能造成的误差,故没有选取这些结构。加之结合有 FAM 的晶体结构仅有 4K8B 一个,故选其用于后续的 MD 模拟。

### 2.2 体系整体的模拟分析

图 2 显示了 HCV NS3/4A PR 和 FAM 的复合物体系在模拟过程中的势能随时间的变化。分子势能在 400 ps 以后达到平衡。经计算,平衡

表 1 HCV NS3/4A PR 结构生物学信息

Table 1 The structural biological information of HCV NS3/4A PR

PDB	Res.(Å)	Polymer	Sequence	Ligand
4U01	2.80	Nonamer	NS3: A1-S182, NS4A: K220-S232	Macrocyclic inhibitor, chloride ion, zinc ion
4K8B	2.80	Dimer	NS3: I3-M179, NS4A: G221-S232	FAM, sulfate ion, zinc ion
4I31	1.93	Dimer	NS3: A1-S182, NS4A: K220-S232	Macrocyclic inhibitor, sodium ion
4I32	2.30	Dimer	NS3: A1-S182, NS4A: K220-S232	Macrocyclic inhibitor, sodium ion
4I33	1.90	Dimer	NS3: A1-S182, NS4A: K220-S232	Macrocyclic inhibitor, sodium ion
3LOX*	2.65	Dimer	NS3: A1-S181, NS4A: K20-K41	Linear inhibitor, beta-mercaptoethanol, zinc ion
3LON*	2.20	Dimer	NS3: A1-S181, NS4A: K20-K41	Linear inhibitor, beta-mercaptoethanol, zinc ion
3OYP*	2.76	Dimer	NS3: A1-M175, NS4A: G221-S232	Peptidomimetic inhibitor, zinc ion
2XCN*	3.02	Dimer	NS3: A1-R180, NS4A: K20-K40	Linear inhibitor, magnesium ion, zinc ion
2OBO*	2.60	Dimer	NS3: M-1-G182, NS4A: K19-K41	Linear inhibitor, beta-mercaptoethanol, zinc ion
2A4Q*	2.45	Dimer	NS3: A1-S181, NS4A: K20-K41	Linear inhibitor, beta-mercaptoethanol, zinc ion
2F9U*	2.60	Dimer	NS3: A1-S183, NS4A: K19-K41	Linear inhibitor, zinc ion
1RTL*	2.75	Dimer	NS3: A1-R180, NS4A: K20-K40	Linear inhibitor, zinc ion
1W3C	2.30	Dimer	NS3: I3-E176, NS4A: G221-G233	Linear inhibitor
1DY8	2.40	Dimer	NS3: A1-M175, NS4A: G221-S232	Linear inhibitor
1DY9	2.10	Dimer	NS3: A1-M175, NS4A: G221-S232	Linear inhibitor, zinc ion
1DXP	2.40	Dimer	NS3: A1-M175, NS4A: G221-S232	Glycerol, zinc ion

Due to the missing residues in the monomer of NS3 or NS4A, the data presented in the table is the relatively complete one.

后的势能平均值为 $-4.26 \times 10^5$  kJ/mol, 标准偏差为 2 000 kJ/mol, 波动范围小于 0.47%。由于模拟体系较大, 共含有 15 614 个原子, 波动对结果的影响不大, 可以忽略。可见, HCV NS3/4A PR 和 FAM 的复合物结构在模拟过程中维持稳定可靠。

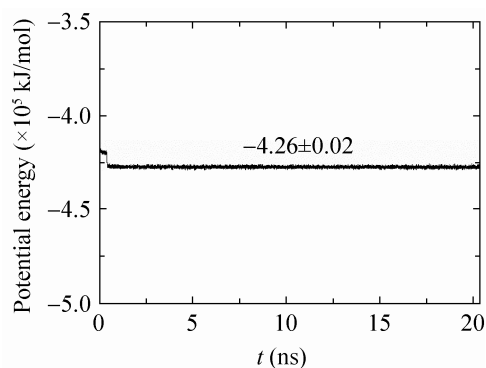


图 2 体系势能随时间的变化

Fig. 2 Potential energy of the system versus simulation time.

图 3A 给出了 PR 分子中 A 链、B 链、C 链和 D 链主链  $C_{\alpha}$  原子的方均根偏差 (Root mean square deviation, RMSD) 随时间的变化。从图 3A 可以看出, 蛋白酶在 400 ps 后基本趋于平稳。A、B 链的 RMSD 高于 C、D 链。经计算, A 链 RMSD 为  $(0.125 \pm 0.020)$  nm、B 链 RMSD 为  $(0.117 \pm 0.017)$  nm、C 链和 D 链分别为  $(0.062 \pm 0.019)$  nm 和  $(0.060 \pm 0.017)$  nm。说明 C、D 链较 A、B 链稳定。这与其结构有关, A、B 链较长, 含 177 个氨基酸残基, 为蛋白酶的主体部分, 大部分暴露在溶剂中, 更易受到溶剂的影响; 而 C、D 链仅含 12 个氨基酸, 分别结合与 A、B 链外侧的结合口袋中, 受到口袋内氨基酸残基的牵制作用, 稳定性较好。

