

DfrA12 蛋白的结构分析及同源建模

刘宗保¹, 张志刚², 石磊^{1,2}

(1. 华南理工大学轻工与食品学院, 广东广州 510640)

(2. 肉食品安全生产技术国家重点实验室, 厦门银祥集团有限公司, 福建厦门 361100)

摘要: 本文通过分析 DfrA12 蛋白的初级结构和空间构型, 为进一步了解该蛋白的结构与功能提供了理论依据。应用 ExPasy、TMHMM2.0、SignalP 4.0、SOPMA、GENO3D 等生物信息学软件对 DfrA12 蛋白的初级和高级结构进行分析和预测。结果显示, DfrA12 蛋白由 165 个氨基酸组成, 富含 α 螺旋和无规则卷曲结构, 不形成跨膜区域, 无信号肽, 定位于细胞内膜。应用同源建模的方法成功构建了该蛋白的三维模型。本研究为深入探讨 DfrA12 蛋白的结构与生物学功能奠定了理论基础。

关键词: DfrA12 蛋白; 生物信息学; 同源建模

文章编号: 1673-9078(2013)7-1505-1508

Structure Analysis and Homology Modeling of DfrA12 Protein

LIU Zong-bao¹, ZHANG Zhi-gang², SHI Lei^{1,2}

(1. College of Light Industry and Food Sciences, South China University of Technology, Guangzhou 510640, China)

(2. State Key Laboratory of Food Safety Technology for Meat Products, Xiamen Yinxiang Group Co., Ltd., Xiamen 361100, China)

Abstract: The aim of this study was to analyze the primary and spatial structure of DfrA12 protein and provide theoretical basis for study on the in-depth research of structure and function. Bioinformatics analysis softwares such as ExPasy, TMHMM 2.0, SignalP 4.0, SOPMA and GENO3D were used to analyse and predict the primary and senior structure of DfrA12 protein. The results showed that DfrA12 protein was composed by 165 amino acids, rich in Alpha Helix and Random coil structure. This protein had no signal peptide and transmembrane region, located in intracellular membrane. The three-dimensional structure of the protein was constructed by homology modeling. This study will lay a foundation for further research on structure and biological function of DfrA12 protein.

Key words: DfrA12 protein; bioinformatics; homology modeling

DfrA12 蛋白是一类二氢叶酸还原酶 (Dihydrofolate reductase, DHFR), DHFR 是细胞内叶酸代谢的关键酶, 可在辅酶 NADPH 的参与下催化二氢叶酸 (Dihydrofolate, DHF) 还原成四氢叶酸 (Tetrahydrofolate, THF)^[1], 它在一碳单位的转运过程中扮演着重要作用^[2]。微生物遗传基因内含有编码 DfrA12 蛋白的 dfrA12 基因通常会使得微生物对甲氧苄氨嘧啶 (Trimethoprim, TMP) 抗生素产生耐药抗性。TMP 配合磺胺类 (Sulfonamides, SAs) 抗生素在畜牧养殖业中被广泛用于预防和治理细菌感染型疾病^[3], 相关研究表明, 畜牧养殖中过量使用抗生素是食源性耐药菌出现和传播的主要因素^[4~5], 这些食源性耐药菌一旦进入食品产业链, 将会对人类的健康产生威胁^[6]。近年来, 大量的科学文献中报道过细菌对包括 TMP 在内的各种

抗生素的耐药情况, 但绝大多数只局限于耐药表型及对耐药菌进行监控的研究^[7~8], 有关耐药机理及相关蛋白的结构功能研究相对较少。截至目前, 没有发现 DfrA12 蛋白三维空间结构模型的相关报道。

