

耐热 β -半乳糖苷酶的结构分析及同源建模

张志刚

(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: 预测耐热 β -半乳糖苷酶的初级结构和高级结构, 为进一步了解该蛋白的结构特征提供理论依据。应用 ProtParam、ProtScale、TMHMM Server v. 2.0、SignalP 4.0 Server、SOPMA、PSORT Prediction、GENO3D 等生物信息学软件对此耐热 β -半乳糖苷酶的初级结构和空间结构进行分析和预测。该耐热 β -半乳糖苷酶由 452 个氨基酸残基组成, 无跨膜区域和信号肽, 定位于细胞质, 二级结构以 α 螺旋和无规则卷曲为主, 应用同源建模的方法, 三维结构分子模型得到了成功构建。为进一步研究耐热 β -半乳糖苷酶的功能特征以及改造奠定了基础。

关键词: 耐热 β -半乳糖苷酶; 生物信息学; 同源建模

文章编号: 1673-9078(2013)4-706-709

Structure Analysis and Homology Modeling of Thermostable

β -galactosidase

ZHANG Zhi-gang

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

Abstract: This study aimed to predict the primary and higher structures of thermostable β -galactosidase, providing theoretical basis for further study on its structural characteristics. Several kinds of bioinformatics softwares, including ProtParam, ProtScale, TMHMM Server v. 2.0, SignalP 4.0 Server, SOPMA, PSORT Prediction and GENO3D, were used to analyse and predict the primary and spatial structure of this enzyme. Result showed that the thermostable β -galactosidase was composed of 452 amino acids residuals. The enzyme protein had no signal peptide and transmembrane region, and located in the cytoplasm. Alpha Helix and Random coil were main structures in its secondary structure. By using homology modeling, three-dimensional structure of the enzyme was constructed.

Key words: thermostable β -galactosidase; bioinformatics; homology modeling

β -半乳糖苷酶(β -galactosidase, EC3.2.1.23)又称乳糖酶(lactase)^[1]、 β -半乳糖半乳糖苷转移酶^[2], 可催化水解半乳糖苷键, 将乳糖水解成为葡萄糖和半乳糖, 同时具有转移半乳糖苷的作用, 在乳糖分子的半乳糖上连接 1-4 个半乳糖, 生成低聚半乳糖(galactooligosaccharide, GOS)^[3-5]。 β -半乳糖苷酶不仅广泛应用于食品、化学、医药等领域, 而且在生物技术领域如基因工程、酶工程、蛋白质工程方面都发挥着重要作用^[6-9]。对于工业应用而言, 良好的耐热性是酶的一个最重要特性之一^[10]。使用耐热 β -半乳糖苷酶进行工业化生产具有适用范围广、转化率高、污染小等优点^[11], 但目前为止, 国内有关于热稳定性 β -半乳糖苷酶的报道很少, 而关于耐热 β -半乳糖苷酶空间结构和功能的

研究尚未见报道。

利用传统方法如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研究蛋白的空间结构和功能不但费时费力, 而且对仪器设备要求较高。随着分子生物学技术的快速发展和应用, 大量新的编码蛋白的DNA序列不断涌现, 传统的方法远远不能满足实验的需求, 因此利用生物信息学软件来预测蛋白的空间结构与功能特性, 对研究蛋白质结构和机理具有重要的理论意义与实用价值。本研究通过应用各种生物信息学软件对耐热 β -半乳糖苷酶的初级结构及高级结构进行了预测和分析, 并在三级结构的基础上进行空间结构的同源建模, 为进一步研究耐热 β -半乳糖苷酶的功能及特征奠定基础。

1 材料与方法

1.1 材料

收稿日期: 2012-11-22

基金项目: “十二五”农村领域国家科技计划课题(2012BAD28B0903)