图 3B 给出了体系整体主链  $C_{\alpha}$  原子、配体 FAM 的 ligand 分子的 RMSD 随时间的变化。与图 3A 对比发现两个小分子的 RMSD 值在模拟

过程中基本趋于平衡,经计算,ligand 的 RMSD 为  $(0.173\pm 0.024)$  nm,无较大波动。表明 FAM 稳定地结合在活性口袋中。而体系整体的 RMSD 数值则维持在较高位置,这可能是由于复合物分子较大,通过两分子 FAM 相互作用,形成二聚体结构,故在模拟过程中分子内有着较为显著的相对运动,4 条链的相对位置及角度有较大变化。

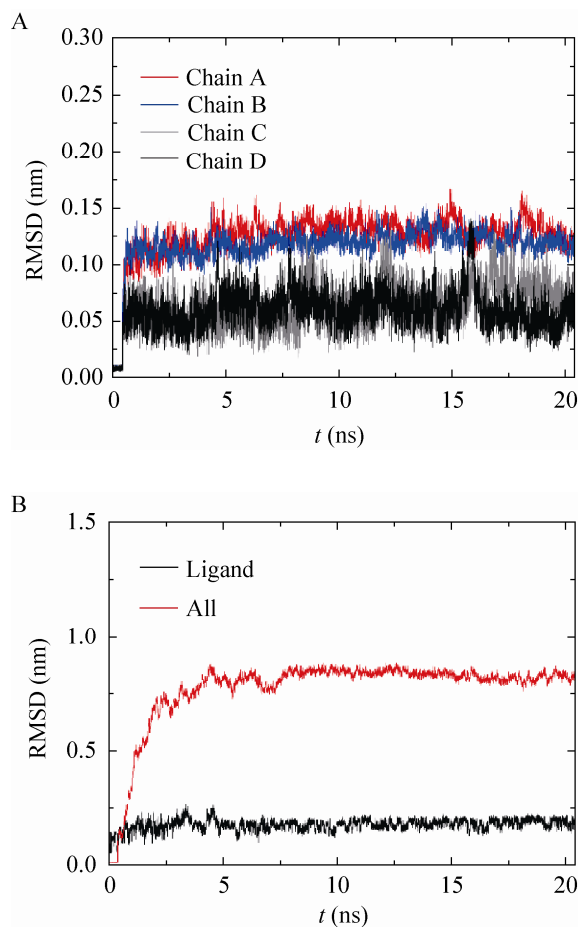


图3 蛋白质体系中  $C_{\alpha}$  原子与 FAM 方均根偏差随时间的变化

Fig. 3 RMSD of the  $C_{\alpha}$  atoms in the proteinic system and FAM versus simulation time.

图 4 给出了复合物分子中蛋白酶 B 链和 D 链的主链  $C_{\alpha}$  方均根涨落 (Root mean square fluctuation, RMSF) 分布。由于体系为二聚体结构,为简化分析运算,此处仅分析 B 链和 D 链。从图 4 中可以看到 4 个区域的涨落较大,分别是 P96-S101、P115-R119、R123-L126、V158-V163。通过 VMD 观察这 4 个区域在水中模拟的运动轨迹,发现 P96-S101 段和 P115-R119 段是两个 loop 区,暴露在水中,与周围的残基作用较弱,所以柔性较大;R123-L126 段和 V158-V163 段则是两个  $\beta$  折叠片,均与 FAM 结合部位相近。较大的运动幅度可能是受到了小分子的影响。D 链仅有较小的涨落出现,说明其柔性较小。

图 5A 给出了晶体结构中蛋白酶 A 链和 B 链温度因子 (B-factor, B 因子) 的自相关性。可以看到, A 链和 B 链的相关性很高,其相关系数为 0.81,这表明模拟分析方法可靠。图 5B 给出了模拟计算得到的 B 因子与 X-ray 实验所得 B 因子的相关性。一般来说,体系较大的 RMSF 和 B 因子均可用于说明相应的残基部分

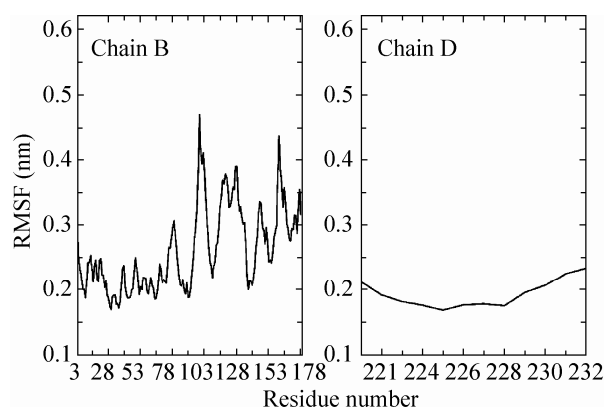


图4 体系 B 链和 D 链  $C_{\alpha}$  原子的方均根涨落分布

Fig. 4 RMSF distribution of the  $C_{\alpha}$  atoms in chain B and chain D.

