



DOI:10.13364/j.issn.1672-6510.20240190

网络首发日期: 2025-10-27; 网络首发地址: <http://link.cnki.net/urlid/12.1355.N.20251027.1128.004>

裂解性多糖单加氧酶的生物信息学分析

白琳, 丁灿灿, 王富豪, 路福平, 刘夫锋

(天津科技大学生物工程学院, 天津 300457)

摘要: 裂解性多糖单加氧酶(lytic polysaccharide monoxygenase, LPMO)能够催化纤维素等多糖中糖苷键的氧化裂解,在秸秆等农业废弃物的资源化利用中起到关键作用。为进一步探究 LPMO 的构效关系,利用生物信息学分析技术系统研究了 UniProtKB 数据库包含的人工校对注释且经过实验验证的 115 个 LPMO。根据其来源统计发现 AA9 家族有 101 个,占比高达 87.8%。依据遗传距离与聚类关系进行分子进化树分析获得 6 个组,所有 AA9 家族 LPMO 都分布在 Group I—Group V 中,其他 3 个家族均聚集在 Group VI 中。分析 34 个来源于曲霉属(*Aspergillus*)的 LPMO,发现这些蛋白质都具有至少 2 个铜离子结合位点,且含有 5 个 100% 高度保守的序列位点。研究来源于曲霉属 LPMO 的基序与结构域发现:模体(motif)组成高度相似的 LPMO 在进化发育关系上较为亲近,但 motif 的排列顺序不同也会影响催化活性;AA9 家族 LPMO 催化结构域主要具有纤维素降解酶活性,而 AA13 家族催化结构域主要具有淀粉降解活性。本研究为进一步探明 LPMO 构效关系提供了理论基础。

关键词: 裂解性多糖单加氧酶; 生物信息学分析; 分子进化; 结构域

中图分类号: Q814

文献标志码: A

文章编号: 1672-6510(2026)02-0001-10

Bioinformatic Analysis of the Lytic Polysaccharide Monoxygenases

BAI Lin, DING Cancan, WANG Fuhao, LU Fuping, LIU Fufeng

(College of Biotechnology, Tianjin University of Science and Technology, Tianjin 300457, China)

Abstract: Lytic polysaccharide monoxygenase(LPMO) can catalyze the oxidative lysis of glycosidic bonds in polysaccharide such as cellulose, and plays a key role in the degradation of biomass such as straw. To further explore the structure-activity relationship of LPMO, 115 experimentally verified LPMO included in UniProtKB database were systematically studied using bioinformatics analysis technology. According to its source statistics, 101 LPMO from AA9 family accounted for as high as 87.8%. Further molecular evolutionary tree analysis based on genetic distance and cluster relationship obtained six groups, all AA9 family LPMO were distributed in Group I-Group V and the other three families were clustered in Group VI. Analysis of 34 LPMO derived from *Aspergillus* revealed that all of these proteins had at least two copper ion binding sites and contained five 100% highly conserved sequence sites. Further study of the motifs and domains of LPMO from *Aspergillus* showed that LPMO with highly similar motifs composition were close in evolution and development, but the different order of motifs can also affect the catalytic activity; the catalytic domain of LPMO of AA9 family mainly had cellulolytic enzyme activity, while the catalytic domain of AA13 family mainly had starch degradation activity. This study lays a theoretical foundation for further investigation of LPMO structure-activity relationship.

Key words: lytic polysaccharide monoxygenase; bioinformatics analysis; molecular evolution; domain

引文格式:

白琳, 丁灿灿, 王富豪, 等. 裂解性多糖单加氧酶的生物信息学分析[J]. 天津科技大学学报, 2026, 41(2): 1-10.

BAI L, DING C C, WANG F H, et al. Bioinformatic analysis of the lytic polysaccharide monoxygenases[J]. Journal of Tianjin university of science and technology, 2026, 41(2): 1-10.

收稿日期: 2024-09-19; 修回日期: 2024-11-01

基金项目: 国家重点研发计划项目(2021YFC2102700); 国家自然科学基金项目(32272269); 天津市科技重大专项与工程项目(25ZXWCSY00160)

作者简介: 白琳(2001—), 女, 山西大同人, 硕士研究生; 通信作者: 刘夫锋, 教授, fufengliu@tust.edu.cn

随着石油等不可再生资源的锐减,人类对可再生能源如木质纤维素的需求增大,如何对其高效利用并维持可持续发展成为研究热点。木质纤维素组成复杂,主要包含木质素、纤维素和半纤维素^[1]。纤维素由葡萄糖聚合而成,而半纤维素是由木糖、阿拉伯糖和半乳糖等聚合而成。虽然已有酶可以降解上述组分,但现有的酶降解效率较低,且上述组分交织在一起,严重降低了酶法降解的效率。裂解性多糖单加氧酶(LPMO)是一类铜离子依赖性的氧化还原酶^[2]。LPMO通过铜离子氧化还原,断裂木质纤维素结晶区的 β -1,4-糖苷键,产生C4氧化产物,从而在木质纤维素的降解过程中发挥着关键作用。自2010年^[3]被发现以来,LPMO在CAZy数据库中被归为辅助活性蛋白(auxiliary activity, AA)家族^[4],目前报道的为以下8类:AA9、AA10、AA11、AA13、AA14、AA15、AA16和AA17家族。

现有研究利用生物信息学方法主要开展挖掘新酶、家族分类和特征表征研究。目前从NCBI等数据库中已挖掘出多种LPMO^[5-9],使用传统神经网络对序列的家族来源进行分类^[10]。为了鉴别其亚型并确定其结构特征,Moses等^[11]对多个LPMO的结构域进行类型特异性表征。Nicolas等^[12]通过构建局部序列比对方法分析了3400个AA9家族LPMO的序列特征,揭示了其N端序列高度变异性与多基因性的原因。Tamburrini等^[13]则分析了27060个不同家族来源的LPMO序列,发现约60%存在无序的C末端区域。

上述研究所用序列大多是没有经过实验验证的序列,因而严重影响其研究结果的可信性。为进一步探明LPMO的结构特点与进化关系,本研究采用生物信息学方法对UniProtKB数据库包含人工校对注释且实验验证的115个LPMO进行系统研究,旨在帮助研究者加深对LPMO的认识,为LPMO相关基因结构和蛋白功能的研究提供理论基础。

1 材料与方法

1.1 LPMO信息收集

以UniProtKB数据库(<https://www.uniprot.org/>)^[14]为主要研究对象,通过检索“lytic polysaccharide monoxygenase”进行数据收集,包括entry ID、蛋白名称、序列和物种来源等信息。

