

黑曲霉 α -L-鼠李糖苷酶 Rha1 的二级结构预测及三级结构模拟

吴喆瑜, 于越, 巩建业, 李利君, 倪辉

(集美大学食品与生物工程学院, 福建厦门 361021)

摘要: 本文研究了酶的结构, 有利于进一步了解酶的功能, 通过对 α -L-鼠李糖苷酶 Rha1 二级结构分析及三级结构模拟, 为该酶进行理性设计提供基础。采用 GORIII、PHD 等九个二级结构预测方法对 Rha1 的二级结构进行分析和预测。结果显示, Rha1 二级结构为 23.51% α -螺旋, 29.77% β -折叠, 46.71% 无规则卷曲, 具有一个与 GH78 家族相似的 $(\alpha/\alpha)_6$ 桶状结构域及两个 β 三明治折叠片结构域。利用 Swiss-Model 以单模板建模的方法构建了 Rha1 的三维结构模型。根据二级结构的结果, 将 Rha1 分为三个结构域, 利用 Modeller 9.15 以多模板分段建模、结构域拼接的方法构建了该酶的三维结构模型, 并对其进行了能量最小化优化。利用 Ramachandran 图以及 Verify_3D 评估对两种方法构建得到的 Rha1 的三维结构模型进行评价, 其中 Swiss-Model 构建得到的模型未通过评价, Modeller 9.15 构建得到的 Rha1 的三维模型结构合理, 通过评价。 α -L-鼠李糖苷酶 Rha1 模型的构建为今后深入研究该酶的结构与功能奠定了基础。

关键词: α -L-鼠李糖苷酶; 生物信息学; 二级结构; 同源建模

文章编号: 1673-9078(2017)6-62-73

DOI: 10.13982/j.mfst.1673-9078.2017.6.010

Secondary Structure Prediction and Three-dimensional Structure

Simulation of the Recombinant α -L-Rhamnosidase Rha1

WU Zhe-yu, YU Yue, GONG Jian-ye, LI Li-jun, NI Hui

(College of Food and Biological Engineering, Jimei University, Xiamen, 361021, China)

Abstract: Exploration of enzyme structure can provide a better understanding of enzyme functions. This study aimed to analyze the secondary and tertiary structures of the α -L-rhamnosidase Rha1 and provide a basis for the rational design of this enzyme. Nine secondary structure prediction methods, including GORIII and PHD, were used to analyze and predict the secondary structure of Rha1. The results showed that Rha1 was composed of 23.51% α -helix, 29.77% β -sheet, and 46.71% random coil. Rha1 included three domains: a $(\alpha/\alpha)_6$ barrel structure, similar to that of the GH78 family, and two β -sandwich domains. The Swiss-Model was used to simulate the three-dimensional structure of Rha1 with a single template. The results showed that Rha1 was divided into three domains. Modeller 9.15 was used to construct the three-dimensional structure of r-Rha1 with multi-template segment modeling and domain splicing. The new simulated r-Rha1 three-dimensional structure was optimized by energy minimization. The two models simulated using Swiss-Model and Modeller 9.15 were evaluated by Ramachandran plot and verify-3D, respectively. The results showed that the model simulated by Swiss-Model was not qualified, whereas the model simulated by Modeller 9.15 had a reasonable structure and passed the assessment. The construction of an α -L-rhamnosidase Rha1 model establishes a foundation for the further study of the structure and biological function of α -L-rhamnosidase Rha1.

Key words: α -L-rhamnosidase; bioinformatics; secondary structure; homology modeling

α -L-鼠李糖苷酶 (EC 3.2.1.40) 是一种特异性水解多糖和糖苷配体末端 α -L-鼠李糖的水解酶^[1], 主要

收稿日期: 2016-10-18

基金项目: 国家自然科学基金项目 (31371751); 厦门市科技计划项目 (3502z20153008)

作者简介: 吴喆瑜 (1992-), 女, 硕士研究生, 研究方向: 生物学

通讯作者: 李利君 (1973-), 女, 副教授, 理学博士, 研究方向: 微生物分子育种及酶的定向进化

在饮料加工行业和药物制备行业具有良好的应用价值。例如: 在饮料加工行业, 该酶可用于对柑橘类果汁进行脱苦处理和澄清处理, 消除橙汁中的橘皮苷晶体, 增加葡萄酒的香气, 提高芦笋汁的抗氧化活性^[2]; 在药物制备行业, α -L-鼠李糖苷酶可以用来制备具有抗炎和抗病毒的作用的普鲁宁^[3]。为了满足 α -L-鼠李糖苷酶在食品及医药工业中的广泛应用, 除了寻找新的酶的来源外, 也可以通过对已发现的 α -L-鼠李糖苷

酶进行定向改造来满足工业上对酶一些性质的要求。

目前,采用理性设计改变酶的性质是计算生物学与酶工程领域中的一个研究热点。与传统的定向进化的方法相比,理性设计具有目的性强、效率高的特点,然而,对酶进行理性设计之前,需要进行相关生物信息学的分析及预测,选择特定的位点进行突变,要求三级结构清楚的酶。目前,比较可靠的分析酶结构的方法有 X 射线晶体衍射的技术、NMR 及冷冻电镜技术。这些方法得到的结构十分准确,但是应用这些技术最大的缺陷是测定方法复杂,价格昂贵,技术难度大,这就限制了这些方法的使用^[4]。随着计算生物学的发展,对于还未得到晶体结构的那些酶,目前可以利用蛋白质结构预测的方法来了解蛋白质结构信息。对于蛋白质三级结构的模拟,目前最常用的是同源建模的方法。同源建模是根据蛋白质二级结构的保守序列进行模板选择,使用与目标序列相似性最高的一个或多个同源蛋白质的三级结构作为模板,对目标序列进行三级结构的模拟。该方法适用范围广,方便快捷,可以有效地对未得到晶体结构的蛋白质进行三级结构的模拟。

由 CAZy 数据库可知, α -L-鼠李糖苷酶主要来源于 GH13, GH78, GH106 这三个糖苷水解酶家族。其中 GH13 家族中多数为其他类型的糖苷水解酶,仅少数具有 α -L-鼠李糖苷酶活性;GH78 家族, GH106 家族只包含 α -L-鼠李糖苷酶,是主要 α -L-鼠李糖苷酶家族。其中, GH78 家族中 α -L-鼠李糖苷酶 GenBank 序列数多达三万多条,但只解出 4 个细菌来源的 α -L-鼠李糖苷酶晶体结构^[5-8]。这四个 α -L-鼠李糖苷酶晶体结构均具有 1 个特征的 $(\alpha/\alpha)_6$ 的桶状催化域结构和不同数量的 β -三明治折叠片结构域。目前有关该酶的结构与功能研究一直还不够清楚,需要进一步深入分析 α -L-鼠李糖苷酶的结构,以此深入解析其多样性。

前期,实验室已从黑曲霉(*Aspergillus niger*)JMU-TS528 中发现一个新的 α -L-鼠李糖苷酶 Rha1^[9],对其进行克隆并表达。为了进一步研究 α -L-鼠李糖苷酶 Rha1 的性质,本研究利用同源建模的方法构建出 Rha1 的完整结构模型,并对该模型自身的合理性进行验证。该模型的建立为深入探究 α -L-鼠李糖苷酶的结构与功能的关系以及利用理性设计进行相应酶学性质的改造提供了必要的基础。

