



噬菌体phi29末端酶gp16蛋白C端结构域在溶液中的构象分析*

武飞飞 蔡汝洁 陈 婷 张云龙 陆昌瑞**

(东华大学化学化工与生物工程学院, 上海 201620)

摘要 噬菌体phi29是一种双链DNA病毒, 它的增殖和成熟过程需要将较大的子代DNA包装入一个空间极其有限的新生病毒衣壳中, 这一步骤由其转运马达完成. 转运马达具有三个部分: 头颈连接器蛋白 (the connector)、pRNA和ATP水解酶蛋白 (gp16). 在结构预测中发现gp16蛋白的C端结构域可能与pRNA或DNA结合, 但尚未有相关文献可以证明这种相互作用, 因此其结构的解析成为研究其功能的关键. 本文利用大肠杆菌SUMO表达系统, 对噬菌体phi29 gp16蛋白的C端结构域进行重组构建并诱导表达后, 通过Ni-NTA亲和层析纯化, 再用分子排阻色谱获得高纯度的目的蛋白质单体, 纯度达到95%左右. 最后用X射线小角散射技术 (small angle X-ray scattering) 对其构象进行分析, 收集小角散射数据, 当目的蛋白的浓度为1.3 g/L、所得到的目的蛋白的三维结构与同源蛋白FtsK的晶体结构高度拟合, 且通过CRY SOL软件反推计算的曲线与实验曲线高度重合. 通过原核表达、纯化及结构分析, 成功获得了该结构域在溶液中的状态信息. 这一研究不仅为后续关于该结构域的功能研究奠定基础, 也为病毒感染的诊断和治疗提供新思路.

关键词 噬菌体phi29, gp16蛋白, C端结构域, X射线小角散射

中图分类号 Q518.2

DOI: 10.16476/j.pibb.2018.0235

双链DNA病毒感染是目前临床治疗中的难点之一, 目前尚未找到有效抑制此类感染的药物. 在抗病毒药物研制过程中, 发现病毒特殊组分与宿主成分的差别是研发药物的关键. 因此, 如果能够找到病毒组装中的关键组分, 并以这些组分为靶点找到它们特异性的抑制剂, 就能为病毒感染的诊断和治疗提供巨大帮助.

双链DNA病毒的增殖和成熟是通过分子马达将子代DNA包装入空间极为有限的新生病毒衣壳中^[1-2]. 噬菌体phi29作为DNA包装马达研究的生物模型^[3-4], 借助分子马达将新合成的基因组DNA包装入二十面体的新衣壳中, 产生具有完整功能的病毒结构. 分子马达中的gp16蛋白 (39ku) 是原衣壳与DNA结合的媒介, 能够协助DNA包装入原衣壳. gp16蛋白由2个结构域组成: N端结构域有194个氨基酸, 包含保守的ASCE ATP酶域, 参与ATP水解; C端结构域有135个氨基酸, 在结构预测中发现具有寡核苷酸/寡糖的结合区域 (OB-fold), 可能参与核酸的包装折叠. 由于gp16蛋白全长不稳

定, 易形成多聚体, 所以很难得到晶体衍射结构. 目前, 只有gp16蛋白N端结构域被解析出来^[5], gp16蛋白C端结构域 (gp16-C) 尚未解析. 该结构域在与pRNA或DNA结合中的作用尚不明确.

gp16-C溶解度较低, 也易形成多聚体, 故难以纯化、极难获得分子构象. 由于SUMO标签蛋白属于泛素肽, 可以增加融合蛋白的可溶性, 所以本实验利用大肠杆菌SUMO克隆表达系统, 构建了gp16-C融合蛋白, 得到高纯度C端结构域的单体. 然后利用X射线小角散射的技术 (SAXS) 成功解析了gp16-C在溶液中的构象, 为后续关于该结构域的功能研究奠定基础, 并能对病毒感染的诊断和治疗提供一种新思路.

* 国家自然科学基金 (31300603), 中央高校基本科研基金 (17D310512, 2232014D3-39) 和励志计划 (18D210501) 资助项目.

** 通讯联系人.

Tel: 021-67792740, E-mail: crlu@dhu.edu.cn

收稿日期: 2018-08-28, 接受日期: 2018-11-26

1 材料与方法

1.1 材料

1.1.1 菌株、表达载体和基因

本实验所选用的菌株是大肠杆菌 (*E.coli*)，感受态细胞 DH5 α 、BL21 (DE3) 均由本实验室自制；使用改良之后含有 SUMO 标签的 pET28a 载体，即 ppSUMO 表达载体，酶切位点是 *BamH* I / *Xho* I；模板 cDNA 序列由生物工程 (上海) 股份有限公司合成。gp16 蛋白 C 端结构域的基因序列按照大肠杆菌密码子的偏好性进行了优化和改进，保留原有野生型蛋白质序列 (GI: 137944)，可以进行后续实验。