1.2 LPMO理化性质分析

通过ExPASy的ProtParam^[15]在线网站([https://](https://web.expasy.org/protparam/)

web.expasy.org/protparam/)对LPMO的理化性质,包括相对分子质量、等电点、分子式、原子总数、脂肪系数、不稳定指数和平均亲水性系数等进行分析。

1.3 LPMO进化分析

采用MEGA 11^[16]的Neighbor-Joining方法绘制分子进化树图谱,通过iTOL^[17]进行美化。

1.4 LPMO糖基化位点预测

N-糖基化、O-糖基化位点预测分别使用NetNGlyc 1.0 Server^[18]和NetOGlyc 4.0 Server^[19],在线服务器网址分别为<https://services.healthtech.dtu.dk/services/NetNGlyc-1.0/>和<https://services.healthtech.dtu.dk/services/NetOGlyc-4.0/>。

1.5 LPMO活性位点分析

通过UniProtKB和NCBI Web CD-Search^[20]工具(<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>)统计LPMO活性位点信息。

1.6 LPMO保守序列信息分析

通过GeneDoc对LPMO进行多序列比对,寻找保守序列信息。

1.7 LPMO基序与结构域分析

采用MEME^[21](<http://meme-suite.org/tools/meme>)预测模体(motif)结构,参数设为10。采用Web CD-Search分析保守结构域。最终结果使用TBtools-II^[22]可视化输出。

1.8 LPMO空间结构分析

采用AlphaFold 2^[23]对LPMO进行三维结构建模分析。

2 结果与讨论

2.1 LPMO信息的检索

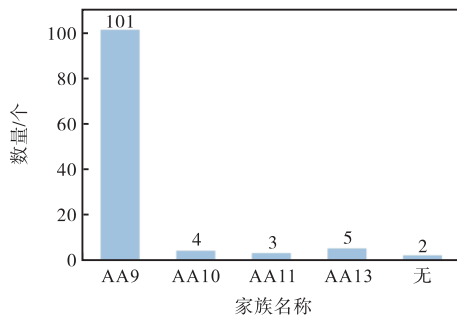
以“lytic polysaccharide monoxygenase”为检索词,在UniProtKB数据库获得人工校对注释且实验验证的数据129条,经手动筛选去除错误信息,精确保留LPMO信息115条。分析LPMO的物种来源,属水平上共有28类,图1(a)清晰显示了各属的分布:占比超过5%的有曲霉属(*Aspergillus*) (29.6%)、葡萄孢属(*Botrytis*) (7.8%)、喜热梭孢壳属(*Thermothelomyces*) (7.8%)、色绒枝霉属(*Malbranchea*) (7%)、柄孢壳属(*Podospora*) (7%)、脉孢菌属(*Neurospora*) (6.1%)、嗜热梭菌属(*Thermothielavioides*) (5.2%)和褐褶菌属(*Gloeophyllum*) (5.2%),其中曲霉属(*Aspergillus*)来源包含34种LPMO,位居第一。对115种酶进行家族来源统计〔图1(b)〕发现,101个

LPMO 来源于 AA9 家族(87.8%), AA10、AA11 和 AA13 家族分别只有 4、3、5 个。新发现的 AA14、AA15、AA16 和 AA17 家族没有收录进该数据库。因此,该研究集中于 AA9 家族的 LPMO,其余家族的

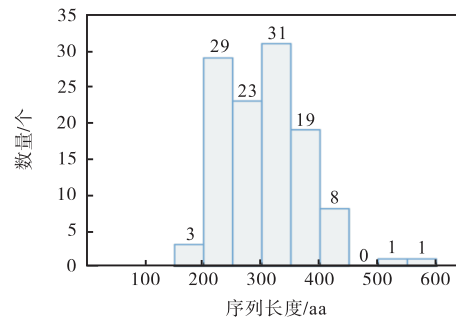
有待将来深入挖掘解析。对 LPMO 的氨基酸序列长度进行统计,结果表明:长度为 200~400 个氨基酸(aa)的 LPMO 最多,没有短于 150 aa 和长于 600 aa 的蛋白质[图 1(c)]。



(a) LPMO 属水平分布图



(b) LPMO 家族分布图



(c) LPMO 氨基酸序列长度分布图

图 1 115 种 LPMO 信息统计

Fig. 1 Statistics of 115 LPMO

2.2 LPMO 的理化性质分析

115 种 LPMO 的理化性质结果如图 2 所示。LPMO 的相对分子质量主要分布在 $2.0 \times 10^4 \sim 4.5 \times 10^4$ 之间。绝大部分 LPMO 的等电点在 4~7 之间,推测大多数 LPMO 可能在酸性条件发挥作用。LPMO 中的 86.1% 属于亲水性蛋白,80% 的稳定性能较好。

2.3 LPMO 的进化分析

115 种 LPMO 的分子进化树如图 3 所示。依据遗传距离与聚类关系以不同颜色将其区分为 6 个组(Group I—Group VI)。发现所有 AA9 家族 LPMO 都分布在 Group I—Group V 中,且其他家族均聚集在 Group VI 中。以蓝、绿和橙色带分别表示 LPMO 在界、门和属水平的分类,只有 4 个 LPMO(深蓝)来源于细菌且聚类在 Group VI,其余 111 个(浅蓝)均为真菌,且只包含子囊菌门(Ascomycota)(浅绿)和担子菌门(Basidiomycota)(深绿)2 个门。突出展示数量最多的来源于曲霉属的 LPMO(深橙),发现除 Group II 外,它们分散在整个进化树中,其中分布在 Group III 和 Group V 中的最多,且涉及除 AA10 外其他所

有家族,故来源于曲霉属的 LPMO 具有一定的样本代表性,可将其应用于后续研究中。将不稳定指数(instability index)、总平均亲水性系数(GRAVY)和理论等电点(theoretical pI)分别以绿色五角星、蓝色圆圈和紫色正方形表示,当不稳定指数 > 40(不稳定蛋白)、GRAVY > 0(疏水蛋白)和等电点 > 7 时表示为实心图形,否则为空心。可知来源于曲霉属的 LPMO 大多为稳定的亲水性蛋白,且在酸性环境发挥作用。

2.4 LPMO 的糖基化位点预测

常见的蛋白质翻译后修饰过程有泛素化、磷酸化、酯基化、甲基化、乙酰化和糖基化等。其中,糖基化是蛋白质最重要的翻译后修饰之一,根据糖基化反应位点主要分为 N-糖基化与 O-糖基化,这两种均在蛋白质折叠、结构形成、稳定性维持和活性发挥等方面起着重要作用。本研究通过 NetNGlyc 1.0 Server 和 NetOGlyc 4.0 Server 对 115 种 LPMO 进行糖基化预测,结果如图 4 所示。由图 4(a)可知,LPMO 的 N-糖基化位点较少,最多不超过 5 个,而具有 0 个、1 个和 2 个的占比最多,分别是 30%、37% 和 21%;来