随着分子生物学技术的不断发展, 并在人类基因组计划的推动下, 通过基因组序列、蛋白质序列的测序, 产生了大量的新的基因及氨基酸序列, 这些新的序列急需综合利用数学、物理、信息科学与技术等进行处理和分析, 在此背景下, 生物信息学应运而生^[9]。生物信息学主要由基因组信息学、蛋白质空间结构模拟以及药物设计构成 3 个重要部分组成^[10], 并在生物技术、生物医学、农业、环境、食品及能源等领域发挥重要的作用^[11]。对于蛋白质结构和功能, 尽管可以通过传统的方法如 X 射线晶体衍射方法、核磁共振技术 (Nuclear Magnetic Resonance, NMR)^[12]、电镜三维重构技术等方法来实现, 但由于传统的蛋白检测技术费时费力且远远跟不上不断涌现的新 DNA 的数量, 因此利用生物信息学软件快速预测蛋白结构与功能特性, 对

收稿日期: 2013-03-30

基金项目: 国家 973 计划项目 (2012CB723705)

作者简介: 刘宗保 (1985-), 男, 博士研究生, 研究方向: 食品安全

通讯作者: 石磊, 男, 教授, 博士生导师

研究蛋白质组尤其是对那些通过实验难以测定结构的蛋白质分析则具有更大的理论意义与实用价值。蛋白质的生物学功能在很大程度上依赖其空间结构,因而对蛋白的结构进行准确预测对了解未知蛋白的功能具有重要的意义^[13]。本文通过应用各种生物信息学软件对DfrA12蛋白的初级结构、二级结构和三级结构进行了预测和分析,并构建了DfrA12蛋白的同源模型,为DfrA12蛋白结构和功能的研究提供理论依据。

1 材料与方法

1.1 材料

以本实验室分离的多重耐药大肠杆菌(*Escherichia coli*)菌株的基因组DNA为模板,以5'-CSGGCATCCAAGCAGCAAG, 3'-CSAAGCAGACTTGACCTGA为引物^[14],应用聚合酶链式反应(Polymerase Chain Reaction, PCR)技术扩增出一段I类整合子耐药基因盒。通过美国生物信息中心(NCBI)网站BLAST程序的同源性比对发现,该基因盒中的一段DNA序列与序列号为JN119853.1、HQ880253.1和HQ880279.1中编码DfrA12蛋白的基因片段均具有100%的同源性,因此可以判断该段基因是一段编码DfrA12蛋白的基因片段,该基因编码的DfrA12蛋白的氨基酸序列为:MNSESVRIYLVAAMGANRVIGNGP NIPWKIPGEQKIFRRLTEGKVVVMGRKTFESIGKPL PNRHTLVISRQANYRATGCVVSTLSHAIALASELGNELYVAGGAEIYTLALPHAHG VFLSEVHQTFEGDA FFPMLNETEFELVSTETIQAVIPYTHSVYARRNG。

1.2 方法

1.2.1 DfrA12蛋白的初级结构分析

利用ExPasy提供的ProtParam程序(<http://expasy.org/tools/protparam.html>)对DfrA12蛋白的氨基酸残基数目、组成、原子组成、蛋白相对分子量、理论等电点及疏水性平均值等参数进行在线分析。

1.2.2 DfrA12蛋白跨膜区域和信号肽预测

应用TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>)对DfrA12蛋白的跨膜区域进行预测,利用SignalP 4.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>)在线服务器对该蛋白的信号肽进行预测。

1.2.3 DfrA12蛋白二级结构分析

利用SOPMA对该蛋白的二级结构进行预测分析(http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)。

1.2.4 DfrA12蛋白的细胞定位

使用PSORT WWW Server提供的PSORT Prediction

工具(<http://psort.hgc.jp/form.html>)对该蛋白的细胞定位进行预测。

1.2.5 蛋白三级结构预测

使用GENO3D (http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html)对该蛋白的三维结构进行预测,并利用Procheck, Errat和Verify_3D (<http://nihserver.mbi.ucla.edu/SAVES/>)对蛋白质结构模型进行评估。使用Rasmol 2.7.5软件展示三级结构模型。

2 结果与讨论

2.1 DfrA12蛋白氨基酸序列的物理性质分析

由ProtParam程序分析结果可知,该DfrA12蛋白由165个氨基酸组成,分子量为18.13 kDa,其分子式为C₈₁₃H₁₂₈₂N₂₂₆O₂₃₄S₅,计算得到的原子总数和等电点分别为2560和7.94。酸性氨基酸残基总数(Asp+Glu)为14,碱性氨基酸残基总数(Arg+Lys)为15,由此可见,该蛋白属于碱性蛋白。疏水性的总平均值(GRAVY)为0.011,为正值,因此该蛋白定义为疏水性蛋白。组成DfrA12蛋白的氨基酸比例见表1。