1.1.2 试剂与仪器

PCR 仪、核酸凝胶成像仪、蛋白质凝胶成像仪均购自 Bio-Rad 公司；FPLC 蛋白质纯化系统购自苏州赛谱仪器有限公司；Pfu DNA 聚合酶购自生工生物工程公司；限制性内切酶 *BamH* I 和 *Xho* I 购自 NEB 公司；SUMO 蛋白酶 (ULP1) 由本实验自制；亲和层析介质 Ni-NTA sefinose TM resin 购自生工生物工程公司；Superdex75 10/300 GL 预装柱购自美国 GE 公司；D2000 蛋白质分子质量标准购自中科瑞泰生物科技有限公司；质粒小量提取试剂盒购自生工。

1.2 实验方法与步骤

1.2.1 SUMO-gp16 蛋白 C 端结构域重组质粒的构建

噬菌体 gp16-C 基因经密码子优化后，可在大肠杆菌中高效表达。以该序列作为 PCR 模板，并使用引物 M1 5'-GCATCTCGAGTTACTGGATACG-3' 和引物 M2 5'-GCATGGATCCAGCGAACGTCGTAAAACCC-3'，在常规 PCR 条件下进行扩增。载体与基因片段经 *BamH* I 和 *Xho* I 双酶切后进行连接和转化，构建后的 SUMO-gp16-C 重组质粒经测序，将其测序结果与设计序列在 NCBI 的 Blast 中进行比对，比对结果完全正确。

1.2.2 SUMO-gp16 蛋白 C 端结构域重组质粒的表达鉴定

将测序正确的 SUMO-gp16-C 重组质粒转化至 *E.coli* BL21 (DE3) 感受态细胞中，经卡那霉素抗性筛选后，挑取阳性单克隆到 5 ml 的 LB 液体培养基中，在 37 $^{\circ}$ C、225 r/min 摇床过夜培养。然后将 5 ml 培养过夜的菌液接种至 1 L 的 LB 液体培养基中 (卡那抗性)，在 37 $^{\circ}$ C、225 r/min 进行培养至吸光度 (*A*) 值达到 0.5 左右。向其中补加卡那霉素，并加入异丙基- β -D-硫代吡喃半乳糖 (IPTG，终浓度 1 mmol/L)，在 18 $^{\circ}$ C、225 r/min 的摇床上诱导表达 12~16 h，留样后收集菌体。

1.2.3 SUMO-gp16 C 端融合蛋白的纯化

本研究所用缓冲液配方见表 1。将诱导培养后的菌液，在 4 $^{\circ}$ C、4 000 r/min 的条件下离心并收集细胞沉淀，加入预冷裂解缓冲液 (lysis buffer) 重悬，使菌液浓度为 20%，并向其中加入 DNase (1:1 000) 以及 MgCl₂ (终浓度 2.5 mmol/L) 和 PMSF (终浓度为 1 mmol/L)，混合均匀；然后用高压细胞破碎机破碎细胞，压力为 1 200 MPa，破菌 3 次。在 4 $^{\circ}$ C、12 000 r/min 离心 1 h，保留上清液。用 10 倍 Ni-NTA 柱体积的 lysis buffer 平衡镍柱，将上清液缓慢倒入柱子中，重复 2 次，使目的融合蛋白与介质充分结合。然后用含不同浓度咪唑的洗涤缓冲液 (wash buffer) 除去杂蛋白质；再用含不同浓度咪唑的洗脱缓冲液 (elute buffer) 洗脱融合蛋白，收集所有的样品 (所有操作在 4 $^{\circ}$ C 冷室进行)。将所有样品留样，用 12% SDS-PAGE 胶和蛋白质印迹法进行鉴定分析，选择 10 000 孔径的浓缩管 (浓缩管的孔径小于融合蛋白) 浓缩含有融合蛋白的洗脱液。将浓缩好的溶液通过咪唑缓冲液 (50 mmol/L Tris-HCl, pH 8.0、400 mmol/L NaCl、5% 甘油、1 mmol/L DTT) 进行缓冲液交换以除去咪唑，在 4 $^{\circ}$ C、4 000 r/min 的条件下离心浓缩融合蛋白至 500 μ l，准备酶切纯化。