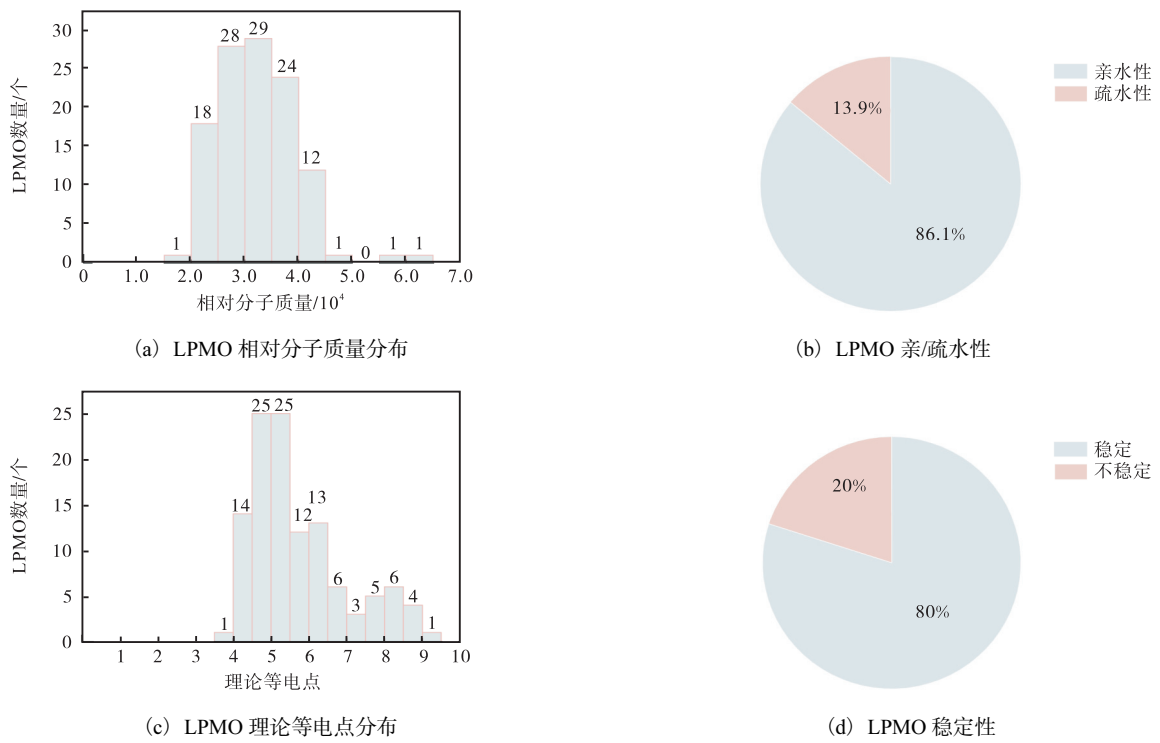


图2 115种 LPMO 的理化性质分析

Fig. 2 Analysis of physicochemical properties of 115 LPMO

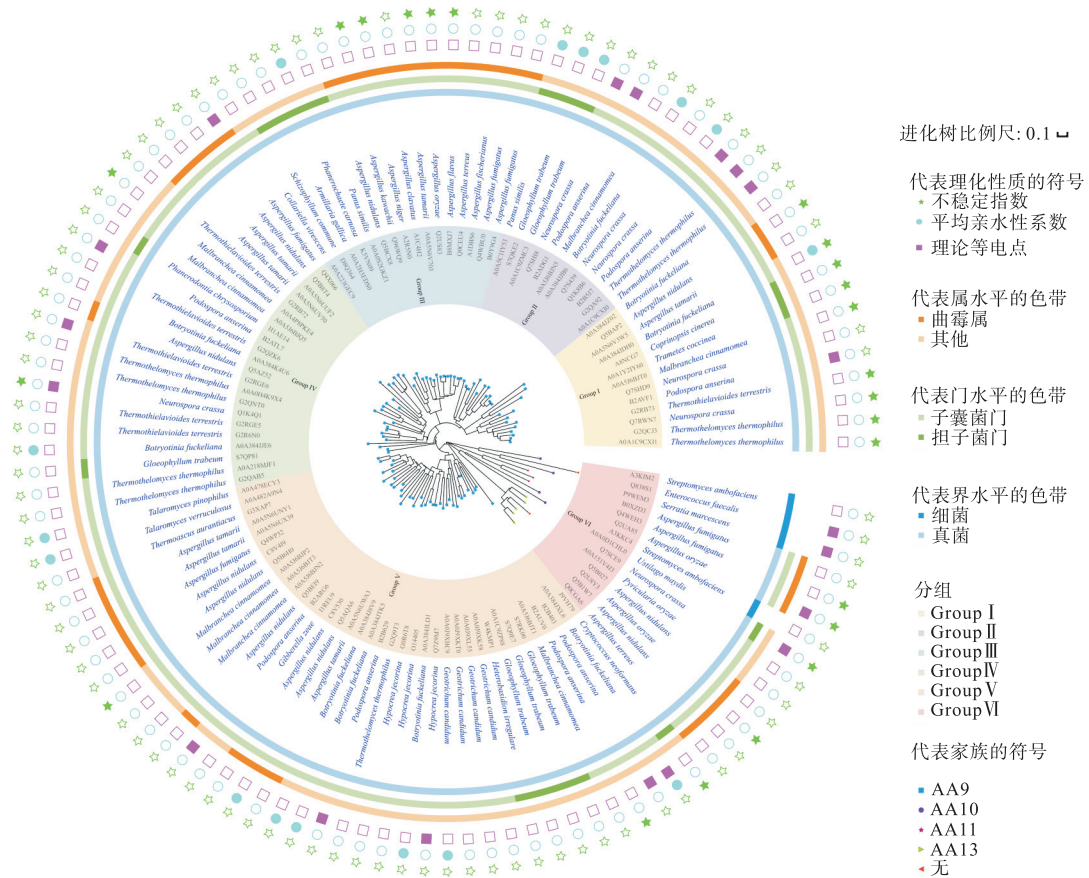
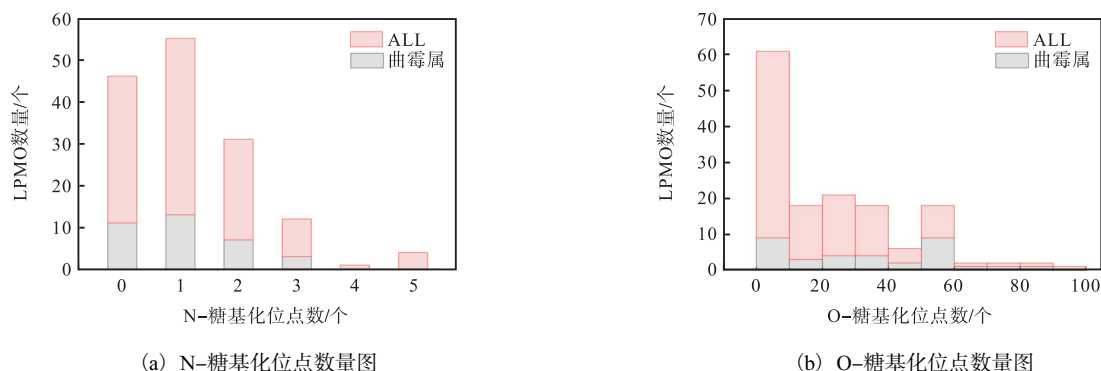


