

DOI: 10.3969/cjz.j.issn.1002-2694.2013.05.010

刚地弓形虫苹果酸脱氢酶基因编码蛋白主要特性与抗原表位的生物信息学分析

刘转转, 杨燕萍, 郑葵阳, 刘宜升

摘要:目的 用生物信息学方法分析刚地弓形虫苹果酸脱氢酶(*TgMDH*)基因编码蛋白的结构与抗原表位,预测该蛋白的免疫原性。方法 利用蛋白分析专家系统(ExpASy)提供的 ProtParam、SOSUI、TMPRED、HNN、MotifScan 和 NCBI 上的 ORF finder、SignalP、Bcepred 等生物信息学在线分析程序,结合 Gene Runner 和 DNAMAN 等生物信息学软件,分析并预测 *TgMDH* 蛋白的开放阅读框、理化性质、可溶性、信号肽、跨膜区、表面可及性、可塑性、亲(疏)水性、翻译后修饰位点、二级结构及抗原表位,通过 CPHmodels 程序预测该蛋白的三维空间结构。结果 *TgMDH* 蛋白由 316 个氨基酸组成,分子式为 $C_{1498}H_{2454}N_{402}O_{440}S_{20}$,分子质量单位为 33 777.5,等电点为 6.01,波长 280 nm 时的吸光度(A)值为 0.364,半衰期为 30 h,有 9 个表面可及性参数 ≥ 1.9 的区域、4 个亲水性参数得分 ≥ 1.9 的区域、7 个柔韧性参数得分 ≥ 2 的区域、14 个翻译后修饰位点和 14 个潜在抗原表位,为可溶性表达蛋白。结论 *TgMDH* 基因编码蛋白为可溶性蛋白,有多个抗原表位,具有免疫原性,可作为弓形虫病疫苗候选抗原。

关键词:刚地弓形虫;苹果酸脱氢酶;生物信息学;抗原表位;免疫原性

中图分类号:R382

文献标识码:A

文章编号:1002-2694(2013)02-0472-04

Bioinformatics analysis of main characteristics and epitopes of the gene encoding malate dehydrogenase protein from *Toxoplasma gondii*

LIU Zhuan-zhuan, YANG Yan-ping, ZHENG Kui-yang, LIU Yi-sheng

(Department of Microbiology and Immunology, Xuzhou Medical College, Xuzhou 221004, China)

ABSTRACT: The immunogenicity of the malate dehydrogenase protein extracted from *Toxoplasma gondii* (*TgMDH*) was predicted and its structure and epitopes were characterized by bioinformatics approaches. The open reading frame, physico-chemical properties, solubility, signal peptide, cross membrane area, surface accessibility, flexibility, hydrophilicity, post-translational modification site, secondary structure, and antigen epitope of *TgMDH* protein were analyzed and predicted by bioinformatics online analysis programs which were provided by expert protein analysis system (ExpASy) including ProtParam, SOSUI, TMPRED, HNN, MotifScan, ORF finder, SignalP and Bcepred, combining with bioinformatics softwares (Gene Runner and DNAMAN). The three dimensional structure of the protein was then processed by CPHmodels. Results indicated that *TgMDH* protein was composed of 316 amino acids with listed characters; the molecular formula was $C_{1498}H_{2454}N_{402}O_{440}S_{20}$, the molecular mass was 33 777.5, the value of theoretical isoelectric point was 6.01, the absorbance (A_{280}) value was 0.364, and the half-life period was 30 h. There were nine zones with surface accessibility ≥ 1.9 , four zones with hydrophilicity ≥ 1.9 , seven zones with flexibility ≥ 2 , fourteen post-translational modification sites and fourteen potential epitopes as soluble protein expression. It's suggested that *TgMDH* protein could be used as a candidate antigen for toxoplasmosis vaccination because of its soluble protein expression and immunogenic property.

