

[文章编号] 1671-587X(2024)03-0739-10

DOI:10.13481/j.1671-587X.20240318

基于幽门螺杆菌 hp0169 基因三维结构的生物信息学分析

林玲辉¹, 李娜¹, 尹晓燕¹, 王晓凌¹, 胡亚平¹, 刘威², 费瑞³, 田新利¹

(1. 邢台医学高等专科学校病原生物学与免疫学教研室, 河北 邢台 054000; 2. 陆军军医大学基础医学院全军免疫学研究所, 重庆 400038; 3. 吉林大学基础医学院细胞生物学系, 吉林 长春 130021)

[摘要] **目的:** 克隆幽门螺杆菌 (*Hp*) hp0169 基因并进行晶体学研究, 分析其二级结构和三维结构。**方法:** 由 UniProt 数据库中检索 *Hp* NCTC26695 菌株 hp0169 基因及其编码蛋白序列, 采用生物信息学方法分析 *Hp* 重组胶原蛋白酶 (HpPrtC) 蛋白理化性质, SOPMA 和 DNASTAR 软件预测 HpPrtC 蛋白二级结构特征, SWISS-MOPEL 软件构建 HpPrtC 蛋白三维结构, IEDB 和 ABCpred 软件预测 HpPrtC 蛋白 B 淋巴细胞抗原表位, SYFPEITHI 网站预测 T 淋巴细胞抗原表位, 专家库 (EP) 算法和随机森林 (RF) 算法预测 HpPrtC 蛋白可结晶性。原核表达 HpPrtC 重组蛋白, 经 Ni²⁺ 亲和层析和分子筛技术纯化蛋白, 结晶试剂盒筛选 HpPrtC 的结晶条件。**结果:** hp0169 基因共包含 1 269 个碱基配对, 编码蛋白全长 422 个氨基酸, 理论等电点为 7.64, 相对分子质量为 47 300。HpPrtC 蛋白为亲水性、可溶性蛋白。HpPrtC 蛋白 α 螺旋的氨基酸数量占全部氨基酸数量百分率为 35.78%, β 片层为 18.72%, β 转角为 6.87%, 无规则卷曲为 38.63%。抗原表位分析, HpPrtC 蛋白含有 B 淋巴细胞的 5 个优势线性表位和 3 个构象表位及多个 T 淋巴细胞潜在优势抗原表位。同源建模, HpPrtC 蛋白呈二聚体, 单体由 β 折叠围成桶状结构, 周围被 α 螺旋和无规则卷曲围绕。HpPrtC 蛋白为中等难度结晶, 且无信号肽和跨膜螺旋, 在 0.2 mol·L⁻¹ 氯化镁、0.1 mol·L⁻¹ 三羟甲基氨基甲烷 (Tris)、3.4 mol·L⁻¹ 乙二醇和 pH 8.5 条件下出现细小成簇的针状晶体。**结论:** HpPrtC 是一种亲水型蛋白, 呈二聚体构造, 在适宜条件下呈细小成簇针状晶体, 其具有 T 淋巴细胞和 B 淋巴细胞优势抗原表位, 可作为 *Hp* 疫苗设计的抗原, 用于构建多价融合疫苗或多表位疫苗, 本研究结果为 *Hp* 的防治提供了实验基础。

[关键词] 幽门螺杆菌; hp0169 基因; 抗原表位; 蛋白结晶; 生物信息学

[中图分类号] R378.99 **[文献标志码]** A

Bioinformatics analysis based on three-dimensional structure of *Helicobacter pylori* hp0169 gene

LIN Linghui¹, LI Na¹, YIN Xiaoyan¹, WANG Xiaoling¹, HU Yaping¹, LIU Wei², FEI Rui³, TIAN Xinli¹

(1. Department of Pathogenic Biology and Immunology, Xingtai Medical College, Xingtai 054000, China;

2. Institute of Immunology, School of Basic Medical Sciences, Army Medical University, Chongqing

400038, China; 3. Department of Cell Biology, School of Basic Medical Sciences, Jilin University,

Changchun 130021, China)

ABSTRACT **Objective:** To clone the *Helicobacter pylori* (*Hp*) hp0169 gene and conduct the

[收稿日期] 2023-06-02

[基金项目] 河北省卫健委医学科学研究项目 (20201591)

[作者简介] 林玲辉 (1984—), 男, 河北省邢台市人, 副教授, 医学硕士, 主要从事原核细胞结构生物学方面的研究。

[通信作者] 田新利, 教授 (E-mail: xtxl66@163.com);

费瑞, 教授, 硕士研究生导师 (E-mail: feirui@jlu.edu.cn)

crystallographic study, and to clarify its secondary and tertiary structures. **Methods:** The hp0169 gene and its encoded protein sequence of the *Hp* NCTC26695 strain were retrieved from the UniProt database. Bioinformatics method was used to analyze the physicochemical properties of the *Hp* recombinant protease (HpPrtC) protein; SOPMA and DNASTar softwares were used to predict the secondary structure characteristics of HpPrtC protein; SWISS-MODEL software was used to construct the tertiary structure of the HpPrtC protein; IEDB and ABCpred softwares were used to predict the antigenic epitopes of the B lymphocytes HpPrtC protein; SYFPEITMI website was used to predict the antigenic epitopes of the T lymphocytes of HpPrtC protein; the expert pool (EP) and random forest (RF) algorithms were used to predict the crystallizability of the HpPrtC protein; the HpPrtC recombinant protein was expressed in the prokaryotic system; the HpPrtC recombinant protein was purified by Ni²⁺ affinity chromatography and size-exclusion chromatography; the crystallization conditions for HpPrtC were screened by crystallization kit. **Results:** The hp0169 gene contained 1 269 base pairs and encoded the protein of 422 amino acids, the theoretical isoelectric point was 7.64 and the relative molecular weight was 47 300. HpPrtC was a hydrophilic and soluble protein. The number of amino acids of alpha helices of HpPrtC accounted for 35.78%, beta sheets 18.72%, beta turns 6.87%, and random coils 38.63%. The antigen epitope analysis results showed that HpPrtC contained five dominant linear epitopes of B lymphocytes, three conformational epitopes, and multiple potential dominant epitopes of T lymphocytes. The homology modeling results showed that HpPrtC formed a dimer, and each monomer displayed a barrel structure surrounded by β sheets, alpha helices, and random coils. HpPrtC was predicted to have moderate crystallizability without signal peptides and transmembrane helices. Small clustered needle-like crystals of HpPrtC were obtained under the conditions of 0.2 mol·L⁻¹ magnesium chloride, 0.1 mol·L⁻¹ tris (hydroxymethyl) amino methane (Tris), 3.4 mol·L⁻¹ hexanediol, and pH=8.5. **Conclusion:** HpPrtC is a hydrophilic protein that forms a dimeric structure and crystallizes into small clustered needle-like crystals under suitable conditions. HpPrtC contains dominant antigenic epitopes of the T lymphocytes and B lymphocytes and can serve as an antigen for the design of *Hp* vaccines to establish the multivalent fusion vaccines or multi-epitope vaccines; the results provide an experimental basis for the prevention and control of *Hp*.

