

HIV-1 整合酶核心区野生型和 F185K 突变型活性和溶解性的比较及分子动力学模拟分析*

何红秋¹⁾ 胡建平²⁾ 刘 斌¹⁾ 陈慰祖¹⁾ 王存新^{1)**}

(¹⁾北京工业大学生命科学与生物工程学院, 北京 100124; (²⁾乐山师范学院化学与生命科学学院, 乐山 614004)

摘要 表达纯化了野生型(WT)及 F185K 突变型 HIV-1 整合酶核心区蛋白(IN_C), 并对二者的溶解性和活性进行了比较. 实验结果表明: F185K 突变后 IN_C 溶解性显著提高, 活性有一定程度降低. 对 WT 和 F185K IN_C 体系进行了 1800 ps 的分子动力学模拟. 模拟结果表明: F185K IN_C 功能 loop 区柔性和蛋白质整体运动性降低, 使蛋白质活性降低, F185K 突变后盐桥网络的变化驱动了 IN_C 局部构象改变, 引起 IN_C 表面的部分疏水残基被包埋, 亲水残基暴露, 相对亲水溶剂可接近面积增大, 同时, 突变后 IN_C 与水之间形成氢键的数量增加, 与水之间作用加强, 以上变化使 IN_C 溶解性提高. 分子动力学模拟与实验结果相吻合. 为理解蛋白质溶解性和对蛋白质进行可溶性改造提供了一定的理论依据.

关键词 HIV-1 整合酶, 活性, 溶解性, 分子动力学模拟

学科分类号 O641, Q7

DOI: 10.3724/SP.J.1206.2009.00126

蛋白质的溶解性对其结构、功能研究以及在医药和工业中的应用具有重要影响^[1]. 溶解性较差的蛋白质表达时易聚集形成包涵体, 为后期纯化带来不便, 对包涵体变性、复性纯化蛋白质则可能造成其错误折叠而影响活性. 因此, 对蛋白质进行改造以增加溶解性, 实现可溶性表达, 是蛋白质工程研究的重要领域. 对蛋白质可溶性改造的主要策略为: 一是预测蛋白质表面的非极性残基, 再用带电残基对其进行替换; 二是用 Lys 系统地替换蛋白质中的疏水残基^[2]. 以上策略已在蛋白质溶解性改造上有成功的应用^[1, 3~7], 其缺点是需要进行大量突变. 近年来, 利用物理化学和生物信息学等方法预测蛋白质溶解性取得了一定的进展^[8~10], 但是对蛋白质进行可溶性改造的问题至今仍在探索中.

HIV-1 整合酶(integrase, IN)由 3 个特定的结构域组成: N 端结构域、核心区(catalytic domain, IN_C)和 C 端结构域. IN 在体内催化 3'加工和链转移, 使病毒 cDNA 与宿主细胞基因组整合. 在体外, IN 可以催化 3'加工、链转移和去整合, 其中, IN_C 是 IN 发挥催化活性的关键区, 具有独立催化去整合的功能^[6]. 野生型(wild type, WT) IN 溶解性较差, 表达时形成包涵体^[6]. Jenkins 等^[6]对 IN_C 的

疏水残基系统地进行了突变, 发现 185 位的 Phe 突变为 Lys (F185K)可极大地提高蛋白质溶解性而其去整合活性不受明显影响. Goldgur 等^[11]构建了 F185K/W131E IN_C, 进一步增加了 IN_C 的溶解性, 并解析得到了晶体结构. Jenkins 等^[12]构建了 F185K/C280S IN, 实现了 IN 全酶的可溶性表达. 目前, 对突变后 IN_C 及 IN 的溶解性和活性定量比较和分析未见报道.

本工作通过实验构建表达了 WT 和 F185K 突变的 IN_C 蛋白, 对其溶解性和功能进行了定量比较. 用分子模拟方法搭建了 WT 和 F185K IN_C 模型, 采用分子动力学(molecular dynamics, MD)模拟方法对二者活性和溶解性方面的差异进行了分析, 在一定程度上解释了突变后 IN_C 活性和溶解性的变化, 为蛋白质可溶性改造提供了一些有价值的信息.

* 国家自然科学基金(30670497)和北京市自然科学基金(5072002)资助项目.

** 通讯联系人.

Tel: 010-67392724, E-mail: cxwang@bjut.edu.cn

收稿日期: 2009-03-05, 接受日期: 2009-05-18

1 材料与方法

1.1 材料

pET28b 载体和 *E. coli* BL21 (DE3) 宿主菌由北京工业大学生命科学与生物工程学院保存; 基因操作的工具酶购自 New England Biolabs 公司; 引物和寡核苷酸链的合成修饰由上海生工生物技术公司完成; 生化试剂和抗体购自 Sigma 公司. 用 Bio-Rad 680 型酶标仪测定吸光度(A_{405}); 测定荧光强度用 Biotek FLX800TBI 荧光分析仪; 蛋白质电泳仪为 Bio-Rad 产品.