KEY WORDS: *Toxoplasma gondii*; malate dehydrogenase; bioinformatics; epitope; immunogenicity

Supported by the Open Projects of Parasitic Disease Prevention and Control High Technology Platform of Jiangsu Province (No. WK009-003), and the Research Program of Xuzhou Medical College (No. 2012KJ08)

江苏省寄生虫病预防与控制高技术平台开放课题(No. WK009-003);徐州医学院科研课题(No. 2012KJ08)联合资助

通讯作者:刘宜升,Email: liuyisheng52@163.com

作者单位:徐州医学院,徐州 221004

Corresponding author: Liu Yi-sheng, Email: liuyisheng52@163.com

刚地弓形虫(*Toxoplasma gondii*, Tg)是一种

专性细胞内寄生原虫,可引起人兽共患弓形虫病,尚无防治本病的理想药物,筛选有效的保护性抗原研制疫苗是关键所在^[1]。

苹果酸脱氢酶(malate dehydrogenase, MDH)是细胞在有氧条件下进行三羧酸循环的重要酶之一^[2-4]。MDH在细胞的多种生理活动中起重要作用,包括线粒体的能量代谢、植物的活性氧代谢及细胞内外物质的转运等^[5-6]。TgMDH通过影响三羧酸循环的氧化供能和机体糖、脂类及蛋白质的代谢,进而干扰弓形虫虫体的存活^[7]。

Ma^[8]等应用蛋白质组学的方法分析刚地弓形虫可溶性速殖子抗原(STAg)组分,结果显示TgMDH为STAg中具有抗原性的成分之一。目前,尚无TgMDH抗原性方面的相关研究报道。特利用生物信息学方法分析TgMDH基因及其编码蛋白的结构特征与抗原表位,预测该蛋白的免疫原性,为研究其免疫保护性提供资料。

1 材料与方法

1.1 基因及蛋白序列 TgMDH基因及蛋白的GenBank登录号分别为AY650028和AAT67462。

1.2 分析方法 利用NCBI上的ORF finder软件获得TgMDH基因的开放阅读框。采用ProtParam(<http://web.expasy.org/protparam/>)程序,预测TgMDH蛋白的分子量、等电点、氨基酸组成及稳定性等理化性质。经SignalP程序(<http://www.cbs.dtu.dk/services/SignalP/>)分析信号肽。TMPRED软件(http://ch.embnet.org/cgi-bin/TMPRED_form_parser)预测蛋白的跨膜区。SOSUI(<http://bp.nuap.nagoya-u.ac.jp/sosui/>)程序分析TgMDH蛋白的可溶性。Bcepred(<http://www.imtech.res.in/raghava/bcepred>)程序分析TgMDH蛋白的表面可及性、柔韧性和亲(疏)水性。MotifScan程序(http://myhits.isb-sib.ch/cgi-bin/motif_scan)预测TgMDH蛋白的翻译后修饰位点。HNN软件(<http://npsa-pbil.ibcp.fr/cgi-bin/npsa-automat.pl>)分析TgMDH蛋白的二级结构。通过DNAMAN软件分析抗原表位,并结合Gene Runner软件分析TgMDH蛋白潜在的B细胞表位。采用CPHmodels程序(<http://www.cbs.dtu.dk/services/CPHmodels/>)预测蛋白质三维空间结构。

2 结果

2.1 TgMDH基因的开放阅读框 利用NCBI上的ORF finder软件分析,TgMDH基因的开放阅读框为951 bp的核苷酸序列。

2.2 TgMDH蛋白的主要理化性质 经ProtParam分析,TgMDH基因编码蛋白由316个氨基酸组成,分子质量单位为33 777.5,分子式为C₁₄₉₈H₂₄₅₄N₄₀₂

O₄₄₀S₂₀,等电点为6.01,酸性氨基酸总数(Asp + Glu)共为36个,碱性氨基酸总数(Arg + Lys)共为34个。假设所有的二硫键形成时,280 nm处的摩尔消光系数为12 295 Mol⁻¹ · cm⁻¹,0.1%浓度(1 g/L)的吸光度(A₂₈₀)为0.364;而当所有的二硫键打开时,280 nm处的摩尔消光系数为11 920 Mol⁻¹ · cm⁻¹,0.1%浓度(1 g/L)的吸光度(A₂₈₀)为0.353。在哺乳动物网状红细胞体外表达的半衰期为30 h,在酵母和大肠埃希菌的体内表达的半衰期分别大于20 h和10 h。在溶液中的不稳定指数为33.21,低于阈值40,表明该蛋白在溶液中性质稳定。脂肪族指数为105.44,总平均亲水性指数0.289。