KEYWORDS *Helicobacter pylori*; hp0169 gene; Crystallization; Antigenic epitope; Bioinformatics

幽门螺杆菌 (*Helicobacter pylori*, *Hp*) 为革兰阴性致病菌, 定植于人类胃黏膜上皮细胞, 已被世界卫生组织列为 I 类致癌因子^[1-2]。目前 *Hp* 感染问题已成为世界性难题, 我国 *Hp* 平均感染率已高达 59%^[3]。其中, 胃窦炎、十二指肠溃疡、胃溃疡和胃腺癌等被证实与 *Hp* 感染存在密切关联。研究^[4-5]显示: 部分毒力基因在 *Hp* 发病机制中发挥重要作用, 其中 *Hp* 胶原酶基因 hp0169 的表达和分泌被证实是细菌在体内定植的先决条件, hp0169 基因编码的 *Hp* 重组胶原蛋白酶 (*Hp* recombinant collagenase, HpPrtC) 为胶原酶, 能够降解细胞外基质中的天然三股螺旋 I 型胶原蛋白并促进细菌黏附^[6]。研究^[7]显示: hp0169 基因被敲除, *Hp* 则无法在沙鼠胃中定植。hp0169 基因在 *Hp* 的定植黏附中起重要作用, 但其致病机制和酶解机制尚不清楚。赵慧琳等^[8]通过构建的 HpPrtC 蛋白原核表达

载体 (pET-22b-HpPrtC) 对 HpPrtC 蛋白进行了初步表达和纯化, 并检测其降解热变性 I 型胶原蛋白和天然 I 型胶原蛋白的活性。但关于 HpPrtC 蛋白结晶的筛选条件、二级结构和免疫原性尚未见相关报道, 且对于 HpPrtC 的跨膜区、信号肽和亲水性等问题尚未完全阐明。本研究对本课题组前期构建的质粒 pET28a-hp0169^[9] 进一步表达和纯化, 以筛选 HpPrtC 的结晶条件。同时, 采用生物信息学方法对 HpPrtC 蛋白进行分析, 以解析其三维结构。可从分子水平和蛋白质空间结构为表位疫苗的研发提供参考, 并根据其三维结构设计黏附拮抗剂从而抑制 *Hp* 定植黏附, 为进一步研究 *Hp* 的防治提供理论依据。

1 材料与方法

1.1 菌株、载体、主要试剂和仪器 *Hp* 标准株购自

英国典型微生物保藏中心 (National Collection of Type Cultures, NCTC), 编号: NCTC 26695。质粒载体 pET-28a、大肠杆菌 DH5 α 化学感受态细胞 (DH5 α super competent cells, *E. coli* DH5 α) 和大肠杆菌 BL21 (DE3) 感受态细胞均由陆军军医大学馈赠。限制性内切酶 *Xho* I、限制性内切酶 *Nco* I 和质粒 DNA 提取试剂盒均购自大连 TaKaRa 公司, 结晶试剂盒购自美国 Hampton Research 公司。ATS 高压均质机购自意大利 ATS 公司, 氨酸标签蛋白层析柱 (HisTrap™ Fast Flow Crude, HisTrap) 和 ÄKTA pure 自动蛋白纯化仪购自美国 GE Healthcare 公司, Gryphon-LCP 蛋白结晶自动化工作站购自美国 Art Robbins Instruments 公司。

1.2 HpPrtC 蛋白理化性质 分别采用美国国家生物技术信息中心 (National Center for Biotechnology Information, NCBI) 网站 (<http://www.ncbi.nlm.nih.gov/gene>) 和蛋白质数据库 UniProt 数据库 (<http://www.uniprot.org/>) 检索 HpPrtC 基因序列及蛋白质信息。采用 ProtParam 软件 (<http://web.expasy.org/protparam/>) 分析 HpPrtC 蛋白的等电点、相对分子质量和功能区等理化性质。

1.3 HpPrtC 蛋白二级结构特征预测 于 TMHMM 网站 (<http://www.cbs.dtu.dk/services/TMHMM>) 预测 HpPrtC 蛋白质膜区。采用 SOPMA^[10] (https://npsa-pbil.ibcp.fr/cgi-bin/secpred_sopma.pl) 和 DNASTar^[11] 软件预测 HpPrtC 蛋白 α 螺旋、 β 折叠、 β 转角和无规则卷曲等二级结构特征。

1.4 HpPrtC 蛋白三维结构构建和抗原表位预测

将 HpPrtC 蛋白氨基酸序列输入 SWISS-MODEL 在线软件^[12] (<https://swissmodel.expasy.org/>), 以来源于甲烷嗜热菌的 U32 家族蛋白酶 MK0906^[13] 为模板进行同源建模, 构建 HpPrtC 蛋白的三维结构。采用 IEDB 软件^[14] (<http://tools.immuneepitope.org/main/>) 和 ABCpred 软件^[15] (https://webs.iitd.edu.in/raghava/abcpred/ABC_method.html) 分析 HpPrtC 蛋白 B 淋巴细胞抗原表位, 采用 SYFPEITHI 网站^[16] (<http://www.syfpeithi.de/bin/mhcserv.dll/epitopeprediction>) 预测 HpPrtC 蛋白 T 淋巴细胞表位。选择等位基因 HLA-A*0201 预测细胞毒性 T 淋巴细胞 (cytotoxic T lymphocyte, CTL) 抗原表位, 选用长度为 8 和 9 个氨基酸, 阈值设定为 22; 以 HLA-DRB1*8.0401 预测 T 淋巴细