1.2 WT 和 F185K IN_C 的表达、纯化以及溶解性和活性比较

以 pNL-IN^[13] (含 F185K/C280S 可溶性突变的 IN 全酶基因) 质粒为模板, 引物 N₃₀, 5' CTGCATATGTCATGGACAAGTAGACTGTAG 3' (下划线为 *Nde* I 酶切位点), C₂₁₂, 5' TATGGATCCTTATTCTTAGTTTGTATGTCTGT 3' (下划线为 *Bam*H I 酶切位点) 进行 PCR 扩增, PCR 产物经 *Nde* I 和 *Bam*H I 酶切、回收, 与 pET28b 连接, 构建重组质粒. 经 DNA 测序鉴定后, 构建正确的 F185K 突变型表达质粒命名为 pIN_CF185K. 通过重叠 PCR 方法将 pIN_CF185K 中 185 位的 Lys 残基突变为 Phe, 得到野生型 IN_C 基因, 参照构建 pIN_CF185K 的方法, 构建 WT IN_C 表达质粒 pIN_CWT. 分别转化 pIN_CWT 和 pIN_CF185K 到 *E. coli* BL21 中, 参照文献[13]的方法进行诱导表达. WT 和 F185K IN_C 从细胞破碎后的沉淀中通过尿素变性、复性和亲和层析纯化^[14]. IN_C 活性的比较采用改进的酶联免疫吸附试验(ELISA)方法, 参照文献[14]进行去整合反应, 反应结束后, 利用亲和素磁珠捕获产物 DNA 方法^[13]代替亲和素包被微孔板的方法, 并用 ELISA 对产物进行定量分析.

1.3 模型和计算

1.3.1 模型搭建. 用同源模建 Jackal 程序包^[15]构建 IN_C 结构. 以 PDB 1QS4^[11] 的 A 链坐标为初始结构, 对缺失的 141~144 号残基, 分别用 1BIS^[16] 中的 B 链同源区补齐 Ile141 和 Pro142, 1QS4 的 B 链同源区补齐 Tyr143 和 Asn144. 由于结构解析时 1QS4 引入了 F185K 和 W131E 两个突变^[11], 用 Jackal 程序包对模建的结构分别进行 E131W/K185F 和 E131W 突变, 得到 WT 和 F185K IN_C 结构. 再用最陡下降法和共轭梯度法分别进行 200 步能量优化, 将其作为 MD 模拟的初始模型.

1.3.2 分子动力学模拟. 用 GROMACS 程序包^[17]对 WT 和 F185K IN_C 在水溶液中进行 MD 模拟. 将 IN_C 质心置于立方体盒子中心, 溶质原子到盒子壁距离 0.7 nm, 采用 SPC 模型在盒子内填充水分子. 为保持体系的总电荷为零, 对 WT 和 F185K IN_C 体系中分别加入 1 和 2 个 Cl⁻ 离子, 除去相应位置的水分子. 采用 GROMACS 全原子力场, 将体系分别以最陡下降法和共轭梯度法进行 200 步能量优化, 然后进行 20 ps 的约束溶质分子 MD 模拟, 最后在 300 K 下进行了 1 800 ps 的无约束恒温 MD 模拟. MD 积分步长取 2 fs, 每隔 100 步记录一次轨迹. 在模拟中, 静电相互作用采用 PME 算法, 范德华相互作用取截断半径为 1.0 nm. 采用 GROMACS 提供的分析程序, 对模拟轨迹的后 1 000 ps 进行分析, 其中溶剂可接近表面积采用 NACCESS 程序包^[18], 探针半径设为 0.14 nm.

2 结果与讨论

2.1 WT 和 F185K IN_C 溶解性和活性比较

IN_C 表达和溶解性分析的 SDS-PAGE 结果如图 1a 所示, 由图 1 可见, 与没有质粒的空宿主菌对照(图 1-J)相比, 转化了 pIN_CF185K 和 pIN_CWT 重组质粒的宿主菌经 IPTG 诱导后有 22 ku 大小的

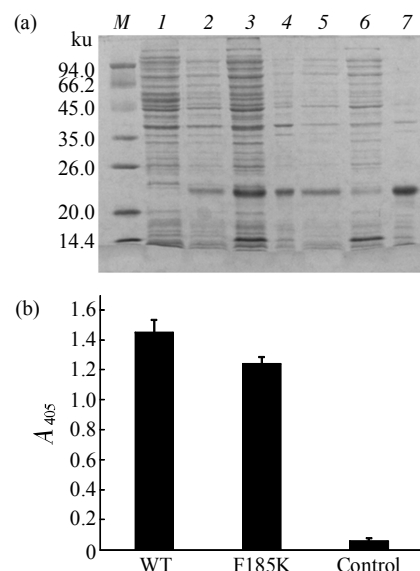


Fig. 1 Solubility and activity comparison of WT and F185K IN_C protein

(a) The SDS-PAGE analysis of the protein expression and solubility for WT and F185K IN_C. 1: Induced *E. coli* BL21 cell control; 2: Induced pIN_CF185K/BL21; 3: Supernatant of induced pIN_CF185K/BL21; 4: Insoluble pellet of induced pIN_CF185K/BL21; 5: Induced pIN_CWT/BL21; 6: Supernatant of induced pIN_CWT/BL21; 7: Insoluble pellet of induced pIN_CWT/BL21; M: Molecular mass marker. (b) Disintegration reaction activity comparison between WT and F185K IN_C.