作者简介: 张志刚(1968-), 男, 高级工程师, 主要从事食品安全与质量控制技术研究

供本文研究的蛋白序列来自美国生物信息中心(NCBI)蛋白质数据库,检索序列号为: YP_003993284.1,该耐热蛋白酶的氨基酸序列为: MSFPKGFLLWGA ATAS YQIEGAWNEDGKGESIWDRFSHQGNILYG HTGDVACDHYHRFEEDVSLMKELGLKAYRFSIAW ARIFPDGYGTVNQKGLEF YDRLINKLVENGIEPVV TIYHWDLPQKLQDIGGWANKEI VNHYFEYAMLLI NRYKDKVKKWITFNEPYCIAFLGHWHG VHAPGIK DFKVAMD VVHNIMLSHFKVVKAVKENNIDVEVGI TLNLTVPVYLQTERLG YKVSEIEREMVNLSSQLDNE LFLDPVLKGSYPQKLLNLYLVQKDLLDSQKAEKM QQEVKENFIFPDFLGINYYTRAVRLYDENS SWIFPI RWEHPAGE YTEMGWEVFPQGLFDLLMWIKEN YP QILYITENGAAYNDKVEDGRVHDQKRVE YLKH HFEAARKAIENG VDLRG YFVWSLMDNFEWAMGYT KRFGIHYVDYETQKRIKDSFYFYQQYIKENS.

1.2 方法

1.2.1 耐热β-半乳糖苷酶的初级结构分析

利用瑞士生物信息学研究所提供的ProtParam程序(<http://web.expasy.org/protparam/>)对耐热β-半乳糖苷酶的相对分子量、氨基酸残基数目、组成、蛋白理论等电点及疏水性平均值等参数进行分析。

1.2.2 耐热β-半乳糖苷酶疏水性、跨膜区域和信号肽预测

应用ProtScale程序(<http://web.expasy.org/protscale/>)对该蛋白酶的疏水性进行分析。分别利用TMHMM Server v. 2.0(<http://www.cbs.dtu.dk/services/TMHMM/>)和SignalP 4.0 Server(<http://www.cbs.dtu.dk/services/signalp/>)在线服务器对耐热β-半乳糖苷酶的跨膜区域和信号肽进行预测。

1.2.3 耐热β-半乳糖苷酶二级结构分析

利用SOPMA(http://npsa-pbil.ibcp.fr/cgi-bin/npsaautomat.pl?page=/NPSA/nps_asopma.html)对该蛋白酶的α螺旋、延伸链、β-转角结构以及无规则卷曲等进行分析预测。

1.2.4 耐热β-半乳糖苷酶的细胞定位

应用PSORT Prediction工具(<http://psort.hgc.jp/form.html>)对该蛋白酶进行细胞定位预测。

1.2.5 蛋白酶三级结构预测

利用GENO3D服务器(http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html)对该蛋白酶的空间结构进行分析预测,并利用Procheck, Errat和Verify_3D(<http://nihserver.mbi.ucla.edu/SAVES/>)对该蛋白酶的结构模型进行评估。使用Rasmol 2.7.5软件展示该蛋白酶的三级结构模型。

2 结果与分析

2.1 耐热β-半乳糖苷酶初级结构分析

通过ProtParam程序分析,结果显示,该耐热β-半乳糖苷酶由452个氨基酸组成,分子量为53.29 kDa,其分子式为C₂₄₅₇H₃₆₆₅N₆₂₅O₆₈₂S₁₃,该蛋白酶的原子总数和理论等电点分别为7442和5.8。酸性氨基酸残基数(Asp + Glu)为63,碱性氨基酸残基数(Arg + Lys)为52,疏水性的总平均值(GRAVY)为-0.428。组成该蛋白酶的氨基酸比例见表1。

表1 耐热β-半乳糖苷酶的氨基酸组成

Table 1 Amino acid composition of thermostable β-galactosidase

氨基酸	数目	摩尔比/%	氨基酸	数目	摩尔比/%
Ala	23	5.1	Leu	37	8.2
Arg	17	3.8	Lys	35	7.7
Asn	25	5.5	Met	11	2.4
Asp	27	6.0	Phe	25	5.5
Cys	2	0.4	Pro	14	3.1
Gln	18	4.0	Ser	17	3.8
Glu	36	8.0	Thr	13	2.9
Gly	31	6.9	Trp	14	3.1
His	15	3.3	Tyr	30	6.6
Ile	31	6.9	Val	31	6.9

