

基于序列比对分析木聚糖 1,4- β -木糖苷酶结构差异和分子进化规律

林燕玲¹, 蔡雨晨¹, 李利君^{1,2*}, 倪辉^{1,2}

(1. 集美大学海洋食品与生物工程学院, 福建厦门 361021)

(2. 福建省食品微生物与酶工程重点实验室, 厦门市食品生物工程技术研究中心, 福建厦门 361021)

摘要: 木聚糖 1,4- β -木糖苷酶 (EC 3.2.1.37) 是一种外切水解酶, 近年来该酶在工业上的使用越来越广泛, 目前对于木聚糖 1,4- β -木糖苷酶结构的研究相对缺少。为了分析木聚糖 1,4- β -木糖苷酶的进化关系以及结构差异, 利用 MEGA X 10.1.8 进行木聚糖 1,4- β -木糖苷酶基因序列多序列比对, 构建系统进化树, 结果分析显示木聚糖 1,4- β -木糖苷酶可分为两大类, 从中筛选出 12 条代表性序列进行氨基酸序列多序列比对, 发现只有 15 个较保守位点, 在进化上呈现不保守的特点。利用生物信息学工具预测分析 12 条序列对应的酶的理化性质、信号肽, 预测结果表明所有酶都表现为亲水性, 有胞内与胞外两种分泌方式。利用 Modeller 9.24 以及 Phyre 2 进行三级结构建模, 根据建模结果木聚糖 1,4- β -木糖苷酶分为 β -折叠桶、碗状结构、(α/β)₈ 桶状结构三种代表性结构, 利用分子对接辅助结合口袋大小计算, 推测底物的大小对其特征结构的形态存在一定程度的影响。基于木聚糖 1,4- β -木糖苷酶结构规律的研究, 可以精确设计点突变, 降低盲目性, 获得需要的理化性质, 为木聚糖 1,4- β -木糖苷酶的改造提供生物信息学指导, 其次为进一步的木聚糖 1,4- β -木糖苷酶结构与功能关系研究提供基础, 进一步拓宽对木聚糖 1,4- β -木糖苷酶在食品、医药等领域应用。

关键词: 木聚糖 1,4- β -木糖苷酶; 分子差异; 进化规律; 多序列比对; 蛋白质建模

文章编号: 1673-9078(2022)08-138-147

DOI: 10.13982/j.mfst.1673-9078.2022.8.0902

Analysis of Structure Difference and Molecular Evolution of Xylan 1,4- β -Xylosidase Based on Sequence Alignment

LIN Yanling¹, CAI Yuchen¹, LI Lijun^{1,2*}, NI Hui^{1,2}

(1.College of Marine Food and Biology Engineering, Jimei University, Xiamen 361021, China)

(2.Fujian Provincial Key Laboratory of Food Microbiology and Enzyme Engineering, Research Center of Food Biotechnology of Xiamen City, Jimei University, Xiamen 361021, China)

Abstract: Xylan 1,4- β -xylosidase (EC 3.2.1.37) is an exohydrolase that has become more widely used in industry in recent years. At present, there has been a relative lack of research on the structure of xylan 1,4- β -xylosidase. In order to analyze the evolutionary relationship and structural differences of xylan 1,4- β -xylosidase, MEGAX10.1.8 was used to compare the multiple sequences of xylan 1,4- β -xylosidase gene sequences and construct a phylogenetic evolution tree. The analysis of results showed that xylan 1,4- β -xylosidase could be divided into two categories, from which 12 representative sequences were screen and selected for the comparison and alignment of multiple amino acid sequences. Only 15 relatively conserved sites were found, showing non-conservative characteristics. The physicochemical properties and signal peptides of the enzymes corresponding to the 12 sequences were predicted and analyzed by bioinformatics tools. The prediction results showed that all enzymes were hydrophilic, and possessed intracellular and extracellular secretion modes. Modeller 9.24 and Phyre2 were used to construct

引文格式:

林燕玲,蔡雨晨,李利君,等.基于序列比对分析木聚糖 1,4- β -木糖苷酶结构差异和分子进化规律[J].现代食品科技,2022,38(8):138-147,+43

LIN Yanling, CAI Yuchen, LI Lijun, et al. Analysis of structure difference and molecular evolution of xylan 1,4- β -xylosidase based on sequence alignment [J]. Modern Food Science and Technology, 2022, 38(8): 138-147, +43

收稿日期: 2021-08-16

基金项目: 国家自然科学基金面上项目 (31871765); 福建省高校产学研合作项目 (2020N5010)

作者简介: 林燕玲 (1998-), 女, 硕士研究生, 研究方向: 生物化学和分子生物学, E-mail: 202011710037@jmu.edu.cn

通讯作者: 李利君 (1973-), 女, 博士, 教授, 研究方向: 食品加工中的生物学问题及相关应用技术, E-mail: ljli@jmu.edu.cn

models for the tertiary structure. According to the modeling results, xylan 1,4- β -xylosidase was divided into three representative structures: β -folded barrel, bowl-like structure, and (α/β)₈ barrel structure. Molecular docking was used to assist the calculation of the size of the binding pocket, and it was speculated that the size of substrate has a certain degree of effect on the morphology of its characteristic structure. Based on the study of the structural pattern of xylan 1,4- β -xylosidase, point mutations can be precisely designed, blindness can be reduced, and desired physicochemical properties can be obtained, to provide bioinformatics guidance for further research on the modification of xylan 1,4- β -xylosidase, then offers a basis for further study on the relationship between the structure and function of xylan 1,4- β -xylosidase, thereby further broadening the application of xylan 1,4- β -xylosidase in food, medicine and other fields.

Key words: xylan 1,4- β -xylosidase; molecular differences; evolutionary laws; multiple sequence alignment; protein modelling

木聚糖 1,4- β -木糖苷酶 (xylan 1,4- β -xylosidase, EC 3.2.1.37) 属于木聚糖酶系^[1], 是一种糖苷水解酶^[2], 具有外切性质^[3]。大多数的木聚糖 1,4- β -木糖苷酶存在于真菌中, 也有部分存在于细菌中^[4], 极少存在于酵母菌中^[5]。因其可以作用于低聚木糖的非还原端释放木糖^[6], 不仅在造纸工业、医药领域应用广泛^[7], 在食品技术中有很大的应用潜能, 如利用木聚糖 1,4- β -木糖苷酶水解低聚木糖的性质制备天然食品甜味剂木寡糖; 在啤酒酿造中代替乳化剂、氧化剂, 降低啤酒的黏度与浊度; 水解小麦、玉米等谷物中的半纤维素, 促进营养的消化吸收等^[4]。