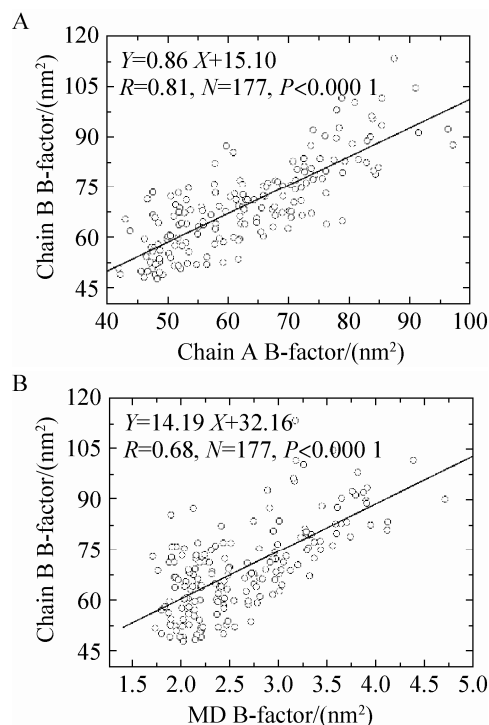


图5 晶体结构 A 链和 B 链的 B 因子相关性以及晶体结构 B 因子与分子模拟 B 因子数据的相关性

Fig. 5 The correlation of B-factors for crystal structure from PDB of chain A with chain B, and that of calculated B-factor values in chain B from MD simulation with B-factors from crystal structure.

柔性较大。经计算，二者相关性为 0.68。考虑到样本数达到了 177 个，两组数据存在着较为

表 2 各能量对结合自由能的贡献 (kJ/mol)

Table 2 The contribution to binding free energy by each kind of energy (kJ/mol)

Item	Protease	FAM	Complex	Delta
$ELE_{IN}$	-6 762.52±156.40	-180.34±5.18	-6 959.93±158.87	-17.08±6.37
$VDW_{IN}$	-1 50.5.78±3.55	17.84±0.94	-1 561.72±2.64	-73.78±5.84
$ELE_{PB}$	-5 081.33±125.82	- 42.89±0.38	-5 073.30±129.40	50.92±4.45
$VDW_{PB}$	101.12±0.64	6.48±0.12	98.89±0.66	-8.70±0.18
$H$	-5 424.77±56.56	-71.95±9.58	-5 545.37±58.18	-48.65±2.64
$TS$	4 066.81±9.64	87.87±0.38	4 137.76±11.01	-16.92±1.00
$\Delta G_{bind}$		-31.73		
$\Delta G_{exp}$		-43.49 <sup>[11]</sup>		

显著的相关性，进而证明用模拟计算分析得到的柔性分布结论较为可靠。

### 2.3 分子识别

表 2 列出了各能量项对 FAM 与 HCV NS3/4A PR 结合自由能的贡献。结合自由能 (Binding free energy) 对药物分子的活性评价起着重要的作用。表 2 中  $ELE_{IN}$  表示真空状态下分子内能的静电能部分； $VDW_{IN}$  表示真空中分子内能的范德华能部分； $ELE_{PB}$  是指通过 PB 算法计算出的体系与溶剂之间的静电能； $VDW_{PB}$  是指通过 PB 算法计算得到的体系与溶剂之间的范德华能。Protease 表示对 HCV NS3/4A PR 单独计算的能量结果；FAM 则表示单独考虑 FAM 的能量结果；Complex 表示对复合物进行模拟计算的能量值；Delta 值则是复合物形成前后的能量变化。

从表 2 可以看出，当 FAM 与 PR 结合成复合物时，体系内能减小。这一过程主要通过两条途径实现：一是体系内能的静电部分减小，释放了 17.08 kJ/mol；二是体系内能的范德华部分的减小，释放了 73.78 kJ/mol，促进了复合物的形成。体系溶剂化能的范德华部分也降低

(-8.70 kJ/mol); 但代替牺牲溶剂化能的静电部分在结合过程中则大幅增大 (50.92 kJ/mol), 逆转了内能静电部分降低引起的促结合结果。

综上, 非极性相互作用为 FAM 与 PR 识别的主要驱动力。此外, 体系总焓变为-48.65 kJ/mol, 总熵变为-16.92 kJ/mol, 求得体系绝对结合自由能为-31.73 kJ/mol。实验文献[12]提到, 与 FAM 结构类型很高的化合物抑制剂的实验结合自由能值  $\Delta G_{\text{exp}}$ 。此化合物仅以 O 取代了 N30, 所以其实验数据有较高的参考价值。经对比, 计算得出的理论值与实验值基本吻合。

氢键的存在可以维持复合物分子的稳定性, 同时也可以增强受体-配体分子之间的特异性识别。为进一步探讨 HCV NS3/4A PR 与 FAM 的识别机制, 表 3 列出了二者之间形成的氢键。氢键采用几何判据<sup>[30]</sup>, 供体原子-氢原子-受体原子之间的夹角大于  $135^\circ$ , 供体原子-受体原子之间的距离小于 0.35 nm。表中 ligand1 和 ligand2 分别表示结合在 A 链和 B 链的 FAM, Freq 表示氢键占有率, 即氢键在整个模拟过程中出现的概率, Freq 越大表示对应的氢键越稳定。

从表 3 可以看出 FAM 与 A157、R155 和 S139 形成了稳定的氢键, 与 X-ray 实验数据<sup>[12]</sup>一致。其中, S139 位于 HCV NS3/4A PR 的催化中心, 是酶发挥生物活性的关键残基之一, R155 和 A157