图3 115种 LPMO 的分子进化树

Fig. 3 Molecular evolutionary tree of 115 LPMO

源于曲霉属的 34 种 LPMO 中,超过半数具有 1~3 个 N-糖基化位点。与 N-糖基化相反,O-糖基化位点数量较多〔图 4(b)〕。例如,A0A0J9XK58、Q5BEI9、Q2UA85 和 Q5AQA6 分别具有 94、82、73

和 61 个,而占比最多的只有 0~9 个位点;曲霉属来源中有约 3/4 的 LPMO 具有 10 个以上 O-糖基化位点,而且约 50% 具有 30 个以上。总而言之,LPMO 中 O-糖基化位点远多于 N-糖基化位点。



注:ALL 表示 115 个 LPMO,曲霉属表示来源于曲霉属的 34 个 LPMO。

图 4 LPMO 的糖基化位点
Fig. 4 Glycosylation sites of 115 LPMO

2.5 来源于曲霉属 LPMO 的活性位点分析

通过 UniprotKB 和 Web CD-Search 对来源于曲霉属 34 种 LPMO 的活性位点进行统计,结果见表 1。由表 1 可知,每个 LPMO 都具有铜离子结合位点,除 B0XZD3、Q4WEH3、Q5AZ52 和 Q5B8T4 具有 2 个 Cu^{2+} 结合位点以外,其余均为 3 个。催化活性位点中组氨酸最多,与 Cu^{2+} 构成组氨酸支架,部分位点如酪氨酸和谷氨酰胺,其位置存在一定规律。除 AA13 家族存在 2 个淀粉结合位点外,其他 LPMO 所具有的化学底物结合位点几乎完全包含其催化活性位点的组氨酸残基,即这些组氨酸残基不仅利于底物结合,也利于 LPMO 催化活性的发挥。

总而言之,存在 Cu^{2+} 的结合位点佐证了 LPMO 是 Cu^{2+} 依赖型酶,特定位置的组氨酸残基与 Cu^{2+} 构成了组氨酸支架,稳定了酶分子的结构,也利于其催化活性的发挥。

2.6 来源于曲霉属 LPMO 的保守序列信息

通过对来源于曲霉属 34 种 LPMO 的氨基酸序列比对,发现了它们的保守序列信息,结果如图 5 所示。具有 100% 高度保守性的序列位点共有 5 个,分别是 H47、G141、G208、Y269 和 Y289。具有 80% 保守性的序列位点涉及 34 个,其中高度保守的有 16 个,分别是 H136、P142、Y146、K166、G171、W185、P202、Y210、R213、E215、H220、Y232、C235 和三联体 PGP292-294,而具有 60% 保守性的序列位点共有 50 个。依据 2.5 节的结果推断可知,H47、H136 和 Y232 为部分 LPMO 的 Cu^{2+} 结合位

点,H220 和 Q230 为 O_2 结合位点,推测 Tyr 和 Gly 等高度保守的氨基酸在 LPMO 结构中同样发挥着关键作用。第 300 位点后不显示高亮区,表明 LPMO 的 C 端序列存在较大差异。总之,来源于曲霉属的 LPMO 存在较保守的 N 端和多样化的 C 端,其结构的多样性导致功能的特异性。

2.7 来源于曲霉属 LPMO 的基序与结构域分析

对来源于曲霉属的 34 种 LPMO 进行基序和结构域分析,结果如图 6 所示。第一组 B0Y9G4、Q4WBU0、A1DBS6、A1C4H2、A0A5N6V703、Q5BCX8、A2R5N0、Q96WQ9、Q0CEU4、B8MXJ7 和 Q2US83 的 motif 组成高度相似,均含有 motif 4、motif 2、motif 6、motif 1、motif 3 和 motif 5。推测它们在进化发育关系上较为亲近,具有更相近的分子功能。同理,虽然第二组 A0A5N6UWA3、Q5AQA6、C8V530、A0A5N6UX39、Q5B6H0、A0A5N6UNY1、Q4WP32、C8V419 和 Q5BEI9 序列差异较大,但均含有 motif 8、motif 7、motif 6 和 motif 1,也具有相近的遗传进化关系和分子功能。此外,第三组 Q0CGA6、Q5B1W7、Q2U8Y3 和 Q5B027 均含有 motif 9、motif 6 和 motif 7,motif 6 和 motif 7 的不同排列组成使其区别于第二组。

分析 34 种 LPMO 的结构域发现:AA9 家族 LPMO 结构域主要为 LPMO_AA9,参与 C1 和/或 C4 和 C6 氧化纤维素链的裂解,其活性包括纤维素单加氧酶的 C1-羟化(EC 1.14.99.54)和 C4-脱氢(EC 1.14.99.56),其特征之一^[24]在于与铜离子的结合位点

符合保守特征残基模式 H-H-[EQ]-[FY], 即 His-His-[Glu/Gln]-[Phe/Tyr], 这与表 1 结果相互印证。另一假设特征是该结构域可能包含化学底物结合位点。例如, Simmons 等^[25]发现绒柄香菇 (*Lentinus similis*) 的 AA9 家族 LPMO A0A0S2GKZ1 在 0.4 nm 处存在与 β -葡萄糖寡糖的接触位点。依据图 6 中结构域 LPMO_AA9 的长度范围与表 1 中化学底物结合位点

的位置, 推断该结构域范围内确实可能存在化学底物结合位点。AA9 结构域显示该类酶具有纤维素降解酶活性, 其特征是底物结合表面是一个扁平的 Ig 样折叠, 这与 AA11 家族 Q2UA85 所具有的 AA9 superfamily 结构域相同。Ig 折叠是指免疫球蛋白多肽链反复折叠形成的桶状结构, 由 2 个 β 片层通过二硫键连接。