胞抗原表位, 选用长度为 15 个氨基酸, 阈值设定为 21。

1.5 HpPrtC 蛋白可结晶性预测 将 HpPrtC 蛋白的氨基酸序列输入 XtalPred 预测软件^[17] (<http://ffas.burnham.org/XtalPred-cgi/xtal.Pl>) 中, 采用专家库 (expert pool, EP) 算法和随机森林 (random forest, RF) 算法进行等级分类, 2 种算法等级分类分数越高, HpPrtC 蛋白可结晶性越小。

1.6 HpPrtC 蛋白表达和提纯 本课题组前期构建了重组质粒 pET-28a-hp0169, 并将重组质粒转化至 *E. coli* BL21 (DE3)^[9]。经异丙基硫代半乳糖苷 (isopropyl β -D-Thiogalactoside, IPTG) 诱导, 表达全基因组蛋白, 离心, 收集菌体, 重悬, 经 ATS 高压均质机破碎, 高速离心, 收集上清液, 经 HisTrap 亲和层析柱和 HiLoad 10/600 Superdex 200 凝胶过滤层析柱提纯, 将表达和纯化的蛋白通过 SDS-聚丙烯酰胺凝胶电泳 (polyacrylamide gel electrophoresis, PAGE), 并根据相对分子质量大小筛选目的蛋白, 并分析蛋白纯度, 采用 Western blotting 法鉴定目的蛋白。

1.7 HpPrtC 蛋白结晶条件筛选 采用考马斯亮蓝法测定蛋白浓度, 将纯化的 HpPrtC 蛋白浓缩至 12 g·L⁻¹, 采用 Hampton 蛋白质结晶试剂盒筛选 HpPrtC 蛋白结晶条件。

2 结果

2.1 HpPrtC 蛋白理化性质 在 NCBI-GenBank 数据库中检索 hp0169 基因, 共包含 1 269 个碱基配对, 编码蛋白全长 422 个氨基酸, 理论等电点为 7.64, 相对分子质量为 47 300。UniProt 数据库检索 HpPrtC 蛋白编号为 P56113.1。

2.2 HpPrtC 蛋白二级结构特征 TMHMM 软件预测 HpPrtC 蛋白跨膜区, 结果显示: HpPrtC 蛋白无紫色区域且不含跨膜螺旋, 表明该蛋白为亲水性、可溶性蛋白。见图 1。SOPMA 软件预测 HpPrtC 蛋白二级结构特征, 结果显示: HpPrtC 蛋白二级结构特征中 α 螺旋的氨基酸数量占全部氨基酸数量百分率为 35.78%, β 片层为 18.72%, β 转角为 6.87%, 无规则卷曲为 38.63%。HpPrtC 蛋白二级结构特征中 β 片层和无规则卷曲所占百分率越高, 其作为抗原表位概率越大。见图 2。DNASTar 软件解析结果显示: HpPrtC 蛋白是亲水性蛋白, 可以产生良好的抗原性。见图 3。

2.3 HpPrtC 蛋白三维结构 HpPrtC 蛋白三维结

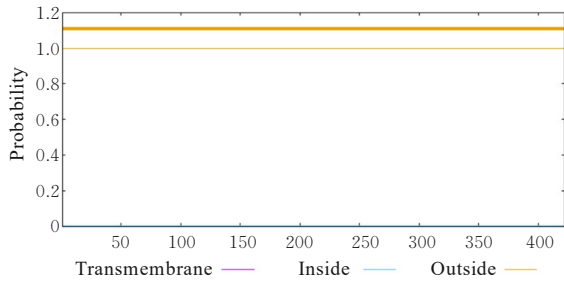
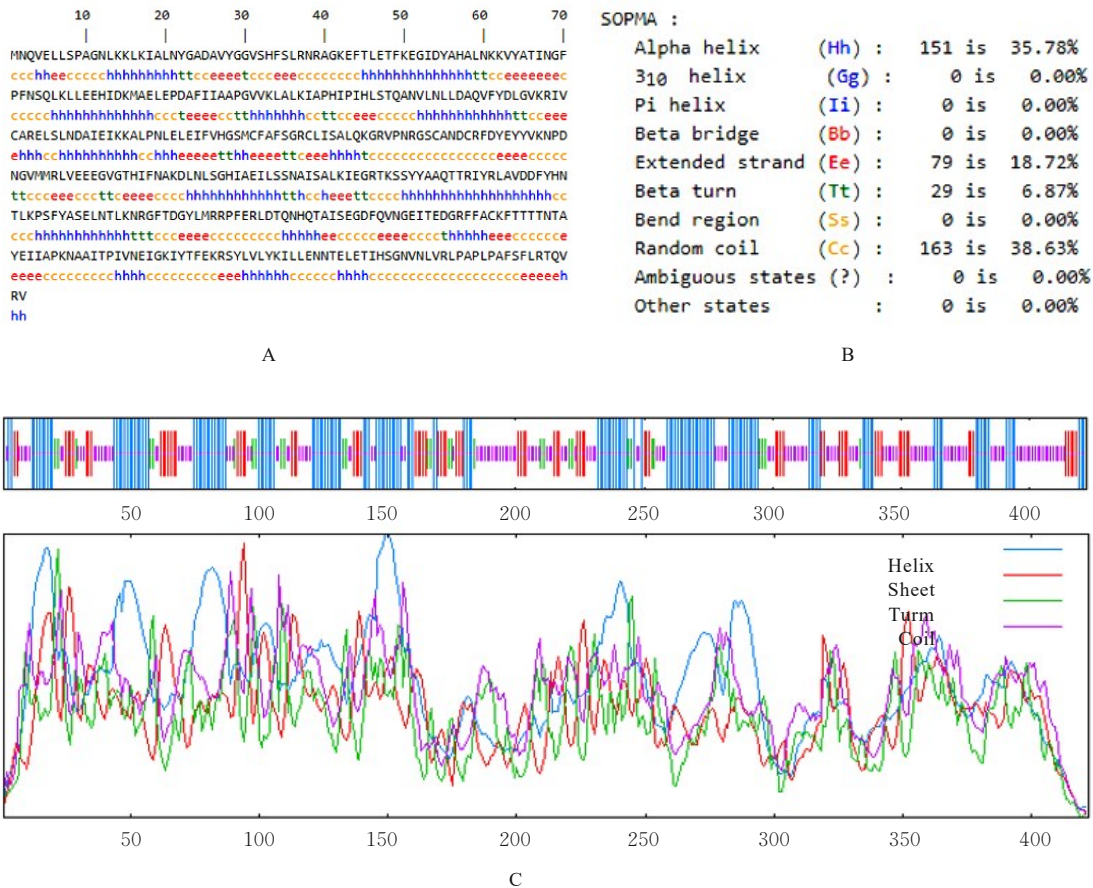