Table 1 Buffers used in this study

Buffers	Details
Lysis Buffer	50 mmol/L Tris, pH 8.0, 500 mmol/L NaCl, 5% glycerol, 1.5 mmol/L DTT
Wash Buffer	50 mmol/L Tris, pH 8.0, 500 mmol/L NaCl, 5% glycerol, 1 mmol/L DTT, 25 mmol/L Imidazole
	50 mmol/L Tris, pH 8.0, 400 mmol/L NaCl, 5% glycerol, 1 mmol/L DTT, 50 mmol/L Imidazole
Elute Buffer	50 mmol/L Tris, pH 8.0, 400 mmol/L NaCl, 5% glycerol, 1 mmol/L DTT, 75 mmol/L Imidazole
	50 mmol/L Tris, pH 8.0, 400 mmol/L NaCl, 5% glycerol, 1 mmol/L DTT, 100 mmol/L Imidazole
	50 mmol/L Tris, pH 8.0, 400 mmol/L NaCl, 5% glycerol, 1 mmol/L DTT, 250 mmol/L Imidazole
Strip Buffer	50 mmol/L Tris, pH 8.0, 400 mmol/L NaCl, 5% glycerol, 1 mmol/L DTT, 500 mmol/L Imidazole

向浓缩好的 SUMO-gp16-C 融合蛋白中加入 SUMO 蛋白酶 (1:10) 以及无咪唑的缓冲液 (1:50), 4℃ 条件下酶切过夜. 将酶切后的样品进行 2 次过柱, 缓慢加入到用裂解缓冲液平衡后的 Ni-NTA 柱中, 酶切样品穿出亲和层析柱, 收集穿出液, 用 12% SDS-PAGE 检测酶切产物. 选择 3 000 孔径的浓缩管在 4℃、4 000 r/min 的条件下离心浓缩至 250 μ l. 测定目的蛋白质的浓度和纯度, 准备进行分子筛纯化.

选用 Superdex75 10/300GL 凝胶分子筛层析柱, 安装到层析系统上. 首先用纯水冲洗 1 个柱体积, 随后用 2 个柱体积的缓冲液 (50 mmol/L Tris, pH 8.0、400 mmol/L NaCl、1 mmol/L DTT、1 mmol/L EDTA) 平衡柱子. 在 0.4 ml/min 的流速, 0.3 MPa 的压力下上样. 收集 UV280 吸收峰处洗脱样品 (所有操作在 4℃ 冷室进行), 用 12% SDS-PAGE 胶检测, 然后将含目的蛋白质的样品混合后浓缩, 加入 5% 的甘油低温保存备用.

1.2.4 数据处理及结构的解析

SAXS 能提供生物大分子及其复合物的低分辨率结构信息且能解析溶液中生物大分子构象状态^[6-8]; 相对 X-ray 晶体衍射技术, SAXS 技术不受蛋白质的相对分子质量大小限制, 操作简捷, 但要求蛋白质样品具有较高的纯度和均一性. 实验主要在溶液中进行, 能够反映接近生理状态下蛋白质的真实形态^[9-11].

本实验利用上海光源 (五站六线) BL19U2^[12] 站线进行散射, 用毛细管为样品池, 通过泵将样品吸入毛细管中进行曝光, 样品台匹配的是 1.5 ml 的 EP 管和小的 PCR 管, 每轮曝光所需的样品量为 60 μ l, 相应的缓冲液需要多次曝光, 所以需要 1 ml 左右. 样品要离心保证没有气泡, 否则影响散射结果. 样品台的温度为 10℃, 为了得到更好的信噪比, 对同一样品连续收集 20 张图像, 通过平均降低噪声. 使用 RAW (16.0.0652.0621) 完成从 2D 图像到 1D 数据的转换. 得到一维散射曲线后, 需要对散射曲线进行强度归一化; 用样品的散射强度减去缓冲液的散射强度, 就得到样品的散射信号; 对散射信号进行 Guinier 分析, 计算分子的回转半径 R_g 以及角度为 0 处的散射强度 $I(0)$, 可以用来判断样品的质量. 然后使用 GNOM 计算分子内部距离分布函数 $P(r)$, 其中最重要的是调整 D_{max} 值, 过大会使尾部出现负值, 过小会使尾部出现陡峰.

得到空间的分布函数后, 利用 Dammin 进行

3D 建模, 来拟合 $P(r)$. 为了验证模型的正确性, 首先使用 MEGA 6.0 软件进行同源蛋白质的序列比对并对其进行进化树分析, 然后将同源蛋白质序列相似性较高的蛋白质的晶体结构拟合到 3D 模型中. 为了进一步验证模型的正确性, 利用 CRYSOLOG 软件将 3D 模型反推计算出 SAXS 曲线, 并分析与 SAXS 实验曲线的重现性.

2 实验结果

2.1 SUMO-gp16 C端重组质粒的构建和表达

为了使目的基因可溶性表达, 将 gp16-C 的基因构建到 ppSUMO 载体中. 经 *Bam*H I / *Xho* I 双酶切后的 ppSUMO 载体大约为 5 600 bp (图 1a), 目的基因片段大约为 405 bp (图 1b).

