

SARS 冠 状 病 毒 主 要 蛋 白 酶 结 构 模 建 及 活 性 部 位 运 动 性 分 析

刘 士 勇* , 裴 剑 锋* , 陈 浩* , 祝 小 雷 , 刘 振 明 , 马 文 喆 , 贺 风 雷 , 来 鲁 华#

(北京 大 学 化 学 与 分 子 工 程 学 院 , 北 京 大 学 理 论 生 物 学 中 心 , 分 子 动 态 与 稳 态 国 家 重 点 实 验 室 , 北 京 100871)

[关 键 词] SARS 冠 状 病 毒 3CL 蛋 白 酶 ; 同 源 模 建 ; 二 聚 体 ; 构 象 变 化

[摘 要] SARS 冠 状 病 毒 3CL 蛋 白 酶 是 病 毒 复 制 过 程 中 的 关 键 酶 , 有 可 能 成 为 针 对 抗 SARS 药 物 研 制 的 重 要 靶 标 。 利 用 同 源 模 建 的 方 法 建 立 了 SARS 冠 状 病 毒 3CL 蛋 白 酶 的 结 构 模 型 , 对 于 二 聚 体 中 蛋 白 质 分 子 间 相 互 作 用 分 析 表 明 该 蛋 白 在 溶 液 中 有 可 能 形 成 二 聚 体 。 利 用 分 子 动 力 学 与 多 正 则 系 综 采 样 相 结 合 的 方 法 对 于 SARS 冠 状 病 毒 3CL 蛋 白 酶 活 性 部 位 的 两 个 环 区 进 行 了 构 象 模 拟 , 找 到 了 两 种 代 表 性 构 象 , 预 示 着 SARS 冠 状 病 毒 3CL 蛋 白 酶 在 催 化 过 程 中 有 可 能 伴 随 着 构 象 变 化 。

Modeling of the SARS coronavirus main proteinase and conformational flexibility of the active site

LIU Shiyong* , PEI Jianfeng* , CHEN Hao* , ZHU Xiaolei , LIU Zhenming , MA Wenzhe , HE Fenglei , LAI Luhua#

(State Key Laboratory of Structural Chemistry of Stable and Unstable Species , College of Chemistry and Molecular Engineering & Center for Theoretical Biology , Peking University , Beijing 100871 , China)

KEY WORDS SARS coronavirus 3CL proteinase ; Comparative modeling ; Homodimer ; Conformational flexibility

SUMMARY SARS coronavirus 3CL proteinase is the key enzyme for virus replication which may serve as the target for drug discovery against SARS. A 3D structure model has been built for SARS coronavirus 3CL proteinase by comparative protein modeling. A homodimer model of the proteinase was also built. Analysis of the dimeric interface suggests the 3CL proteinase may have dimer form in solution. The conformational flexibility of the active site has been simulated by molecular dynamics combined with multi-canonical sampling. The active site loops have two typical conformations which may be related to the conformational movement associated with the enzymatic reaction.

自从发现变种的冠状病毒是导致严重急性呼吸道综合征(SARS)的主要元凶以来^[1,2],目前已测定了18个不同来源SARS病毒的全基因组序列^[3-6]。对于这些基因组序列的比较分析发现SARS病毒基因组有一定变异,但变异速度不快。冠状病毒的主要蛋白酶或称3CL蛋白酶是病毒复制中的关键蛋白,其主要功能是水解病毒所表达的两个多聚蛋白^[7]。序列分析表明3CL蛋白酶在已测定的SARS基因组中是完全保守的,有可能是药物设计的关键靶标之一。我们对于SARS冠状病毒3CL蛋白酶的空间结构进行了同源模建,分析了其底物结合部位,研究了结合部位采取多种构象的可能性,并对于可能的二聚体界面进行了分析。