1 材料与方法

1.1 材料

供试材料 α -L-鼠李糖苷酶 Rha1 cDNA 序列(2062

bp, NCBI 登录号为 KC750908.1), 氨基酸序列 (655AA)。

供试生物信息数据库包括 NCBI (核酸数据库 <http://www.ncbi.nlm.nih.gov>)、Expasy (蛋白质数据库 <http://www.expasy.org>)、蛋白质序列网络分析 Network Protein Sequence @analysis (<http://www.bioinf.manchester.ac.uk/dbbrowser/bioactivity/NPS2.html>)、同源建模模板搜索服务器 Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>)、HHpred (<http://toolkit.genzentrum.lmu.de/hhpred>)、Verify-3D 在线评分服务器 The Structure Analysis and Verification Server version 4(<http://services.mbi.ucla.edu/SAVES/>)及 Protein Structure Validation Software suite (http://psvs-1_4-dev.nesg.org/)。

供试生物信息软件包括多重序列比对软件 DNAMAN、三级结构建模软件 Swiss-Model (<http://www.expasy.ch/swissmod/SWISS-MODEL>)和 Modeller9.15、生物大分子三级结构优化及分析软件 Discovery Studio 2.5。

1.2 方法

1.2.1 α -L-鼠李糖苷酶 Rha1 二级结构分析

利用法国里昂 Institut de Biologie et Chimie des Protéines 提供的蛋白质序列网络分析 NPS@(<http://www.bioinf.manchester.ac.uk/dbbrowser/bioactivity/NPS2.html>)进行。

通过将 α -L-鼠李糖苷酶 Rha1 的氨基酸序列提交至蛋白质序列网络分析 NPS@,选择九个二级结构在线分析方法 DPM, DSC, GOR I, GORIII, HNN, MLRC, PHD, SOPM, Predator, 分别对 α -L-鼠李糖苷酶 Rha1 的二级结构成分,包括 α -螺旋、 β -折叠、无规则卷曲及延伸链等结果进行输出并比较分析。

1.2.2 Swiss-Model 建模

三级结构建模是利用瑞士日内瓦 Glaxo Smith Kline 中心提供的自动蛋白分子建模服务器(homology modeling) Swiss-Model (<http://www.expasy.ch/swissmod/SWISS-MODEL>)进行。将 α -L-鼠李糖苷酶 Rha1 的氨基酸序列提交至 Swiss-Model 服务器,进行单模板建模获得该蛋白质的三级结构,同源建模返回的结构(pdb 格式)可在 Discovery Studio 可视化即得到 α -L-鼠李糖苷酶的初始的三级结构。

1.2.3 Modeller9.15 建模

1.2.3.1 α -L-鼠李糖苷酶 Rha1 同源建模模板搜索

①利用 Blast 对 α -L-鼠李糖苷酶 Rha1 同源建模模板进行搜索

模板搜索是利用美国国家生物技术信息中心(National Center of Biotechnology Information)中心提供的 Blast(<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)进行。将 α -L-鼠李糖苷酶 Rha1 的序列提交至 Blast 数据库,选择 Protein Blast,在数据库搜索一栏选择 Protein Data Bank proteins (PDB),通过搜索可得到与 α -L-鼠李糖苷酶 Rha1 具有同源性的模板。

②利用 Phyre2 对 α -L-鼠李糖苷酶 Rha1 同源建模模板进行搜索

模板搜索是利用英国伦敦帝国理工学院的生物信息小组提供的 Protein Homology Recognition Engine V 2.0 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>)进行。在 Phyre2 在线数据库填写基本信息,将 α -L-鼠李糖苷酶 Rha1 的序列提交至 Phyre2 数据库,在 modelling mode 选项中选择 intensive,利用 phyre search 进行搜索,通过搜索可得到与 α -L-鼠李糖苷酶 Rha1 具有同源性的模板。

③利用 HHpred 对 α -L-鼠李糖苷酶 Rha1 同源建模模板进行搜索

模板搜索是利用德国 Max-Planck 发育生物学研究所提供的 Homology detection & structure prediction by HMM-HMM comparison (<http://toolkit.tuebingen.mpg.de/hhpred>)进行。将 α -L-鼠李糖苷酶 Rha1 的序列提交至 HHpred 数据库,在数据库搜索一栏选择最新的 PDB 数据库,通过搜索可得到与 α -L-鼠李糖苷酶 Rha1 具有同源性的模板。

1.2.3.2 三维模型的建立

利用 Modeller9.15 中的单模板建模及多模板建模进行三维结构模拟。首先选择已搜索得到的多个蛋白质结构作为 α -L-鼠李糖苷酶 Rha1 同源建模的各部分模板,通过 model_singal.py 等脚本进行单模板建模,通过脚本 multiple_template/salign.py, multiple_template/align2d_mult.py, multiple_template/model_mult.py 等脚本进行多模板建模,在 Windows 系统下,通过对建模模板比对及建模条件的约束,构建具有完整结构的三级结构。

1.2.3.3 模型的优化

在 Modeller 9.15 中,通过脚本 loop_refine.py 实现对该三级结构各区域 loop 的优化及整体 loop 的优化。

在 Discovery Studio2.5 中的 Minimizer 模块下对模型进行能量最小化,首先将模型导入 Discover Studio 可视化界面中,将模型选中,选择 chemistry 中的 hydrogens,选择 Add 加 H;选择 Tool 中的 Simulated Structure 中选择 Apply Forcefield 给模型添加立场;选择 Protocols 中的 Minimization;在 Minimization 中的

Algorithm 中选择 Stepest descent, Max Steps 选择 100,在一次优化的基础上再次选择 Algorithm 中 Conjugate Descent, MaxSteps 选择 200, RMS Gradient 均为 0.1。通过两次模型的优化,使模型处在能量最低,结构最合理的情况。

1.2.4 模型的评价

1.2.4.1 利用 Ramachandran 图对优化之后的得到的模型进行评价

利用 Discovery Studio2.5 对优化之后的得到的模型进行 Ramachandran 图分析,首先将模型导入 Discovery Studio 可视化界面中,将模型选中,选择 chart 中的 Ramachandran plot 进行评价。

1.2.4.2 利用 Verify-3D 对优化之后的得到的模型进行评价

Verify-3D 评价是利用 The Structure Analysis and Verification Server version 4(<http://services.mbi.ucla.edu/SAVES/>)及 Protein Structure Validation Software suite (http://psvs-1_4-dev.nesg.org/)进行。

将模型提交至 SAVES 在线服务器中,根据提示输入验证码,即可进行模型 Verify-3D 评价。打开 PSVS 在线服务器,选择 run PSVS,按要求填写所需要的信息,type of structure 选择 homology model,将模型提交至 PSVS 在线服务器,即可进行模型 Verify-3D 评价。

2 结果与讨论

2.1 α -L-鼠李糖苷酶 Rha1 的二级结构预测

二级结构在线预测的主要原理是通过一定的算法来计算局部残基的相互影响,同时再额外结合多重序列比对的思想,综合分析评价后给出蛋白质最可能形成的二级结构。蛋白质二级结构数据库蛋白质二级结构归属数据库(Database of Secondary Structure assignments for all Protein entries, DSSP)中对蛋白质二级结构采用八态定义,如表 1 所示。