表 1 来源于曲霉属 34 种 LPMO 的活性位点
Tab. 1 Active sites of 34 LPMO from *Aspergillus*

ID	催化活性位点	化学底物结合位点	ID	催化活性位点	化学底物结合位点
A0A5N6UNY1	Cu ²⁺ :H22, H107, Y196 O ₂ :H185, Q194	H22, L62, E102, M103, H107, H185, S186, Y233	Q2U8Y3	Cu ²⁺ :H47, H137, Y270 O ₂ :无	无
A0A5N6UUF2	Cu ²⁺ :H18, H101, Y186 O ₂ :H176	H18, Y46, V95, G96, H101, H176, G177, Y224	Q2UA85	Cu ²⁺ :H20, H79, E93 O ₂ :无	无
A0A5N6UV50	Cu ²⁺ :H21, H96, Y182 O ₂ :H172	H21, Y47, Y90, G91, H96, H172, R173, Y220	Q2US83	Cu ²⁺ :H21, H102, Y180 O ₂ :H169	H21, N50, H88, D89, H102, H169, E170, Y218
A0A5N6UWA3	Cu ²⁺ :H22, H108, Y197 O ₂ :H186, Q195	H22, N63, P103, M104, H108, H186, S187, Y234	Q4WBU0	Cu ²⁺ :H20, H102, Y180 O ₂ :H169	H20, N49, H88, D89, H102, H169, E170, Y218
A0A5N6UX39	Cu ²⁺ :H20, H104, Y191 O ₂ :H180, Q189	H20, L59, D99, M100, H104, H180, S181, Y228	Q4WEH3	Cu ²⁺ :H22, H82 O ₂ :无	无
A0A5N6V3W5	Cu ²⁺ :H17, H99, Y186 O ₂ :H174, Q184	H17, T46, A85, H86, H99, H174, A175, Y224	Q4WP32	Cu ²⁺ :H22, H107, Y196 O ₂ :H185, Q194	H22, L62, T102, M103, H107, H185, S186, Y233
A0A5N6V703	Cu ²⁺ :H21, H102, Y180 O ₂ :H169, Q178	H21, N50, H88, D89, H102, H169, E170, Y218	Q4X066	Cu ²⁺ :H18, H101, Y187 O ₂ :H176	H18, Y46, V95, G96, H101, H176, G177, M224
A1C4H2	Cu ²⁺ :H20, H102, Y180 O ₂ :H169	N49, H88, T89, H102, H169, E170, Y218	Q5AQA6	Cu ²⁺ :H17, H102, Y191 O ₂ :H180, Q189	H17, N56, A97, M98, H102, H180, S181, Y228
A1DBS6	Cu ²⁺ :H20, H102, Y180 O ₂ :H169	H20, N49, H88, D89, H102, H169, E170, Y218	Q5AZ52	Cu ²⁺ :H19, Y176 O ₂ :H165, Q174	H19, N47, T88, N89, H92, H165, G166, Y214
A2R5N0	Cu ²⁺ :H21, H103, Y183 O ₂ :H172	H21, N50, H89, S90, H103, H172, E173, Y221	Q5B027	Cu ²⁺ :H19, H109, Y242 O ₂ :无	无
B0XZD3	Cu ²⁺ :H19, H89 O ₂ :无	无	Q96WQ9	Cu ²⁺ :H21, H103, Y183 O ₂ :H172	H21, N50, H89, S90, H103, H172, E173, Y221
B0Y9G4	Cu ²⁺ :H20, H102, Y180 O ₂ :H169	H20, N49, H88, D89, H102, H169, E170, Y218	Q5B6H0	Cu ²⁺ :H22, H107, Y197 O ₂ :D51, H186, Q195	H22, N62, T102, M103, H107, H186, S187, Y334
B8MXJ7	Cu ²⁺ :H21, H102, Y180 O ₂ :H169	H21, N50, H88, D89, H102, H169, E170, Y218	Q5B8T4	Cu ²⁺ :H18, Y186 O ₂ :H176	H18, Y46, V95, G96, H101, H176, G177, M224
C8V4I9	Cu ²⁺ :H20, H105, Y194 O ₂ :H183, Q192	H20, T60, T100, M101, H105, H183, S184, M231	Q5BAP2	Cu ²⁺ :H17, H99, Y186 O ₂ :H174, Q184	H17, T46, A85, Y86, H99, H174, A175, Y225
C8V530	Cu ²⁺ :H23, H110, Y199 O ₂ :H188, Q197	H23, N63, P103, M106, H110, H188, G189, Y236	Q5BCX8	Cu ²⁺ :H19, H101, Y179 O ₂ :H168, Q177	H19, N48, H87, N88, H101, H168, E169, Y217
Q0CEU4	Cu ²⁺ :H20, H102, Y180 O ₂ :H169	H20, N49, H88, N89, H102, H169, E170, Y218	Q5BEI9	Cu ²⁺ :H20, H105, Y194 O ₂ :H183, Q192	H20, L60, E100, M101, H105, H183, S184, Y231
Q0CGA6	Cu ²⁺ :H18, H108, Y241 O ₂ :无	淀粉结合位点 1: S320, G325, HSTD327-330, H374 淀粉结合位点 2: FVTTMH306-311, A337, GYTA339-342, PL345-346	Q5B1W7	Cu ²⁺ :H19, H109, Y242 O ₂ :无	淀粉结合位点 1: S305, G310, MSTS312-315, M359 淀粉结合位点 2: LVTTTY291-296, A322, KYSS324-327, PL330-331

注:A为丙氨酸(Ala),C为半胱氨酸(Cys),D为天冬氨酸(Asp),E为谷氨酸(Glu),F为苯丙氨酸(Phe),G为甘氨酸(Gly),H为组氨酸(His),I为异亮氨酸(Ile),K为赖氨酸(Lys),L为亮氨酸(Leu),M为甲硫氨酸(Met),N为天冬酰胺(Asn),P为脯氨酸(Pro),Q为谷氨酰胺(Gln),R为精氨酸(Arg),S为丝氨酸(Ser),T为苏氨酸(Thr),V为缬氨酸(Val),W为色氨酸(Trp),Y为酪氨酸(Tyr)。

AA13 家族结构域有 AA13_LPMO-like 和 AA13_LPMO-like superfamily, 此类酶含有淀粉降解

活性, 作用于糖苷键时有与其他家族 LPMO 完全不同的结合表面, 即 AA13 家族蛋白与底物以不同的方

式相互作用。其活性位点包含一个氨基末端的组氨酸连接的单核铜,除铜离子外,LPMO 也与锌离子结合,如 Q2U8Y3 即是米曲霉 AA13 家族 LPMO 与锌离子在 0.4 nm 处结合^[26],而锌被用于结构研究是因为它比铜更容易获得晶体。这种酶从回生淀粉中生成终止醛酸的低麦芽糖,并显著促进 β -淀粉酶将这种顽固性底物转化为麦芽糖。CBM20_glucoamylase 是指葡萄糖淀粉酶 C 末端的 CBM20 结构域,也具有淀粉降解活性,它折叠为具有 2 个淀粉结合位点的反平行 β -桶结构^[27],位点 1 充当初始淀粉识别位点,位点 2 参与淀粉适当区域的特异性识别。