目前木聚糖 1,4- β -木糖苷酶的研究集中在筛选、纯化、表征和克隆表达上, 缺乏对结构的系统研究。对木聚糖 1,4- β -木糖苷酶的结构进行系统的分类研究, 有助于今后研究其结构与功能之间的关系, 进一步拓展其在食品等领域的应用。多序列比对 (Multiple sequence alignment)^[8]是将三个及三个以上的生物学序列进行比对, 比对结果作为探究序列同源性的依据, 揭示序列保守性, 与此同时多重序列比对在构建系统进化树、预测蛋白质二级结构以及三级结构、确定未知序列家族等方面也有广泛应用。系统进化树^[9]是研究生物进化过程的基本框架, 三维结构叠合可以反映蛋白质立体结构差异。分子对接可以预测两个或两个以上已知结构结合方式的计算机算法, 广泛应用与分子识别研究^[10]。本研究利用生物信息学方法, 分析木聚糖 1,4- β -木糖苷酶的基因序列, 对基因序列进行多重序列比对, 构建系统进化树, 同时预测其理化性质, 并对该酶的三维结构进行建模、结构叠合与分子对接, 进而分析木聚糖 1,4- β -木糖苷酶分子进化规律以及结构差异, 为木聚糖 1,4- β -木糖苷酶功能的进一步研究提供生物信息学参考。

1 材料与方法

1.1 数据获取

国际生物技术信息中心 NCBI ([http://www.ncbi.](http://www.ncbi.nlm.nih.gov/)

<http://www.ncbi.nlm.nih.gov/>)。

1.2 实验方法

登录 NCBI 网站, 在搜索栏中输入关键词 “xylan 1,4- β -xylosidase”, 收集所有木聚糖 1,4- β -木糖苷酶的基因序列、氨基酸序列以及相关信息, 以 fasta 格式保存。利用 MEGA X 10.1.8^[11]软件构建系统进化树进入 ClustalW 2.0^[12]软件, 设置 Gap opening penalty 为 10.00, Gap extension penalty 为 0.20, Delay divergent sequence 为 30%, 其他参数默认, 对蛋白质序列进行多重序列比对, 序列比对结果用 Esprict 3^[13] (<http://esprict.ibcp.fr/ESPrict/ESPrict/>)与 WebLogo^[14] (<http://weblogo.berkeley.edu/logo.cgi>) 在线工具进行显示。通过 ExPASy-ProtParam (<https://web.expasy.org/protparam/>)、ExPASy-ProtScale^[15] (<https://web.expasy.org/protscale/>)在线分析工具分析氨基酸序列的理化性质、疏水性, 采用 TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>)、PSIPRED^[16] (<http://bioinf.cs.ucl.ac.uk/psipred/>)、Signal P 5.0^[17] (<http://www.cbs.dtu.dk/services/SignalP-5.0/>)在线分析工具寻找蛋白质跨膜区域以及跨膜区域位置、分析信号肽。蛋白质的三维结构建模采用了两种方法, 第一种为同源建模法, Modeller 9.24^[18]软件, 穿针引线法选用了 Phyre2.0^[19] (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) 在线服务器进行建模。建模结构用 SAVES 5.0 (<https://servicesn.mbi.ucla.edu/SAVES/>)进行模型质量评估。使用 UCSF Chimera^[20] 1.14 软件进行结构叠合, 分析结构差异。利用 Autodock Vina^[21]工具进行分子对接, 结合 Proteins Plus (<https://proteins.plus/>) 计算结合口袋大小, 探讨结构进化规律。

2 结果与分析

2.1 木聚糖 1,4- β -木糖苷酶基因序列系统进化树构建以及代表序列筛选

利用 NCBI 网站收集木聚糖 1,4-β-木糖苷酶的基因序列，除去不完整序列，共获得来自 GH3、GH39、GH43 家族的 72 条完整序列。72 条序列来源广泛，包括原核生物链霉菌、单胞菌，真核生物曲霉、酵母菌、镰刀菌，以及真核动、植物。将获得的 72 条序列进行多序列比对构建系统进化树（图 1）。从进化树的结果来看，木聚糖 1,4-β-木糖苷酶可分成两大类。第一类

含有 38 条序列，进一步分为三小类，编号为 I、II、III，第一小类包括 8 条真菌来源序列，第二小类包括 10 条细菌与真菌来源序列，第三小类包括 20 条真菌来源与真核动物来源序列。剩余序列构成第二类，同样分为三小类，编号为 IV、V、VI。第四小类的 12 条序列都源于细菌，第五小类仅有 1 条细菌来源与 3 条真核动物来源的序列，第六类则全为细菌来源。