位于的扭曲片段也是 PR 抑制剂的重要结合部位<sup>[1]</sup>。FAM 与 PR 之间较为稳定的 3 个氢键为维持复合物的稳定性, 并有助于促进抑制剂的特异性识别, 发挥较高的抑制活性 ( $K_i=0.02 \mu\text{mol/L}$ )。

表 4 列出了 HCV NS3/4A PR 与底物 FAM 识别的关键残基能量分解结果。考虑到体系为同源二聚体结构, 并结合两分子 FAM, 为方便分析, 此处仅研究了与 B 链结合的 ligand2。表中的 Resid 表示关键残基的残基号, ligand1 表示与 A 链结合的 FAM;  $E_{\text{vdw}}$  为真空范德华结合能;  $E_{\text{ele}}$  为真空静电结合能;  $E_{\text{gb}}$  为极性溶剂化结合能,  $E_{\text{gbsur}}$  表示非极性溶剂化结合能。能量分解显示, 小分子 ligand1 在结合过程中起着重要作用, 说明复合物中 FAM 以双分子形式存在是其发挥药效的重要形式。B 链的 F154'-A157' 段和 R130' -Y134' 段也在结合中起着关键的作用, 与有关文献[1]报道的 HCV NS3/4A 蛋白酶抑制剂结合部位基本一致。此外, V55' 也在 FAM 的结合中作出了较大的贡献。Welsch 等<sup>[31]</sup>曾探究过 V55 突变后产生酰胺基抵抗的现象, 可见, V55' 是 FAM 发挥抑制作用的关键残基。由此可以断定: HCV NS3/4A 蛋白酶的 F154-A157、R130-Y134 为 FAM 结合的主要区域, V55 与另一分子的 FAM 为维持复合物结构稳定, 发挥抑制作用提供了重要保证。

表 3 FAM 和 HCV NS3/4A PR 之间所形成的氢键

Table 3 The hydrogen bonds between FAM and HCV NS3/4A protease

Donor	Acceptor	Distance (nm)	Angle ( $^\circ$ )	Freq (%)
Ligand1-N28-H11	A157-O	0.2854±0.008	160.17±8.83	82.86
Ligand1-N30-H1	A157-O	0.2902±0.007	153.90±7.74	32.28
Ligand1-N41-H40	R155-O	0.2896±0.007	164.63±8.57	19.63
Ligand2-N41-H40	A157'-O	0.2858±0.008	149.76±9.85	79.03
Ligand2-N30-H1	A157'-O	0.2863±0.008	149.05±10.35	74.38
S139'-OG-HG	ligand2-O42	0.2803±0.011	157.01±10.87	30.95



图 6A 给出了在势能最低构象中,以 ligand2 (FAM) 为中心,半径为 0.4 nm 范围内的接触残基。图中, D79'-D81'、V132'-S139'、F154'-D168'、用黄色 solid-ribbon 模式表示, A 链的关键残基用红色 stick 模式表示, B 链的关键残基用蓝色 stick 模式表示, FAM ligand1 用

line 模式表示。ligand2 (FAM) 用 stick 模式表示。从图中可以看出位于酶催化中心的 H57'的咪唑环与 FAM 的喹啉环通过  $\pi$ - $\pi$  作用结合在一起。表明蛋白酶此处活性口袋平坦, FAM 的喹啉环为使其在活性口袋中结合地更加紧密,有利于提高抑制活性。图 6B 主要展示了 FAM 在复合

表 4 HCV NS3/4A 蛋白酶与底物 FAM 结合密切相关残基的能量分解结果 (kJ/mol)

Table 4 The energy decomposition of the key residues in HCV NS3/4A protease significantly concerned with binding ligand FAM. (kJ/mol)

No.	Resid	$E_{vdw}$	$E_{ele}$	$E_{gb}$	$E_{gbsur}$	$E_{tot}$
1	Ligand1	-9.82±1.57	-1.05±1.01	4.28±0.40	-1.05±0.08	-7.65±1.13
2	R155'	-1.65±0.95	-6.00±0.61	3.91±0.18	-0.19±0.00	-3.94±0.24
3	V55'	-3.76±0.55	-1.22±0.43	1.97±0.26	-0.38±0.04	-3.39±0.37
4	F154'	-2.36±0.54	-0.92±0.67	0.77±0.23	-0.26±0.02	-2.77±0.16
5	I153'	-2.61±0.12	-0.49±0.67	1.31±0.50	-0.16±0.02	-1.94±0.21
6	A156'	-1.30±0.16	-0.29±0.03	0.28±0.13	-0.09±0.02	-1.41±0.10
7	G152'	-1.01±0.26	0.33±0.10	-0.19±0.05	-0.04±0.01	-0.92±0.30
8	Y134'	-1.60±0.22	-1.53±0.72	2.40±0.94	-0.16±0.03	-0.89±0.57
9	S133'	-0.94±0.21	0.18±0.23	-0.03±0.30	-0.01±0.00	-0.80±0.14
10	A157'	-0.91±0.23	0.35±0.25	0.01±0.25	-0.12±0.03	-0.68±0.15
11	T76	-0.67±0.40	0.07±0.07	0.10±0.06	-0.12±0.06	-0.61±0.42
12	T54	-0.65±0.33	-0.29±0.19	0.54±0.32	-0.15±0.04	-0.56±0.23
13	G137'	-0.93±0.47	-2.01±1.81	2.52±1.37	-0.09±0.01	-0.51±0.65
14	R130'	-0.65±0.70	-0.13±0.25	0.41±0.41	-0.14±0.02	-0.50±0.86