2.2 SUMO-gp16 C端融合蛋白的表达与纯化

融合蛋白含有 His-SUMO 蛋白标签, 可以通过 Ni-NTA 亲和层析进行纯化. 融合蛋白大小约为 34 ku (图 2a, b), 符合理论分子质量大小. SDS-PAGE 胶检测结果如图 2a 所示, 泳道 3 中几乎不含目标蛋白质, 表明融合目的蛋白质几乎全部与介质结合, 且通过不同咪唑浓度的洗涤缓冲液漂洗后, 除掉了大部分杂蛋白质 (泳道 4~5), 再用不同咪唑浓度的洗脱缓冲液将融合目的蛋白质洗脱下来 (6~12 泳道), 由图 2a 可知蛋白质的纯度较高. 通过蛋白质印迹法对融合目的蛋白质进行定性检测, 由图 2b 可以看出融合蛋白的表达量较高.

然后, 通过可特异性识别并切除 SUMO 蛋白标签的 ULP1 酶对融合蛋白进行酶切, 获得目的蛋白质. 酶切之后的目的蛋白质大小约为 16 ku (图 2c), 与理论值相符. 亲和纯化后的 SUMO-gp16 C 端融合蛋白在 4℃ 过夜酶切条件下仍保持稳定, 且酶切效率极高, 约达 95% 以上. 酶切之后对蛋白质进行 2 次过柱, 使目的蛋白质与 SUMO 标签分开; 与此同时, 第一次过柱时随 elute buffer 一同洗脱下来的杂蛋白质, 经过二次过柱后与亲和层析介质结合, 使得目的蛋白质得到进一步纯化.

为了得到高纯度的目的蛋白质, 采用 Superdex75 10/300GL 凝胶过滤层析柱对 gp16-C 进一步纯化. 目的蛋白质在出峰处被洗脱下来, 峰形单一且对称, 说明所获得的 gp16-C 纯度很高 (95% 以上), 具有良好的均一性, 符合 SAXS 对样品的要求; 目的蛋白质洗脱峰的洗脱体积对应的分子质量介于 14~17 ku 之间, 根据预装柱标准蛋白质洗脱体积, 目的蛋白质的大小与理论值相符合

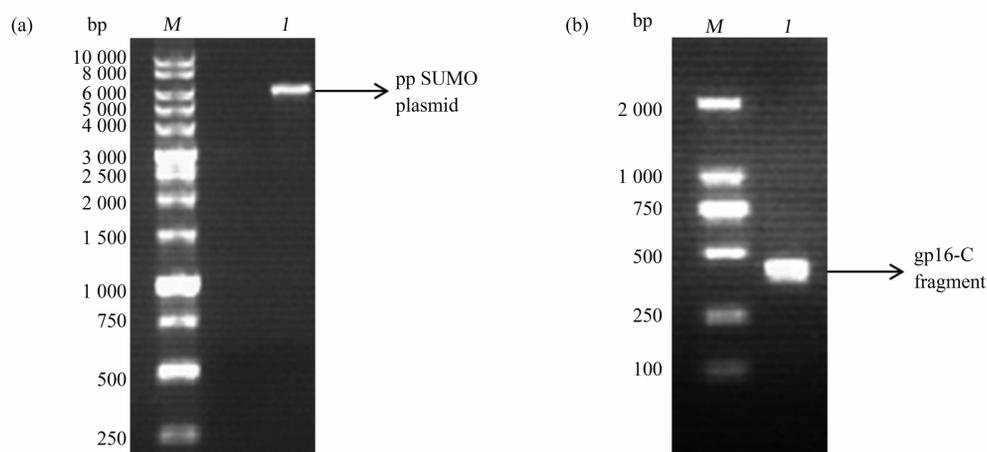


Fig. 1 Agarose gel electrophoresis of ppSUMO plasmid and gp16-C fragment prior to ligation

(a) *M*: DNA Marker 1 kb; *I*: ppSUMO plasmid after *Bam*H I/*Xho* I digestion. (b) *M*: DNA Marker D2000; *I*: gp16-C fragment.