2.2 耐热β-半乳糖苷酶疏水性、跨膜区域和信号肽预测

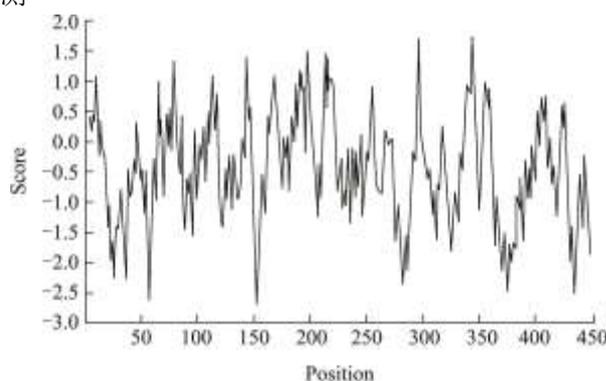


图1 耐热β-半乳糖苷酶的疏水性氨基酸分布图

Fig.1 The distribution of hydrophobic amino acids in thermostable β-galactosidase

对耐热β-半乳糖苷酶的疏水性分析如图1,根据SAPS数据库中的统计结果,该蛋白酶的疏水性氨基酸数目为186个。相关研究表明蛋白质的耐热性与其疏水内核直接相关,蛋白内部疏水内核增加则耐热性提高,反之则低。通常耐热性强的蛋白质其疏水氨基酸含量较高,主要是因为疏水性氨基酸有助于形成刚性蛋白内核。

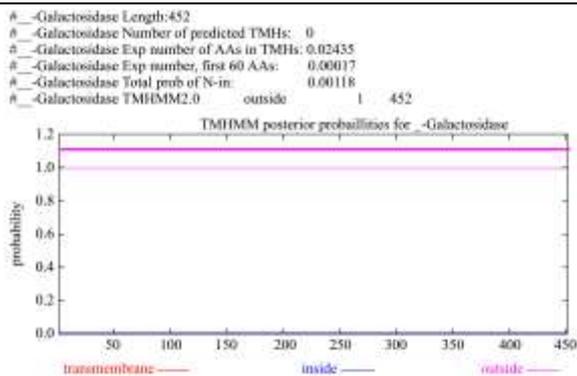


图2 耐热β-半乳糖苷酶跨膜区域预测

Fig.2 Transmembrane regional prediction of thermostable β-galactosidase

通过TMHMM Server v. 2.0在线软件对该蛋白酶的跨膜区域的预测可知,该蛋白酶的跨膜区域为零,该蛋白在跨膜区域中的氨基酸序列期望值为0.00118,存在跨膜区域时该数值应当大于18,因此该耐热β-半乳糖苷酶没有跨膜区域(图2)。应用SignalP 4.0 Server软件对该蛋白酶的信号肽进行预测,结果显示(图3),该蛋白酶不存在信号肽。

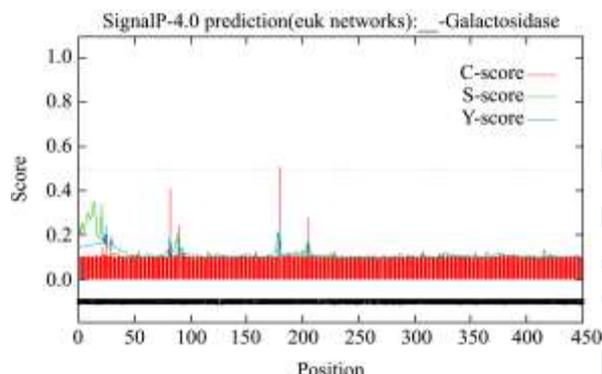


图3 耐热β-半乳糖苷酶信号肽预测

Fig.3 Signal peptide prediction of the thermostable β-galactosidase

2.3 耐热β-半乳糖苷酶二级结构预测

通过SOPMA服务器对该蛋白酶的二级结构进行分析得知,在该蛋白酶的二级结构中:α螺旋结构(Alpha helix, h)为36.73%,扩展长链(Extended strand, e)为14.60%,β-转角结构(Beta turn, t)为9.29%,无规则卷曲结构(Random coil, c)为39.38(如图4)。

2.4 耐热β-半乳糖苷酶的细胞定位

对该蛋白酶的细胞定位预测使用PSORT WWW Server程序提供的PSORT Prediction工具进行,预测结果表明,该蛋白存在于细胞质中的可能性为11%,而在细胞周质间隙以及内外膜中都无法确定是否有该蛋白的存在。