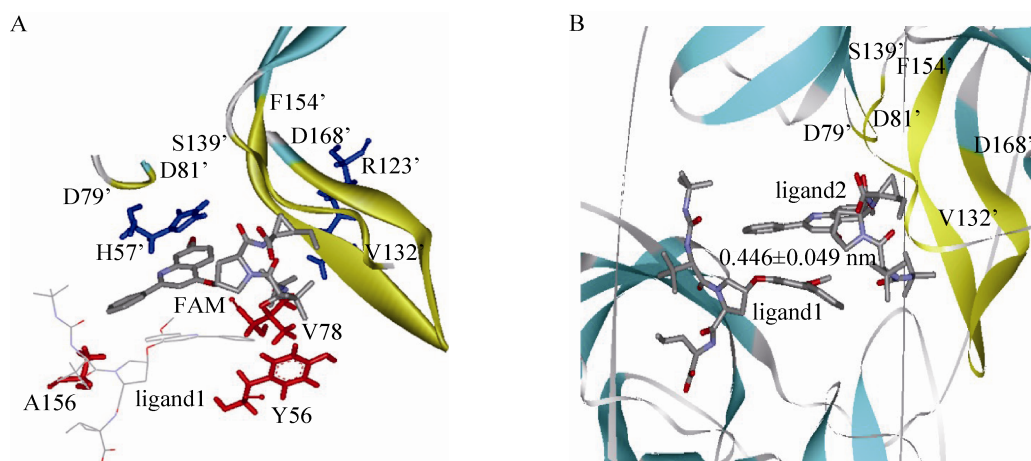


图 6 HCV NS3/4A PR 与底物 FAM 的结合模式图

Fig. 6 The map of the binding mode between HCV NS3/4A protease and FAM.

物中的双分子结合模式。从图中可以看出, ligand2 分子的苯环与 ligand1 分子的喹啉环相互平行。经计算, 平面间的距离为  $(0.446 \pm 0.049)$  nm, 符合产生  $\pi$ - $\pi$  作用的条件。通过 VMD 跟踪观察, 也发现 FAM 在模拟过程中一直保持这一结构。结合表 3 的能量分解分析, 可以进一步确定, 这种结合模式非常稳定, 在分子识别和维持抑制活性方面作出了重要的贡献。

## 2.4 耐药性突变

有报道<sup>[32-35]</sup>显示, HCV NS3 蛋白 R155K、D168E/V 和 V170T 突变株对包括 Faldaprevir 等多种 HCV NS3/4A PR 可逆抑制剂产生耐药。为了进一步探究这些突变残基对 FAM 分子识别的影响, 对复合物晶体结构相关氨基酸进行了定点突变, 并对突变后与 FAM 形成复合物的结合自由能变进行了分析。图 7 给出了 HCV NS3/4A PR 氨基酸定点突变后关键残基对结合自由能贡献的变化。图中蓝色表示野生型 (Wild type, WT) HCV NS3/4A PR 关键残基的结合自由能; 红色表示 R155K 突变株 NS3/4A PR 关键残基的结合自由能; 淡蓝色表示 D168E 突变株 NS3/4A PR 关键残基的结合自由能, 而紫色和绿色则分别表示 D168V 和 V170T 突变株 NS3/4A PR 关键残基的结合自由能。

经计算, 形成 FAM-R155K 突变株复合物的结合自由能变为  $-11.55$  kJ/mol, FAM-D168E 突变株复合物为  $-17.48$  kJ/mol, 而形成 FAM-D168V 突变株复合物与 FAM-V170T 突变株复合物的结合自由能变则分别为  $-10.91$  kJ/mol 和  $-12.14$  kJ/mol。与野生型 FAM-NS3/4A PR 复合物的  $-31.73$  kJ/mol 相比不难看出, R155、D168 和 V170 的突变对 FAM 的结合能力造成了不同程度的下降。说明 FAM 对突变型的 HCV NS3/4A PR 结合能力较弱。

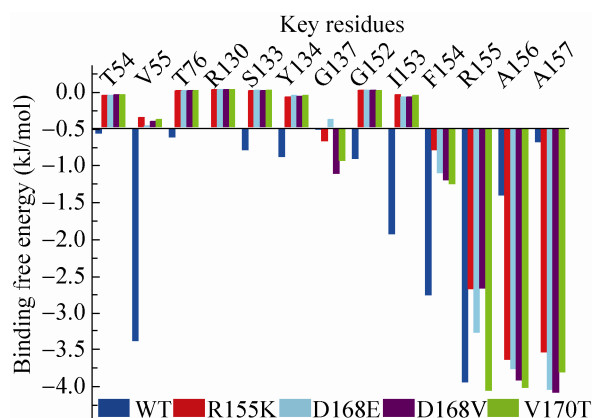


图 7 HCV NS3/4A PR 氨基酸定点突变后关键残基对结合自由能贡献的变化

Fig. 7 The change of the decomposition of key HCV NS3/4A PR residues after the site-directed amino acid mutation.