图1 TMHMM软件预测HpPrtC蛋白跨膜螺旋
Fig.1 Transmembrane helix of HpPrtC protein predicted by TMHMM software

构与甲烷嗜热菌的U32家族蛋白酶MK0906序列一致性为27%，覆盖率为49%，呈二聚体构造(图4A)，呈7个同向平行β片层形成的桶状结构，周围有6个α螺旋和多个无规则卷曲(图4B)。

2.4 HpPrtC蛋白B淋巴细胞抗原表位 长度为

16个氨基酸且阈值>0.55的B淋巴细胞相关抗原表位共40个，其中阈值≥0.85的优势抗原表位12个。见表1。HpPrtC蛋白的抗原表位分别为：线性表位、β-转角、表面可及性、骨架区柔韧性、抗原指数和亲水性等。见图5。结合ABCpred软件分析结果，排除α-螺旋、β-折叠和疏水区等不易形成表位的区域，确定5个B淋巴细胞优势表位，分别为217~232位(LVEEEGVGTHIFNAKD)、236~251位(SGHIAEILSSNAISAL)、252~267位(KIEGRTKSSYYAAQTT)、262~277位(YAAQTTRIYRLAVDDF)和344~359位(TTTTNTAYEIIAPKNA)。IEDB软件ElliPro^[18]预测不连续B淋巴细胞表位，阈值范围为0.523~0.835，共9个不连续的B淋巴细胞表位，构象表位的大小为3~87个氨基酸残基。见表2。以上结果均表明HpPrtC蛋白具有较好的免疫原性。



A: Amino acid sequence of HpPrtC protein; B: Proportion of secondary structure components of HpPrtC protein; C: Secondary structure map of HpPrtC protein.

图2 SOPMA软件预测HpPrtC蛋白二级结构特征图

Fig.2 Feature maps of secondary structure of HpPrtC protein predicted by SOPMA software

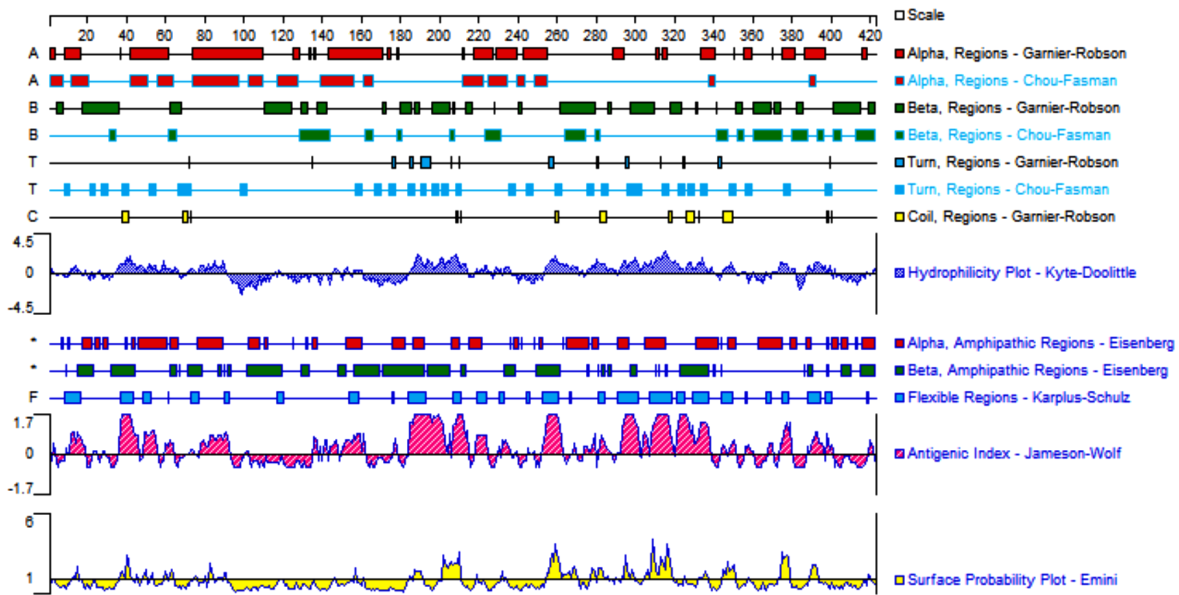


图3 DNASTAR 软件预测 HpPrtC 蛋白二级结构特征和抗原表位

Fig. 3 Secondary structure features and antigenic epitopes of HpPrtC protein predicted by DNASTAR software

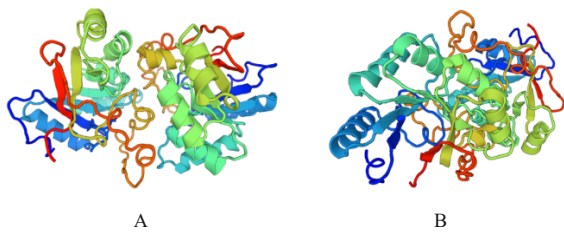
A: Dimer structure of HpPrtC protein; B: β -barrel of HpPrtC protein.

图4 HpPrtC 蛋白氨基酸序列三维结构

Fig. 4 Three-dimensional structures of amino acid sequences of HpPrtC protein

表1 ABCpred 软件预测 HpPrtC 蛋白 B 淋巴细胞抗原表位
Tab. 1 Antigenic epitopes of B lymphocytes of HpPrtC protein predicted by ABCpred software

Number	Amino acid sequence	Start postion	Score
1	GGVSHFSLRN RAGKEF	29	0.89
2	NGEITEDGRFFACKFT	329	0.88
3	KIEGR TKSSYYAAQTT	252	0.88
4	RGSCANDCRFDY EYYV	191	0.88
5	HIDKMAE LEPDAFIIA	82	0.87
6	TTTTNTAYEIIAPKNA	344	0.86
7	QNHQTAISEGDFQVNG	315	0.86
8	YAAQTTRIYRLAVDDF	262	0.86
9	SGHIAEILSSNAISAL	236	0.86
10	KEGIDYAHALNKKVYA	50	0.85
11	AGKEFTLET FKEGIDY	40	0.85
12	LVEE EGVGTHIFNAKD	217	0.85