1 SARS 冠 状 病 毒 3CL 蛋 白 酶 序 列 分 析

蛋白质同源模建方法是进行蛋白质结构预测较为可靠的方法,所得到的结构模型可用于蛋白质结构与功能分析及药物设计等^[8]。我们利用Blastp程序^[9]针对SARS冠状病毒的基因组序列对于蛋白质晶体结构数据库PDB^[10]中全部蛋白质结构所对应的蛋白质序列数据进行了搜索,发现SARS冠状病毒3CL蛋白酶与猪传染性肠胃炎病毒(TGEV)3CL蛋白酶有显著的同源性,与PDB中其它蛋白只有较短片段的同源性。利用Clustal W程序^[11]对于所有冠状病毒的3CL蛋白酶进行多重序列比对,见图1。TGEV与SARS冠状病毒3CL蛋白酶的氨基酸序

基金项目 国家自然科学基金(20173001,90103029),教育部、科技部和北京大学资助。supported by NSFC(20173001,90103029), the Ministry of Education of China, the Ministry of Science and Technology of China, and Peking University.

* These authors contribute equally to this work

Corresponding author, email: lhlai@pku.edu.cn

列同源性为 43% ,序列相似性为 57% ,只在一个表面环区上有单点插入。由于 3CL 蛋白酶是半胱氨酸蛋白水解酶 ,活性部位的氨基酸较为保守。参照

TGEV 3CL 蛋白酶等的活性部位^[7] ,可以推测 SARS 冠状病毒 3CL 蛋白酶参与催化反应的残基为 Cys145 ,His41。

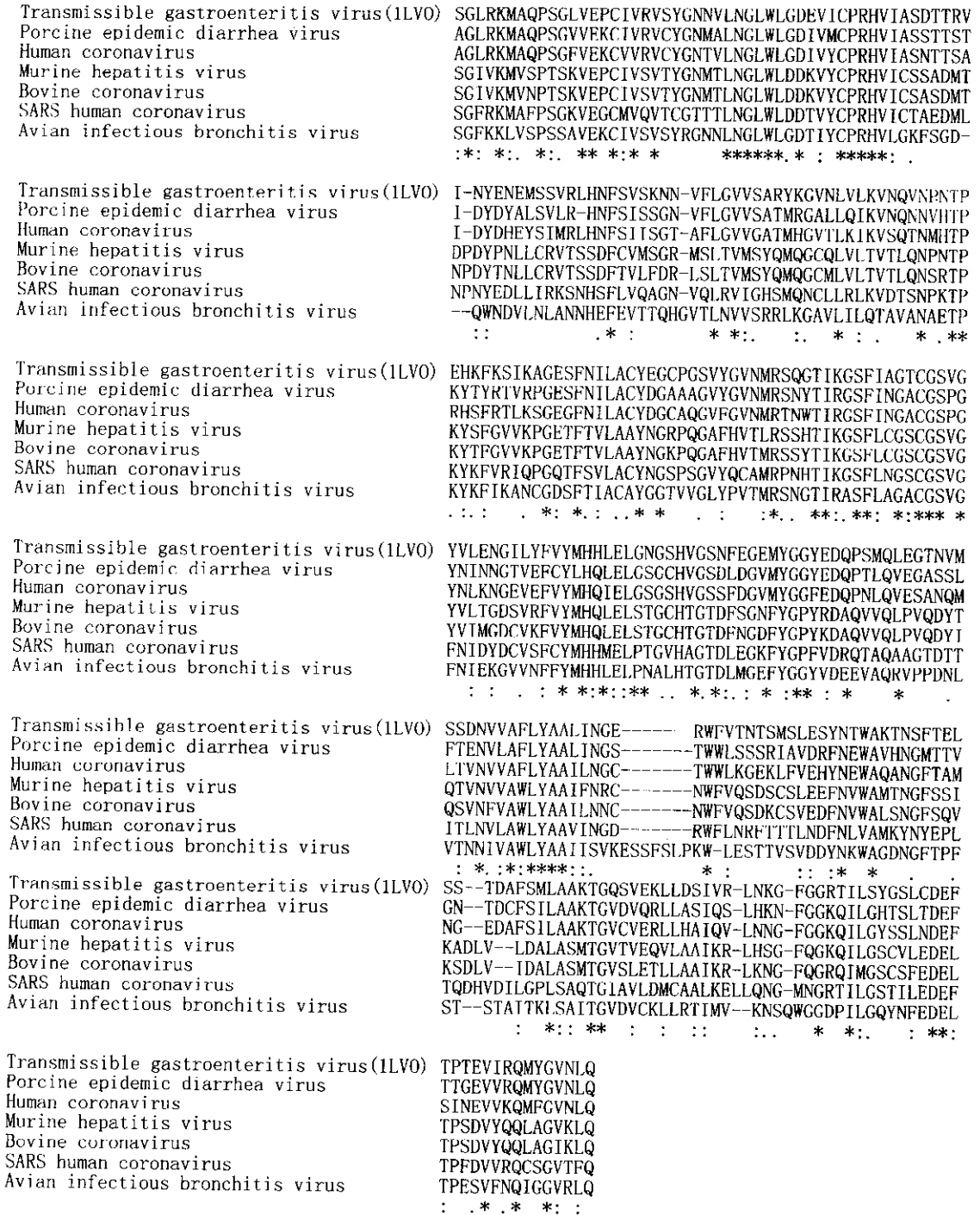


图 1 7 种冠状病毒 3CL 蛋白酶的序列比对。其中相同的残基以 * 表示 相似的残基以 . 及 : 表示

Figure 1 Sequence alignment of 3C-like proteinase from seven coronavirus. * represents conserved residues ,. and. represent similar residues