表 1 蛋白质二级结构八态定义

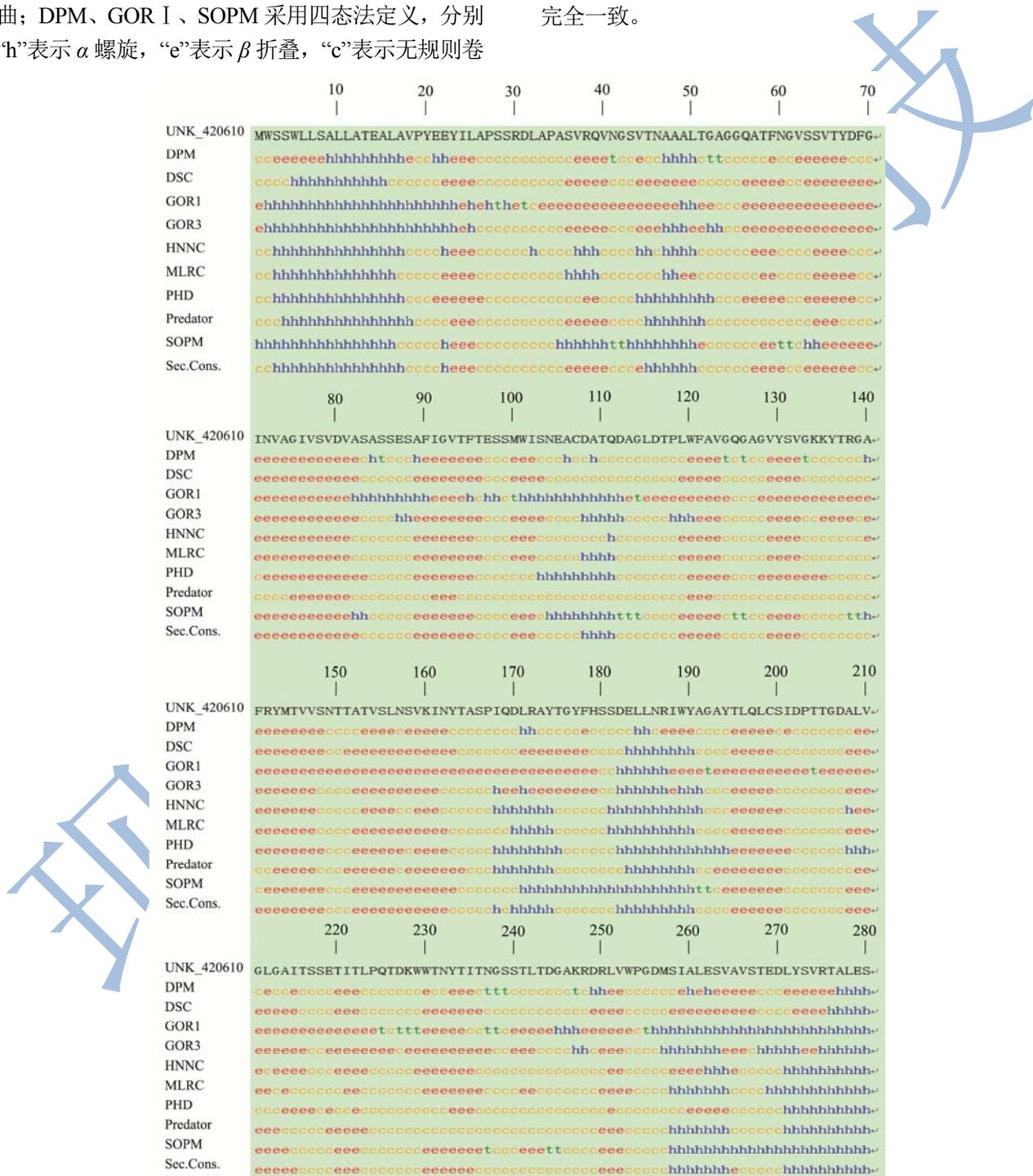
Table 1 Eight descriptions of protein secondary structure

序号	八态定义	主要内容
1	h	α 螺旋
2	g	310-螺旋
3	i	π -螺旋
4	e	β 折叠
5	b	β 连接
6	t	β 转角
7	s	β 弯曲
8	c	无规则卷曲

目前，已开发几十种蛋白质二级结构预测方法。在众多的二级结构预测方法中，根据文献所报导的准确性及重要性选择了九个有代表性的蛋白质二级结构预测方法 DPM、DSC、GOR I、GORIII、HNN、MLRC、PHD、Predator 和 SOPM。而上述所选择的 9 种方法对二级结构的定义各有不同。其中 DSC、GORIII、HNN、MLRC、PHD 和 Predator 采用三态法定义，分别用“h”表示 α 螺旋，“e”表示 β 折叠，“c”表示无规则卷曲；DPM、GOR I、SOPM 采用四态法定义，分别用“h”表示 α 螺旋，“e”表示 β 折叠，“c”表示无规则卷

曲，“t”表示转角。鉴于这种二级结构定义的不一致，为了统一对比标准，本研究在评价时采用了三态法即 α 螺旋、 β 折叠和无规卷曲来进行相关的比较分析。

将 α -L-鼠李糖苷酶 Rha1 的序列提交至上述九个在线二级结构预测服务器，得到的结果见图 1。由图 1 预测结果可得，由于不同的预测方法做选用的算法或运算矩阵不同，或者多重序列比对时所选取的数据库有所差别，最终会导致不同方法预测出的二级结构不完全一致。



依据所选用算法,这九种预测方法可分为三种不同类型:第一类为统计经验算法,其中包括基于经验统计规则的 MLRC 法、基于双重预测方法的 DPM 法和基于信息论算法的 GOR I 法、GORIII法,预测结果见表 2;第二类为物理化学算法,其中包括基于蛋

白质结构的物理化学原理的 Predator 法、DSC 法和 SOPM 法,预测结果见表 3;第三类为机器学习算法,其中包括基于神经网络算法的 HNN 法和 PHD 法,预测结果见表 4。

表 2 基于统计经验算法对 Rha1 的二级结构预测结果

Table 2 Results of Rha1 secondary structure prediction based on empirical algorithms

方法	α 螺旋(h)	β 折叠(e)	无规则卷曲(c)
MLRC	24.89% (163AA)	25.04% (164AA)	50.08% (328AA)
DPM	11.60% (76AA)	32.06% (210AA)	49.31% (323AA)
GOR I	25.50% (167AA)	54.05% (354AA)	11.45% (75AA)
GORIII	28.85% (189AA)	41.68% (273AA)	29.47% (193AA)

表 3 基于物理化学算法对 Rha1 的二级结构预测结果

Table 3 Results of Rha1 secondary structure prediction based on physical chemistry algorithms

方法	α 螺旋(h)	β 折叠(e)	无规则卷曲(c)
DSC	9.47% (62AA)	43.97% (288AA)	46.56% (305AA)
Predator	20% (131AA)	17.86% (117AA)	62.14% (407AA)
SOPM	29.62% (194AA)	29.01% (190AA)	9.31% (61AA)

表 4 基于神经网络算法对 Rha1 的二级结构预测结果

Table 4 Results of Rha1 secondary structure prediction based on neural network algorithms

方法	α 螺旋(h)	β 折叠(e)	无规则卷曲(c)
HNN	26.41% (73AA)	23.82% (156AA)	49.77% (326AA)
PHD	29.01% (190AA)	28.55% (187AA)	42.44% (278AA)