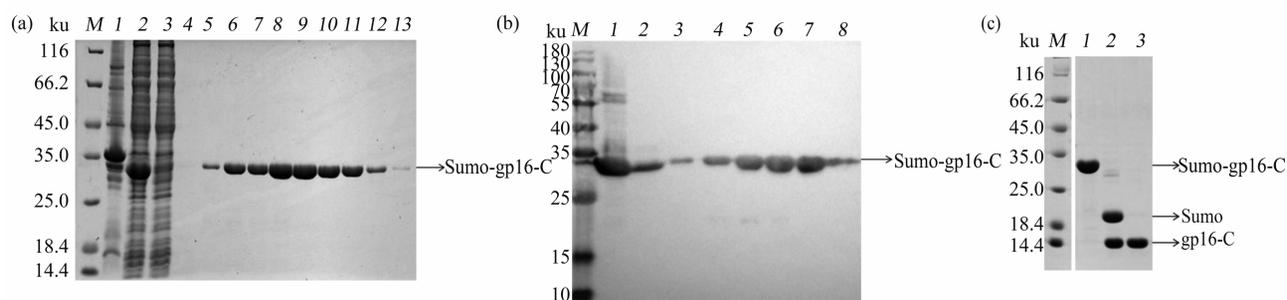


Fig. 2 SUMO-gp16 C terminal fusion protein purified

(a) SDS-PAGE analysis of SUMO-gp16 C terminal fusion protein purified by Ni affinity-chromatography. *M*: Protein marker 26610; *I*: Precipitate; 2: Supernatant; 3: Flow-through; 4: Wash (25 mmol/L imidazole); 5: Wash (50 mmol/L imidazole); 6: Eluted fraction (75 mmol/L imidazole); 7-9: Eluted fraction (100 mmol/L imidazole); 10-12: Eluted fraction (250 mmol/L imidazole); 13: Strip fraction (500 mmol/L imidazole); (b) Western blot analysis of SUMO-gp16 C terminal fusion protein purified by Ni-affinity chromatography. *M*: Protein marker; 1: Supernatant; 2: Flow-through; 3: Wash (25 mmol/L imidazole); 4: Wash (50 mmol/L imidazole); 5: Eluted fraction (75 mmol/L imidazole); 6: Eluted fraction (100 mmol/L imidazole); 7: Eluted fraction (250 mmol/L imidazole); 8: Strip fraction (500 mmol/L imidazole); (c) SDS-PAGE analyses of ULP1 protease digestion. *M*: Protein marker 26610; 1: Before digestion; 2: After digestion; 3: Ni-NTA resin flow-through.

(图3a)。最后用SDS-PAGE胶鉴定目的条带大小大约为16 ku (图3b), 说明gp16-C蛋白在生理条件下以稳定的单体结构存在。

2.3 gp16 C端的结构分析

首先对gp16-C的样品和相应缓冲液进行数据平均(图4a); 然后对数据进行扣减, 从扣除缓冲液后的散射结果可以看出是一条平缓的曲线(图4b), 结果较为理想; Guinier分析得到蛋白质的 R_g 值为19.32。由图4c可看出散射曲线和理论值可以

重合, 样品的质量较好; 进行Dammin分析并调整不同的 D_{max} 值, 结果显示实验值与理论值叠加后吻合度较高(图4d)。同时, 可以看出计算得到的 $P(r)$ 值在大 r 区域与横坐标结合处较平缓, 既没有出现负值, 尾部也没有出现陡峰, 这说明得到的数据比较可靠。

为验证实验结果的正确性, 对其进行了生物信息学分析。将ASCE^[5]家族的phi29 gp16、FtsK、T4 gp17、S6f、phi12 p4同源蛋白质序列进行了比

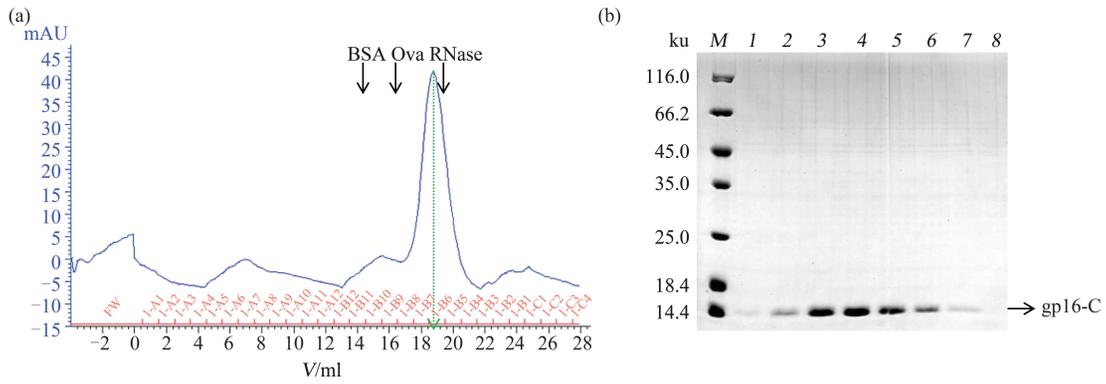


Fig. 3 gp16 C terminal gel filtration chromatography

(a) The size exclusion chromatogram of gp16 C terminal, Superdex 7510/300 GL, 0.4 ml/min, buffer (50 mmol/L Tris, pH 8.0, 400 mmol/L NaCl, 1 mmol/L DTT, 1 mmol/L EDTA). BSA 67 Ku, Ovalbumin 43 Ku, Ribonuclease 13.7 Ku. (b) SDS-PAGE analysis of the peak fractions. M: Protein marker 26610; 2~7: Collected fractions.