2.5 HpPrtC 蛋白 T 淋巴细胞抗原表位 选择等位基因 HLA-A*0201 预测 MHC-I 类分子提呈的 CTL 细胞抗原表位, 长度为 8 和 9 个氨基酸, 阈值设定为 23, 氨基酸序列为 NLSGHIAEI 和 ALNKKVYATI 等。选择等位基因 HLA-DRB1*0401 预测 MHC-II 类分子提呈的 Th 细胞抗原表位, 长度为 15 个氨基酸, 阈值设定为 26, 氨基酸序列为 DAVYGGVSHFSLRNR 等。表明 HpPrtC 蛋白具有 T 淋巴细胞潜在优势抗原表位。见表 3。

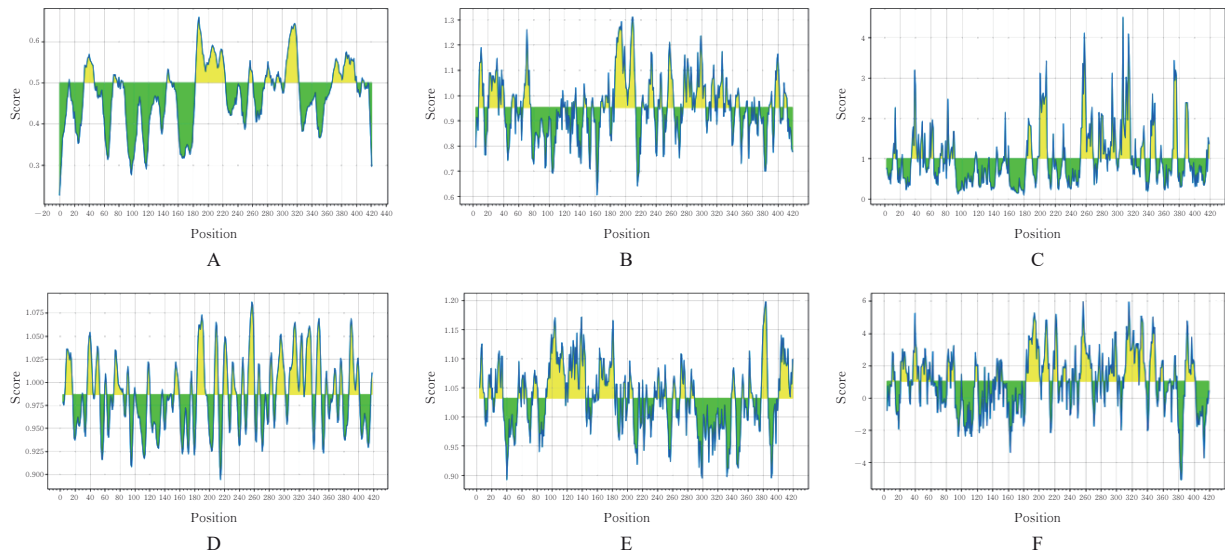
2.6 HpPrtC 蛋白可结晶性 EP 法和 RF 法等级分类分数分别为 3 及 8, 均表明 HpPrtC 蛋白为中等难度结晶, 且无信号肽和跨膜螺旋。见图 6。

2.7 HpPrtC 重组蛋白诱导表达和纯化 取 67.4 和 76.3 mL 洗脱峰进行 SDS-PAGE 分析, 结果显示: 2 个洗脱峰蛋白相对分子质量相同, 约为 47 000, 表明 HpPrtC 重组蛋白为单体和二聚体的动态平衡状态, 且以单体结构为主, 该蛋白为目的蛋白。见图 7。

2.8 HpPrtC 蛋白结晶条件 在 $0.2 \text{ mol}\cdot\text{L}^{-1}$ 氯化镁、 $0.1 \text{ mol}\cdot\text{L}^{-1}$ 三羟甲基氨基甲烷 [tris (hydroxymethyl) laminomethane, Tris]、 $3.4 \text{ mol}\cdot\text{L}^{-1}$ 乙二醇和 pH 8.5 条件下出现细小成簇的针状晶体, 其他条件下的液滴未见明显结晶。表明 HpPrtC 蛋白结晶以蛋白沉淀为主。见图 8。

3 讨论

Hp 作为普通的微需氧细菌, 是目前唯一能够



A: Bepipred linear epitope; B: β -turn; C: Emini surface accessibility; D: Karplus & schulz flexibility; E: Kolaskar & tongaonkar antigenicity; F: Parker hydrophilicity.

图5 IEDB软件预测HpPrtC蛋白B淋巴细胞线性表位

Fig. 5 Linear epitopes of B lymphocytes of HpPrtC protein predicted by IEDB software

表2 IEDB软件预测HpPrtC蛋白构象表位

Tab. 2 Construct epitopes of HpPrtC protein predicted by IEDB software

Number	Residues	Number of residues	Score
1	A:R306, A:P308, A:F309	3	0.835
2	A:F200, A:D201, A:Y202, A:E203, A:Y204, A:Y205, A:V206, A:K207, A:N208, A:P209, A:D210, A:N211, A:G212, A:V213, A:M214, A:M215, A:R216, A:L217, A:V218, A:E219, A:E220, A:E221, A:G222, A:V223, A:G224, A:T225, A:H226, A:N229, A:Q315, A:N316, A:H317, A:Q318, A:T319, A:A320	34	0.808
3	A:T333, A:D335, A:G336, A:R337, A:F338, A:K342, A:F343, A:T344, A:T345, A:T346, A:T347, A:N348, A:I349, A:A350, A:Y351, A:K357, A:N358, A:A359, A:A360, A:I361, A:T362, A:P363, A:I364, A:V365, A:N366, A:E367, A:I368, A:G369, A:K370, A:I371, A:Y372, A:T373, A:S378, A:Y379, A:L380, A:V381, A:L382, A:Y383, A:K384, A:I385, A:L386, A:L387, A:E388, A:N389, A:N390, A:T391, A:E392, A:L393, A:E394, A:T395, A:I396, A:H397, A:S398, A:G399, A:N400, A:V401, A:N402, A:L403, A:V404, A:R405, A:L406, A:P407, A:A408, A:P409, A:L410, A:P411	66	0.742
4	A:V420, A:E421, A:S422, A:K423, A:N424, A:G425, A:V426	7	0.738
5	A:Q3, A:V31, A:S32, A:H33, A:F34, A:K50, A:D54, A:H57, A:G69, A:F70, A:P71, A:F72, A:N73, A:S74, A:Q75, A:L76, A:K77, A:L78, A:L79, A:E80, A:E81, A:H82, A:I83, A:Y84, A:K85, A:A87, A:E88, A:L89, A:E90, A:P91, A:A98, A:P99, A:G100, A:V101, A:V102, A:K103, A:L104, A:A105, A:L106, A:K107, A:I108, A:A109, A:P110, A:H111, A:I112, A:V122, A:L123, A:N124, A:L125, A:L126, A:D127, A:A128, A:Q129, A:V130, A:F131, A:Y132, A:D133, A:L134, A:G135, A:V136, A:K137, A:L145, A:S146, A:L147, A:N148, A:D149, A:A150, A:I151, A:E152, A:I153, A:K154, A:K155, A:A156, A:L157, A:P158, A:N159, A:L160, A:A240, A:E241, A:S244, A:S245, A:N246, A:A247, A:I248, A:F277, A:Y278, A:L304	87	0.663
6	A:D276, A:H279, A:N280, A:T281, A:L282, A:K283, A:P284, A:S285	8	0.661
7	A:R307, A:E310, A:D313	3	0.613
8	A:C178, A:L179, A:I180, A:A182, A:L183, A:Q184, A:K185, A:G186, A:R187	9	0.596
9	A:F44, A:T45, A:L46, A:E47	4	0.523