表1 DfrA12蛋白的氨基酸组成分析

Table 1 Amino acid composition of DfrA12 protein

氨基酸	数目	摩尔比/%	氨基酸	数目	摩尔比/%
Ala	15	9.1	Leu	13	7.9
Arg	10	6.1	Lys	5	3.0
Asn	9	5.5	Met	4	2.4
Asp	1	0.6	Phe	7	4.2
Cys	1	0.6	Pro	8	4.8
Gln	4	2.4	Ser	10	6.1
Glu	13	7.9	Thr	11	6.7
Gly	14	8.5	Trp	1	0.6
His	6	3.6	Tyr	6	3.6
Ile	11	6.7	Val	16	9.7

2.2 DfrA12蛋白跨膜区域和信号肽预测

应用TMHMM Server v. 2.0对DfrA12蛋白的跨膜区域进行分析,结果显示该蛋白的跨膜区域为零,在跨膜区域中的氨基酸序列期望值为0.18955,当该结果大于18时才有跨膜区域,因此该蛋白没有跨膜区域(图1)。SignalP 4.0 Server对DfrA12蛋白的信号肽预测结果显示(图2),该蛋白不存在信号肽。

2.3 DfrA12蛋白二级结构预测

利用SOPMA服务器对DfrA12蛋白的二级结构进行预测分析,结果显示,在DfrA12蛋白的二级结构中:32.12%为 α 螺旋结构(Alpha helix, h),20%为扩展长链(Extended strand, e),5.45%为 β -转角结构(Beta turn, t),

42.47%为无规则卷曲结构(Random coil, c)(如图3)。

```

# DfrA12 Length: 165
# DfrA12 Number of predicted TMHs: 0
# DfrA12 Exp number of AAs in TMHs: 0.42644
# DfrA12 Exp number first 69 AAs: 0.04743
# DfrA12 Total prob of N-in: 0.18955
DfrA12 TMHMM2.0 outside | 165