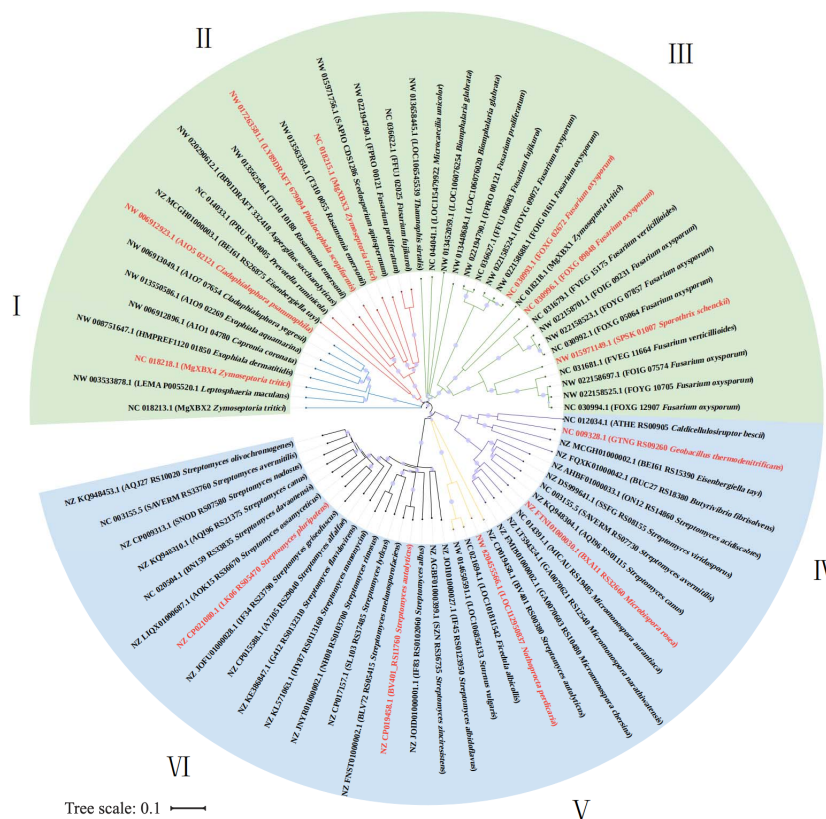


图 1 72 条木聚糖 1, 4-β-木糖苷酶构建的系统进化树

Fig.1 Phylogenetic tree based on 72 xylans 1,4-β-xylosidase sequences

注：图中圆点大小表示 Bootstrap 大小，距离标尺表示序列间差异数值的单位长度。

表 1 12 条代表性序列对应的蛋白登录号

Table 1 Protein accession numbers of the 12 representative sequences

Genebank	基因符号	蛋白登录号	来源
NC_018218	MgXBX4	XP_003857132.1	<i>Zymoseptoria tritici</i>
NC_018215	MgXBX3	XP_003853364.1	<i>Zymoseptoria tritici</i>
NZ_CP014485	SPSK_01007	XP_016584170.1	<i>Sporothrix schenckii</i>
NC_030996	FOXG_09848	XP_018247270.1	<i>Fusarium oxysporum</i>
NC_030993	FOXG_02672	XP_018236340.1	<i>Fusarium oxysporum</i>
NC_009328.1	GTNG_RS09260	WP_008880070.1	<i>Geobacillus thermodenitrificans</i>
NW_020455566	LOC112950837	XP_025899255.1	<i>Nothoprocta perdicaria</i>
NW_006912923	A1O5_02121	XP_007740927.1	<i>Cladophialophora psammophila</i>
NZ_CP019458.1	BV401_RS11760	WP_079257095.1	<i>Streptomyces autolyticus</i>
NZ_FTNI01000030.1	BXA11_RS32660	WP_204053806.1	<i>Microbispora rosea</i>
NZ_CP021080.1	LK06_RS05470	WP_039651703.1	<i>Streptomyces pluripotens</i>
NW_017263581	LY89DRAFT_679094	XP_018078127.1	<i>Phialocephala scopiformis</i>

	20	70	80	90	100	130	180	300	310	360	390	400	430	450
BV401_RS11760	GVATLAL	FTHTQYSADQGTESGRK	..AAAAL	LSG	HIMGW	MRKTG	LAGKIAR	IPDAFGATRKFRDVT	RWL	AQGGTA	GGRA	GRFA	DDDAVL	GRSFG
LK06_RS05470	GAVVLVL	FTHTQYSADQGTIAAIR	..RVEQL	LRK	HIMGW	FVRASG	LAATVAR	LPDEFAATGKLRVSE	VWV	AEGGAD	GGA	SRFA	SRSTVL	GKRFD
GTNG_RS09260	GRGLAL	FTYIDRIFDSYLALNIR	..PFIEL	GFM	QTVFY	FVERYG	ICGGSD	VALHSIPKPTFHTFM	FFN	VPFSS	GRPR	ETLR	KNEVTL	.TYIG
BXA11_RS32660	GRFDLAL	FTYVDQVIDAYLDLGA	..PFVEL	GFM	QTVFY	LVDRYG	ISPGSD	LTHRQIKKPTYHLYAF	MA	IPVGS	GRPA	DALR	RHEVTL	RRLLG
MgXBX3	FTIEEKI	FPQPILMGAAFDDDLIK	..DVAT	IST	KDSRW	LQGG.D	LNGIPTC	LPLAITKDKIAVIGD	WA	VQMG	GGVA	GRLP	FADMKI	G.EHG
LY89DRAFT_679094	FTLEELI	FPQPILSIAGLNRSLIN	..QIGSI	IST	RAPIW	LQGGVD	VNGVPS	LPLASTVKS.IALIGP	WA	VLQMG	GGAA	GRLV	FVNFTA	TTNAG
MgXBX4	SDPKPLI	MADYHVLSMGEIGGKVT	..DHGKI	FGL	..QLW	IGVAIG	LVDDDG	KPLAADDHRRFFEA	AWM	VQGS.	GKNH	AKIR
A1O5_02121	MAPQPLV	MADYHVLSLSEPGGPVT	..DHGV	LRA	..QLW	IGVAVG	FVDDDG	ALLAADDHRRFFEA	AWM	VLGWT.	GINH	VKMR
LOC112950837	LSPGEMV	FSHELVRVANATATEV	..RAKHNY	FSS	RHPLW	LQGG.H	INGVPAC	LPLREPAGRIAVVGP	PFA	LLLFNA	AGLA	AGRL	NLSVSV	.WEQP
FOXG_09848	LVSGQTI	YASKDLINWRHIGNALV	..RQSQ	MPKM	PTLRY	YVTTSN	WNGTGR	YTDAEHIDFSSGIP	SSFV	LQHLSL	GIDT	GSMI	QASGNI	NGGKG
SPSK_01007	YSDVKKK	WDHRRRAGGPQNRMTF	AGIIL	QEFDPK	ALALT	AVTLAR	ETAIQPA	TPLAPGVP	LPATGKVR	LAL	RLLA	HGVD	VQGGA	GGREG
FOXG_02672	VVLFQPV	YHSKDLNSWTLKSHA	LRPEQL	LYGT	GRFWL	LFNWS	LPHNSTA	VPRQWDDFSRAKLD	SSW	RNKIGI	TELN	SGLA	SIASSN	ATGDG

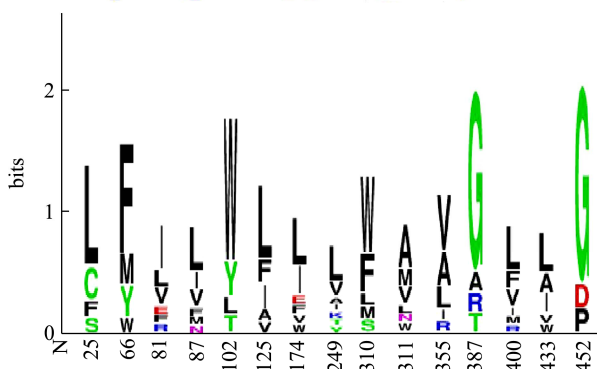


图 2 12 条木聚糖 1,4-β-木糖苷酶代表性序列保守位点

Fig.2 12 Representative sequence conserved sites of xylan 1,4-β-xylosidase

图 1 的进化分类结果表明木聚糖 1,4-β-木糖苷酶的进化进程不仅存在细菌与细菌、真菌与真菌之间的基因转移，同时也存在细菌与真菌之间、细菌与真核动物、真菌与真核动物之间的基因转移。分支代表进化谱系随时间的变化，分支长度越长，代表序列的变化越大，进化树中细菌来源的序列分支长度较短，而真菌以及真核生物来源的序列分支长度相对长，说明在木聚糖 1,4-β-木糖苷酶进化过程中，真菌来源与真核动物来源更易发生变化，这可能与原核生物与真核生物细胞以及基因组之间的差异有关。以进化树的聚类为样本单位，按照分层抽样的原则，从进化树中挑选 12 条木聚糖 1,4-β-木糖苷酶代表性序列 (表 1)，进行进一步的结构分类研究。