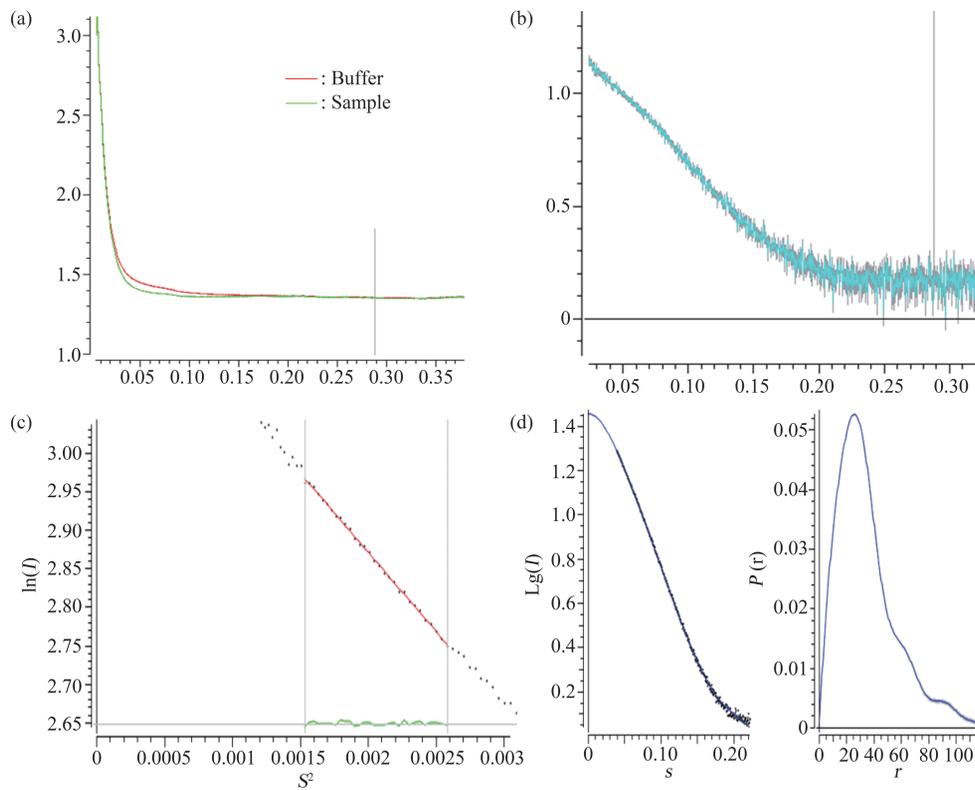


Fig. 4 SAXS data processing for gp16-C

(a) Averaged SAXS data of gp16-C solution. (b) The scattering results of gp16-C excluding buffer. (c) Guinier analysis of gp16-C. The analysis was conducted from point 85 to point 112 and the R_g value was calculated as 19.32. (d) GNOM analysis of gp16-C. The D_{max} value is adjusted to coincide the scattering curve with the theoretical curve, and then the $P(r)$ value was obtained from residual evaluation, system error, stability and other aspects.

对分析. 结果发现噬菌体phi29的gp16和FtsK蛋白序列相似度最高. 通过构建进化树 (图5), 推测

phi29的gp16蛋白和FtsK蛋白可能具有相似的结构.

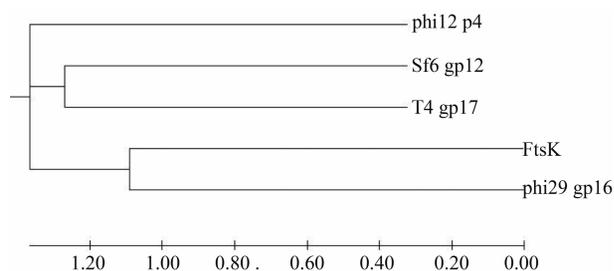


Fig. 5 Evolutionary tree of gp16 homologous

最后利用 Dammin 对 gp16-C 进行 3D 建模 (图 6a), SAXS 解析了 gp16-C 在溶液中的构象. 经生物信息学分析, 发现 FtsK^[13] 蛋白质序列 (409~544) 的晶体结构 (图 6b) 与 gp16-C 结构拟合度较高 (图 6c), 推测二者可能具有相似的功能, 这一结果为今后的研究提供了新的方向, 为 gp16-C 蛋白质功能的研究提供新思路.