2 SARS 冠状病毒 3CL 蛋白酶同源结构建模及二聚体分析

我们以猪传染性肠胃炎病毒 3CL 蛋白酶的晶

体结构(PDB 代码 1LVO)中的 A 链为模板^[7] ,利用 Modeller 程序^[8,12]进行了 SARS 冠状病毒 3CL 蛋白酶同源建模 ,选择所给出的最优结构进行能量极小化(图 2) 。利用 Procheck 程序^[13]检查所得到结构

模型的合理性,发现 99.6% 的残基主链二面角都在允许区域内。TGEV 3CL 蛋白酶的单体结构由 3 个结构域构成,其中第 1、2 个结构域紧密堆积成糜蛋白酶折叠类型,第 3 个结构域由 5 个 α 螺旋构成,较为独立。由于 TGEV 3CL 蛋白酶在晶体结构以二聚体形式存在,且二聚体界面部份的氨基酸序列在冠状病毒 3CL 蛋白酶中具有保守性,因此 SARS 冠状病毒的 3CL 蛋白酶也有可能以二聚体形式存在。参照 TGEV 3CL 蛋白酶二聚体的结构,我们搭建了 SARS 冠状病毒 3CL 蛋白酶的二聚体结构(图 3)。二聚体界面所埋藏的面积 $1\,311\text{ \AA}^2$,其中疏水面积为 990 \AA^2 。从对于蛋白质同多聚体界面的统计分析来看,多数蛋白质同二聚体界面的埋藏面积在 368 \AA^2 到 $4\,746\text{ \AA}^2$ 之间,且随分子量的变化有线性增长^[15]。对于分子量 3 万左右的蛋白质分子,二聚体的埋藏面积在 1500 \AA^2 左右。稳定的蛋白质聚合体一般有较大的疏水埋藏面积。从二聚体的埋藏面积及疏水性来看,SARS 冠状病毒的 3CL 蛋白酶在水溶液中有可能存在二聚体形式,但其生物活性状态是单体还是二聚体形式还有待进一步实验验证。在我们完成模型构建后,德国科学家 2003 年 5 月 13 日在 Science 上发表的文章中对于 SARS 冠状病毒的 3CL 蛋白酶进行了同源建模,给出了其单体的结构模型^[15]。他们对于人冠状病毒 229E 3CL 蛋白酶在水溶液中的聚合状态研究表明该蛋白质在水溶液中以二聚体及单体混和物的状态存在,其中二聚体的含量大约为 60%。