表3 SYFPEITHI 软件预测 HpPrtC 蛋白 T 淋巴细胞抗原表位

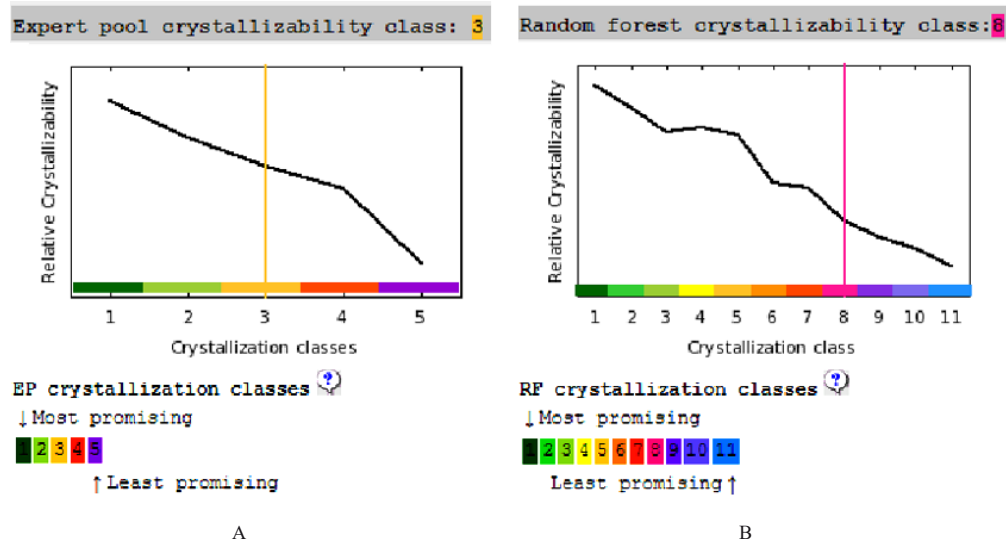
Tab.3 Antigenic epitopes of T lymphocytes of HpPrtC protein predicted by SYFPEITHI software

Start postion	Amino acid sequence	Length	Score
25	DAVYGGVSHFSLRNR	15	28
52	GIDYAHALNKKVYAT	15	28
258	KSSYYAAQTTRIYRL	15	28
284	PSFYASELNTLKNRG	15	28
103	KLALKIAPHIPIHLS	15	26
159	NLELEIFVHGSMCFA	15	26
236	SGHIAEILSSNAISA	15	26
248	ISALKIEGRTKSSYY	15	26
271	RLAVDDFYHNTLKPS	15	26
390	NTELETIHSGNVNLV	15	26
234	NLSGHIAEI	8	28
122	VLNLLDAQV	8	25
58	ALNKKVYAT	8	24
156	ALPNLELEI	8	24
320	AISEGDFQV	8	24
379	YLVLYKILL	8	23
58	ALNKKVYATI	9	25
107	KIAPHIPIHL	9	23
125	LLDAQVFDL	9	23
395	TIHSGNVNLV	9	23

在人类胃部存活的细菌，主要定植于人胃黏膜层。*Hp*在胃黏膜上黏附和定植后可通过损害人体的屏障体系及引起机体炎症反应等方式破坏人体组织，

从而引发胃炎，提高胃腺癌发生风险^[19-21]。研究^[22]显示：引起*Hp*的发病因素主要包括定植因素和细胞毒力因素，其中HpPrtC为定植因子，对于*Hp*的定植黏附起重要作用。本研究结果显示：HpPrtC蛋白既无跨膜区，也无典型的信号肽，是一种亲水性蛋白质。HpPrtC蛋白为全基因组蛋白，原核表达和纯化过程中蛋白主要存在于胞浆中，可能通过其他途径将HpPrtC蛋白酶分泌至胞外。本研究结果显示：经Ni-NTA琼脂糖亲和层析和凝胶过滤层析纯化获取的目的蛋白与赵慧琳等^[8]对HpPrtC原核表达并经Ni-NTA琼脂糖亲和层析的蛋白相对分子质量一致，且经凝胶过滤层析纯化的目的蛋白更加精纯，对胶原蛋白的酶解活性可能更高。HpPrtC蛋白结构较为稳定，这可能与HpPrtC蛋白酶解胶原蛋白的活性及其介导*Hp*的定植黏附有关^[23]。