2.3 TgMDH蛋白的信号肽和跨膜区 应用SignalP 4.0 Server分析显示TgMDH基因表达的蛋白不存在信号肽。TMPRED分析结果显示,TgMDH蛋白不存在跨膜区结构。

2.4 TgMDH蛋白的可溶性 采用SOSUI程序分析,TgMDH蛋白的平均疏水性系数为0.218 671,为可溶性蛋白。

2.5 TgMDH蛋白的表面可及性、柔韧性和亲(疏)水性 采用Bcepred程序分析,TgMDH蛋白具有4个亲水性参数得分≥1.9的区域(61~67、85~91、204~210和213~221位氨基酸),7个柔韧性参数得分≥2的区域(1~5、56~62、80~89、210~219、244~251、288~294和296~302位氨基酸),9个表面可及性参数得分≥1.9的区域(1~7、37~43、59~67、82~91、108~114、149~160、204~220、244~255和289~306位氨基酸)。

2.6 TgMDH蛋白的翻译后修饰位点 亚细胞定位的预测通过翻译后修饰位点分析,理论推测该蛋白含有14个翻译后修饰位点,分别为:1个N-糖基化位点(162~165位氨基酸)、5个酪蛋白激酶II磷酸化位点(54~57、88~91、164~167、204~207和291~294位氨基酸)、4个N-豆蔻酰化位点(7~12、15~20、146~151和239~244位氨基酸)和4个蛋白激酶c磷酸化位点(2~4、150~152、158~160和304~306位氨基酸)。

2.7 TgMDH蛋白的二级结构 预测TgMDH蛋白的二级结构由316个氨基酸组成,其中125个(占39.56%)可能形成α螺旋,61个(占19.3%)可能形成β折叠片,130个(占41.14%)可能形成无规卷曲。

2.8 TgMDH蛋白的B细胞抗原表位 采用Gene Runner软件分析TgMDH的杰弗森-沃夫(Jame-

son-Wolf)抗原指数,预测 TgMDH 潜在的抗原决定簇,结果见图 1,横轴为蛋白质的氨基酸位置,纵轴为杰弗森-沃夫(Jameson-Wolf)抗原指数。位于横轴以上连续的峰值区间即为潜在的抗原表位,显

示 TgMDH 存在较多的抗原表位。

采用 DNAMAN 软件对 TgMDH 蛋白的氨基酸序列中潜在的抗原表位做进一步分析,显示有 14 个潜在的抗原表位,见表 1。

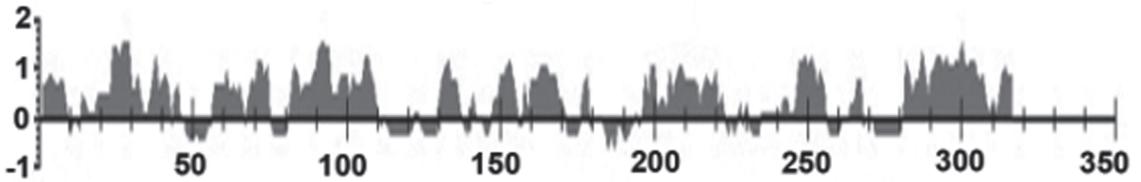


图 1 TgMDH 蛋白的抗原表位分析

Fig. 1 Analysis of the antigenic index of TgMDH

表 1 TgMDH 氨基酸序列中 14 个潜在抗原表位

Tab. 1 Fourteen potential epitopes of amino acid sequence of TgMDH

顺序 Order	氨基酸位置 Position of amino acids	潜在抗原序列 Potential sequence of antigen	分值 Score
1	267—280	KDMYVGVPVIGAG	1.237
2	250—261	RKRVLPCAAAYLN	1.200
3	16—59	ATLALLSAVKELGDVVMFDVVQDLPPQGKCLDLYQLTPISGVDVR	1.188
4	104—151	GEAIKQYCPNAFVICITNPLDVMVYILREKCGLPHPKVCGMAGVLDSA	1.173
5	189—198	VGGIPLPELV	1.145
6	65—84	DYSVLKDADVIHTAGVPRK	1.131
7	221—229	EIVSLLKTG	1.127
8	301—313	SVESVKTLAAAP	1.121
9	282—291	VEKIVELDLT	1.114
10	158—174	SERLNVSVDIHALVMG	1.110
11	231—246	AFFAPAAAGVLMAEAY	1.109
12	179—187	TMVPLPRFT	1.091
13	205—213	QQEVDIVQ	1.077
14	91—96	DLLAIN	1.069