另外, 从图 7 可以看出突变株的 V55、I153 和 F154 能量下降最明显。这可能是由于 R-K、D-E、D-V 以及 V-T 突变后, 改变了口袋内部的疏水性造成的; 此外, 侧链体积增加, 使得口袋内部的原子堆积, 从而使得口袋发生了形变, 降低了 FAM 的结合自由能。虽然突变株的 A156、A157 的结合自由能较野生型高, 但由于突变后影响了 FAM 识别的大部分关键残基, 故复合物体系总体自由能显著降低。再次证明了 V55、I153 和 F154 对 FAM 的识别十分重要, 与前文分析结果一致。

## 2.5 构象变化

### 2.5.1 自由能曲面分析

图 8 给出了  $26.85$  °C 下体系自由能曲面图。图中的颜色越深, 相应区域的自由能越低。由图 8 可知, 体系在模拟过程中存在 4 个相对独立的低自由能区域, 即左侧的 M1、右上部的 M2、中部的 M3 以及右下部的 M4。为了进一步分析图中出现低自由能区的原因, 我们进行了聚类分析。这 4 个区域分别与图 9 中的 4 类能量最低构象相对应。

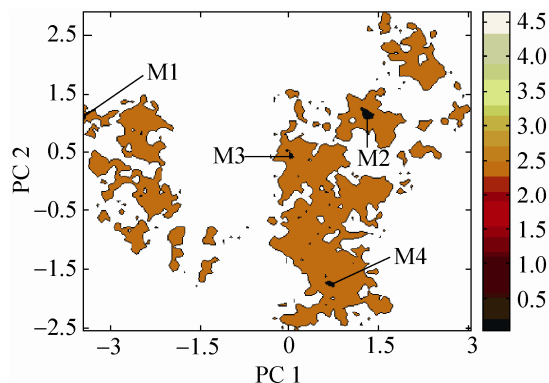


图 8 体系的自由能曲面图

Fig. 8 The free energy landscape map of the system.

### 2.5.2 聚类分析

图 9 给出了 HCV NS3/4A 蛋白酶与 FAM 所构成的复合物聚类分析的结果。RMSD 域值为 0.31 nm。从图 9 中可以看出经过 MD 模拟所得的 20 400 个体系构象共 4 类, 1.0 ns 以前为第 1 类; 2.76–4.41 ns、4.62–5.92 ns、6.03–6.04 ns、6.16–6.19 ns 之间为第 2 类; 4.42–4.61 ns、1.62–1.76 ns、5.93–6.15 以及 6.20–20.4 ns 之间为第 3 类; 剩余构象合为第 4 类, 它们所占的概率分别是 5.34%、15.63%、70.67%和 8.36%。聚类分析结果表明, 在 MD 模拟的 20.4 ns 时间内, HCV NS3/4A 蛋白酶和配体 FAM 形成的复合物存在 4 个优势构象; 各类构象之间有一定的结构跃迁, 这也部分地解释了 RMSD 中出现的涨落。

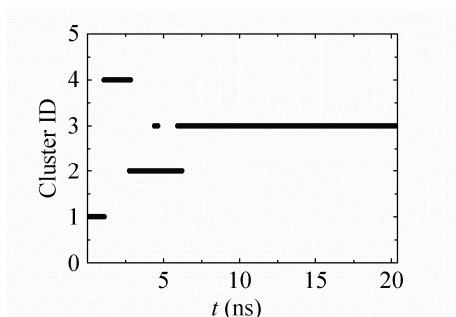


图 9 基于 HCV NS3/4A 蛋白酶 RMSD 的聚类分析  
Fig. 9 Cluster analysis based on RMSD of the HCV NS3/4A protease.

图 10 给出了复合物聚类分析得到的各类代表性构象。图中结合于 A 链的 FAM 用 stick 模式表示; 另一分子 FAM 和蛋白酶催化中心 (H57、D81 和 S139) 用紫红色 stick 表示; 蓝色片段表示能量分解中有较大贡献的 F154–A157 和 R130–Y134。从图 10 可以看到, 在 4 类代表性构象中, 口袋区的结合模式变化不大, 仅是 ligand2 (FAM) 分子的苯环与 ligand1 分子的喹啉环空间相对位置略有变化, 这点在图 6 也有提到。结合模式方面, FAM 的吡咯环主体通过 C22 和 N23 位取代侧链, 稳定附着于 F154–A157 段, 为其喹啉环侧链和乙烯基侧链顺利嵌入催化中心提供了保证。由于此 3 处催化中心位于蛋白酶内部, 在模拟过程中, 没有发生明显的相对位移, 使得 FAM 更有效地占据催化中心, 发挥抑制活性。