第一类基于统计经验算法的四种二级结构预测方法中较经典的是 GOR 法,该方法是依据 Garnier 等在 1978 年借用了信息论原理和贝叶斯统计学的方法,首先提出的 GOR I 的预测方法。这种方法根据已知的各氨基酸构象结构及其在序列中的位置,先构建出一个信息量表,然后通过这一信息量表,对应预测任意位置的任意氨基酸可能形成的二级结构。GOR 方法具有物理意义明确,数学上较严格,程序易编写的特点,也是所有算法中理论基础最好的方法^[14]。这种方法的主要局限性在于仅考虑了单个氨基酸残基之间的影响,没有考虑到不同残基之间的相互作用,这导致了预测结果的正确率不高,GOR I 的正确率仅为 56%。后来 Giba 等又把残基对之间的相互作用考虑进去,还添加了相邻残基对预测位置氨基酸构象的影响,改进为 GOR III,其最终的预测正确率提高至 63%。

第二类基于物理化学算法的三种二级结构预测方法中,近年来较为流行的是混合预测法 Predator 法,该方法是用多种方法分别进行预测后,确定不同方法的权重以合并预测结果,同时该方法的局限性在于混合方法在权重上很难把握,根据相关文献^[14]表明,利用该方法进行预测的正确率为 68%。

第三类二级结构预测方法是基于神经网络法,神

经网络方法在 1988 年首次被应用于蛋白质的二级结构预测中,但是在早期应用时的准确率并没有很高,仅有 64.3%。经过长时间的发展,基于这一原理的二级结构预测方法已经达到 800 多种,而其中最受欢迎的就是在本研究中所用到的 PHD 方法,其利用高精度的“黑盒”预测法,是基于复杂的非线性算法网络,采用了三层前馈神经网络对蛋白质二级结构进行预测,并且利用多重序列比对获取更多的同源序列信息,还有结合结构类的预测结果,最终使得预测准确率大为提高。这种方法是一种非参数方法,具有很强的学习能力和适应能力,应用方便,准确率较高。但是它的预测方法并没有考虑到蛋白质本身的理化性质,有时会出现难以理解的预测结果。目前为止,PHD 被认为是二级结构预测的最可靠的方法,正确率可达 72.2%^[10]。

总的来说,到目前为止,二级结构预测的准确性仍是未能完全解决的问题。现有的算法一般对于 α -螺旋预测精度较好,对 β -折叠较差,而对除 α -螺旋和 β -折叠等之外的无规则二级结构则效果很差。为了尽可能准确地获得 α -L-鼠李糖苷酶的二级结构,一般采用基于不用算法的多个预测方法进行预测,本研究同时选取了三种不同算法中具有代表性的九种预测方法进

行预测, 由蛋白质序列网络分析 NPS@运算分析, 由图 1 中 Sec.Cons 一栏可知, α -L-鼠李糖苷酶的二级结构为: α 螺旋(h)占 23.51 (154AA), β 折叠(e)占 29.77% (195AA), 无规则卷曲(c)占 46.71% (306AA), 且具有一个(α/α)₆ 的桶状结构域, 这与 GH78 家族中, α -L-鼠李糖苷酶具有(α/α)₆ 的桶状特征结构域相符合。

2.2 利用 Swiss-Model 建立 α -L-鼠李糖苷酶 Rha1 的三维模型

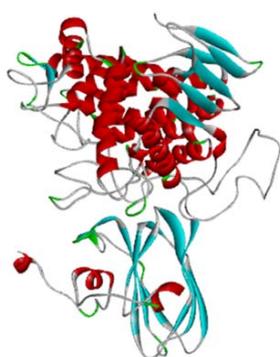


图 2 α -L-鼠李糖苷酶 Rha1 Swiss-Model 建模

Fig.2 α -L-rhamnosidase Rha1 modelling by Swiss-Model

目前最常用的三维结构预测的方法是 Swiss-

Model 在线同源建模法^[11], 该方法方便快捷。本研究通过 Swiss-Model 在线服务器对 α -L-鼠李糖苷酶 Rha1 进行单模板建模。Swiss-Model 服务器根据 Rha1 氨基酸序列选择覆盖率最广, 相似度最高的 PDB: 2OKX 作为模板进行建模, 并由此构建了 α -L-鼠李糖苷酶 Rha1 的三级结构模型, 建模结构如图 2 所示。最终得到的模型中模拟了 606 个氨基酸, 而 α -L-鼠李糖苷酶 Rha1 氨基酸全长 656 个氨基酸, 前 17 个氨基酸为信号肽, 在模拟结构时去掉信号肽部分的氨基酸后, 序列中还剩下 639 个氨基酸因此模型有 33 个氨基酸没有被模拟出来, 这将会影响后续蛋白质结构分析。利用 Swiss-Model 进行同源建模, 该方法更适用于数据库中相似度高 (>30%) 且具有晶体结构的同源蛋白作为单一模板进行预测的蛋白^[12]。本研究中所选择的单模板相似度只有 22%, 利用单模板建模没能得到完整的 Rha1 三级结构, 故进一步利用 Modeller9.15 采用分段建模结构域拼接的方法模拟其三级结构。

2.3 利用 Modeller 9.15 建立 α -L-鼠李糖苷酶 Rha1 的三维模型

2.3.1 Blast 模板搜索结果

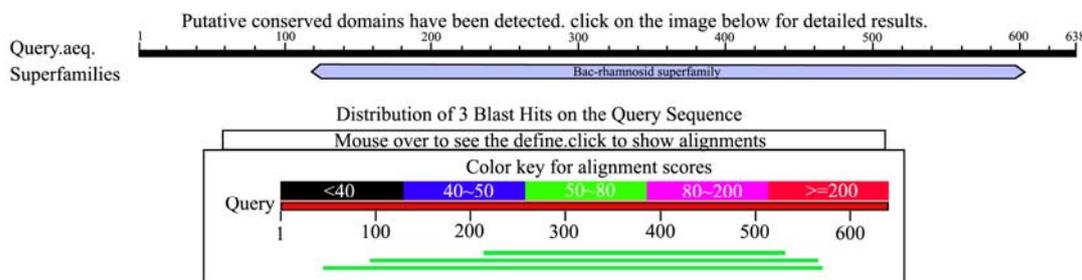


图 3 使用 Blast 进行 Rha1 模板搜索的结果

Fig.3 Results of Rha1 template searching using BLAST

通过 NCBI 中的 Protein Blast 功能, 在数据库搜索一栏选择 Protein Data Bankproteins (PDB), 从而进行对 α -L-鼠李糖苷酶的晶体模板进行搜索, 得到如图 3 的搜索结果。从图 3 结果可得: 共搜索到 3 个晶体模板, 其 PDB 号分别为 4XHC、2OKX 和 3CIH。

2.3.2 Phyre2 模板搜索结果

表 5 使用 Phyre2 进行 Rha1 模板搜索的结果

Table 5 Results of Rha1 template searching using Phyre2

序号	SCOP 编号	酶家族	序列长度
1	c3w5mA	α -L-鼠李糖苷酶	652
2	c2okxB	α -L-鼠李糖苷酶	628
3	c3cihA	α -L-鼠李糖苷酶	652
4	c4xhcB	α -L-鼠李糖苷酶	531

5	c4zlgA	磷酸化酶	378
6	c1v7wA	壳二糖磷酸化酶	402
7	c2cqtA	纤维二糖磷酸化酶	402
8	d1v7wa1	糖基转移酶	478
9	c3c67B	糖苷合成酶	339
10	c4ufcA	磷酸化酶	473
11	d2jg0a1	糖苷酶	310
12	c2jg0A	海藻糖酶	310

通过 Phyre2 服务器一共搜索出 12 个与 α -L-鼠李糖苷酶结构相似的模板, 从表 5 中可以发现其中前 4 个为 α -L-鼠李糖苷酶现有的四个晶体结构。与本研究中的 α -L-鼠李糖苷酶同属于一个家族。除了与 GH78 家族的相似度较高以外, 还会与一些磷酸化酶和其它