2.5 耐热β-半乳糖苷酶三维结构预测

将该蛋白酶的氨基酸序列在线提交Geno3D服务器后,从返回同源序列中选择同源性高、并且有三维

结构模型的pdb2jalB、pdb2wc3B和pdb3ahxA三个序列作为模板。提交模板后返回Model-1、Model-2、Model-3三个模型数据文件,将模型文件经RasMol2.7.5软件展示,结果如图5所示。对各个模型经过综合评价后发现,Model-1的能量最低,为-8866.41 kcal/mol,结合Procheck和Errat软件对三个模型的评价结果,发现Model-1更接近于该蛋白酶的真实的空间构象。

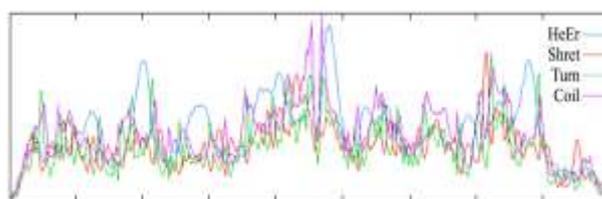
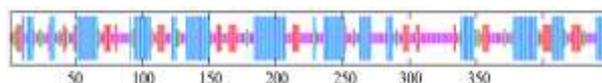


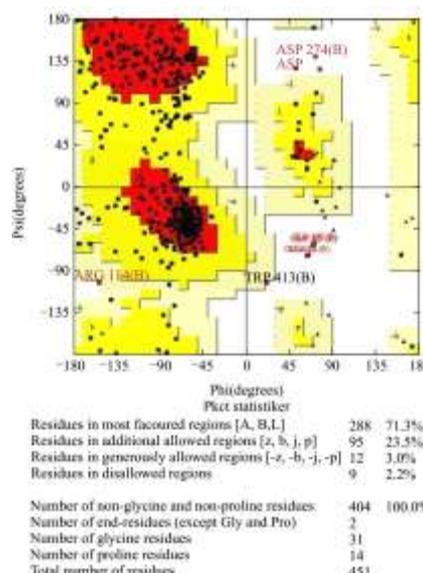
图4 耐热β-半乳糖苷酶二级结构预测

Fig.4 Secondary structure prediction of thermostable β-galactosidase



图5 不同耐热β-半乳糖苷酶的同源模型结构比较

Fig.5 Comparison of some homology models for thermostable β-galactosidase



氨基酸残基为71.3%，另外有3.0%处于最大允许区内，只有2.2%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。通过Verify_3D软件对Model-1模型的评估显示，88.05%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验证。

3 讨论

生物信息学是一门新兴的交叉学科，它主要从事对生物信息的获取、加工、储存、分配、分析和解读，综合运用数学、物理、信息科学与生物学工具，以达到解析数据中生物学含义的目的^[13]。从大量数据中提取新知识是生物信息学最重要的任务，数据库是生物信息学的主要内容，是一切生物信息学工作的基础。当前，最主要的三大数据库GenBank核酸序列数据库，SWISSPROT蛋白质数据库和PDB生物大分子结构数据库，已成为全世界从事生命科学研究的科研人员获取序列、结构和其他信息的基本来源^[14~15]。当前，对蛋白质分子空间结构与功能的研究被认为是分子生物学领域中最具挑战性的课题之一^[16]。由于传统的方法如晶体衍射、核磁共振等技术不但费时费力，而且成本较高，因此，从理论上对一个已知序列的蛋白质的空间结构进行预测的问题变得日益紧迫和重要。