2.9 TgMDH 蛋白的三维空间结构 利用 CPH-models 程序为 TgMDH 蛋白寻找模板,通过氨基酸序列比对,得到 1 个与 TgMDH 蛋白同源性为 65.5% 的蛋白质。该蛋白属于氧化还原酶家族,称为布鲁氏菌乳酸苹果酸脱氢酶 A 链,在蛋白质结构数据库中的登录号为 3GVH. A。以该序列的结构信息为模板建立 TgMDH 的蛋白模型,其结构用 Ras Mol 软件显示,见图 2。

3 讨论

利用数据库资源,结合生物信息学技术,应用生物信息学软件,为进一步实验研究提供了重要信息和依据^[9-10]。

B 细胞表位是抗原分子中与抗体结合的特异性位点,通常由 5~15 个氨基酸组成,其可以诱导机体产生保护性抗体并介导体液免疫应答反应。本研究



图 2 TgMDH 的空间结构模型

Fig. 2 Model of steric structure of TgMDH

采用的亲水性方案、表面可及性方案、可塑性方案、二级结构方案和抗原性方案等多参数综合预测 TgMDH 蛋白的抗原表位。

亲水性代表氨基酸残基在水溶液的暴露程度,

亲水性强表明氨基酸残基易与水分子接近,倾向于暴露在蛋白表面,易被特异性抗体识别^[11], *TgMDH* 蛋白有 4 个亲水性参数得分 ≥ 1.9 的区域,说明抗原表位可能存在于这 4 个氨基酸区段。表面可及性反映蛋白质抗原中氨基酸残基被溶剂分子接触的可能性, *TgMDH* 蛋白有 9 个表面可及性参数得分 ≥ 1.9 的氨基酸区段,这些区段极有可能位于蛋白表面,且易被溶剂分子接触,说明抗原表位可能存在于这 9 个氨基酸区段。可塑性好的区域具有一定的柔韧性,易形成抗原表位,容易与抗体进行嵌合, *TgMDH* 蛋白有 7 个柔韧性参数得分 ≥ 2 的氨基酸区段,说明抗原表位可能存在于这 7 个氨基酸区段。

在蛋白质的二级结构中, α 螺旋、 β 折叠结构规则,主要作为蛋白质骨架起稳定作用,而蛋白质的转角及无规卷曲比较松散,易于扭曲、盘旋并展示在蛋白的表面,成为抗原表位的可能性较大^[12], *TgMDH* 蛋白的无规卷曲 130 个,所占比例达 41.14%,提示可能具有多个抗原表位。

在 B 细胞表位抗原性预测的诸多方法中,Jameson-Wolf 法是最常用的方法之一。本研究用 Gene Runner 软件分析 *TgMDH* 蛋白的 Jameson-Wolf 指数,预测 *TgMDH* 蛋白具有多个 B 细胞抗原表位,提示其具有免疫原性,进一步用 DNAMAN 软件分析,结果显示 *TgMDH* 蛋白具有 14 个抗原表位。在此基础上,采用 CPHmodels 程序对蛋白的三维结构进行模拟,可以寻找抗原被抗体结合的准确部位。

蛋白质的磷酸化、酰基化和糖基化是蛋白质生物合成后活性调解的一种化学修饰,对蛋白具有重要的生物学意义^[13]。 *TgMDH* 蛋白具有 1 个 N-糖基化位点、5 个酪蛋白激酶 II 磷酸化位点、4 个 N-豆蔻酰化位点和 4 个蛋白激酶 c 磷酸化位点,这些结合位点主要参与和调节蛋白质的功能、代谢、翻译后的修饰和加工,也可能作为潜在的抗原表位。通过抗原表位的分析获得优势抗原肽段,可用于弓形虫抗原保护性的研究。