家族的水解酶有一定的相似性, 由于建模是以同源建模为基础, 找到多个模板对于模拟 α -L-鼠李糖苷酶的三维模型提供了更多的参考。

2.3.3 HHpred 模板搜索结果

通过 HHpred 的模板搜索, 共搜索出 98 个 PDB

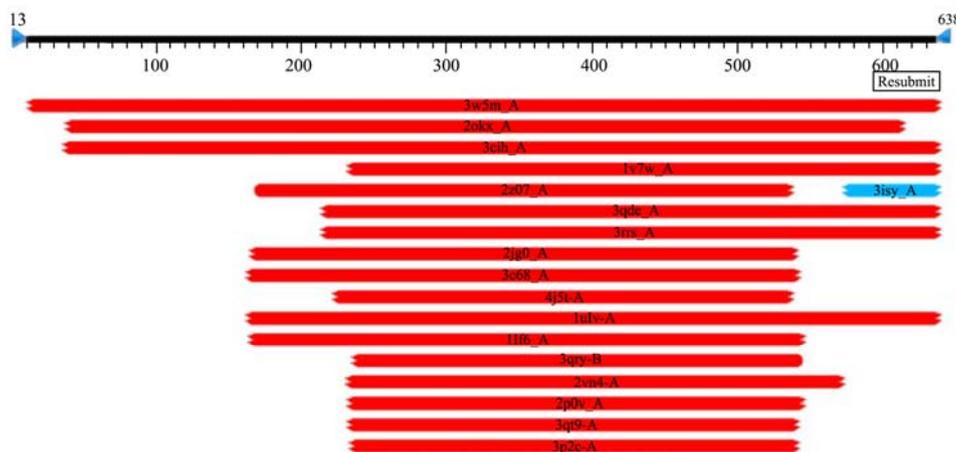


图4 使用 HHpred 进行 Rha1 模板搜索的结果

Fig.4 Results of Rha1 template searching using HHpred

2.3.4 利用多模板建模的方法建立 α -L-鼠李糖苷酶 Rha1 的三维模型

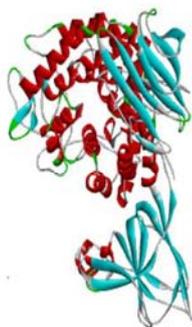


图5 α -L-鼠李糖苷酶 Rha1 Modeller9.15 建模

Fig.5 α -L-rhamnosidase Rha1 modelling by Modeller 9.15

利用 Modeller 9.15^[13]多模板建模的功能, 选择多个模板预测 α -L-鼠李糖苷酶 Rha1 的三级结构。综合模板的搜索结果和已知的 4 个 α -L-鼠李糖苷酶晶体结构, 我们发现 GH78 家族中 α -L-鼠李糖苷酶除了都有一个 $(\alpha/\alpha)_6$ 的特征桶状结构外, 其余结构域都是由不同个数的 β -折叠形成的 β 三明治折叠片构成。因此, 根据 α -L-鼠李糖苷酶 Rha1 二级结构的预测结果及已知的晶体模板结构, 我们将 Rha1 分为三个不同结构域: 第一个结构域为 1~180 位氨基酸, 主要是 β -折叠及无规则卷曲组成的 β 三明治折叠片; 第二个结构域为 181~530 位氨基酸, 主要由 6 段 α 螺旋组成的 $(\alpha/\alpha)_6$ 桶状结构; 第三个结构域为 531~639 位氨基酸, 主要是 β -折叠及无规则卷曲组成的 β 三明治折叠片。通过对这三个片段的模板搜索、比对和分析, 分别选择

的模板, 图 4 只显示了该软件评分为 100% 的模板, 其中, 相似性最高, 覆盖率最大的是前三个晶体模板, 分别为 3W5M、2OKX 和 3CIH, 这三个模板同属于 GH78 家族 α -L-鼠李糖苷酶。

PDB 序列号为 4HXC 蛋白作为 α -L-鼠李糖苷酶 Rha1 的 1~180 位氨基酸序列的模板、PDB 序列号为 2OKX 蛋白作为 α -L-鼠李糖苷酶 Rha1 的 181~530 位氨基酸序列的模板、PDB 序列号为 3W5M 蛋白作为 α -L-鼠李糖苷酶 Rha1 的 531~639 位氨基酸序列的模板。通过序列比对, 确定选择的各部分序列的相似性均高于 30%。确定好模版后, 利用 Modeller9.15 进行多模板建模, 将各部分建好的模型通过结构域拼接的方法构建完整的 α -L-鼠李糖苷酶 Rha1 的三级结构, 模型见图 5。从图 5 可以看出, α -L-鼠李糖苷酶 Rha1 是由 2 个 β -三明治折叠片结构域和 1 个 $(\alpha/\alpha)_6$ 桶状催化结构域组成。

2.3.5 α -L-鼠李糖苷酶 Rha1 的模型优化

Modeller 9.15 建立模型是初始的 Rha1 三维结构^[4], 为了获得质量优良的模型, 因此需要对 α -L-鼠李糖苷酶 Rha1 初始的三维结构进行修正。利用 Modeller 9.15 中的 loop refine 对构建好的模型进行 loop 区域的优化, 将优化之后的模型在 Discovery Studio 中的 minimizer 模块进行能量最小化, Discover Studio 能够进行分子力学的初步优化, 完成分子的动力学模拟, 可以进行最陡下降、共轭梯度等多种分子力学模拟。本研究先后利用最陡下降法在 CVFF 力场下进行 100 步的结构松弛计算及 200 步的共轭梯度法进行结构优化。利用第一种方法最陡下降法对模型进行优化, 从表 6 可得, 该模型最终的能量值为 -24029.95904 kcal/mol; 再利用第二种方法共轭梯度法对方法一得到的模型再次进行优化, 从表 7 可得, 该模型最终的能

量值为-35773.41719 kcal/mol。模型的最终能量值越低则表明该模型构象越稳定^[12]。利用共轭梯度法优化后的模型与原始的模型进行对比如图 6，其中优化较为明显的区域是 190~195 位氨基酸所形成的 loop 区域（图 7a）及 580~590 位所形成的 loop 区域（图 7b）

如。利用能量最小化对模型的 loop 区域进行合理的优化以此来稳定其 α -螺旋及 β -折叠，使 α -L-鼠李糖苷酶的结构处于能量最低的构象，维持蛋白质整体的稳定性。

表 6 α -L-鼠李糖苷酶 Rha1 模型利用最陡下降法优化结果

Table 6 Optimization results of α -L-rhamnosidase Rha1 modelling using the steepest descent method

名字	α -L-鼠李糖苷酶模型
分子力场	CHARMm
初级势能/(kcal/mol)	276576.98
势能/(kcal/mol)	-24029.95
范德瓦耳能量/(kcal/mol)	-2255.68
静电能量/(kcal/mol)	-26445.77
初始 RMS 能量梯度/[kcal/(mol·Angstrom)]	14020.85
最终 RMS 能量梯度/[kcal/(mol·Angstrom)]	2.37
能量最小化标准	循环数>能量最小化循环至 100 步之后退出
能量最小化循环数	100

表 7 α -L-鼠李糖苷酶 Rha1 利用共轭梯度法优化结果

Table 7 Optimization results of α -L-rhamnosidase Rha1 modelling using the conjugate gradient method