目的条带(图 1-2, 图 1-5), 与 IN 核心区大小相符, WT IN_C 以包涵体形式存在于细胞沉淀中(图 1-7), 上清液(图 1-6)中只有痕量蛋白存在, F185K IN_C 虽然也有部分在沉淀中(图 1-4), 但有较大部分以可溶的形式出现在上清液中(图 1-3). 通过对目的条带初定量, F185K IN_C 上清液中目的蛋白是 WT IN_C 中的 20 倍, 与文献报道的溶解性差别相当^[6].

采用 ELISA 去整合方法, 用 A_{405} 值对反应产物定量以检测 IN_C 的催化活性, 结果如图 1b 所示. WT IN_C 催化去整合活性是 F185K 的 1.2 倍, 突变后 IN_C 活性有一定程度降低. 这与 Jenkins 等^[6]报道的 F185K 突变后 IN_C 活性不受影响甚至稍有提高不太相符. 为此, 表达、纯化了 WT 和 F185K/C280S IN 全酶, 并分别采用荧光法^[19]和 ELISA 方法^[13]对二者的活性进行了比较, 结果如图 2 所示. 由图 2 可见, WT IN 的 3'加工和链转移反应活性都明显高于 F185K/C280S IN, 前者催化反应的结果信号分别为后者的 1.4 和 1.3 倍, 相比较于两种 IN 蛋白核心区的活性差别(1.2 倍), 全酶的差别更大, 可能与全酶中引入了第二个突变(C280S), 对 IN 的活性也有一定的影响有关. 因此, 可溶性突变后 IN 和 IN_C 活性有一定的降低, 但降低的程度不大. 文献报道^[6, 12]突变后 IN_C 和 IN 全酶活性不受影响甚至稍有提高, 可能是由于文献

中对野生型 IN_C 和 IN 蛋白采用包涵体变性、复性后纯化, 而对可溶性突变型蛋白则从细胞破碎后上清液中直接纯化, 两种纯化工艺不同所致. 变性、复性操作后蛋白质由于错误折叠或不折叠等原因, 其活性会受到一定的影响. 本文中直接对野生型和突变型蛋白都采用包涵体变性、复性的办法纯化, 纯化工艺的一致性使对两种蛋白质活性的定量更准确, 结果更可靠.

2.2 MD 模拟结果分析

对 MD 模拟平衡后的 1 000 ps 采样数据, 进行了分析(图 3). 首先分析了两个体系势能随时间的变化, 从图 3a 可看出, WT 和 F185K IN_C 体系势能的平均值分别是 -395536 kJ/mol 和 -396686 kJ/mol, 标准偏差分别为 464 kJ/mol 和 469 kJ/mol, 涨落幅度都在 0.15% 内, 体系势能基本平稳. 图 3b 给出了体系 C _{α} 原子的均方根偏差(root mean square deviations, RMSD)随模拟时间的变化: WT 和 F185K IN_C 的 RMSD 均值分别为 0.24 nm 和 0.17 nm, F185K 的 RMSD 值较小, 表明 F185K IN_C 的整体运动性较 WT 小. 图 3c 给出了蛋白质各氨基酸残基 C _{α} 原子均方根涨落(root mean square fluctuations, RMSF), 可以看出两个体系均存在两个柔性较大的区域, 残基序号分别为 138~149 和 185~195. 两个区域均为 loop 区, 这与前人对整合酶 MD 模拟的结果相似^[20~23]. 185~195 区与第 185 突变残基毗邻, 柔性最大; 138~149 loop 区的柔性稍小, 能和 IN_C 的催化核心(Asp64-Asp116-Glu152)一起参与结合 DNA 并发挥催化功能^[21~23]. WT IN_C 分子的整体和功能 loop 区的 RMSF 均大于 F185K. 图 3d 给出了功能 loop 区 C _{α} 原子的 RMSD 随时间的变化, 由图 3 可见, WT 在 850~950 ps 之间 RMSD 有一个较大的跃迁, 此后趋于稳定, 在 0.19 nm 附近涨落, 略大于 F185K 的 0.18 nm, 与图 3c 中 RMSF 所体现出的两种蛋白质运动性差别一致. Greenwald 等^[23]通过实验发现功能 loop 区的运动性与 IN 的催化活性相关联. Brigo 等^[21]和 Barreca 等^[22]通过对 WT、耐药突变型 IN_C 以及它们与抑制剂的复合物进行了 MD 模拟, 得出功能 loop 区的柔性大小与 IN 的功能相关, 抑制剂通过抑制功能 loop 区的柔性而抑制 IN 的活性. WT IN_C 的功能 loop 区柔性及蛋白质整体运动性略大于 F185K. 因此, WT IN 的 3'加工和链转移活性均略高于 F185K/C280S IN, WT IN_C 去整合活性略高于 F185K IN_C.