蛋白质理化分析表明, *TgMDH* 蛋白由 316 个氨基酸组成,分子量为 33 777.5,等电点为 6.01,在原核和真核表达系统及水溶液中的稳定性较高,适合基因工程表达。该蛋白为可溶性蛋白,有利于进行蛋白纯化。根据所获得的 *TgMDH* 的生物信息,我们将通过实验进一步分析、验证其生物学活性,为寻找弓形虫保护性抗原和研制相关疫苗提供实验依据。

(感谢山西医科大学寄生虫学教研室殷国荣教授和王芬老师对本文的修改。)

参考文献:

- [1]Kur J, Holec-Gasior L, Hiszczyńska-Sawicka E. Current status of toxoplasmosis vaccine development[J]. *Expert Rev Vaccines*, 2009, 8(6): 791-808. DOI: 10.1586/erv.09.27
- [2]Goward CR, Nicholls DJ. Malate dehydrogenase: a model for structure, evolution, and catalysis[J]. *Protein Sci*, 1994, 3(10): 1883-1888.
- [3]Pradhan A, Tripathi AK, Desai PV, et al. Structure and function of *Plasmodium falciparum* malate dehydrogenase: role of critical amino acids in co-substrate binding pocket[J]. *Biochimie*, 2009, 91(11-12): 1509-1517. DOI: 10.1016/j.biochi.2009.09.005
- [4]Wrenger C, Muller IB, Butzloff S, et al. Crystallization and preliminary X-ray diffraction of malate dehydrogenase from *Plasmodium falciparum*[J]. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 2012, 68 (Pt 6): 659-662. DOI: 10.1107/S1744309112014571
- [5]Zheng N, Xu J, Wu Z, et al. *Clonorchis sinensis*: molecular cloning and functional expression of novel cytosolic malate dehydrogenase [J]. *Exp Parasitol*, 2005, 109(4): 220-207.
- [6]Nava G, Laclette JP, Bobes R, et al. Cloning, sequencing and functional expression of cytosolic malate dehydrogenase from *Taenia solium*: Purification and characterization of the recombinant enzyme[J]. *Exp Parasitol*, 2011, 128(3): 217-224. DOI: 10.1016/j.exppara.2011.03.008
- [7]Shen CJ, He A, Zheng XY, et al. Cloning and sequence analysis of the lactate dehydrogenase-like malate dehydrogenase cDNA from *Toxoplasma gondii* [J]. *Chin J Zoonoses*, 2005, 21(6): 461-466, 490. (in Chinese)
申川军, 何嵩, 郑小英, 等. 刚地弓形虫 LDH-Like MDH cDNA 的克隆和序列分析[J]. *中国人兽共患病杂志*, 2005, 21(6): 461-466, 490.
- [8]Ma GY, Zhang JZ, Yin GR, et al. *Toxoplasma gondii*: proteomic analysis of antigenicity of soluble tachyzoite antigen[J]. *Exp Parasitol*, 2009, 122(1): 41-46. DOI: 10.1016/j.exppara.2009.01.011
- [9]Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network[J]. *Proteins*, 2006, 65(1): 40-48.
- [10]Wang F, Wang HL, Yin LT, et al. Bioinformatics analysis of the structure and epitope of the gene encoding phosphoglycerate mutase 2 protein from *Toxoplasma gondii* [J]. *J Pathog Biol*, 2010, 5(11): 840-843. (in Chinese)
王芬, 王海龙, 殷丽天, 等. 刚地弓形虫磷酸甘油酸酯变位酶 2 基因编码蛋白主要特性与抗原表位的生物信息学分析[J]. *中国病原生物学杂志*, 2010, 5(11): 840-843.
- [11]Hopp TP, Woods KR. Prediction of protein antigenic determinants from amino acid sequences[J]. *Proc Natl Acad Sci USA*, 1981, 78(6): 3824-3828.
- [12]Lv YB, Wan Y, Wu YZ. Prediction of the B cell epitope for the S protein of SARS coronavirus[J]. *Acta Academiae Medicinae Militaris Tertiae*, 2004, 26(2): 101-103. (in Chinese)
吕燕波, 万瑛, 吴玉章. SARS 病毒 S 蛋白的 B 细胞表位预测 [J]. *第三军医大学学报*, 2004, 26(2): 101-103.
- [13]Delom F, Chevet E. Phosphoprotein analysis: from proteins to proteomes[J]. *Proteome Sci*, 2006, 4: 15. DOI: 10.1186/1477-5956-4-15