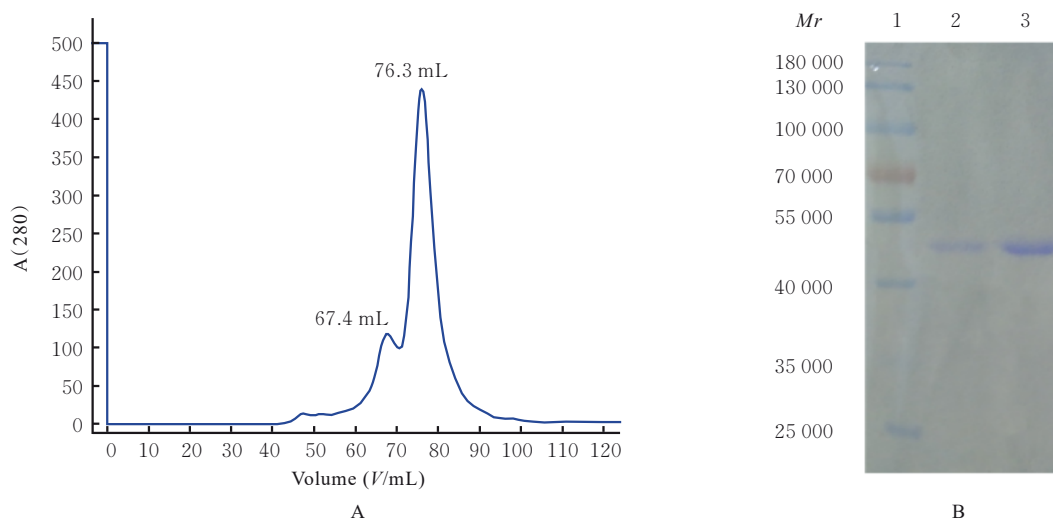
有效的抗原表位是诱导T淋巴细胞和B淋巴细胞产生免疫反应的基础，其与蛋白质二级结构存在密切关联， β 转角和无规则卷曲可判断其能否成为抗原表位^[24]。本研究结果显示：在HpPrtC蛋白中 β 转角为6.87%，无规则卷曲为38.63%，表明该蛋白可能具有明显的抗原表位。预测T淋巴细胞表位的目的是CD4+或CD8+T淋巴细胞表面T淋巴细胞抗原受体（T lymphocyte antigen receptor, TCR）识别的抗原肽，此过程中最关键的是主要组织相容性复合体（major histocompatibility complex, MHC）分子与抗原肽结合和提呈。HLA的生物多态性能够协助人类抵御多种疾病入侵，而



A:EP; B: RF.

图6 EP和RF算法预测HpPrtC蛋白结晶情况

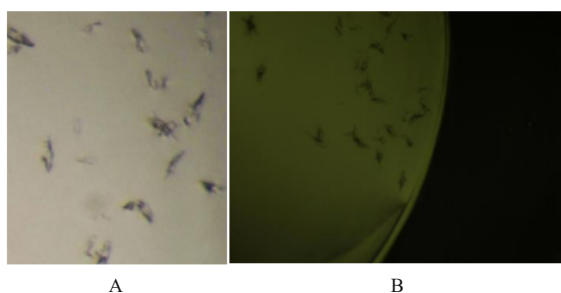
Fig.6 Crystallization of HpPrtC protein predicted by EP and RF algorithms



A: Superdex-200 purification of HpPrtC protein; B: SDS-PAGE analysis of elution peak (Lane 1: Marker; Lane 2: Elution peak at 67.4 mL; Lane 3: Elution peak at 76.3 mL).

图7 HpPrtC蛋白凝胶过滤层析和SDS-PAGE分析

Fig. 7 Gel filtration chromatography and SDS-PAGE analysis on HpPrtC protein



A: Bright; B: Dark.

图8 倒置显微镜观察HpPrtC蛋白晶体($\times 100$)

Fig. 8 HpPrtC protein crystals observed under microscope($\times 100$)

不同类型的HLA可将抗原肽提呈至不同的细胞,以启动不同类型的免疫应答^[25]。然而,HLA等位基因的分布频数因地区和族群不同而存在差异,其中HLA-A*0201型限制性抗原表位在我国汉族人群中最为常见^[26]。本研究结果显示:HpPrtC具有10个T淋巴细胞潜在优势抗原表位。目前,常用于预测B淋巴细胞表位的软件包括BCpred、ABCpred、LBtope、SVMTrip和IEDB等,其中ABCpred和IEDB软件在预测表位方面应用最多^[27-29]。本研究结果显示:HpPrtC蛋白具有5个B淋巴细胞优势抗原表位,阈值 ≥ 0.85 ,长度为16个氨基酸,其中有9个线性表位分数大于0.569,3个构象表位分数大于0.67。上述结果提示HpPrtC蛋白具有T淋巴细胞和B淋巴细胞优势抗原表位,可作为Hp疫苗设计的抗原,用于构建多价融合疫

苗或多表位疫苗。

X射线晶体学可测定蛋白质的三维结构,衍射数据与晶体品质有密切关联。本研究基于HpPrtC蛋白的结晶条件,由500个结晶条件中仅筛选出1个针簇状晶体的出品条件,表明重组HpPrtC蛋白也具有可结晶性,但较为困难,可能是由于蛋白表面具有较多的柔性无规则卷曲所致,与XtalPred软件的预测结果相似,可结晶性为中等困难^[30-31]。本研究结果显示:HpPrtC蛋白氨基酸序列在第294~321位之间有较长的无规则卷曲。郭玲等^[32]研究显示:Hp黏附素HpaA蛋白氨基酸中有一段较长的无规则卷曲,蛋白表达后不易结晶,去除较长的无规卷曲区域后蛋白则更易结晶。本研究中重组HpPrtC蛋白结晶为针状蛋白晶体,较难进行X射线衍射。因此,在后续研究中可通过表达一段更稳定的HpPrtC蛋白序列以获得高质量晶体,收集衍射数据,解析HpPrtC蛋白的三维结构,为研发Hp的新型防治方法提供理论依据。

综上所述,HpPrtC是一种亲水型蛋白,呈二聚体构造,在适宜条件下呈细小成簇针状晶体,其具有T淋巴细胞和B淋巴细胞优势抗原表位,可作为Hp疫苗设计的抗原,用于构建多价融合疫苗或多表位疫苗,为Hp的防治提供实验基础。

利益冲突声明:

所有作者声明不存在利益冲突。

作者贡献声明:

林玲辉参与数据收集整理和论文撰写, 李娜和尹晓燕参与实验操作及研究过程, 王晓凌和胡亚平参与数据整理及分析, 刘威参与指导论文撰写, 费瑞参与论文审校, 田新利参与实验设计及指导。

[参考文献]