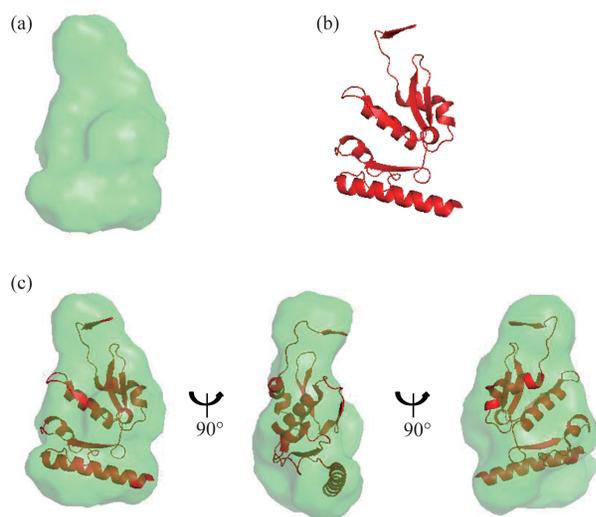


Fig. 6 gp16 structural model and fit by SAXS

(a) gp16 C terminal envelop calculated from SAXS; (b) Crystal structure of FtsK (PDB ID 2iuu); (c) Best overall fit of FtsK into gp16 C terminal domain SAXS envelop.

为了进一步验证模型的正确性, 用模型反推理论曲线, 然后与实验值作对比. 将 gp16-C 的 3D 模型利用 CRY SOL 软件, 反推计算出理论小角散射曲线. 可以看出 3D 模型反推出的曲线与实际实验得出的曲线高度拟合 (图 7), 计算出的方差为 1.8.

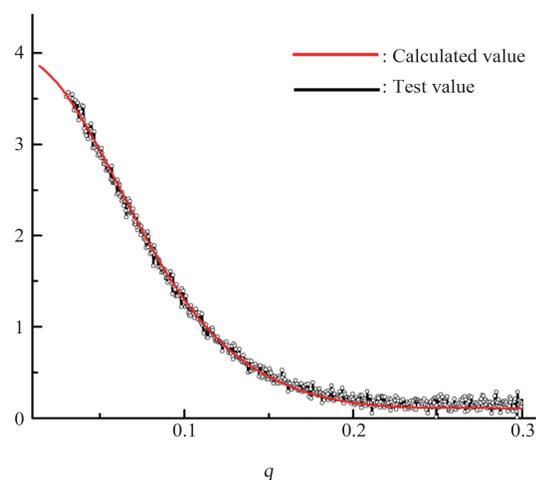


Fig. 7 The inverse curve of gp16 C terminal domain model

The red is the calculated CRY SOL curve; The black is the curve of the experiment.

3 讨 论

双链线性 DNA 病毒, 例如疱疹病毒、天花病毒^[14-15] 等严重危害人类健康, 而如何诊断和治疗病毒感染仍是当今生物医学界的难点之一. 病毒须将基因组包装到蛋白质外壳中之后才能成熟, 但目前对这种组装机制的研究还不透彻^[16]. 组装机制的研究将为人们在预防病毒感染及治疗方面提供重要依据及理论支持.

噬菌体 phi29 的包装机制存在特殊性. 通过对其机制的研究, 可以为研制靶向性的药物提供可能, 为抗病毒治疗开辟了新的研究途径. 噬菌体 phi29 包装分子马达中的 gp16-C 可能与 pRNA 或 DNA 结合, 在核酸包装中发挥重要作用. 因此, 本实验构建了 gp16-C 的融合蛋白, 增加了蛋白质的可溶性, 简化了后续的纯化工作. 首先通过 Ni-NTA 亲和层析柱进行纯化, 使用不同咪唑浓度的缓冲液, 提高了目的蛋白质的纯化效率. 经过 ULP1 酶切之后, 目的蛋白质仍然保持稳定, 且酶切之后的二次亲和层析使目的蛋白质进一步纯化. 然后通过分子筛层析对目的蛋白质进行纯化, 发现目的蛋白质以稳定的单体存在. 最后用 SAXS 技术结构进行初步解析, 采用浓度梯度法, 对不同浓度样品进行了 SAXS 实验, 并最终确定了当样品浓度

为1.3 g/L时, 所得SAXS数据最为理想. 为了验证SAXS结果的准确性, 对其进行了生物信息学分析, 在同源蛋白FtsK中找到了一段与gp16-C相似性较高的序列, 并发现它们的结构也高度拟合, 推测它们可能具有相似的功能. 紧接着将所得到的模型利用CRY SOL进行反推, 计算出来的曲线与实际实验的曲线高度拟合, 进一步验证了实验结果的准确性.

综上所述, 本课题旨在利用大肠杆菌SUMO表达系统对噬菌体phi29的gp16蛋白C端结构域进行表达和纯化, 并通过X射线小角散射技术获得了该结构域在溶液中的状态信息, 初步解析出gp16-C的三维结构, 为后续关于该结构域的功能研究提供帮助, 也为病毒感染的诊断和治疗提供新思路.