2.2 木聚糖 1,4-β-木糖苷酶氨基酸序列分析

2.2.1 序列保守位点分析

利用 ClustalW 2.0 软件对 12 条代表性氨基酸序列进行多重序列比对，构建保守序列谱^[22]，用 Esprint 3 在线工具显示 12 条序列的保守位点 (图 2)。木聚糖 1,4-β-木糖苷酶存在 15 个较保守位点，位于以 BV401_RS11760 为基准的 25、66、81、87、102、125、174、294、310、311、355、387、400、433、452 号位点。虽然保守位点分布分散，但在进化过程中的保守性相对较好，推测可能是与酶的结构或功能密切相关的氨基酸位点。没有发现超保守位点，说明已选择的木聚糖 1,4-β-木糖苷酶代表性序列进化过程相对独

立，可以从中总结其进化及结构规律。

2.2.2 氨基酸序列系统进化树分析

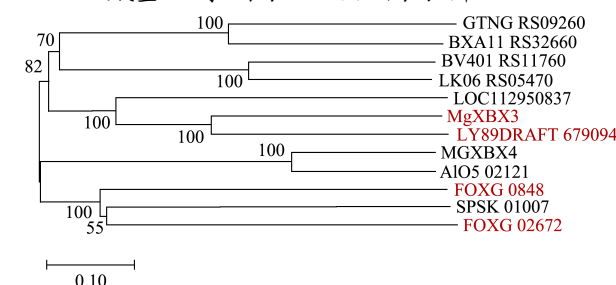


图 3 12 条氨基酸序列系统进化树

Fig.3 Phylogenetic tree of 12 amino acid sequences

利用 MEGA X 软件，对 12 条代表性序列的氨基酸序列进行系统进化树的构建。氨基酸序列系统进化树的分析结果 (图 3) 将 12 条代表性序列分成三大类，GTNG_RS09260、BXA11_RS32660、BV401_RS11760、LK06_RS05470 为第一类，属于 GH39 家族，GH3 家族的 LOC112950837、MgXBX3、LY89DRAFT_679094 为第二类，FOXG_09848、FOXG_02672、SPSK_01007、MgXBX4、A1O5_02121 为第三类，属于 GH43 家族。系统进化树的分类与其家族相对应，且与基因序列进化的分类具有一致性，表示木聚糖 1,4-β-木糖苷酶的序列的家族划分是以序列相似性为基础。利用 Signal P 5.0 预测氨基酸序列的信号肽 (表 2)，发现只有 LY89DRAFT_679094、MgXBX3、FOXG_09848、FOXG_02672 四条序列存在信号肽，均匀的分布在 GH3 家族与 GH43 家族的聚

类上,说明木聚糖 1,4-β-木糖苷酶在进化过程中分出了胞外酶与胞内酶,但信号肽的有无并不能体现其结构进化规律。

2.2.3 氨基酸序列理化性质分析

用 ExPASy-ProtParam 在线工具对 12 条代表性氨基酸序列进行理化性质的预测^[23],结果如表 3 所示。代表序列的氨基酸数在 324~829 之间,分子质量(Mr)最大为 89769.18,最小为 36301.39,相差较大。除了 LOC112950837、BV401_RS11760 的预测等电点分别为 7.95、8.81,大于 7.40 为偏碱性蛋白质外,其余的都为酸性蛋白质。LOC112950837 源于无尾线虫,而 BV401_RS11760 源于自溶链霉菌,对应物种都能较好的适应偏碱性的环境,符合酶的特征与环境相互适应的关系。比较负电荷残基总数(Asp+Glu)与正电荷残基总数(Arg+Lys)的数量,从整体上来看,几乎所有的序列都带负电荷较多,说明其大多为酸性蛋白,这一预测结果与等电点预测结果相符。序列的总原子数相差较大,极差为 7514,脂肪系数的范围为 65.68~84.59,序列的总平均亲水性为-0.04~-0.64,皆为负数,表现为亲水性蛋白,易溶于水,这为木聚糖 1,4-β-木糖苷酶表现出水解性质提供基础条件。

2.2.4 氨基酸序列疏水性分析

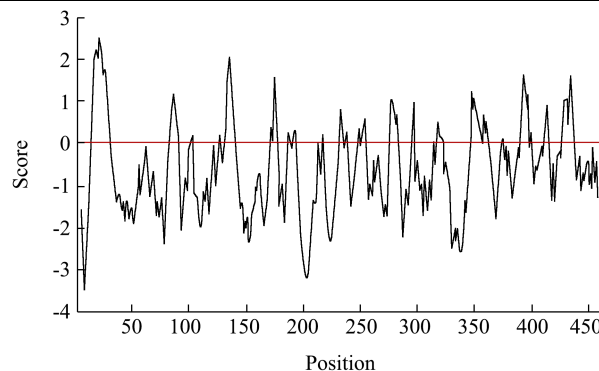


图 4 BV401_RS11760 疏水性预测图

Fig.4 Prediction chart of BV_401RS 11760 hydrophobicity

使用 ExPASy-ProtScale 在线工具对 12 条代表性氨基酸序列进行疏水性分析(图 4)。分析结果以 0 位基准,在 0 以下的峰为亲水峰,以上的峰则为疏水峰,根据亲水峰和疏水峰的数量比较来判断蛋白质的疏水性。木聚糖 1,4-β-木糖苷酶 12 氨基酸序列疏水性表明,所有的氨基酸序列亲水峰的数量都明显多于疏水峰的数量,表现为亲水性,与 ExPASy-ProtParam 对氨基酸序列疏水性的一致,进一步证明木聚糖 1,4-β-木糖苷酶为亲水性蛋白,说明在蛋白肽链折叠的过程中,亲水性氨基酸更倾向于在蛋白外部,使蛋白具有可溶性。

表 2 信号肽预测结果

Table 2 Signal peptide prediction results

基因符号	蛋白登录号	裂解位点	概率
LY89DRAFT_679094	XP_018078127.1	23 and 24: TAA-QD	0.68
MgXBX3	XP_003853364.1	18 and 19: ATA-QL	0.81
FOXG_09848	XP_018247270.1	17 and 18: VSG-QT	0.66
FOXG_02672	XP_018236340.1	21 and 22: AQA-IW	0.94