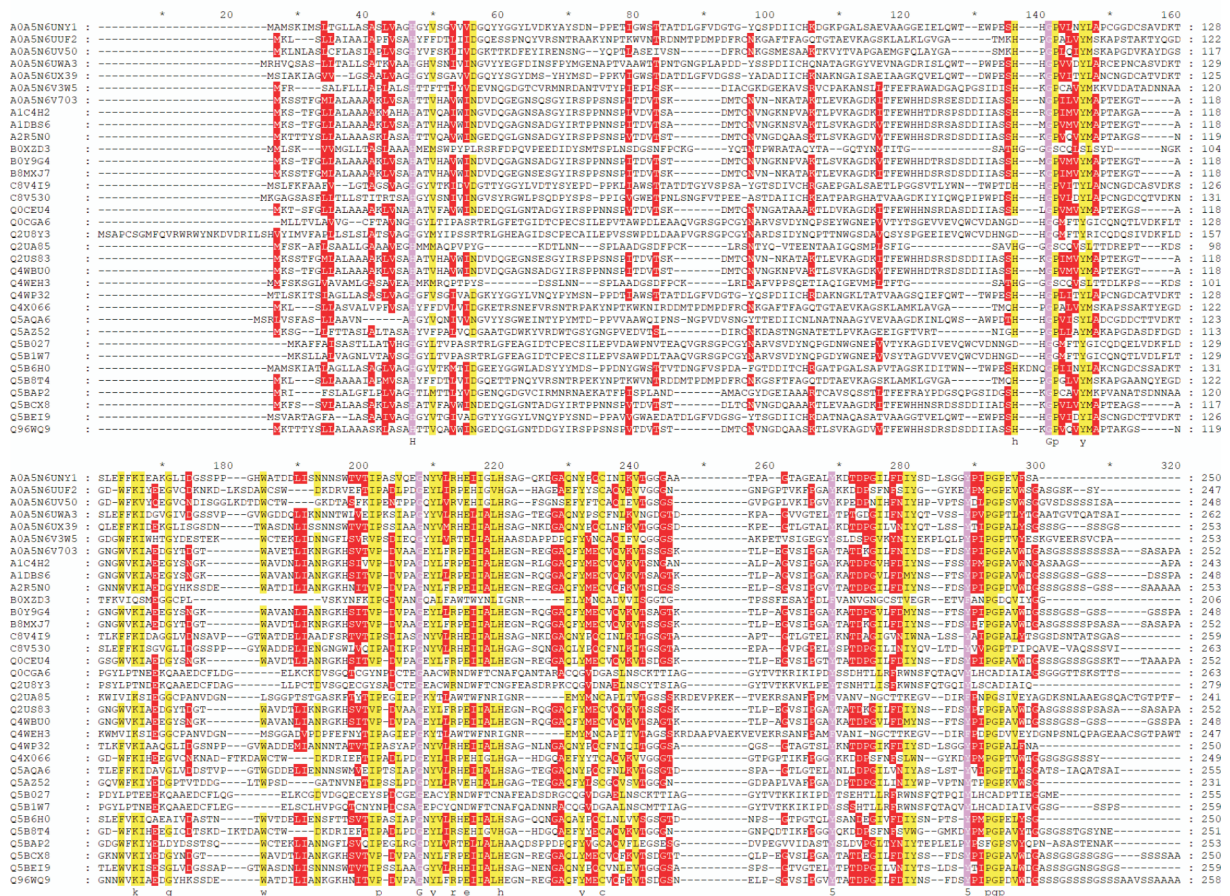
AA11 家族的 B0XZD3 具有 LPMO_10 superfamily 结构域,其与纤维素结合结构域相关,参与纤维素降解过程。此外,结构域 fCBD、CBM_1 是真菌型纤维素结合域,也具有纤维素降解活性。Q5BE19 具有较为独特的结构域 motB superfamily^[28],是指鞭毛运动蛋白 motB 结构域,参与菌体细胞的运动。

2.8 LPMO 的空间结构分析

从图 6 中可以看出,B0Y9G4 具有经典的 AA9 家族 LPMO 的序列特征,它不仅与部分 LPMO 相似的 motif 组成与排列,还具有最多的 motif,因此将其作为进一步分析空间结构的代表。

AlphaFold 因其具有出色的预测准确率成为蛋白质结构预测的重要工具^[23]。本研究采用 AlphaFold 2 预测并以 Cartoon 方式展示 B0Y9G4 的三维结构,其中 loop 环、 α -螺旋、 β -片层和辅因子 Cu^{2+} 的结合位点分别用灰色、蓝绿色、黄色和粉色进行区分,结果如图 7 所示。B0Y9G4 具有 3 个 Cu^{2+} 结合位点,其附近存在的大量 β 片层构成了核心结构域 LPMO_AA9,C 端的少量 β 片层构成了结构域 fCBD,两者均发挥纤维素降解活性,有利于纤维素的降解。

为了验证上述预测结构的准确性,选择同家族已知晶体结构的且来源于 *Lentinus similis* 的 LPMO A0A0S2GKZ1 进行研究,结构如图 8 所示。