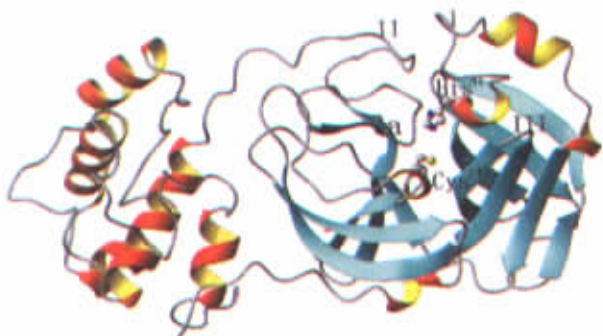


图 2 SARS 冠状病毒 3CL 蛋白酶结构模型

Figure 2 Structure model of the SARS coronavirus 3CL proteinase

3 SARS 冠状病毒 3CL 蛋白酶活性部位分析

SARS 冠状病毒 3CL 蛋白酶活性部位的催化残基为 Cys145, His41, 其附近的 β 折叠段 a 及环区 I、II、III 构成了底物结合部位(图 2)。酶与底物或抑制剂作用时常伴随构象变化。为了研究 SARS 冠状病毒 3CL 蛋白酶的生物活性构象,我们利用分子动

力学与多正则系综采样相结合的方法模拟了环区 I、II 的可能构象。传统的分子动力学模拟所得的构象能量分布服从 Boltzmann 分布,跨越能垒的能力比较弱,因而很难发生大的构象重排,对构象空间的采样能力也比较弱。Hansman^[16], Nakajima, Bartels^[17]和 Karplus 等^[18]将多正则方法引入了分子动力学模拟框架,大大提高了采样效率。多正则方法的主要思想在于使得模拟最终构象的势能分布为一定值。起始构象为同源建模结果经过能量极小化后的构象,模拟中只允许残基 45-53(I)和 184-193(II)这两段环区可动,其他部分原子坐标固定。力场采用隐式溶剂力场 ACS^[19],模拟温度为 1000 K。模拟时间步长为 1 fs,总模拟时间 10 ns,每 50 ps 输出能力和保存构象。能量采样范围为 $-16\,500\text{ kcal}\cdot\text{mol}^{-1} \sim -17\,500\text{ kcal}\cdot\text{mol}^{-1}$ 。计算使用 CHARMM(版本 28b1)完成。对于采样所得构象,我们对势能 $< -17\,500\text{ kcal}\cdot\text{mol}^{-1}$ 且为局部极小的构象共 29 520 个进行了聚类分析。当设定截断值为 1.5 \AA 时共得到 4 大类构象。图 4 为两个特征的代表构象。可以看出环区 I、II 有构象开合的现象,预示着在酶的反应过程中有可能伴随着构象变化。



图 3 SARS 冠状病毒 3CL 蛋白酶二聚体结构

Figure 3 The dimeric structure of the SARS coronavirus 3CL proteinase

4 结论

我们利用同源建模方法建立了 SARS 冠状病毒 3CL 蛋白酶的单体及二聚体三维结构模型,二聚体的研究表明二聚体界面有较大的埋藏面积及疏水性,因此该蛋白质在水溶液中有可能存在二聚体形式。针对构成底物结合部位的两个表面环区的多正则分子动力学模拟表明酶的活性部位附近在反应过程中有可能发生一定的构象变化。

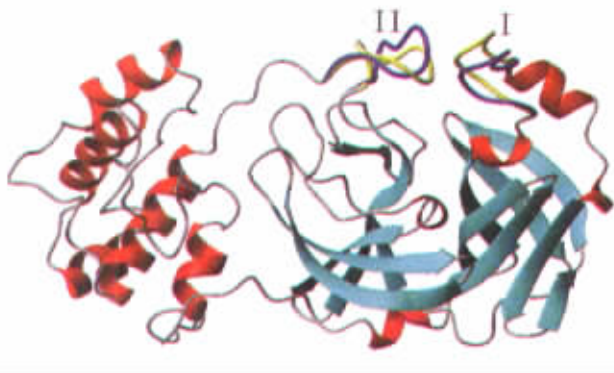


图 4 SARS 冠状病毒 3CL 蛋白酶环区 I、II 的两个代表性构象

Figure 4 Two typical conformations of the active site loops in the SARS coronavirus 3CL proteinase