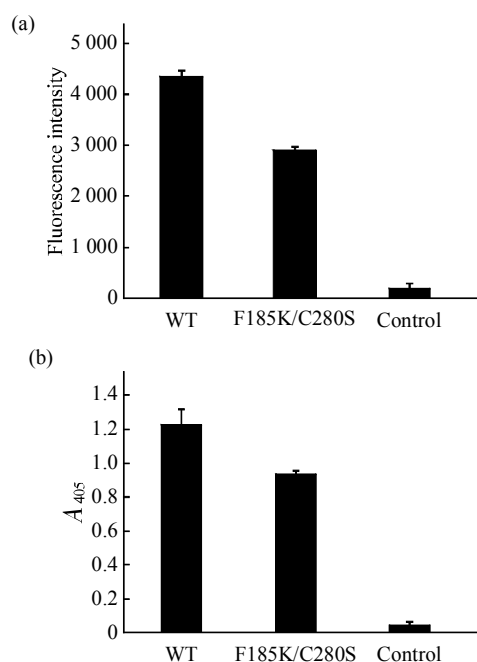


Fig. 2 Activities comparison between WT and F185K/C280S full IN protein

(a) 3'-processing reaction activity comparison. (b) Strand transfer reaction activity comparison.

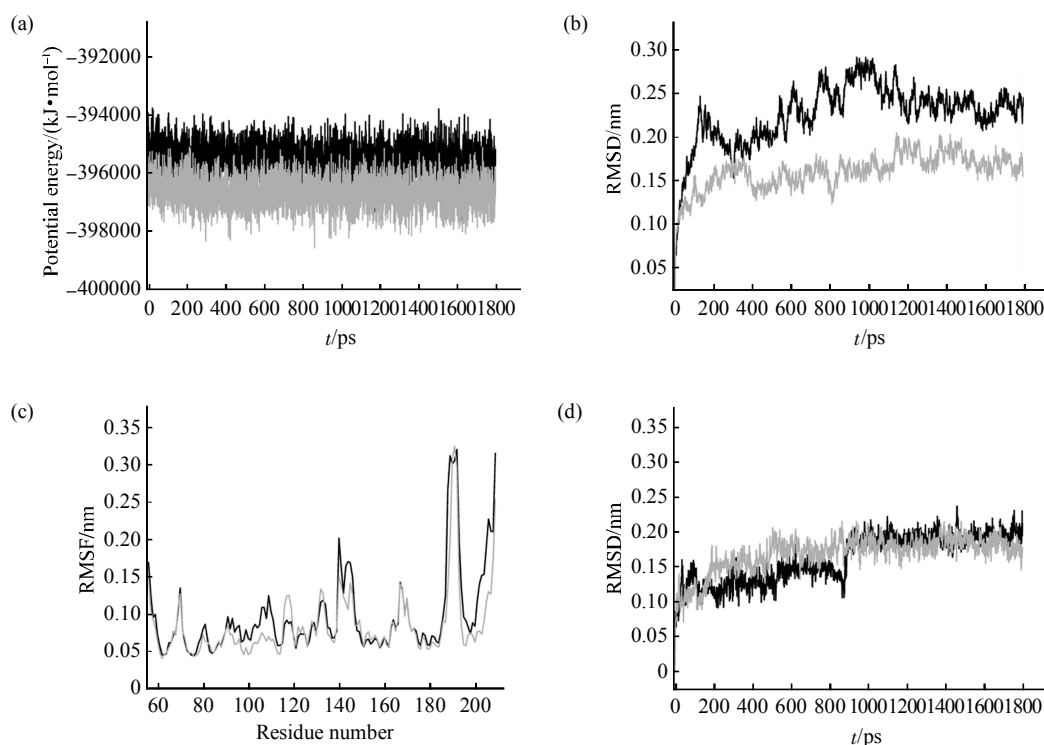


Fig. 3 Analysis of the MD trajectory of WT and F185K IN_c as a function of simulation time

(a) The potential energy. (b) The RMSD of C_α atoms. (c) The RMSF of C_α atoms. (d) The RMSD of the C_α atoms of the functional loop (138~149). —: WT; —: F185K.

IN_c 活性中心 3 个保守残基 Asp64、Asp116 和 Glu152 (DDE 基序)在空间结构中形成一个催化三角, 与 IN 的功能密切相关, DDE 中任何一个残基突变都使得 IN 失去活性^[24]. 表 1 给出了活性中心 D64、D116 和 E152 残基质心之间的距离和标准偏差. 从表 1 可看出, 两个体系在 MD 模拟过程中 DDE 基序残基相互之间的距离保持稳定, 涨落幅度也较小, 表明两个体系中起催化作用的关键残基动力学特性基本一致, 这与 F185K 突变后 IN_c 仍能保持较好活性有关.

Table 1 The average distances(nm) between centers of mass of residues in DDE motif

MD system	D64-D116	D116-E152	D64-E152
WT	0.66 ± 0.03	1.28 ± 0.06	0.93 ± 0.04
F185K	0.62 ± 0.03	1.24 ± 0.05	0.92 ± 0.04

2.3 溶剂可接近面积与氢键分析

图 4 给出了 WT 和 F185K IN_c 溶剂可接近面积(solvent accessible surface area, SASA)以及 IN_c 与溶剂(水)间形成氢键数量在 MD 模拟过程中随时间变化的趋势. 从图 4a 中可看出, 模拟达到平衡

后, F185K 和 WT IN_c 的总溶剂可接近面积(total)均值分别为 82.2 nm² 和 84.4 nm², F185K 比 WT 小 2.2 nm², 表明 F185K IN_c 的表面残基被包埋更多. 从图 4b 和 4c 中可看出, F185K IN_c 的非极性溶剂可接近面积(non-polar) 51.9 nm² 比 WT 的 53.9 nm² 小 2.0 nm², 而两个体系的极性溶剂可接近面积(polar)无显著差异, 分别为 30.5 nm² 和 30.3 nm². 因此, F185K 突变后 IN_c 部分残基被包埋, 溶剂可接近面积减小, 但是被包埋的部分基本是非极性的疏水残基, 而亲水的极性溶剂可接近面积保持不变, 其总效应为相对的亲水面积增大, 说明蛋白质的极性增强.