表 3 12 条代表性木聚糖 1,4-β-木糖苷酶理化性质

Table 3 Physicochemical properties of 12 representative xylan 1,4-β-xylosidase

基因符号	氨基 酸数	分子 质量	等电 点 pI	(Asp+Glu)	(Arg+Lys)	分子式	总原 子数	脂肪 系数	总平均 亲水性
MgXBX4	324	36301.39	4.80	52	13	C ₁₆₄₃ H ₂₄₁₅ N ₄₂₃ O ₄₉₃ S ₁₀	4984	65.68	-0.50
MgXBX3	785	85737.15	4.70	83	49	C ₃₈₃₅ H ₅₈₃₅ N ₁₀₀₃ O ₁₁₈₀ S ₂₇	11880	79.32	-0.22
FOXG_09848	554	60979.59	6.17	48	44	C ₂₇₅₂ H ₄₁₆₉ N ₇₁₇ O ₈₂₅ S ₁₅	8478	70.52	-0.31
FOXG_02672	559	63268.35	6.39	52	49	C ₂₈₉₇ H ₄₃₀₃ N ₇₅₉ O ₈₂₀ S ₁₂	8791	74.67	-0.40
LY89DRAFT_679094	797	85855.01	4.35	77	34	C ₃₈₂₉ H ₅₇₉₆ N ₁₀₀₀ O ₁₂₂₁ S ₁₅	11861	81.59	-0.16
LOC112950837	736	79125.36	7.95	64	66	C ₃₅₅₄ H ₅₄₉₇ N ₁₀₀₃ O ₁₀₀₃ S ₂₅	11082	84.02	-0.04
SPSK_01007	829	89769.18	5.85	108	89	C ₃₉₅₄ H ₆₁₅₁ N ₁₁₈₉ O ₁₁₉₄ S ₁₀	12498	82.64	-0.42
GTNG_RS09260	506	58544.73	5.69	66	54	C ₂₆₇₈ H ₄₀₄₆ N ₆₉₂ O ₇₅₆ S ₁₆	8188	81.42	-0.30
BV401_RS11760	463	50806.79	8.81	58	63	C ₂₂₃₇ H ₃₄₃₄ N ₆₆₈ O ₆₆₇ S ₁₄	7020	60.17	-0.64
BXA11_RS32660	519	57974.33	5.71	70	54	C ₂₆₀₁ H ₃₉₈₃ N ₇₄₇ O ₇₄₈ S ₉	8088	84.59	-0.30
LK06_RS05470	471	52006.31	6.08	59	53	C ₂₃₁₉ H ₃₅₂₀ N ₆₆₀ O ₆₈₂ S ₁₄	7195	69.41	-0.45
A1O5_02121	331	37091.36	5.05	48	31	C ₁₆₇₈ H ₂₄₈₂ N ₄₄₀ O ₄₉₉ S ₉	5108	69.27	-0.45

表 4 12 条代表性氨基酸序列建模结果

Table 4 Results of modeling 12 representative amino acid sequences

基因符号	Genebank	Templates	Verify 3D 得分
MgXBX4	NC_018218	5glk	100.00
MgXBX3	NC_018215	5ae6	95.31
FOXG_09848	NC_030996	5jow	93.14
FOXG_02672	NC_030993	5joe	96.97
BXA11_RS32660	NZ_FTNI01000030.1	2bfg	94.61
LOC112950837	NW_020455566	5z87	93.17
A1O5_02121	NW_006912923	5gln	98.19
GTNG_RS09260	NC_009328.1	2bfg	91.90
SPSK_01007	NZ_CP014485	5zqj	93.90
LY89DRAFT_679094	NW_017263581	5ae6	95.03
BV401_RS11760	NZ_CP019458.1	5bxa	84.23
LK06_RS05470	NZ_CP021080.1	5bxa	82.38

表 5 结构分类表

Table 5 Structural classification

分类类别	基因符号
第一大类	GTNG_RS09260
	BXA11_RS32660
	BV401_RS11760
	LK06_RS05470
第二大类	LOC112950837
	MgXBX3
	LY89DRAFT_679094
第三大类第一小类	A1O5_02121
	MgXBX4
第三大类第二小类	FOXG_09848
	FOXG_02672
	SPSK_01007

三条序列,属于 GH3 家族,由 $(\alpha/\beta)_8$ TIM 折叠桶、 $(\alpha/\beta)_6$ 三明治结构、FnIII型 β 折叠片层三个结构域共同组成, $(\alpha/\beta)_8$ TIM 折叠桶、 $(\alpha/\beta)_6$ 三明治结构形成了一个上方开口大,下方开口小的碗状结构;A1O5_02121、MgXBX4、FOXG_09848、FOXG_02672 和 SPSK_01007 五条 GH43 家族序列共同组成了第三类(图 7),这一类的木聚糖 1,4- β -木糖苷酶都存在一个五叶的 β 折叠桶。此外,通过结构叠合,发现 C 末端延伸的 β 折叠片层结构可将第三类更为细致的划分成两个小类,A1O5_02121、MgXBX4 为第一小类(图 8a),只包含五叶 β 折叠桶这一基本结构,第二小类(图 8b)的 FOXG_09848、FOXG_02672 和 SPSK_01007 则在 C 末端连接一个 β 折叠片层,与 Rohman 等^[24]对木糖苷酶的结构分类结果一致,有文献报道这种 C 端的 β 折叠片层对木聚糖 1,4- β -木糖苷酶的催化活性至关重要^[25]。这一分类与 GH43 家族氨基酸序列进化树聚类相符,暗示 C 端延伸的 β -折叠片层可以反映 GH43 家族木聚糖 1,4- β -木糖苷酶的结构规律。

2.3 木聚糖 1,4- β -木糖苷酶结构分析

利用 Modeller 9.24 软件以及 Phyre2 在线服务器对 12 条代表性氨基酸序列进行三维结构建模,利用 SAVES 5.0 对建模结果进行 Verify 3D 得分计算,评估模型质量。根据 SAVES 5.0 提供的 Verify 3D 得分,木聚糖 1,4- β -木糖苷酶皆在 80% 以上(表 4),说明模型质量良好,可进行进一步的结构分析。

借助 Chimera 1.14 软件进行可视化以及结构叠合,叠合结果显示,依据木聚糖 1,4- β -木糖苷酶的特征结构,可分成三个大类。糖苷水解酶 GH39 家族的 GTNG_RS09260、BXA11_RS32660、BV401_RS11760、LK06_RS05470 四条序列组成第一类(图 5),此类结构都拥有一个 $(\alpha/\beta)_8$ TIM 折叠桶结构;第二类(图 6)包含 LOC112950837、MgXBX3、LY89DRAFT_679094