- [1] 张秋月, 李培杰, 董全江, 等. 胃癌相关幽门螺杆菌毒力和粘附基因进化分析[J]. 中国病原生物学杂志, 2021, 16(2): 161-166.
- [2] 覃艳春, 黄衍强, 陆 钢, 等. 幽门螺杆菌感染慢性胃炎模型小鼠肠道各区域的菌群分布特征及其机制[J]. 吉林大学学报(医学版), 2023, 49(2): 289-297.
- [3] 中华医学会消化病学分会幽门螺杆菌学组. 第六次全国幽门螺杆菌感染处理共识报告(非根除治疗部分)[J]. 中华消化杂志, 2022, 42(5): 289-303.
- [4] FITZGERALD R, SMITH S M. An overview of helicobacter pylori infection [J]. *Methods Mol Biol*, 2021, 2283: 1-14.
- [5] DIACONU S, PREDESCU A, MOLDOVEANU A, et al. *Helicobacter pylori* infection: old and new[J]. *J Med Life*, 2017, 10(2): 112-117.
- [6] ZHAO H L, JI X F, CHEN X X, et al. Functional study of gene hp0169 in *Helicobacter pylori* pathogenesis[J]. *Microb Pathog*, 2017, 104: 225-231.
- [7] BERNEGGER S, JARZAB M, WESSLER S, et al. Proteolytic landscapes in gastric pathology and cancerogenesis[J]. *Int J Mol Sci*, 2022, 23(5): 2419.
- [8] 赵慧琳, 吴玉龙, 荣倩玉, 等. 幽门螺杆菌定植因子HpPrtC胶原蛋白酶的异源表达及其功能活性研究[J]. 中国病原生物学杂志, 2018, 13(9): 958-962.
- [9] 田新利, 刘 东, 郭 玲, 等. 幽门螺杆菌胶原蛋白酶HpPrtC的表达、纯化及结晶尝试[J]. 中国病原生物学杂志, 2018, 13(7): 719-723.
- [10] GEORJON C, DELÉAGE G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments [J]. *Bioinformatics*, 1995, 11(6): 681-684.
- [11] THAWNG C N, SMITH G B. A transcriptome software comparison for the analyses of treatments expected to give subtle gene expression responses [J]. *BMC Genomics*, 2022, 23(1): 452.
- [12] 董 浩, 江海圳, 李 超, 等. 小鼠NR1D1基因过表达载体的构建及其生物信息学分析[J]. 吉林大学学报(医学版), 2022, 48(1): 94-103.
- [13] SCHACHERL M, MONTADA A A, BRUNSTEIN E, et al. The first crystal structure of the peptidase domain of the U32 peptidase family[J]. *Acta Crystallogr D Biol Crystallogr*, 2015, 71(Pt 12): 2505-2512.
- [14] MENDES M, MAHITA J, BLAZESKA N, et al. IEDB-3D 2.0: structural data analysis within the Immune Epitope Database[J]. *Protein Sci*, 2023, 32(4): e4605.
- [15] AHMAD MALIK A, OJHA S C, SCHADUANGRAT N, et al. ABCpred: a webserver for the discovery of acetyl- and butyryl-cholinesterase inhibitors [J]. *Mol Divers*, 2022, 26(1): 467-487.
- [16] SCHULER M M, NASTKE M D, STEVANOVIĆ S. SYFPEITHI: database for searching and T-cell epitope prediction[J]. *Methods Mol Biol*, 2007, 409: 75-93.
- [17] SLABINSKI L, JAROSZEWSKI L, RYCHLEWSKI L, et al. XtalPred: a web server for prediction of protein crystallizability[J]. *Bioinformatics*, 2007, 23(24): 3403-3405.
- [18] ZHENG D D, LIANG S D, ZHANG C. B-cell epitope predictions using computational methods [J]. *Methods Mol Biol*, 2023, 2552: 239-254.
- [19] 金 昭, 王儒帅, 王朝阳, 等. 梅花草提取物对幽门螺杆菌小鼠胃黏膜定植的影响及其机制[J]. 吉林大学学报(医学版), 2022, 48(5): 1131-1138.
- [20] 常明珠, 李雨澎, 母润红, 等. 幽门螺旋杆菌检测方法及其应用价值的研究进展[J]. 吉林大学学报(医学版), 2023, 49(1): 253-260.
- [21] SAXENA A, MUKHOPADHYAY A K, NANDI S P. *Helicobacter pylori*: perturbation and restoration of gut microbiome[J]. *J Biosci*, 2020, 45(1): 110.
- [22] ZAGARI R M, FRAZZONI L, MARASCO G, et al. Treatment of *Helicobacter pylori* infection: a clinical practice update[J]. *Minerva Med*, 2021, 112(2): 281-287.
- [23] BOON L, UGARTE-BERZAL E, VANDOOREN J, et al. Protease propeptide structures, mechanisms of activation, and functions [J]. *Crit Rev Biochem Mol Biol*, 2020, 55(2): 111-165.
- [24] CROWCROFT N S, KLEIN N P. A framework for research on vaccine effectiveness[J]. *Vaccine*, 2018, 36(48): 7286-7293.
- [25] PALMER W H, NORMAN P J. The impact of HLA polymorphism on herpesvirus infection and disease [J]. *Immunogenetics*, 2023, 75(3): 231-247.
- [26] 韩 雪, 寸怡娜, 朱兰芳, 等. 甲型流感病毒T细胞表位预测及HLA-A2限制性抗原肽筛选[J]. 贵州医科大学学报, 2022, 47(5): 530-535, 541.
- [27] RAMANA J, MEHLA K. Immunoinformatics and epitope prediction[J]. *Methods Mol Biol*, 2020, 2131: 155-171.

- [28] 张舒阳, 毕研贞, 刘守胜, 等. 慢性乙型肝炎与HBV相关慢加急性肝衰竭患者血浆外泌体差异miRNA的生物信息学分析[J]. 临床肝胆病杂志, 2023, 39(8): 1848-1856.
- [29] MENDES M, MAHITA J, BLAZESKA N, et al. IEDB-3D 2.0: structural data analysis within the Immune Epitope Database[J]. Protein Sci, 2023, 32(4): e4605.
- [30] 柏晓辉, 刘雪, 朱雯培, 等. 肺炎链球菌耐药相关蛋白Sp_0010生物信息学分析及结晶尝试[J]. 中国病原生物学杂志, 2020, 15(11): 1268-1271, 1276.
- [31] LIU H, ZHAO Y, SUN J. Heterogeneous nucleation in protein crystallization[J]. Biomimetics, 2023, 8(1): 68.
- [32] 郭玲, 章金勇, 刘东, 等. 幽门螺杆菌黏附素HpaA重组蛋白的表达、纯化及晶体生长[J]. 免疫学杂志, 2015, 31(11): 986-990.