参考文献

- 1 Peiris JSM, Lai ST, Poon LLM, *et al.* Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet*, 2003, 361 : 1319-1330
- 2 Ksiazek TG, Erdman DVM, Goldsmith PHC, *et al.* A novel coronavirus associated with severe acute respiratory syndrome, *N Engl J Med*, 2003 348 :1947-1958
- 3 <http://www.ncbi.nlm.nih.gov/genomes/SARS/SARS.html>
- 4 Mara MA, Jones SJM, Astell CR, *et al.* The genome sequence of the SARS-associated coronavirus. *Science*, 2003 (in press)
- 5 Rota PA, Oberste MS, Monroe SS, *et al.* Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, 2003 (in press)
- 6 Qin ED, Zhu QY, Man Yu, *et al.* A complete sequence and com-

- parative analysis of a SARS-associated virus (isolate BJ01). *Chi Sci Bull*, 2003, 48 :941-948
- 7 Anand K, Palm GJ, Mesters JR, *et al.* Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain. *EMBO J*, 2002, 21 :3213-3224
- 8 Sali A & Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 1993 234, 779-815
- 9 Altschul SF, Madden TL, Schaffer AA, *et al.* Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. *Nucleic Acids Res*, 1997, 25 :3389-3402
- 10 Berman HM, Westbrook J, Feng Z, *et al.* The Protein Data Bank. *Nucleic Acids Research*, 2000, 28 :235-242
- 11 Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 1994, 22 :4673-4680
- 12 Fiser A, Do RK, Sali A. Modeling of loops in protein structures. *Protein Science*, 2000, 9 :1753-1773.
- 13 Laskowski RA, MacArthur MW, Moss DS & Thornton JM, Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J App Cryst*, 1993, 26 :283-291
- 14 Jones S, Thornton JM, Principles of protein-protein interactions. *Proc Natl Acad Sci USA*, 1996, 93 :13-20
- 15 Anand K, Ziebuhr J, Wadhwani P, *et al.* Coronavirus main proteinase structure : basis for design of anti-SARS drugs. at <http://www.scienceexpress.org/2003-05-13> *Science*, 2003 (in press)
- 16 Hansmann UHE, Okamoto Y, Eisenmenger F. Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem Phys Letters*, 1996, 259 :321-330
- 17 Nakjima N, Nakamura H, Kidera A. Multicanonical ensemble generated by molecular dynamics simulations for enhanced conformational sampling of peptides. *J Phys Chem*, 1997, 101 :817-824
- 18 Bartels C, Karplus M. Multidimensional adaptive umbrella sampling : applications to main chain and side chain peptide conformations. *J Comput Chem*, 1997, 18 :1450-1462
- 19 Schaefer M, Karplus M. A Comprehensive Analytical Treatment of Continuum Electrostatics. *J Phys Chem*, 1996, 100 :1578-1599
(2003-05-22 收稿)
(本文编辑 周传敬)

· 消息 ·

北京大学 3 项 SARS 研究项目获国家自然科学基金专项资助

近日接到国家自然科学基金委员会通知,北京大学 3 项关于“非典型性肺炎(SARS)”基础研究的项目申请获得资助,总经费 85 万元。这 3 个项目的负责人分别是:生命科学学院的邓宏魁教授,北京大学第一医院的王广发教授和医学部基础医学院的王月丹。

这次防治“非典型性肺炎”基础研究项目是基金委适应形势需要、采取超常规方法紧急启动的专项项目,首期拨出专款 600 万元。从提出立项建议、发出申请指南,到受理申请、专家评审,最后确定资助项目,总共不到一个月的时间。

基金委希望能利用自身在资助基础科学研究方面较好的条件和基础,重点资助非典型肺炎有关病原体的生物学特性的研究,力求尽快在病原的传染途径、病原的变异规律、病原的致病机理、免疫的预防和免疫的机制等方面获得重大突破,为防治非典型肺炎工作做出贡献。

北京大学科研人员在这次非典攻坚战中,快速响应、积极应战。在基金委遴选资助的总共 22 个项目中,我校共获得 3 项。

(北京大学科学研究部 蔡晖)