名字	α -L-鼠李糖苷酶模型
分子力场	CHARMm
初级势能/(kcal/mol)	276576.97
势能/(kcal/mol)	-35773.41
范德瓦耳能量/(kcal/mol)	-3574.21
静电能量/(kcal/mol)	-37961.12
初始 RMS 能量梯度/[kcal/(mol·Angstrom)]	14020.85
最终 RMS 能量梯度/[kcal/(mol·Angstrom)]	1.43
能量最小化标准	循环数>能量最小化循环至 200 步之后退出
能量最小化循环数	200

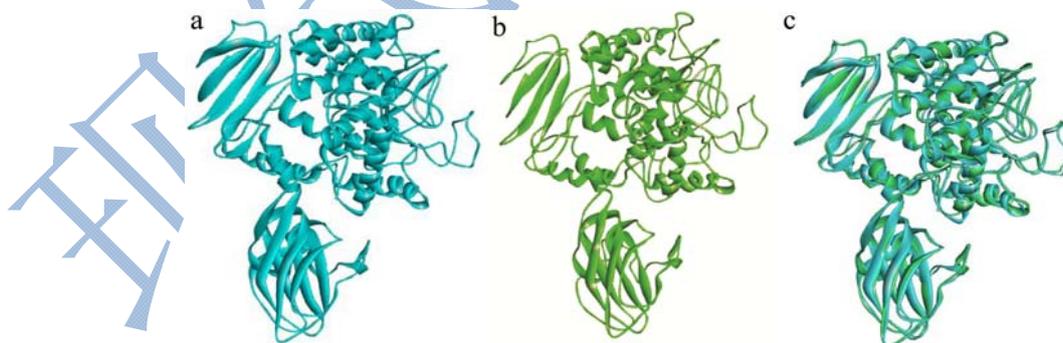


图 6 Modeller9.15 模拟的 α -L-鼠李糖苷酶 Rha1 优化前后的模型图

Fig.6 Original α -L-rhamnosidase Rha1 model and the model optimized by Modeller 9.15. The blue domain is the original model, and the green domain is the optimized model

注：a，原始型 Rha1 的三维模型；b，Rha1 优化后的三维模型；c，优化前后模型的叠合。

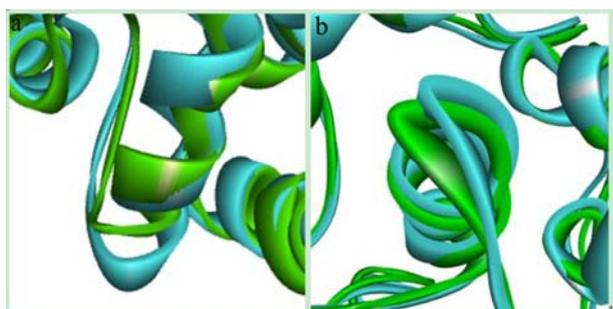


图7 Modeller9.15 模拟的 α -L-鼠李糖苷酶 Rha1190-195 位氨基酸区域的优化 (a) 及 580-590 位氨基酸区域的优化 (b)

Fig.7 The 190-195 amino acid loop refinement of the original α -L-rhamnosidase model by Modeller 9.15 (a), and the 580-590 amino acid loop refinement of the optimized α -L-rhamnosidase model by Modeller 9.15 (b)

注: 蓝色结构域为原始模型; 绿色结构域为优化后的模型。红色的箭头代表优化的区域。

2.4 α -L-鼠李糖苷酶 Rha1 的三维模型评价

2.4.1 Ramachandran 图评价

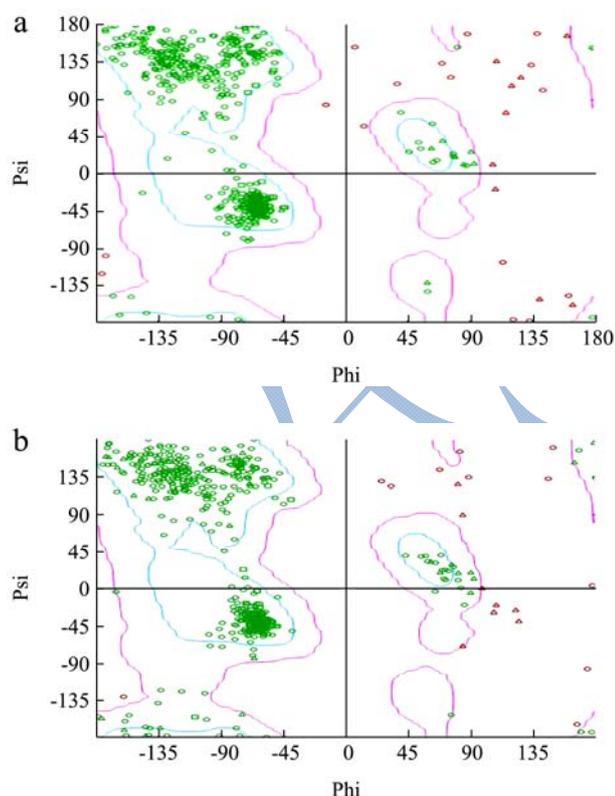


图8 利用 Swiss-Model (a) 和 Modeller9.15 (b) 对 α -L-鼠李糖苷酶 Rha1 建模模型的 Ramachandran 图

Fig.8 Ramachandran plot of α -L-rhamnosidase Rha1 by Swiss-Model (a) and Modeller 9.15 (b)

注: 蓝色区域代表最合适的区域; 紫色区域代表允许区域; 红色的点代表不合理区域。

Ramachandran 图是目前评价三维模型最常见的方法, 主要描述蛋白质或肽立体结构中肽键内 α 碳原子和氮原子间的键的旋转度 (φ) 和 α 碳原子和羰基碳原子间的键的旋转度 (ψ), 同时指出蛋白质或肽类中氨基酸的允许和不允许的构象。当大于或等于 90% 的 φ 、 ψ 二面角分布在允许区的时候, 则为正常分布。利用 Discovery Studio 对 Swiss-Model 和 Modeller 9.15 预测模拟得到的 Rha1 的模型进行结构合理性评价, 对蛋白质结构的立体化学性质, 其主链、侧链立体化学稳定性进行评估, 计算得到该模型的 Ramachandran 图 (图 9)。图中蓝色区域为“最适区”, 该区域含有的氨基酸个数越多, 结构越可信; 紫色区域为“允许区”; 其他区域的点 (红色) 为 φ - ψ 构象不合理的氨基酸^[14]。

从图 8 对比可看出, 利用两个软件建立出来的 α -L-鼠李糖苷酶三维结构大部分都在允许区的范围内, Swiss-Model 建立出来的三维模型中蓝色区域和紫色区域为 95.9%, 红点区域为 4.1%; Modeller 9.15 建立出来的三维模型中蓝色区域和紫色区域为 97.1%; 红点区域为 2.9%, 即 Swiss-Model 建立得到的三维模型 φ - ψ 不合理区域多于 Modeller 9.15 建立出来的模型。在同源建模中, 最信任的区域是结构性质中主要的参考指标, 预期其大于 90% 为正常分布^[14]。由评价图最终得到结果 Swiss-Model 模拟的预期值为: 95.9%, 而 Modeller 9.15 建立出来的模型预期值为 97.1%。从模型的完整性及预期值两者综合得出 Modeller 9.15 建立出来的模型更为可靠。