## 3 结论

对 HCV NS3/4A 蛋白酶与底物 FAM 的复合物进行 20.4 ns 的显含水分子动力学模拟。基于 RMSF 计算值找出了分子的柔性区域, 并与实验 B 因子值进行了相关性分析, 良好的相关性验证了模拟方法的可靠性。FAM 的取代尿素侧链上的 NH 与蛋白酶 A157, 以及乙烯基侧链的 O 与 S139 形成较为稳定的氢键。结合自由能计算值与类似化合物的实验值吻合较好, 其中范德华相互作用为复合物形成的主要驱动力。通过能量分解分析发现, FAM 的双分子结合模式为发挥药效关键形式, V55 在形成复合物的关键残基, 对现有实验数据做了有益补充。随后的耐药性突变实验也证明了 V55、I153 和 F154 对 FAM 识别重要作用。最后, 对复合物体系的构象变化做了简要分析, 得到了模拟过程中的 4 类代表性稳定构象。模拟结果为以 HCV NS3/4A

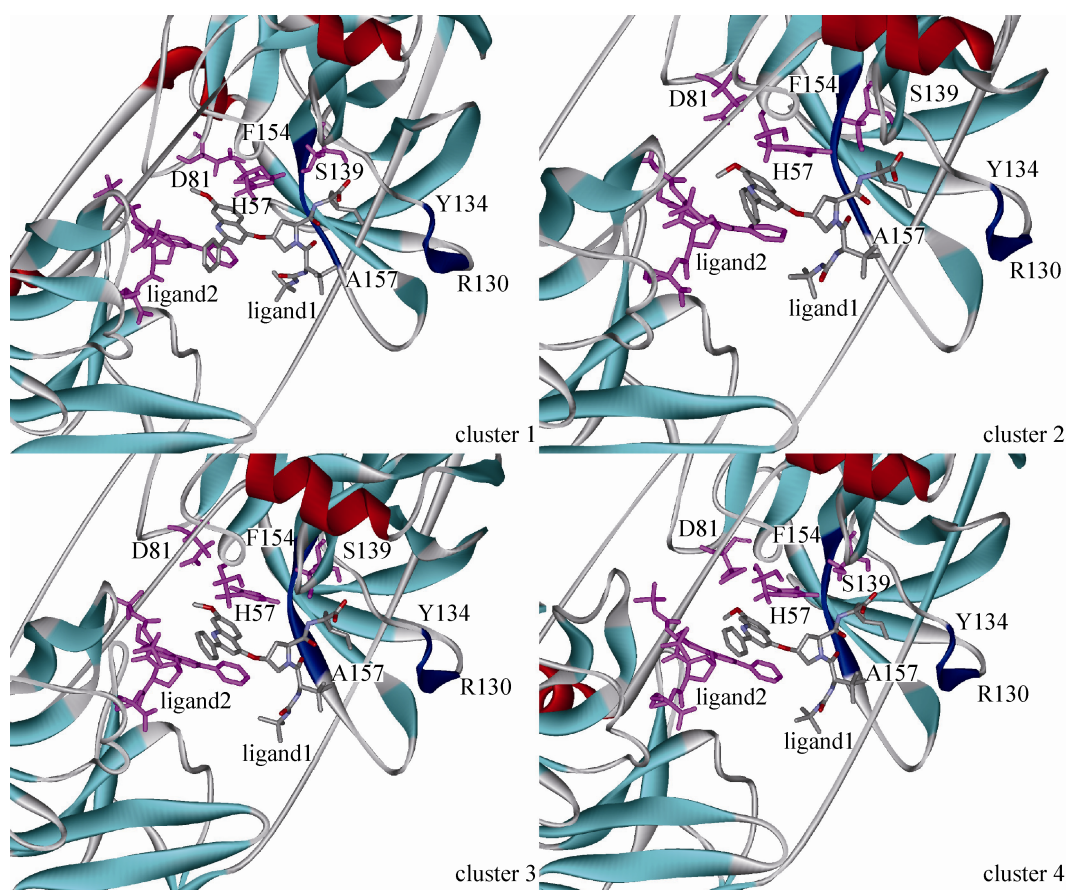


图 10 聚类分析的代表构象

Fig. 10 The representative conformation of cluster analysis.

蛋白酶为靶点的药物分子设计和药物分子的作用机理研究提供了一定的理论指导。

## REFERENCES

- [1] Guan Y, Sun HY, Li YY, et al. The competitive binding between inhibitors and substrates of HCV NS3/4A protease: a general mechanism of drug resistance. *Antiviral Res*, 2014, 103 (3): 60–70.
- [2] Lavanchy D. Evolving epidemiology of hepatitis C virus. *Clin Microbiol Infect*, 2011, 17(2): 107–115.
- [3] Shepard CW, Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis*, 2005, 5(9): 558–567.
- [4] Chen SL, Morgan TR. The natural history of hepatitis C virus (HCV) infection. *Int J Med Sci*, 2006, 3(2): 47–52.
- [5] Poordad F, Dieterich D. Treating hepatitis C: current standard of care and emerging direct-acting antiviral agents. *J Viral Hepat*, 2012, 19(7): 449–464.
- [6] Zeuzem S, Poordad F. Pegylated-interferon plus ribavirin therapy in the treatment of CHC: individualization of treatment duration according to on-treatment virologic response. *Curr Med Res Opin*, 2010, 26(7): 1733–1743.
- [7] Pawlotsky JM. New hepatitis C therapies: the toolbox, strategies, and challenges. *Gastroenterology*, 2014, 146(5): 1176–1192.
- [8] Kim DW, Gwack Y, Han JH, et al. C-terminal



- domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. *Biochem Biophys Res Commun*, 1995, 215(1): 160–166.
- [9] Baril M, Racine ME, Penin F, et al. MAVS dimer is a crucial signaling component of innate immunity and the target of hepatitis C virus NS3/4A protease. *J Virol*, 2009, 83(3): 1299–1311.
- [10] Li K, Foy E, Ferreon JC, et al. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc Natl Acad Sci USA*, 2005, 102(8): 2992–2997.
- [11] Foy E, Li K, Wang CF, et al. Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease. *Science*, 2003, 300(5622): 1145–1148.
- [12] LaPlante SR, Nar H, Lemke CT, et al. Ligand bioactive conformation plays a critical role in the design of drugs that target the hepatitis C virus NS3 protease. *J Med Chem*, 2014, 57(5): 1777–1789.
- [13] Wang JM, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. *J Comput Chem*, 2000, 21(12): 1049–1074.
- [14] Wang JM, Wolf RM, Caldwell JW, et al. Development and testing of a general amber force field. *J Comput Chem*, 2004, 25(9): 1157–1174.
- [15] Jorgensen WL, Chandrasekhar J, Madura JD, et al. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*, 1983, 79(2): 926–935.
- [16] Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J Comput Phys*, 1977, 23(3): 327–341.
- [17] Kollman PA, Massova I, Reyes C, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res*, 2000, 33(12): 889–897.
- [18] Wang W, Donini O, Reyes CM, et al. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu Rev Biophys Biomol Struct*, 2001, 30(1): 211–243.
- [19] Simonson T. Macromolecular electrostatics: continuum models and their growing pains. *Curr Opin Struct Biol*, 2001, 11(2): 243–252.
- [20] Bashford D, Case DA. Generalized born models of macromolecular solvation effects. *Annu Rev Phys Chem*, 2000, 51(1): 129–152.
- [21] Still WC, Tempczyk A, Hawley RC, et al. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc*, 1990, 112(16): 6127–6129.
- [22] Weiser J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J Comput Chem*, 1999, 20(2): 217–230.
- [23] Maisuradze GG, Liwo A, Scheraga HA. Relation between free energy landscapes of proteins and dynamics. *J Chem Theory Comput*, 2010, 6(2): 583–595.
- [24] Chen LY. Exploring the free-energy landscapes of biological systems with steered molecular dynamics. *Phys Chem Chem Phys*, 2011, 13(12): 6176–6183.
- [25] Nguyen PH, Derrenbaumaux P. Configurational entropy: an improvement of the quasiharmonic approximation using configurational temperature. *Phys Chem Chem Phys*, 2012, 14(2): 877–886.
- [26] Hegger R, Altis A, Ngeuyen PH, et al. How complex is the dynamics of peptide folding? *Phys Rev Lett*, 2007, 98(2): 028102.
- [27] Barlow PN, Steinkasserer A, Norman DG, et al. Solution structure of a pair of complement modules by nuclear magnetic resonance. *J Mol Biol*, 1993, 232(1): 268–284.
- [28] Feig M, Karanicolas J, Brooks III CL, et al. MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graphics Modell*, 2004, 22(5): 377–395.
- [29] Daura X, Gademann K, Jaun B, et al. Peptide folding: when simulation meets experiment. *Angew Chem Int Ed*, 1999, 38(1/2): 236–240.
- [30] Hu JP, Gong XQ, Su JG, et al. Study on the molecular mechanism of inhibiting HIV-1 integrase by EBR28 peptide *via* molecular modeling approach. *Biophys Chem*, 2008, 132(2/3): 69–80.

- 
- [31] Welsch C, Schweizer S, Shimakami T, et al. Ketoamide resistance and hepatitis C virus fitness in Val55 variants of the NS3 serine protease. *Antimicrob Agents Chemother*, 2012, 56(4): 1907–1915.
- [32] Berger KL, Lisetté L, Triki I, et al. Viral resistance in hepatitis C virus genotype 1-infected patients receiving the NS3 protease inhibitor Faldaprevir (BI 201335) in a phase 1b multiple-rising-dose study. *Antimicrob Agents Chemother*, 2013, 57(10): 4928–4936.
- [33] Berger KL, Triki I, Cartier M, et al. Baseline hepatitis C virus (HCV) NS3 polymorphisms and their impact on treatment response in clinical studies of the HCV NS3 protease inhibitor faldaprevir. *Antimicrob Agents Chemother*, 2014, 58(2): 698–705.
- [34] zeuzem S, buggisch P, agarwal K, et al. The protease inhibitor, GS-9256, and non-nucleoside polymerase inhibitor tegobuvir alone, with ribavirin, or pegylated interferon plus ribavirin in hepatitis C. *Hepatology*, 2012, 55(3): 749–758.
- [35] Soumana DI, Ali A, Schiffer CA. Structural analysis of asunaprevir resistance in HCV NS3/4A protease. *ACS Chem Biol*, 2014, 9(11): 2485–2490.

(本文责编 郝丽芳)