表 2 列出了 IN_c 中各 loop 区溶剂可接近面积的大小. F185K IN_c 较 WT 的 loop 区总溶剂可接近面积减少 0.47 nm², 其中非极性溶剂可接近面积减少 1.62 nm², 极性溶剂可接近面积却增加了 1.15 nm². 因此, F185K 突变后, IN_c 的 loop 区通过构象变化, 使疏水的残基被包埋, 亲水的残基被暴露, 增加了蛋白质的亲水性. loop 区共 53 个氨基酸残基, 约占 IN_c 残基数 32%, 但其溶剂可接近面积却占整个蛋白质的 43%, loop 区在蛋白质溶解性方面的贡献起到了较关键的作用.

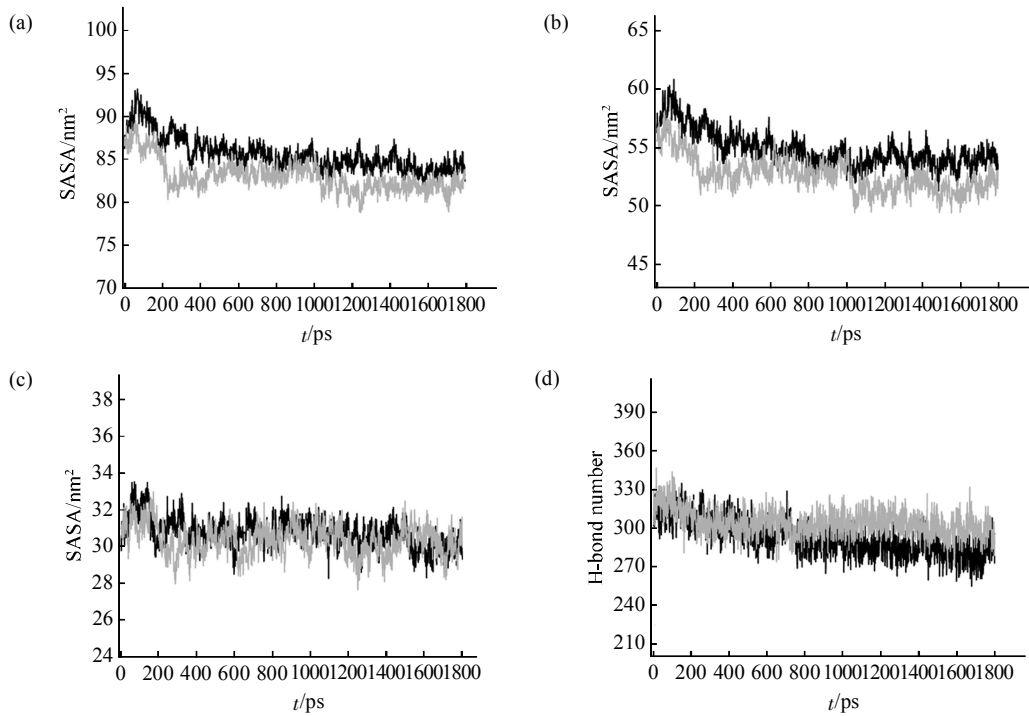


Fig. 4 Time evolution of the solvent accessible surface area (SASA) and hydrogen bond number of the WT and F185K IN_c

(a) The total SASA. —: WT total; —: F185K total. (b) The non-polar SASA. —: WT npolar; —: F185K npolar. (c) The polar SASA. —: WT polar; —: F185K polar. (d) Hydrogen bond number between IN_c and water. —: WT; —: F185K.

Table 2 The solvent accessible surface area (SASA) of loop regions in WT and F185K IN_c

Name and sequence	SASA/nm ²					
	Total		Non-polar		Polar	
	WT	F185K	WT	F185K	WT	F185K
Loop1 (080~094)	5.37	5.22	3.18	2.88	2.19	2.34
Loop2 (116~124)	4.74	4.41	2.59	2.35	2.15	2.06
Loop3 (138~149)	10.73	10.23	7.50	6.88	3.23	3.35
Loop4 (167~172)	5.61	5.53	3.06	2.99	2.55	2.54
Loop5 (185~195)	9.48	10.07	6.31	5.92	3.17	4.15
Sum	35.93	35.46	22.64	21.02	13.29	14.44

图 4d 给出了 MD 模拟过程中 IN_c 与水之间形成氢键数量随时间变化的趋势. 由图 4 可见, 随着模拟达到平衡, F185K IN_c 与水形成氢键的数量明显多于 WT, 其均值分别为 302 和 285 个, F185K 平均多出 17 个氢键, 表明 IN_c 与水之间的相互作用增强. 蛋白质表面残基与水分子之间形成氢键网络是蛋白质在水中溶解的一个至关重要的因素^[25], 因此, F185K IN_c 与水之间氢键个数的增加也是其溶解性提高的原因之一.