在本研究中，我们应用了一系列生物信息学软件对耐热 β -半乳糖苷酶的初级和高级结构进行了分析和预测，在对该耐热 β -半乳糖苷酶的初级结构研究后发现，该蛋白酶由452个氨基酸组成，其理论等电点为5.80，酸性氨基酸残基总数多余碱性氨基酸残基总数，由此断定，该耐热 β -半乳糖苷酶属于酸性蛋白酶。该蛋白酶的疏水性的总平均值为-0.428，为负值，因此该耐热 β -半乳糖苷酶应该为水溶性蛋白酶。从TMHMM Server v. 2.0和SignalP 4.0 Server软件的预测结果显示，该蛋白酶不存在跨膜区域和信号肽。应用PSORT Prediction工具对该耐热 β -半乳糖苷酶进行细胞定位分析后发现，该蛋白酶存在于细胞质中的可能性为11%，而在细胞周质间隙以及内外膜中都无法确定是否有该蛋白的存在。在该蛋白酶的二级结构中， α 螺旋结构(Alpha helix)和无规则卷曲结构(Random coil)分别占据36.73%和39.38%，为主要的结构形式。在此次研究中，我们应用Geno3D服务器对该耐热 β -半乳糖苷酶进行三级结构的预测和分析，将该蛋白酶的氨基酸序列在线提交Geno3D服务器后，从返回同源序列中选择同源性高、并且有三维结构模型的pdb2jalB、pdb2wc3B和pdb3ahxA三个序列作为模板。提交模板后，服务器返回Model-1、Model-2、Model-3

三个模型数据文件，经过综合评价这三个模型后发现，Model-1更符合该耐热 β -半乳糖苷酶真实的空间构象。应用Verify_3D软件对Model-1模型评估后发现，有88.05%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耐热 β -半乳糖苷酶同源模型。

4 结论

本研究通过应用一系列的生物信息学软件对耐热 β -半乳糖苷酶的一级结构、亲疏水性、跨膜区、信号肽、细胞定位以及二级结构等进行了分析。并在三级结构的基础上进行空间结构的同源建模，为进一步研究耐热 β -半乳糖苷酶的功能特征以及改造奠定基础。

参考文献

- [1] Reuter S, Rusborg Nygaard A, Zimmermann W. β -Galactooligosaccharide synthesis with β -galactosidases from *Sulfolobus solfataricus*, *Aspergillus oryzae*, and *Escherichia coli* [J]. Enzyme and Microbial Technology, 1999, 25(6):509-516
- [2] 李德荣, 郑海艳, 谢庆武. β -半乳糖苷酶产生真菌的筛选及酶学性质研[J]. 广西轻工业, 2010, 5:3-4
- [3] 秦燕. 乳糖酶及其合成低聚半乳糖的研究[D]. 广州: 华南理工大学, 2001
- [4] 归莉琼, 魏东芝. 生物活性物质—低聚半乳糖[J]. 中国乳品业, 1998, 26(4):9-12
- [5] 郑建仙, 李漩. 低聚半乳糖的制备[J]. 食品与机械, 1999, 1:20-21
- [6] 秦燕, 宁正祥, 胡新宇. β -半乳糖苷酶的应用研究进展[J]. 沈阳农业大学报, 2000, 31(6):595-599
- [7] Jung Min Lee, Dae Kyun Chung, Joo Heon Park, et al. Cloning and nucleotide sequence of the β -galactosidase gene from *Lactis spp. Lactis* ATCC7962 [J]. Biotechnology Letters, 1997, 19(2): 179-183
- [8] Ranwala A P, Suematsu C, Masuda H. The role of β -galactosidase in the modification of cell wall components during muskmelon fruit ripening [J]. Plant Physiol, 1992, 100: 1318-1325
- [9] Romanos M A, Clare J J, Beesley K M, et al. Recombinant *Bordetella pertussis* pertactin (P69) from the yeast *Pichia pastoris*: high-level production and immunological properties [J]. Vaccine, 1991, 9: 901-906
- [10] Bin Liu, Yiqian Wang, Xiaobo Zhang. Characterization of a recombinant maltogenic amylase from deep sea thermophilic *Bacillus* sp. WPD616 [J]. Enzyme and Microbial Technology, 2013, 53(1): 1-7

- 2006, 39:805-810
- [11] 葛佳佳,陈卫,张灏.耐热 β -半乳糖苷酶基因在大肠杆菌中的表达[J].无锡轻工大学学报,2002,21(3):296-298
- [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] 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
- [14] 秦惠基,胡家荣.生物信息学的兴起与展望[J].广东药学院学报,1999,1:28-29
- [15] 任秀艳,高永闯,李美茹,等.HrpNEa蛋白的结构与功能分析[J].安徽农业科学,2010,38(5):2254-2256
- [16] 张建平,李峰,李荣,等.THE蛋白的结构分析与功能预[J].湖南文理学院学报(自然科学版),2006, 18(1): 24-26

现代食品科技