2.4.2 Verify-3D 评价

目前评价三维模型另一种较常见的方法为 Verify-3D, 是基于“穿线”法的模型评估程序, 采用 3D-1D 的打分函数来检测所构建模型与自身氨基酸序列的匹配度关系。该模块是通过统计分析得出的结果, 来评估蛋白质的三维结构与其氨基酸序列是否相容, 分数越高, 说明同源建模的可信度越大^[4]。利用该评价方法对 Swiss-Model 和 Modeller9.15 模拟得到的 Rha1 模型进行评价^[15], 分别以每个氨基酸的相容性得分来代表氨基酸与周围的环境的情况, 如果相容性得分为负, 则表示该三维结构可能不正确; 若结果中至少有 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 检测且认为该结构合理^[16]。

将 Swiss-Model 建立的模型及 Modeller9.15 建立的模型分别放入 SAVE 评价软件中, 结果如图 9~10。Swiss-Model 建立的模型 Verify-3D 分数为 79.4%, 低于 80%, 在评价中显示为警告; Modeller9.15 建立出来的模型评分为 84.2%, 高于 80%, 通过评价。

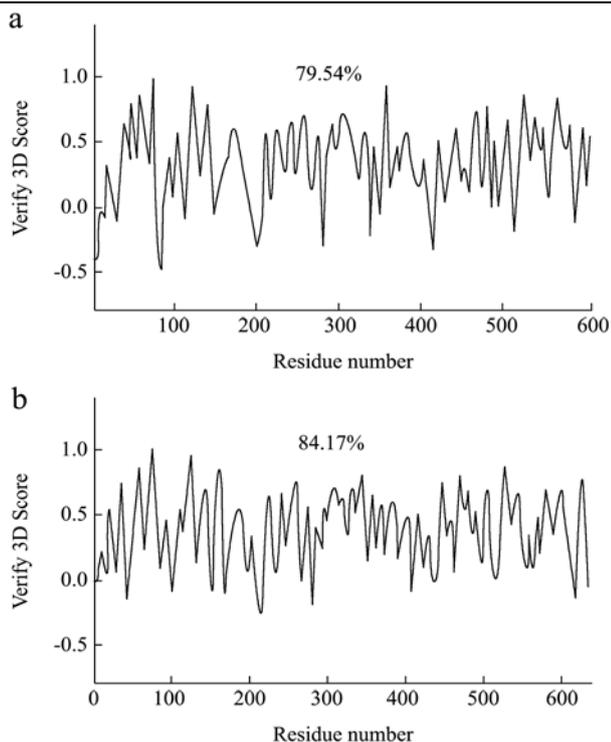


图9 Swiss-Model (a) 及 Modeller 9.15 (b) α -L-鼠李糖苷酶 Rha1 建模模型 Verify-3D 评分

Fig.9 Verify3D score of α -L-rhamnosidase Rha1 by Swiss-Model (a) and Modeller 9.15 (b)

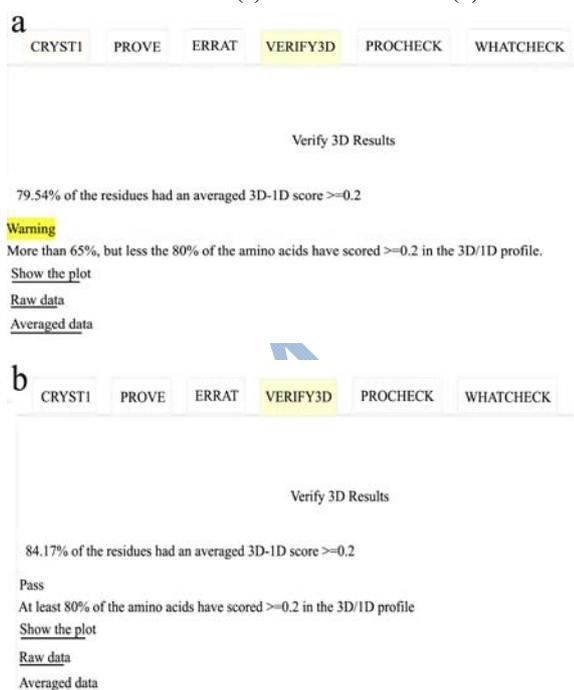


图10 Swiss-Model (a) 及 Modeller 9.15 (b) α -L-鼠李糖苷酶 Rha1 建模模型 Verify-3D 评分结果

Fig.10 Verify3D score of α -L-rhamnosidase Rha1 for Swiss-Model (a) and Modeller 9.15 (b) models

综合以上两种评价体系的评定结果, 利用 Swiss-Model 建立的 α -L-鼠李糖苷酶 Rha1 三维结构模型评

价结果均低于利用 Modeller 9.15 建立的模型, 故 Modeller 9.15 建立的模型有较高的可靠性, 可利用该模型进行后续的研究。

Swiss-Model 建模的原理是采用单一的相似度最高的目标蛋白进行单模板建模, 当目标蛋白与模板序列同源性低于 30% 时, 采用单模板进行三维结构模拟得到的模型较难通过评价^[11]。为了更好地解决模板同源性较低的问题, 本研究中利用结构域划分的方法将去除信号肽的 Rha1 分为三个片段, 再分别进行模板比对, 通过序列比对, 结果显示 Rha11-200 氨基酸与 PDB:4XHC 相似度为 35.5%, 200-530 氨基酸与 PDB:2OKX 相似度为 32.6%, 561-639 氨基酸与 PDB:3W5M 相似度为 31.8%, 最终采用多模板建模的方法模拟得到 Rha1 的三维结构并通过评价。利用多模板分段建模的方法, 为序列同源性低的蛋白质进行三维结构模拟提供了新的参考。

2.5 α -L-鼠李糖苷酶 Rha1 的三维模型的结构特征

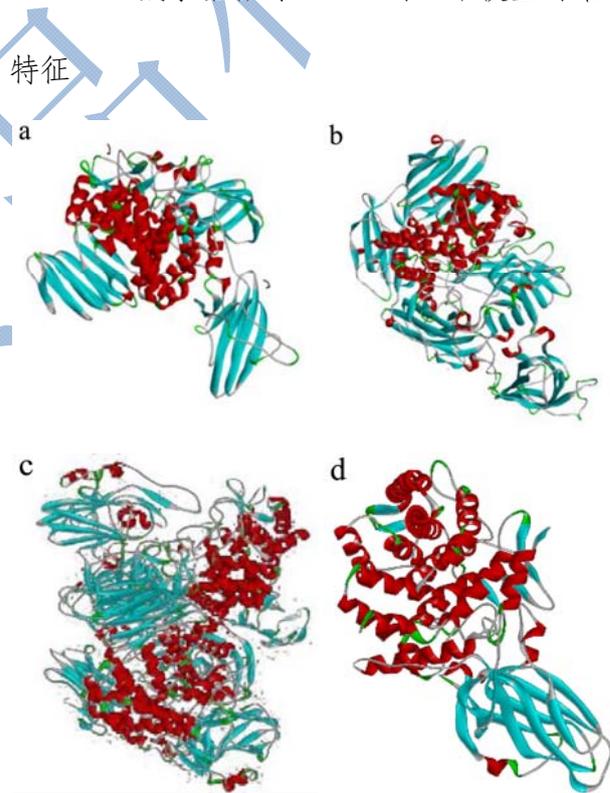


图 11 四个已知的 α -L-鼠李糖苷酶的晶体结构

Fig.11 Four known crystal structures of α -L-rhamnosidase.