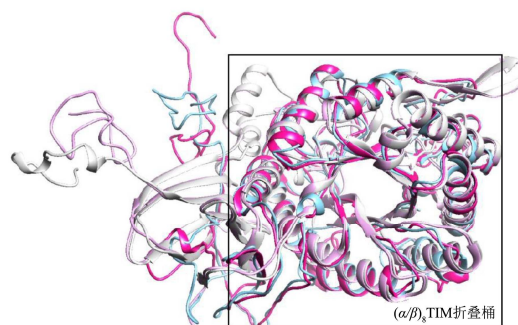


图 5 木聚糖 1,4- β -木糖苷酶第一大类叠合图

Fig.5 Overlay of the first major class of xylan 1,4- β -xylosidase

注:白色:GTNG_RS09260;紫红色:LK06_RS05470;粉红色:BXA11_RS32660;浅蓝色:BV401_RS11760。

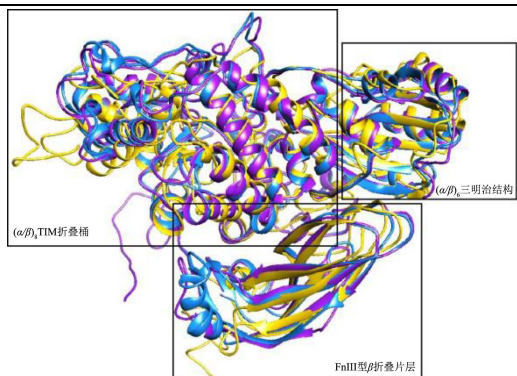


图 6 木聚糖 1,4-β-木糖苷酶第二大类叠合图

Fig.6 Overlay of the second major class of xylan 1,4-β-xylosidase

注：黄色：LOC112950837；紫色：MgXBX3；蓝色：LY89DRAFT_679094。

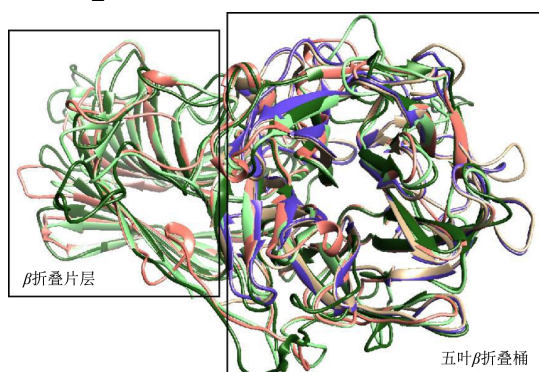


图 7 木聚糖 1,4-β-木糖苷酶第三大类叠合图

Fig.7 Overlay of the third major class of xylan 1,4-β-xylosidase

注：米黄色：A1O5_02121；蓝紫色：MgXBX4；橙色：FOXG_09848；浅绿色：FOXG_02672；深绿色：SPSK_01007。

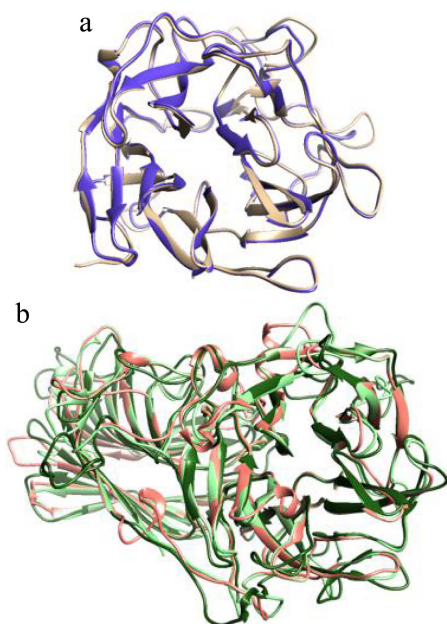


图 8 木聚糖 1,4-β-木糖苷酶第三大类的两个小类叠合图

Fig.8 Overlay of two subclasses of the third major class of xylan 1,4-β-xylosidases

注：a 为第一小类叠合图，b 为第二小类叠合图。

目前已有的研究报道中，木聚糖 1,4-β-木糖苷酶的底物谱较广，不同家族的酶对应水解不同的底物。表 6 列举了不同家族木聚糖 1,4-β-木糖苷酶的酶学性质，GH3 家族的木聚糖 1,4-β-木糖苷酶无法水解低聚木糖，但可以作用于天然糖苷类底物以及多种人工底物，如 pNPX（对硝基苯基-β-D-吡喃木糖苷）、pNPA（对硝基苯基-α-L-阿拉伯呋喃苷）、pNPG（对硝基苯基-β-D-吡喃葡萄糖苷）；GH39 家族木聚糖 1,4-β-木糖苷酶表现出广泛的底物特异性，不仅能水解天然糖苷类底物，对低聚木糖和多聚木糖同样也有水解作用；GH43 家族木聚糖 1,4-β-木糖苷酶仅能水解人工底物 pNPX、pNPG 以及低聚木糖，需要内切木糖苷酶的协同作用下才能水解多聚木糖。

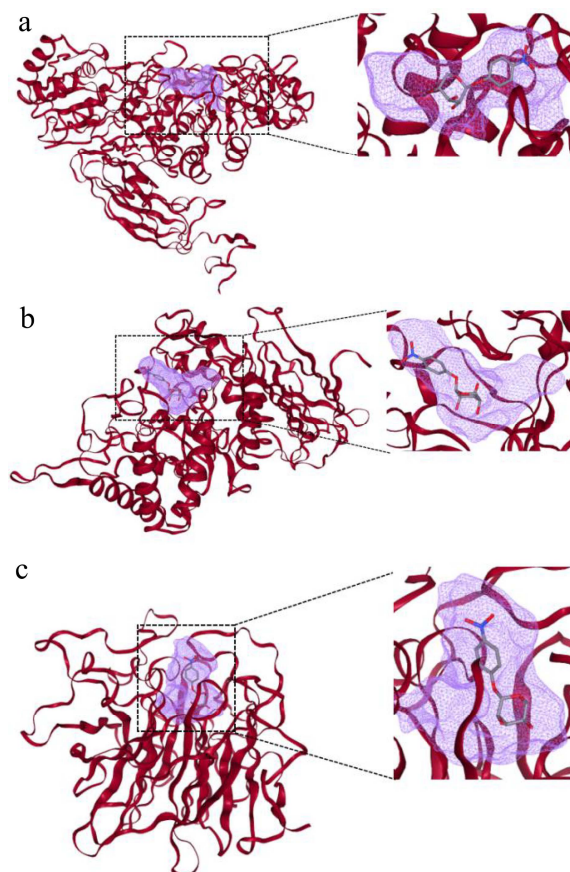


图 9 不同分类木聚糖 1,4-β-木糖苷酶结合口袋

Fig.9 Binding pockets of xylan 1,4-β-xylosidases from different classifications

注：a：LOC112950837；b：BXA11_RS32660；c：A1O5_02121。

以所有木聚糖 1,4-β-木糖苷酶的共有底物 pNPX 为配体，与属于不同分类的木聚糖 1,4-β-木糖苷酶 LOC112950837、BXA11_RS32660、A1O5_02121 进行分子对接确定结合口袋位置，计算结合口袋大小(图 9，表 7)。计算结果显示 GH3 家族的结合口袋宽且浅

(图 9a), 适合分子量大的天然糖苷类底物如人参皂苷等与之结合; 此外 GH3 家族木聚糖 1,4- β -木糖苷酶蛋白质三维结构与 α -阿拉伯呋喃糖苷酶以及 β -葡萄糖苷酶结构相似度极高, 可以水解 pNPX、pNPA、pNPG, 表现出三酶性质^[26]。GH39 家族木聚糖 1,4- β -木糖苷酶结合口袋大、开口开放 (图 9b), 这种大且开的口袋