```

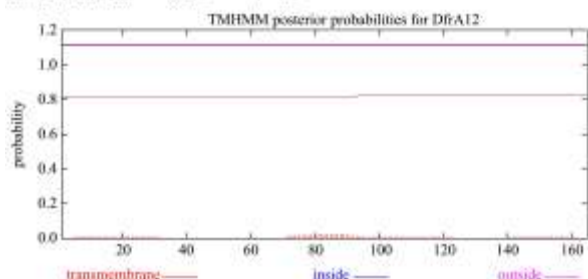


图1 DfrA12蛋白跨膜区域预测

Fig.1 Transmembrane regiona prediction of DfrA12 protein

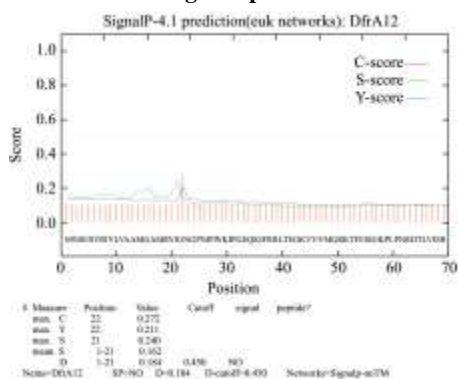


图2 DfrA12蛋白信号肽预测

Fig.2 Signal peptide prediction of DfrA12 protein

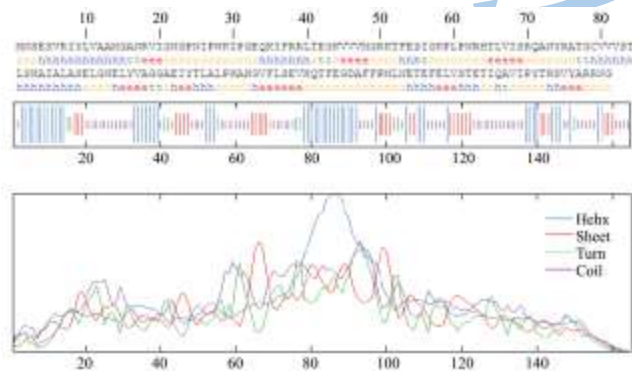


图 3 DfrA12蛋白二级结构预测

Fig.3 Secondary structure prediction of DfrA12 protein

2.4 DfrA12蛋白的细胞定位

应用 PSORT WWW Server 程序中的 PSORT Prediction工具对该蛋白的细胞定位进行分析, 预测结果表明, 该蛋白存在于细胞内膜中的可能性为11.9%, 而在细胞质、细胞壁膜间隙以及外膜中都无法确定是否有该蛋白的存在。

2.5 DfrA12蛋白三维结构预测

随着生物信息的发展, 使得同源建模的准确性得到了提高。当前免费提供蛋白质三维结构预测的服务器主要有Swiss-Model、CPH-models和Geno3D, 其中Geno3D功能比较齐全, 允许选择建模参数和多条模板,

并可生成多个可供选择的模型, 具有较好的灵活性。我们将DfrA12蛋白的氨基酸序列在线提交Geno3D 服务器后, 返回一些相似序列, 在PDB(Protein Data Bank)中选择同源性高、并且有三维结构模型的pdb3jw3A、pdb3fl8A和pdb3sa1A这三个蛋白序列作为模板。将选择的模板在线提交给Geno3D后返回Model-1、Model-2、Model-3三个模型数据文件, 将这些文件经RasMol2.7.5软件展示, 得到如图4 所示的同源结构模型图。从能量最低化角度以及利用Procheck 和 Errat对三个模型进行综合评价, 结果发现Model-1更接近于DfrA12蛋白真实的空间构象。



Model-1 Model-2 Model-3

图 4 不同 DfrA12 蛋白的同源模型结构比较

Fig.4 Comparison of some homology models for DfrA12 protein

通过Model-1的Procheck全面分析结果(图5)可以看出, 在159个主要氨基酸残基中, 75.7%的残基处在核心允许区, 2.2%处于最大允许区内, 只有1.5%的氨基酸残基处于不可信区域, 这一结果满足立体化学结构中对二面角的分布要求。应用Verify_3D对Model-1模型进行评估, 结果显示, 至少有80%的氨基酸残基3D-1D的相容性分值大于0.2, 说明这个模型通过了Verify_3D检测。