2.4 IN_c 蛋白的结构分析

回转半径 (radius of gyration, R_g) 被用来衡量 MD 过程中蛋白质结构的紧凑度和蛋白质的整体形状^[17, 26]. 体系在 MD 模拟过程中 R_g 随时间的变化如图 5a 所示, WT 和 F185K IN_c 的平均值分别为 1.53 和 1.52 nm, F185K 的结构更紧凑, 这一结果与前面关于突变型比野生型蛋白的 RMSD 和溶剂可接近面积的平均值要小的结果相吻合. 图 5b~d 分别显示了 MD 模拟过程中蛋白质在空间各方向

上 R_g 的变化情况. 由图 5 可见, WT 和 F185K IN_C 的 R_{gx} 没有差别, 平均值均为 1.39 nm; WT 的 R_{gy} 均值为 1.23 nm, 明显高于 F185K 的 1.18 nm, 说明在该方向 F185K IN_C 的结构更紧凑; WT 和

F185K IN_C 的 R_{gz} 平均值分别为 1.12 和 1.13 nm, 表明 F185K 在该方向较 WT 是一个稍为膨胀的状态. 总体而言, F185K IN_C 的结构比 WT 紧凑, 其局部发生较大的构象变化.

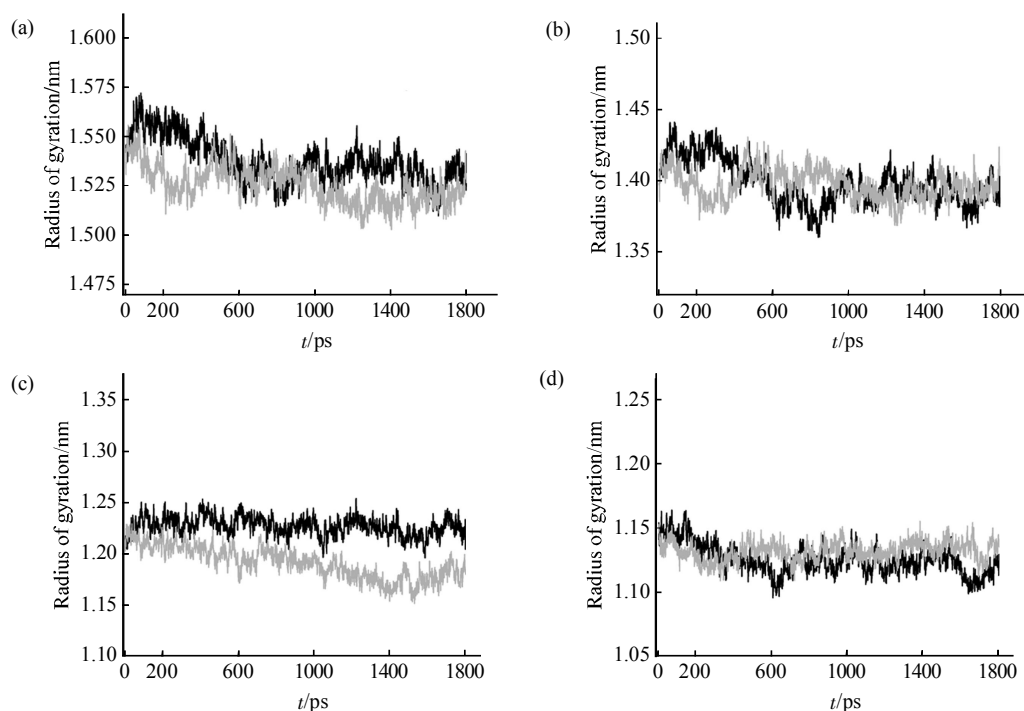


Fig. 5 Time dependence of the radius of gyration (R_g)

(a) The general R_g . —: R_g WT. —: R_g F185K. (b) R_{gx} . —: R_{gx} WT. —: R_{gx} F185K. (c) R_{gy} . —: R_{gy} WT. —: R_{gy} F185K. (d) R_{gz} . —: R_{gz} WT. —: R_{gz} F185K.

为进一步分析 WT 和 F185K IN_C 在构象上的差别, 将 MD 模拟轨迹后 1 000 ps 的平均构象叠落 (图 6). 结果表明, 两个蛋白质整体叠落较好, RMSD 为 0.218 nm, 但 loop 3、loop 5 和 C 端的 α 螺旋有较大的构象区别, 与前面分析这两个 loop 区的柔性比较大相吻合. F185K IN_C 的两个 loop 区及 C 端的 α 螺旋均向质心收缩, 其中 loop 5 和 C 端 α 螺旋整体向质心有较大的收缩, loop 3 收缩相

对较小. 这一结果与前面分析的回转半径变小和溶剂可接近面积减少的结果相吻合.

盐桥在蛋白质柔性、分子识别、结构域的运动和蛋白质构象变化等方面发挥着重要作用^[27, 28]. 表 3 列出了 F185K IN_C 相对于 WT 的盐桥变化. 以 1.4 nm 为截断距离, WT 和 F185K IN_C 分别有蛋白质内盐桥 97 和 101 个. F185K 突变后, IN_C 内部的盐桥发生了明显变化, 特别是 Lys185 和 loop 5、C 端 α 螺旋及其相邻区域的带电残基间形成了一些较强的盐桥, 包括正电 - 正电排斥和正电 - 负电吸引作用, 这些盐桥之间的相互作用作为驱动力, 使 loop 5 和 C 端 α 螺旋有明显的运动和较大的构象变化, 同时, 由于这些盐桥的形成和 WT IN_C 中原有的部分盐桥被破坏, 蛋白质内部的盐桥网络发生了较大变化, 蛋白质的整体构象也随之变化, 盐桥可能是 F185K 突变后引起 IN_C 构象变化的主要驱动力.