赋予了酶容纳大分子量底物与木糖分支链的能力, 水解底物囊括人工底物 pNPX、天然糖苷、低聚木糖与多聚木糖, 拥有宽广的底物谱^[27]。GH43 家族木聚糖 1,4- β -木糖苷酶的结合口袋相较于 GH39 家族木聚糖 1,4- β -木糖苷酶显得窄、深, 仅能支持与只有 2~7 个木糖分子的低聚木糖结合 (图 9c)。

表 6 部分来源木聚糖 1,4- β -木糖苷酶学性质表Table 6 List of enzymatic properties of xylan 1,4- β -xylosidase from some sources

来源	GH	最适温度/°C	最适 pH	底物特异性	Km/(mmol/L)	
1	<i>Aspergillus niger</i> ^[28]	3	35.00	2.50	pNPX	3.10
					人参皂苷 Rb3	1.55
					人参皂苷 C-Mx	1.04
2	<i>Dictyoglomus turgidum</i> ^[29]	3	75.00	5.00	pNPX	0.83
					pNPA	2.01
					pNPG	2.39
					7-木糖苷-10-脱乙酰基紫杉醇	0.03
3	<i>Thermotoga petrophila</i> ^[30]	3	90.00	6.00	pNPX	0.13
					pNPA	6.17
					人参皂苷 Rd	ND
4	<i>C. cellulans</i> sp. ^[31]	3	45.00	8.50	pNPX	4.00
					pNPA	ND
					pNPG	ND
5	<i>Aspergillus niger</i> ^[32]	3	60.00	4.00~5.00	pNPX	0.66
					pNPA	4.54
					pNPG	4.01
					3,4-DNP-Xyl	0.39
					木二糖	30.80
6	<i>Dictyoglomus thermophilum</i> ^[33]	39	75.00	6.00	pNPX	1.66
					pNPG	ND
					木二糖	ND
					三七皂苷 R1	ND
7	<i>Geobacillus</i> sp. ^[34]	39	70.00	6.50	pNPX	2.38
					低聚木糖	ND
8	<i>Caulobacter crescentus</i> ^[35]	39	55.00	6.00	pNPX	9.30
					白桦木聚糖	ND
					甘蔗渣	ND
9	<i>Massilia</i> sp. ^[36]	43	50.00	6.50	pNPX	2.27
					pNPA	4.64
10	<i>Bacteroides ovatus</i> ^[37]	43	35.00	7.00	pNPX	1.71
					低聚木糖	ND
11	<i>Clostridium</i> sp. ^[38]	43	40.00	5.00~6.00	pNPX	2.71
					低聚木糖	ND
12	<i>Paecilomyces thermophila</i> ^[39]	43	55.00	7.00	pNPX	4.50
					pNPA	ND
					玉米芯 (低聚木糖)	ND

表7 不同分类木聚糖 1,4- β -木糖苷酶结合口袋Table 7 Binding pockets of xylan 1,4- β -xylosidases from different classifications

基因符号	结合自由能	体积(\AA^3)	表面积(\AA^2)
LOC112950837	-4.60	462.24	676.11
BXA11_RS32660	-7.00	484.23	493.21
A1O5_02121	-8.70	511.42	513.17

综上所述,木聚糖 1,4- β -木糖苷酶的底物特异性与结构中结合口袋的大小与位置呈现了一定的相关性,说明酶在进化过程中为了适应底物,可能会对结构进行调整以获得不同的底物结合口袋。木聚糖 1,4- β -木糖苷酶的进化规律不仅体现在三级结构上,水解底物的大小对其结构进化也产生了一定程度的影响。

3 结语

本文利用 NCBI 数据库,收集了木聚糖 1,4- β -木糖苷酶的基因序列,成功构建系统进化树后,从中筛选了 12 条代表性序列,对其进行多重序列比对以及一级结构、二级结构、三级结构建模等分析。多序列比对结果显示木聚糖 1,4- β -木糖苷酶的保守位点较少,进化过程并不保守,在长期进化的过程中这些位点能始终存在,说明其可能对于结构或功能的意义很大。蛋白质理化性质以及疏水性预测性都表明了这 12 条代表性序列都是亲水蛋白质,信号肽预测结果提示木聚糖 1,4- β -木糖苷酶分泌方式可分为胞内分泌与胞外分泌;根据蛋白质结构叠合的结果,可以将木聚糖 1,4- β -木糖苷酶分成三个大类,第一大类共有一个 $(\alpha\beta)_8$ 折叠桶;第二大类的结构较为复杂,由三个结构域共同组成,从 N 端到 C 端依次是 $(\alpha\beta)_8$ TIM 折叠桶、 $(\alpha\beta)_6$ 三明治结构、FnIII型 β 折叠片层;第三大类是一个五叶 β 折叠桶状结构,C 端 β 折叠片层的有无将第三大类划分成两个小类,这三种结构共同构成了木聚糖 1,4- β -木糖苷酶的代表性结构。结构分类结果与氨基酸序列进化树分类结果一致,说明木聚糖 1,4- β -木糖苷酶的进化规律可以体现在三级结构上,每种类型的酶对应底物不同,说明底物大小能在一定程度上影响木聚糖 1,4- β -木糖苷酶的进化。本文通过分析木聚糖 1,4- β -木糖苷酶的 12 条代表性氨基酸序列,说明了木聚糖 1,4- β -木糖苷酶的进化规律,并总结归纳其结构特征,为木聚糖 1,4- β -木糖苷酶的定向改造、酶结构与功能关系研究奠定生物信息学基础。

参考文献

[1] 李娜,张蕊,黄遵锡,等. β -木糖苷酶的生物活性物质转化功

能研究进展[J].微生物学通报,2020,47(7):2290-2299

LI Na, ZHANG Rui, HUANG Zunxi, et al. Research progress in bioactive substances transformation by β -xylosidase [J]. Microbiology China, 2020, 47(7): 2290-2299

[2] César Rafael Fanchini Terrasan, José Manuel Guisan, Eleonora Cano Carmona. Xylanase and β -xylosidase from *Penicillium janczewskii*: Purification, characterization and hydrolysis of substrates [J]. Electronic Journal of Biotechnology, 2016, 23: 54-62

[3] Yihan Liu, Lin Huang, Dong Zheng, et al. Biochemical characterization of a novel GH43 family β -xylosidase from *Bacillus pumilus* [J]. Food Chemistry, 2019, 295: 653-661

[4] 汤勇,蔡俊. β -木糖苷酶的研究进展[J].中国酿造,2018,37(10):14-19

TANG Yong, CAI Jun. Research progress of β -xylosidase [J]. China Brewing, 2018, 37(10): 14-19

[5] Khuanjarat Choengpanya, Siriphan Arthornthurasuk, Pakorn Wattana-amorn, et al. Cloning, expression and characterization of β -xylosidase from *Aspergillus niger* ASKU28 [J]. Protein Expression and Purification, 2015, 115: 132-140