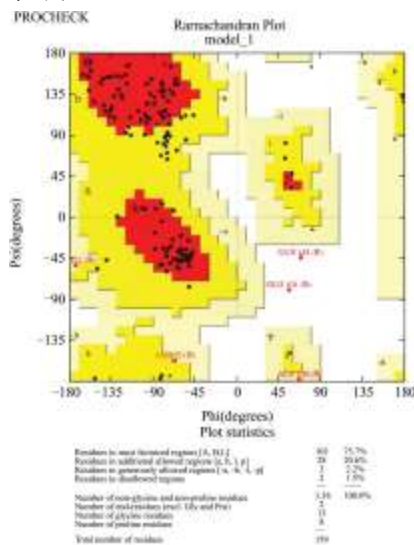


图 5 Model-1 的 Procheck 评价

Fig.5 Procheck analysis of Model-1

3 结论

3.1 在本次实验中, 我们对DfrA12蛋白的初级结构和高级结构进行了预测, 通过分析DfrA12蛋白一级结构

发现, 该蛋白的主肽链由165个氨基酸组成, 其理论等电点为7.94, 酸性和碱性氨基酸残基总数分别为14和15, 由此可见, 该蛋白属于碱性蛋白。该蛋白的疏水性的总平均值(GRAVY)为0.011, 为正值, 因此该蛋白定义为疏水性蛋白。应用TMHMM Server v. 2.0和SignalP 4.0 Server软件对该蛋白的跨膜区域和信号肽进行预测发现, 该蛋白不存在跨膜区域和信号肽。通过PSORT Prediction工具对该蛋白的细胞定位进行分析表明, 该蛋白存在于细胞内膜中的可能性较高。分析DfrA12蛋白的二级结构发现, 该蛋白富含 α 螺旋和无规则卷曲结构。我们将DfrA12蛋白的氨基酸序列在线提交Geno3D服务器, 选择pdb3jw3A、pdb3fl8A和pdb3sa1A蛋白序列作为模板, 得到了Model-1、Model-2、Model-3三个模型数据文件。从能量最低化角度以及利用Procheck和Errat对三个模型进行综合评价, 结果发现Model-1更接近于DfrA12蛋白真实的空间构象。通过Verify_3D对Model-1模型的评估发现, 至少有80%的氨基酸残基3D-1D的相容性分值大于0.2, 说明我们构建出了合理的DfrA12蛋白同源模型。

3.2 本研究通过应用一系列的生物信息学软件对DfrA12蛋白的一级结构、跨膜区域、信号肽、细胞定位、二级和三级结构进行了预测和分析, 并应用Geno3D服务器成功构建了DfrA12蛋白的同源模型, 为DfrA12蛋白结构和功能的进一步研究提供了理论依据。

参考文献

- [1] Dao-Jing Li, Zuo-Jun Shen. Dihydrofolate reductase and tumor resistance [J]. CHEMISTRY OF LIFE, 2010, 30(6): 923-926
- [2] Mishra P J, Humeniuk R, Mishra P J, et al. A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance [J]. Proc Natl Acad Sci USA, 2007, 104: 13513-13518
- [3] Pawin Padungtod, Chanwit Tribuddharat, Rungtip Chuanchuen. Widespread presence of DfrA12 and its association with DfrA12-aadA2 cassette in Salmonella enterica isolates from swine [J]. Southeast Asian J Trop Med Public Health, 2011, 6(42): 1471-1476
- [4] S Koike, IG Krapac, H D Oliver, et al. Monitoring and Source Tracking of Tetracycline Resistance Genes in Lagoons and Groundwater Adjacent to Swine Production Facilities over a 3-Year Period [J]. Appl. Environ. Microbiol, 2007, 73(5): 4813-4823
- [5] Gellin G, B E Langlois, K A Dawson, et al. Antibiotic resistance of gram-negative enteric bacteria from pigs in three herds with different histories of antibiotic exposure [J]. Appl. Environ. Microbiol, 1989, 55: 2287-2292
- [6] Van den Bogaard A E, E E Stobberingh. Epidemiology of resistance to antibiotics. Links between animals and humans [J]. Int. J. Antimicrob. Agents, 2000, 14: 327-335
- [7] Bywater R, H Deluyker, E Deroover, et al. A European survey of antimicrobial susceptibility among zoonotic and commensal bacteria isolated from food-producing animals [J]. J. Antimicrob. Chemother, 2004, 54: 744-754
- [8] De Jong A, R Bywater, P Butty, et al. A pan-European survey of antimicrobial susceptibility towards human-use antimicrobial drugs among zoonotic and commensal enteric bacteria isolated from healthy food-producing animals [J]. J. Antimicrob. Chemother, 2009, 63: 733-744
- [9] 王朝霞. 生物信息学: 一门前沿交叉学科 [J]. 安徽教育学院学报, 2002, 20(6): 56-57
- [10] JONATHAN PEVSNER. 生物信息学与功能基因组学 [M]. 孙之荣, 译. 北京: 化学工业出版社, 2006
- [11] 孟双, 徐冲, 陈丽媛, 等. 生物信息学在生物学研究领域的应用 [J]. 微生物学杂志, 2011, 31(1): 78-81
- [12] 殷志祥. 蛋白质结构预测方法的研究进展 [J]. 计算机工程与应用, 2004, 20: 54-57
- [13] 徐建华. 生物信息学在蛋白质结构与功能预测中的应用 [J]. 医学分子生物学杂志, 2005, 2(3): 227-232
- [14] Leverstein-van Hall MA, Paaauw A, Box ATA, et al. Presence of integron-associated resistance in the community is widespread and contributes to multidrug resistance in the hospital [J]. J Clin Microbiol, 2002, 40: 3038-304