Red domains are $(\alpha/\alpha)_6$ barrel catalytic structures

注: a 表示 3CIH; b 表示 3W5M; c 表示 2OKX; d 表示 4XHC。

构建得到的 α -L-鼠李糖苷酶 Rha1 由 3 个结构域组成, 包括 1 个典型的 GH78 家族的 $(\alpha/\alpha)_6$ 桶状结构域和 2 个 β -三明治折叠片结构域。结合二级结构分析可

得, N-端结构域是一个由 Val1-Arg170 组成的 β -三明治折叠片结构域, Ile171-Arg571 组成的是 1 个 $(\alpha/\alpha)_6$ 的桶状结构域, C-端结构域是一个由 Leu572-Val630 组成的 β -三明治折叠片结构域。在四个已知的晶体结构中, 都具有一个 $(\alpha/\alpha)_6$ 的桶状结构域及不同数目的 β -三明治折叠片结构域(图 11)。其中 3CIH 具有 3 个 β -三明治折叠片结构域^[7], 3W5M 具有 4 个 β -三明治折叠片结构域^[6], 2OKX 具有 5 个 β -三明治折叠片结构域^[5], 4XHC 具有 1 个 β -三明治折叠片结构域^[8]。与这四个已知晶体结构不同的是 Rha1 具有 2 个 β -三明治折叠片结构域。

自 1973 年以来, α -L-鼠李糖苷酶已经研究了 30 多年。在 30 多年的研究中发现, α -L-鼠李糖苷酶来源较广泛, 存在于植物、动物及微生物中; 性质较多样, 具有宽泛的 pH 范围及广泛的底物适应性。通过对比已有的 α -L-鼠李糖苷酶的序列和结构特征发现, 四个晶体间的序列同源性低, 均小于 27%, 结构差异较大, 不同结构域的功能还没人研究, 缺乏对酶的结构与功能之间的关系的深入研究。因此, 通过对不同来源的 α -L-鼠李糖苷酶进行三维结构的构建和相关性质分析, 有助于明确该酶的结构与功能之间的关系, 推动这方面研究工作的开展。Rha1 模型虽然通过了评价, 但在今后的使用过程中一方面应根据实验结果和其它的计算生物学手段对其模型继续进行优化, 使之尽可能接近酶的实际情况; 另一方面仍需继续研究其晶体结构, 最终为深入研究该酶的结构与功能的关系奠定结构基础。

3 结论

3.1 随着蛋白质定向进化技术的发展, 越来越多的学者采用理性设计的方法改变蛋白质的性质。采用理性设计的方法进行试验首先需要明确蛋白质的三维结构, 对于无法通过试验手段得到三维结构的蛋白质, 可利用同源建模的方法获得。在实际操作过程中, 一些目标蛋白很难找到同源性高的模板蛋白, 而利用同源性低的模板蛋白进行单模板建模得到的模型通常较难通过模型评价。本研究表明, 对于序列相似性较低难以通过单模板建模得到通过评价的蛋白质, 可以进行序列比对划分结构域, 采用多模板分段建模、结构域进行拼接的方法模拟其三维结构。这种利用多模板分段建模的方法, 为序列同源性低的蛋白质进行三维结构模拟提供了一个可参考的思路。

3.2 α -L-鼠李糖苷酶的结构与功能的关系目前尚未清楚。本研究利用生物信息学的方法, 确定了 α -L-鼠李糖苷酶 Rha1 的二级结构和三级结构, Rha1 具有与

已知的 GH78 来源的四个 α -L-鼠李糖苷酶的晶体结构相似, 都具有一个 $(\alpha/\alpha)_6$ 桶状结构域。利用同源建模的方法获得通过评价的 Rha1 的三维模型, 为进一步研究 α -L-鼠李糖苷酶 Rha1 的结构与功能的关系提供了模型依据, 也有利于进行理性设计改造该酶的热稳定性、催化活性及底物适应性, 以满足工业应用的需求。

参考文献

- [1] Qian S, Wang H, Zhang C, et al. Isolation and characterization of dioscin- α -l-rhamnosidase from bovine liver [J]. Journal of Molecular Catalysis B Enzymatic, 2013, 97(8): 31-35
- [2] Yadav S, Yadav R S S, Yadav K D S. An alpha-L-rhamnosidase from *Aspergillus awamori* MTCC-2879 and its role in debittering of orange juice [J]. International Journal of Food Science & Technology, 2013, 48(5): 927-933
- [3] Ni H, Xiao A F, Cai H N, et al. Purification and characterization of *Aspergillus niger* α -L-rhamnosidase for the biotransformation of naringin to prunigen [J]. African Journal of Microbiology Research, 2012, 6(24): 5276-5284
- [4] 李军. 纤维素酶 E4 的同源建模和分子对接研究[D]. 广州: 华南理工大学, 2012
- [5] LI Jun. Study on homology modeling and molecular-docking simulation of cellulase E4 [D]. Guangzhou: South China University of Technology, 2012
- [6] Cui Z, Maruyama Y, Mikami B, et al. Crystal structure of glycoside hydrolase family 78 α -L-rhamnosidase from *Bacillus*, sp. GL1 [J]. Journal of Molecular Biology, 2007, 374(2): 384-98
- [7] Fujimoto Z, Jackson A, Michikawa M, et al. The structure of a *Streptomyces avermitilis* α -L-rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement [J]. Journal of Biological Chemistry, 2013, 288(17): 12376-12385
- [8] Grandits M, Michlmayr H, Sygmund C, et al. Calculation of substrate binding affinities for a bacterial GH78 rhamnosidase through molecular dynamics simulations [J]. Journal of Molecular Catalysis B Enzymatic, 2013, 92(4): 34-43
- [9] O'Neill E C, Stevenson C E M, Paterson M J, et al. Crystal structure of a novel two domain GH78 family α -l-rhamnosidase from *Klebsiella oxytoca* with rhamnose bound [J]. Proteins Structure Function & Bioinformatics, 2015, 83(9): 1742-1749
- [9] Li LJ, Yu Y, Zhang X, et al. Expression and biochemical

- characterization of recombinant α -L-rhamnosidase r-Rha1 from *Aspergillus niger* JMU-TS528 [J]. International Journal of Biological Macromolecules, 2016, 85: 391-399
- [10] Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy [J]. Journal of Molecular Biology, 1993, 232(2): 584-99
- [11] Schwede T, Kopp J, Guex N, et al. SWISS-MODEL: An automated protein homology-modeling server [J]. Nucleic Acids Research, 2003, 31(13): 3381-3385
- [12] Daga P R, Patel R Y, Doerksen R J. Template-based protein modeling: recent methodological advances [J]. Current Topics in Medicinal Chemistry, 2010, 10(1): 84-94
- [13] Webb B, Sali A. Comparative protein structure modeling using Modeller [J]. Current Protocols in Bioinformatics, 2016, 86: 2.9.1-2.9.37
- [14] Kleywegt G J, Jones T A. Phi/psi-chology: Ramachandran revisited [J]. Structure, 1996, 4(12): 1395-1400
- [15] 李军,曹以诚,葛洪.基于 Discovery Studio 平台纤维素酶 E4 同源建模[J].广东农业科学,2012,39(9):136-139
LI Jun, CAO Yi-cheng, GE Hong. Homology modeling of cellulose E4 based on discovery studio [J]. Guangdong Agricultural Sciences, 2012, 39(9): 136-139
- [16] 刘宗保,张志刚,石磊.DfrA12 蛋白的结构分析及同源建模 [J].现代食品科技,2013,29(7):1505-1508
LIU Zong-bao, ZHANG Zhi-gang, SHI Lei. Structure analysis and homology modeling of DfrA12 protein [J]. Modern Food Science and Technology, 2013, 29(7): 1505-1508