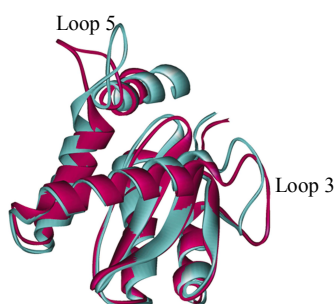


Fig. 6 Superimposed average structures from the MD simulations of the WT (red) and F185K IN_C (blue)

Table 3 Changes of salt bridges of F185K IN_C as compared with WT IN_C

Increased salt bridges	Distance/nm	Reduced salt bridges	Distance/nm
R107-D185	1.33 ± 0.13	R107-D202	1.23 ± 0.12
E069-E092	1.39 ± 0.13	R107-Q209	1.39 ± 0.17
K085-K185	1.16 ± 0.11	R107-E198	1.29 ± 0.11
E157-R187	1.35 ± 0.09	R199-D207	0.77 ± 0.16
K156-K188	1.31 ± 0.12	H078-E198	1.24 ± 0.08
K185-R187	0.85 ± 0.18	H078-K186	1.25 ± 0.10
K185-R199	1.31 ± 0.17		
K185-E198	0.83 ± 0.18		
K185-K186	1.09 ± 0.12		
K185-K188	1.27 ± 0.11		

Cutoff = 1.4 nm.

3 结 论

通过构建、表达、纯化 WT 和 F185K IN_C 蛋白, 比较其活性和溶解性, 结果表明, F185K 突变后 IN_C 溶解性显著提高, 活性有一定程度的降低. 通过对 WT 和 F185K IN_C 两个体系的 MD 模拟, 对比分析结果发现: a. 突变后, 体系的功能 loop 区柔性有一定程度的降低, 整体运动性略有减小, 催化核心 DDE 基序残基之间距离无显著变化, 因此, F185K IN 催化活性减小, 但活性受影响程度不大. b. F185K 突变对 IN_C 内部的盐桥改变驱动了蛋白质构象的变化, 特别是 loop 区构象变化明显, 引起蛋白质表面的部分疏水残基被包埋, 亲水残基被暴露, IN_C 的相对极性溶剂可接近面积增大. 同时, 蛋白质与水之间形成的氢键数量有明显的增加, 这些变化使 IN_C 溶解性显著提高. MD 模拟与实验结果相吻合. 本工作为蛋白质可溶性改造提供了一些有价值的信息和理论依据.

参 考 文 献

- Japrun D, Chusacultachai S, Yuvaniyama J, *et al.* A simple dual selection for functionally active mutants of plasmodium falciparum dihydrofolate reductase with improved solubility. *Protein Eng Design & Selection*, 2005, **18**(10): 457~464
- Malissard M, Berger E G. Improving solubility of catalytic domain of human β -1,4-galactosyltransferase 1 through rationally designed amino acid replacements. *Eur J Biochem/FEBS*, 2001, **268** (15): 4352~4358
- Wetzel R, Perry L J, Veilleux C. Mutations in human interferon gamma affecting inclusion body formation identified by a general immunochemical screen. *Biotechnology*, 1991, **9**(8): 731~737
- McElroy H E, Sisson G W, Schoettlin W E, *et al.* Studies on engineering crystallizability by mutation of surface residues of human thymidylate synthase. *J Crystal Growth*, 1992, **122**(1~4): 265~272
- Dale G E, Broger C, Langen H, *et al.* Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase. *Protein Eng*, 1994, **7**(7): 933~939
- Jenkins T M, Hickman A B, Dyda F, *et al.* Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc Natl Acad Sci USA*, 1995, **92**(13): 6057~6061
- Das D, Georgiadis M M. A directed approach to improving the solubility of moloney murine leukemia virus reverse transcriptase. *Protein Sci*, 2001, **10**(10): 1936~1941
- Simialowski P, Maetin-Galiano A J, Mikolajka A, *et al.* Protein solubility: sequence based prediction and experimental verification. *Bioinformatics*, 2007, **23**(19): 2536~2542
- Tjong H, Zhou H X. Prediction of protein solubility from calculation of transfer free energy. *Biophys J*, 2008, **95**(6): 2601~2609
- Idicula-Thomas S, Kulkarni A J, Kulkarni B D, *et al.* A support vector machine-based method for predicting the propensity of a protein to be soluble or to form inclusion body on overexpression in *Escherichia coli*. *Bioinformatics*, 2006, **22**(3): 278~284
- Goldgur Y, Craigie R, Cohen G H, *et al.* Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. *Proc Natl Acad Sci USA*, 1999, **96**(23): 13040~13043
- Jenkins T M, Engelman A, Ghirlando R, *et al.* A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J Biol Chem*, 1996, **271** (13): 7712~7718
- He H Q, Ma X H, Liu B, *et al.* A high-throughput format assay for HIV-1 integrase strand transfer reaction using magnetic beads. *Acta Pharmacol Sin*, 2008, **29** (3): 397~404
- Gao K, Wong S, Bushman F. Metal binding by the D, DX35E motif of human immunodeficiency virus type 1 integrase: selective rescue of Cys substitutions by Mn²⁺ *in vitro*. *J Virol*, 2004, **78**(13): 6715~6722
- Petrey D, Xiang Z X, Tang C L, *et al.* Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins: Struct Funct Genet*, 2003, **53**(Suppl 6): 430~435
- Goldgur Y, Dyda F, Hickman A B, *et al.* Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc Natl Acad Sci USA*, 1998, **95**(16): 9150~9154
- David V S, Erik L, Berk H. Gromacs User Manual, Version 3.0, Nijenborgh 4, 9747 AG Groningen, The Netherlands, 2001
- Hubbard S J, Thornton J M. NACCESS Computer Program, Version 2.1.1, Dept. of Biochemistry and Molecular Biology. UK: University College London, 1993
- He H Q, Ma X H, Liu B, *et al.* High-throughput real-time assay based on molecular beacons for HIV-1 integrase 3' -processing reaction. *Acta Pharmacol Sin*, 2007, **28**(6): 811~817
- Brigo A, Lee K W, Mustata G I, *et al.* Comparison of multiple