注:高亮区紫色、黄色、红色分别表示序列具有 100%、80%、60% 保守性;因第 300 位点后不显示高亮区,故而不进行展示。

图 5 来源于曲霉属 34 种 LPMO 的保守序列信息

Fig. 5 Conserved sequence information for 34 LPMO from *Aspergillus*

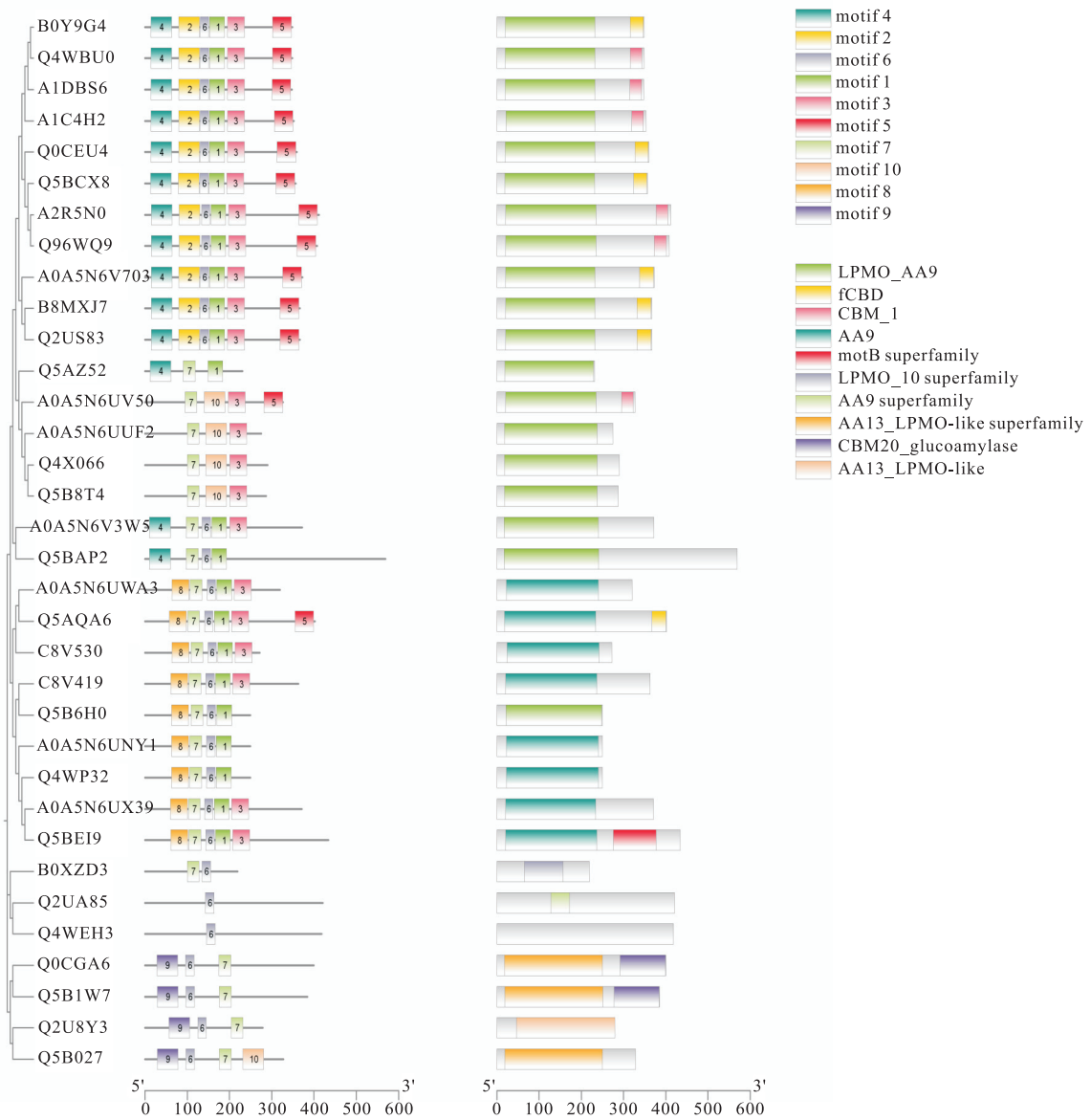


图6 来源于曲霉属 34种 LPMO的基序与结构域

Fig. 6 Motifs and domains of 34 LPMO from *Aspergillus*

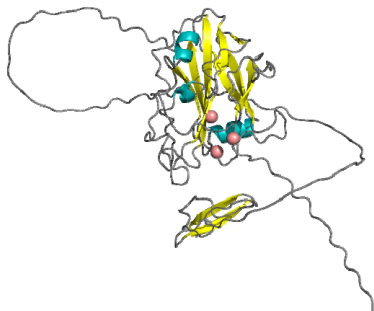


图7 B0Y9G4的三维结构

Fig. 7 Three-dimensional structure of B0Y9G4

A0A0S2GKZ1 与 B0Y9G4 的共同特征包括均含有少量的 α -螺旋、由 β -片层构成的核心结构域以及相同数量的 Cu^{2+} 结合位点。两者之间的主要差异在于 Cu^{2+} 结合位点的具体位置 (A0A0S2GKZ1 :

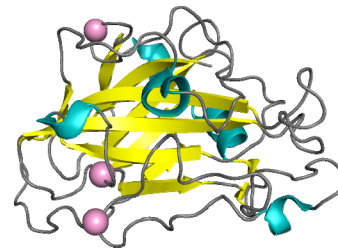


图8 A0A0S2GKZ1的三维结构

Fig. 8 Three-dimensional structure of A0A0S2GKZ1

H20/H97/Y183; B0Y9G4: H20/H102/Y180)、 α -螺旋的分布位置以及无规则卷曲区域的大小。尽管存在这些结构上的差异,但是 2 个 LPMO 都具备催化裂解 β -1,4-糖苷键的功能。

3 结论

本研究获得 UniProtKB 数据库中人工校对注释且经过实验验证的 LPMO 数据 115 条。分析其理化性质,亲水性蛋白占 86.1%,稳定性蛋白占 80%,相对分子质量主要分布在 $2.0 \times 10^4 \sim 4.5 \times 10^4$,绝大多数等电点在 4~7 之间,氨基酸序列长度以 200~400 aa 为主。分析其物种来源,属水平上共有 28 类,其中来源于曲霉属的有 34 个,占比 29.6%。此外,115 个 LPMO 中来自 AA9 家族的有 101 个,占比 87.8%。

分子进化树依据遗传距离与聚类关系区分为 6 个组(Group I—Group VI),所有 AA9 家族 LPMO 都分布在 Group I—Group V 中,其他家族均聚集在 Group VI 中。界水平,只有 4 个 LPMO 来源于细菌且聚类在 Group VI,其余 111 个均为真菌;门水平,真菌只包含 Ascomycota 和 Basidiomycota;属水平,来源于曲霉属的 LPMO 数量最多,除 Group II 外分散在整个进化树中,以 Group III 和 Group V 居多。

LPMO 的 N-糖基化位点较少,最多不超过 5 个,来源于曲霉属的 LPMO 中约 50% 具有 1~3 个;尽管大多数 LPMO 只有 0~9 个 O-糖基化位点,但个别如 A0A0J9XK58 的 O-糖基化位点高达 94 个,来源于曲霉属的 LPMO 中约 50% 具有 30 个以上。此外,来源于曲霉属的 LPMO 都具有至少 2 个铜离子结合位点,其余以组氨酸结合位点居多,除 AA13 家族存在 2 个淀粉结合位点外,其他 LPMO 所具有的化学底物结合位点几乎完全包含其催化活性位点中的组氨酸结合位点。来源于曲霉属的 LPMO 具有 5 个 100% 高度保守的序列位点:H47、G141、G208、Y269 和 Y289,其中 H47 是部分 LPMO 的 Cu^{2+} 结合位点,相较于各家族 LPMO 的 N 端具有保守性,C 端更具多样性。

motif 组成高度相似的 LPMO 在进化发育关系上较为亲近,但 motif 的排列顺序不同也会影响分子功能。来源于曲霉属的 LPMO 中 AA9 家族结构域主要为 LPMO_AA9 和 AA9,均具有纤维素降解酶活性;AA13 家族结构域有 AA13_LPMO-like 和 AA13_LPMO-like superfamily,均含有淀粉降解活性。LPMO_10 superfamily、fCBD 和 CBM_1 结构域均参与纤维素降解过程,CBM20_glucoamylase 参与淀粉降解过程。