[6] Ghulam Mustafa, Sumaira Kousar, Muhammad Ibrahim Rajoka, et al. Molecular cloning and comparative sequence analysis of fungal β -xylosidases [J]. AMB Express, 2016, 6(1): 1-14

[7] 许淑敏.嗜碱芽孢杆菌 β -木糖苷酶的分子改造、固定化及在枯草杆菌中的表达[D].沈阳:沈阳农业大学,2018

XU Shumin. The modification, immobilization of β -xylosidase and its expression in *Bacillus subtilis* [D]. Shenyang: Shenyang Agricultural University, 2018

[8] Tianwen Wang, Chen Liang, Yajing Hou, et al. Small design from big alignment: engineering proteins with multiple sequence alignment as the starting point [J]. Biotechnology Letters, 2020, 42(8): 1305-1315

[9] Kapli Paschalia, Yang Ziheng, Telford Maximilian J. Phylogenetic tree building in the genomic age [J]. Nature Reviews Genetics, 2020, 21(7): 428-444

[10] Sheng-You Huang, Xiaoqin Zou. Advances and challenges in protein-ligand docking [J]. IJMS, 2010, 11(8): 3016-3034

[11] Vipin Kumar Sohal, Apurba Dey, Amarpal Singh. MEGA biocentric software for sequence and phylogenetic analysis: a review [J]. Int. J. of Bioinformatics Research and Applications, 2010, 6(3): 230-240

[12] Hung Jui-Hung, Weng Zhiping. Sequence alignment and

- homology search with BLAST and ClustalW [J]. Cold Spring Harbor Protocols, 2016, 11
- [13] Patrice Goueta, Xavier Roberta, Emmanuel Courcelleb. ESPript/ENDscript: sequence and 3D information from protein structures [J]. Acta Crystallographica, 2005, 61(a1): 42-43
- [14] Gavin E Crooks, Gary Hon, John-Marc Chandonia, et al. WebLogo: A sequence logo generator [J]. Cold Spring Harbor Laboratory Press, 2004, 14(6): 1188-1190
- [15] Christine Hoogland, Khaled Mostaguir, Ron D Appel, et al. The World-2DPAGE constellation to promote and publish gel-based proteomics data through the ExPASy server [J]. Journal of Proteomics, 2008, 71(2): 245-248
- [16] Buchan Daniel W A, Jones David T. The PSIPRED protein analysis workbench: 20 years on [J]. Nucleic Acids Research, 2019, 47(W1): W402-W407
- [17] José Juan Almagro Armenteros, Konstantinos D Tsirigos, Casper Kaae Sønderby, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks [J]. Nature Biotechnology: The Science and Business of Biotechnology, 2019, 37(4): 420-423
- [18] Bitencourt-Ferreira Gabriela, de Azevedo Walter Filgueira. Homology modeling of protein targets with MODELLER [J]. Methods in Molecular Biology (Clifton, N.J.), 2019, 2053: 231-249
- [19] Lawrence A Kelley, Stefans Mezulis, Christopher M Yates, et al. The Phyre2 web portal for protein modeling, prediction and analysis [J]. Nature Protocols, 2015, 10(W1): 845-858
- [20] Zheng Yang, Keren Lasker, Dina Schneidman-Duhovny, et al. UCSF Chimera, MODELLER, and IMP: An integrated modeling system [J]. Journal of Structural Biology, 2012, 179(3): 269-278
- [21] Oleg Trott, Arthur J Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading [J]. Journal of Computational Chemistry, 2010, 31(2): 455-461
- [22] Rajamanikandan Sundarraj, Nihar Nalini Mohanty, Revanaiah Yogisharadhya, et al. Comparative sequence, structure and functional analysis of Skp protein, a molecular chaperone among members of Pasteurellaceae and its homologues in Gram-negative bacteria [J]. Meta Gene, 2020, 24(24): 100680-100690
- [23] 谢勇,洪晓昆,鄢仁祥,等.重组琼胶酶rAgaN3基因的生物信息学分析[J].生物信息学,2017,15(1):16-26
- XIE Yong, HONG Xiaokun, YAN Renxiang, et al. Bioinformatics analysis of recombinant rAgaN3 gene of agatase [J]. Chinese Journal of Bioinformatics, 2017, 15(1): 16-26
- [24] Ali Rohman, Bauke W Dijkstra, Ni Nyoman Tri Puspaningsih. β -Xylosidases: structural diversity, catalytic mechanism, and inhibition by monosaccharides [J]. International Journal of Molecular Sciences, 2019, 20(22): 5524-5550
- [25] Ontañon Ornella M, Ghio Silvina, Marrero Díaz de Villegas Rubén, et al. EcXyl43 β -xylosidase: molecular modeling, activity on natural and artificial substrates, and synergism with endoxylanases for lignocellulose deconstruction [J]. Appl Microbiol Biotechnol, 2018, 102(16): 6959-6971
- [26] Boyce Angela, Walsh Gary. Purification and characterisation of a thermostable β -xylosidase from *Aspergillus niger* van Tieghem of potential application in lignocellulosic bioethanol production [J]. Applied Biochemistry and Biotechnology, 2018, 186(186): 712-730
- [27] Mirjam Czjzek, Alon Ben David, Tsafir Bravman, et al. Enzyme-substrate complex structures of a GH39 β -xylosidase from *Geobacillus stearothermophilus* [J]. Journal of Molecular Biology, 2005, 353(4): 838-846
- [28] 朱铃,王雨辰,赵江源,等.重组 β -木糖苷酶转化人参皂苷Rb₃及C-Mx[J].高等学校化学学报,2020,41(5):1010-1017
- ZHU Ling, WANG Yuchen, ZHAO Jiangyuan, et al. Transformation of ginsenoside Rb₃ and C-Mx by recombinant β -xylosidase [J]. Chemical Journal of Chinese Universities, 2020, 41(5): 1010-1017
- [29] Qi Li, Yujie Jiang, Xinyi Tong, et al. Cloning and characterization of the β -xylosidase from *Dictyoglomus turgidum* for high efficient biotransformation of 10-deacetyl-7-xylositaxol [J]. Bioorganic Chemistry, 2020, 94(94): 103357-103367
- [30] Zhang Shanshan, Xie Jingcong, Zhao Linguo, et al. Cloning, overexpression and characterization of a thermostable β -xylosidase from *Thermotoga petrophila* and cooperated transformation of ginsenoside extract to ginsenoside 20(S)-Rg3 with a β -glucosidase [J]. Bioorganic Chemistry, 2019, 85(85): 159-167
- [31] Ye Yuan, Yanbo Hu, Han Zhang, et al. Characterization of a recombinant multifunctional glycoside hydrolase family 3 β -xylosidase/ α -l-arabinofuranosidase/ β -glucosidase from *Cellulosimicrobium cellulans* sp. 21 [J]. Journal of Molecular Catalysis. B, Enzymatic, 2016, 131(131): 65-72