- molecular dynamics trajectories calculated for the drug-resistant HIV-1 integrase T66I/M154I catalytic domain. *Biophys J*, 2005, **88** (5): 3072~3082
- 21 Brigo A, Lee K W, Fogolari F, *et al.* Comparative molecular dynamics simulations of HIV-1 integrase and the T66I/M154I mutant: binding modes and drug resistance to a diketo acid inhibitor. *Proteins: Struct Funct Bioinform*, 2005, **59**(4): 723~741
 - 22 Barreca M L, Lee K W, Chimirri A, *et al.* Molecular dynamics studies of the wild-type and double mutant HIV-1 integrase complexed with the 5CITEP inhibitor: mechanism for inhibition and drug resistance. *Biophys J*, 2003, **84**(3): 1450~1463
 - 23 Greenwald H, Le V, Butler S L, *et al.* The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry*, 1999, **38**(28): 8892~8898
 - 24 Engelman A, Craigie R. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*. *J Virol*, 1992, **66**(11): 6361~6369
 - 25 D'Arcy A, Stihle M, Kostrewa D, *et al.* Crystal engineering: a case study using the 24 kDa fragment of the DNA gyrase B subunit from *Escherichia coli*. *Acta Cryst D*, 1999, **55**(pt 9): 1623~1625
 - 26 Duan Y, Wang L, Kollman P A. The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation. *Proc Natl Acad Sci USA*, 1998, **95** (17): 9897~9902
 - 27 Kumar S, Wolfson H J, Nussinov R. Protein flexibility and electrostatic interactions. *IBM J Res & Dev*, 2001, **45**(3): 499~512
 - 28 Manoj N, Srinivas V R, Surolia A, *et al.* Carbohydrate specificity and salt bridge mediated conformational change in acidic winged bean agglutinin. *J Mol Biol*, 2000, **302**(5): 1129~1137

Activity, Solubility Comparison and Molecular Dynamics Simulation Analysis of Wild Type and F185K Mutant Type HIV-1 Integrase Catalytic Domain*

HE Hong-Qiu¹⁾, HU Jian-Ping²⁾, LIU Bin¹⁾, CHEN Wei-Zu¹⁾, WANG Cun-Xin^{1)**}

¹⁾ College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China;

²⁾ College of Chemistry and Life Science, Leshan Normal University, Leshan 614004, China)

Abstract Wild type (WT) and F185K mutant type of HIV-1 integrase catalytic domain (IN_C) were expressed and purified, and their solubility and activity were compared. The experiment results show that the solubility of F185K mutated IN_C was dramatically increased, whereas the activity was reduced to some extent. Subsequently, 1 800 ps molecular dynamics (MD) simulations for the WT and F185K type of IN_C in water were performed. The MD simulation results demonstrate that the flexibility of the catalytic loop region and the total mobility of F185K IN_C was reduced, which causes the decrease of activity. After the F185K mutation, changes of the salt bridge network drove the conformational change of IN_C, resulted in the burying of some hydrophobic residues and exposure of some other hydrophilic residues on the protein surface. Therefore, the relative hydrophilic solvent accessible surface of IN_C was increased. Moreover, the F185K mutation increased the hydrogen number between the IN_C protein and water molecule, as a consequence, the protein-water interaction was enhanced. These above changes contribute to the solubility increase of IN_C. It is found that the results obtained from MD simulation are in good agreement with the experiment data. The above mentioned results provides valuable insight for the understanding of protein solubility and will be helpful in protein engineering for increasing the solubility of proteins.

Key words HIV-1 integrase, activity, solubility, molecular dynamics simulation

DOI: 10.3724/SP.J.1206.2009.00126

*This work was supported by grants from The National Natural Science Foundation of China (30670497) and Beijing Natural Science Foundation (5072002).

**Corresponding author. Tel: 86-10-67392724, E-mail: cxwang@bjut.edu.cn

Received: March 5, 2009 Accepted: May 18, 2009