来源于曲霉属的 LPMO B0Y9G4 具有 3 个 Cu^{2+} 结合位点,其附近存在大量的 β -片层构成了核心结构域 LPMO_AA9,C 端的少量 β -片层构成了结构域 fCBD,该空间结构利于纤维素的降解。与同家族已知晶体结构的且来源于 *Lentinus similis* 的 LPMO A0A0S2GKZ1 相比,尽管存在一些序列结构上的差异,但都具备催化裂解 β -1,4-糖苷键的功能。

综上所述,本研究为后续对 LPMO 结构与功能的改造研究提供了理论基础。

参考文献:

- [1] WANG W, LEE D J. Lignocellulosic biomass pretreatment by deep eutectic solvents on lignin extraction and saccharification enhancement: a review[J]. Bioresource technology, 2021, 339: 125587.
- [2] MEIER K K, JONES S M, KAPER T, et al. Oxygen activation by Cu LPMOs in recalcitrant carbohydrate polysaccharide conversion to monomer sugars[J]. Chemical reviews, 2018, 118(5): 2593–2635.
- [3] VAAJE K G, WESTERENG B, HORN S J, et al. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides[J]. Science, 2010, 330(6001): 219–222.
- [4] DRULA E, GARRON M L, DOGAN S, et al. The carbohydrate-active enzyme database: functions and literature[J]. Nucleic acids research, 2022, 50(D1): 571–577.
- [5] GUO X, AN Y J, JIANG L Y, et al. The discovery and enzymatic characterization of a novel AA10 LPMO from *Bacillus amyloliquefaciens* with dual substrate specificity[J]. International journal of biological macromolecules, 2022, 203: 457–465.
- [6] MEKASHA S, TUVENG T R, ASKARIAN F, et al. A trimodular bacterial enzyme combining hydrolytic activity with oxidative glycosidic bond cleavage efficiently degrades chitin[J]. The journal of biological chemistry, 2020, 295(27): 9134–9146.
- [7] FERREIRA J A, HORTA M A C, SANTOS C A, et al. Integrative genomic analysis of the bioprospection of regulators and accessory enzymes associated with cellulose degradation in a filamentous fungus (*Trichoderma harzianum*) [J]. BMC Genomics, 2020, 21: 757.
- [8] GUO X, SANG J C, CHAI C C, et al. A lytic polysaccharide monooxygenase from *Myceliophthora thermophila* C1 and its characterization in cleavage of glycosidic chain of cellulose[J]. Biochemical engineering journal,

- 2020, 162: 107712.
- [9] ZHANG M J, FU X P, GU R R, et al. A novel starch-active lytic polysaccharide monoxygenase discovered with bioinformatics screening and its application in textile desizing[J]. BMC Biotechnology, 2024, 24: 2.
- [10] SRIVASTAVA P A, HEGG E L, FOX B G, et al. PreDSLpmo: a neural network-based prediction tool for functional annotation of lytic polysaccharide monoxygenases[J]. Journal of biotechnology, 2020, 308: 148–155.
- [11] MOSES V, HATHERLEY R, BISHOP O T, et al. Bioinformatic characterization of type-specific sequence and structural features in auxiliary activity family 9 proteins[J]. Biotechnology for biofuels, 2016, 9: 239.
- [12] LENFANT N, HAINAUT M, TERRAPON N, et al. A bioinformatics analysis of 3400 lytic polysaccharide oxidases from family AA9[J]. Carbohydrate research, 2017, 448: 166–174.
- [13] TAMBURRINI K C, TERRAPON N, LOMBARD V, et al. Bioinformatic analysis of lytic polysaccharide monoxygenases reveals the pan-families occurrence of intrinsically disordered C-terminal extensions[J]. Biomolecules, 2021, 11: 1632.
- [14] The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021[J]. Nucleic acids research, 2021, 49(D1): 480–489.
- [15] DUVAUD S, GABELLA C, LISACEK F, et al. ExPasy, the Swiss Bioinformatics Resource Portal, as designed by its users[J]. Nucleic acids research, 2021, 49(1): 216–227.
- [16] TAMURA K, STECHER G, KUMAR S. MEGA11: molecular evolutionary genetics analysis version 11[J]. Molecular biology and evolution, 2021, 38(7): 3022–3027.
- [17] LETUNIC I, BORK P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool[J]. Nucleic acids research, 2024, 52(W1): 78–82.
- [18] GUPTA R, BRUNAK S. Prediction of glycosylation across the human proteome and the correlation to protein function[J]. Pacific symposium on biocomputing, 2002, 7: 310–322.
- [19] STEENTOFT C, VARKHRUSHEV S Y, JOSHI H J, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology[J]. The EMBO journal, 2013, 32(10): 1478–1488.
- [20] WANG J, CHITSAZ F, DERBYSHIRE M K, et al. The conserved domain database in 2023[J]. Nucleic acids research, 2023, 51(D1): 384–388.
- [21] BAILEY T L, ELKAN C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers[C]//International Society for Computational Biology. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. California: AAAI Press, 1994: 28–36.
- [22] CHEN C J, WU Y, LI J W, et al. TBtools-II: a “one for all, all for one” bioinformatics platform for biological big-data mining[J]. Molecular plant, 2023, 16(11): 1733–1742.
- [23] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold[J]. Nature, 2021, 596: 583–589.
- [24] SPAN E A, SUESS D L M, DELLER M C, et al. The role of the secondary coordination sphere in a fungal polysaccharide monoxygenase[J]. ACS Chemical biology, 2017, 12(4): 1095–1103.
- [25] SIMMONS T J, FRANDBSEN K E H, CIANO L, et al. Structural and electronic determinants of lytic polysaccharide monoxygenase reactivity on polysaccharide substrates[J]. Nature communications, 2017, 8: 1064.
- [26] FRANDBSEN K E, POULSEN J C, TOVBORG M, et al. Learning from oligosaccharide soaks of crystals of an AA13 lytic polysaccharide monoxygenase: crystal packing, ligand binding and active-site disorder[J]. Acta crystallographica section D: structural biology, 2017, 73: 64–76.
- [27] SORIMACHI K, LE G M F, WILLIAMSON G, et al. Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to beta-cyclodextrin[J]. Structure, 1997, 5(5): 647–661.
- [28] SANTIVERI M, ROA E A, KUHNE C, et al. Structure and function of stator units of the bacterial flagellar motor[J]. Cell, 2020, 183(1): 244–257.

责任编辑: 郎婧