参 考 文 献

- [1] Aathavan K, Politzer A T, Kaplan A, *et al.* Substrate interactions and promiscuity in a viral DNA packaging motor. *Nature*, 2009, **461**(7264): 669-673
- [2] Earnshaw W C, Casjens S R. DNA packaging by the double-stranded DNA bacteriophages. *Cell*, 1980, **21**(2): 319-331
- [3] Cao S, Saha M, Zhao W, *et al.* Insights into the structure and assembly of the bacteriophage ϕ 29 double-stranded DNA packaging motor. *Journal of Virology*, 2015, **88**(8): 3986-3996
- [4] Grimes S, Jardine P J, Anderson D. Bacteriophage phi 29 DNA packaging. *Advances in Virus Research*, 2002, **58**(4): 255-294
- [5] Mao H, Saha M, Reyes-Aldrete E, *et al.* Structural and molecular basis for coordination in a viral DNA packaging motor. *Cell Reports*, 2016, **14**(8): 2017-2029
- [6] Chen Y, Pollack L. SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdisciplinary Reviews RNA*, 2016, **7**(4): 512-526
- [7] Skou S, Gillilan R E, Ando N. Synchrotron-based small-angle X-ray scattering (SAXS) of proteins in solution. *Nature Protocols*, 2014, **9**(7): 1727-1739
- [8] Lipfert J, Doniach S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu Rev Biophys Biomol Struct*, 2007, **36**(1): 307-327
- [9] Ke A, Doudna J A. Crystallization of RNA and RNA - protein complexes. *Methods*, 2004, **34**(3): 408-414
- [10] Sonda P, Yeckel A, Daoutidis P, *et al.* Complex dynamics within the vertical. *Bridgman Crystal Growth Process*. 2001, **34**(25): 517-522
- [11] Rambo R P, Tainer J A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, 2013, **496**(7446): 477-481
- [12] Li N, Li X, Wang Y, *et al.* The new NCPSS BL19U2 beamline at the SSRF for small-angle X-ray scattering from biological macromolecules in solution1. *Journal of Applied Crystallography*, 2016, **49**(Pt5): 1428-1432
- [13] Massey T H, Mercogliano C P, Yates J, *et al.* Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. *Molecular Cell*, 2006, **23**(4): 457-469
- [14] Jang J, Lee Y J, Kim J S, *et al.* Epidemiological correlation between fecal adenovirus subgroups and pediatric intussusception in Korea. *Journal of Korean Medical Science*, 2017, **32**(10): 1647-1656
- [15] Olson V A, Shchelkunov S N. Are we prepared in case of a possible smallpox-like disease emergence? *Viruses*, 2017, **9**(9): 242
- [16] Chemla Y, Aathavan K, Michaelis J, *et al.* Mechanism of force generation of a viral DNA packaging motor. *Cell*, 2005, **122**(5): 683-692

Solution Analysis of Phage phi29 gp16 Protein C terminal Domain*

WU Fei-Fei, CAI Ru-Jie, CHEN Ting, ZHANG Yun-Long, LU Chang-Rui**

(College of Chemistry, Chemical Engineering and Biotechnology, Donghua University, Shanghai 201620, China)

Abstract Phage phi29, a double-stranded DNA virus, assembles its genome into its protein capsid to near-crystalline density by a highly efficient molecular motor. The motor contains the connector, prohead RNA and ATP hydrolytic enzyme protein gp16. Currently, no structure-function information is available regarding the C-terminal domain of gp16. Our research aims to understand its role and interaction with pRNA or DNA. This study provides a method for recombinant gp16 protein synthesis and purification by the *E. coli* SUMO expression system, and here we report the structural envelope of the C terminal domain of gp16 obtained through small angle X-ray scattering (SAXS) technology. The gp16 C terminal domain gene was recombinantly expressed using the SUMO tag. The gp16 gene was flanked by an N-terminal 6His-SUMO tag and purified by Ni-NTA affinity chromatography. High purity target protein monomer was obtained by size exclusion chromatography. This brings the purity of the target protein to about 95%. Then the concentration of the target protein to 1.3 g/L, and the SAXS data were collected at Shanghai Synchrotron Radiation Facility (SSRF) beamline BL19U. The resulting molecular envelop highly resembles the homologous protein FtsK, and the curves calculated by the CRY SOL software are highly coincident with the experimental curves. Through recombinant expression, purification and structure analysis, we obtained the solution structure of the gp16. Structural information derived from this study laid the foundation for future structural and functional research of gp16 C terminal, ultimately targeting the viral mechanisms for infection.

Key words phage phi29, gp16 protein, C terminal, small angle X-ray scattering

DOI: 10.16476/j.pibb.2018.0235

* This work was supported by grants from The National Natural Science Foundation of China (31300603), Fundamental Research Funds for the Central Universities (17D310512, 2232014D3-39) and Motivational plan(18D210501).

** Corresponding author.

Tel: 021-67792740, E-mail: crlu@dhu.edu.cn

Received: August 28, 2018 Accepted: November 26, 2018