基因组学与应用生物学，2021 年，第 40 卷，第 9-10 期合刊，第 3106−3114 页

研究报告
Research Report

分子通讯与木聚糖酶耐热性的关系研究

王馨颖　李心羽　冯一诚　胡凯惠　丁彦蕊*
江南大学, 人工智能与计算机学院, 理学院, 无锡, 214122
* 通信作者, yr_ding@jiangnan.edu.cn

摘　要　复杂网络是研究木聚糖酶耐热机制的有效方法。本研究构建来自嗜热子囊菌的耐热型木聚糖酶(*Thermoascus aurantiacus xylanase*, Xyna_Theau)和来自变铅青链霉菌的常温型木聚糖酶(*Streptomyces lividans xylanase*, Xyna_Strli)的残基相互作用网络，并利用 Dijkstra 算法与基于图的频繁模式挖掘算法进行频繁路径挖掘。结果表明：同温度下，Xyna_Strli 的通讯较 Xyna_Theau 更为活跃，受到温度升高的影响更大；Xyna_Strli 的频繁路径集中在 β1、β2 与 loop8、β7 区域；而 Xyna_Theau 的频繁路径则紧密联系了 β1、β2、β3 区域，并且也发现两蛋白质部分区域间通讯路径相同；此外，对比发现温度升高时 Xyna_Strli 的通信受到 C 端，turn 转折以及热敏感区域影响较大。

关键词　频繁路径挖掘; 耐热性; 木聚糖酶

The Relationship Between Molecular Communication and the Heat Resistance of Xylanase

Wang Xinying　Li Xinyu　Feng Yicheng　Hu Kaihui　Ding Yanrui*
School of Artificial Intelligence and Computer Science, School of Science, Jiangnan University, Wuxi, 214122
* Corresponding author, yr_ding@jiangnan.edu.cn
DOI: 10.13417/j.gab.040.003106

Abstract　The complex network is an effective method to study the heat resistant mechanism of xylanase. The residue interaction networks of thermostable xylanase from *Thermoascus aurantiacus xylanase* and mesophilicxylanase from *Streptomyces lividans xylanase* were constructed, and frequent path mining was carried out using Dijkstra algorithm and graph-based frequent pattern mining algorithm.The results showed that at the same temperature, the communication of Xyna_Strliwasmore active than that of Xyna_Theau, and was more affected by the increased temperature. The frequent pathways of Xyna_Strliare concentrated in the β1, β2 and loop8, β7 regions, while the frequent pathways of Xyna_Theauare closely related to the β1, β2, β3 regions, and the two proteins were found to have the same pathways in some regions. In addition, comparison showed that Xyna_Strli communication was greatly affected by C terminal, turn and heat-sensitive region when the temperature rised.

Keywords　Frequent path mining; Heat resistance; Xylanase

木聚糖酶作为酶制剂，可破坏细胞结构以提高果蔬出汁率，可水解木聚糖从而降低麦汁粘度，可水解阿拉伯糖进而降低饲料粘稠度，在食品、酿酒、饲料等行业得到广泛应用。其中耐热木聚糖酶应用于饲料造粒过程等需要高温环境的环节(Collins et al., 2010)。相关研究表明：芳香族相互作用可通过促使二聚体的形成提升木聚糖酶的耐热性(Georis et al.,2000; 杨浩萌等, 2005)，在氨基酸末端或 α 螺旋中引入二

基金项目：本研究由国家自然科学基金项目(21541006)资助
引用格式：Wang X.Y., Li X.Y., Feng Y.C., Hu K.H., and Ding Y.R., 2021, The relationship between molecular communication and the heat resistance of xylanase, Jiyinzuxue yu Yingyong Shengwuxue (Genomics and Applied Biology), 40 (9-10 combined issue): 3106-3114. (王馨颖, 李心羽, 冯一诚, 胡凯惠, 丁彦蕊, 2021, 分子通讯与木聚糖酶耐热性的关系研究, 基因组学与应用生物学, 40(第9-10 期合刊): 3106-3114.)

硫键也有益于热稳定性(Jeong et al., 2007; Wang et al., 2012)；盐桥对热稳定性较为重要(Zhao et al., 2017)。表面精氨酸(Bai et al., 2014)、脯氨酸与甘氨酸(Ling et al., 2014)也均对木聚糖酶的热稳定性有着重要作用；此外，还发现木聚糖酶存在热稳定区域(Ding and Cai, 2013)。

随着复杂网络理论的发展，人们开始通过网络角度研究蛋白质耐热机制。将蛋白质编码为残基相互作用网络，为识别蛋白质关键残基与功能性残基以及变构过程中信号传导提供了新的思路与方法。Amitai 等(2004)以残基为节点、残基之间作用力为边构建残基相互作用图，识别出乙酰胆碱酯酶和枯草菌素的底物结合位点等多个酶的功能残基，并且发现结合位点、配体结合和进化保守残基可直接或间接与蛋白质所有残基进行通信。Brinda 和 Vishveshwara (2005)通过构建蛋白质残基相互作用网络识别出由芳香族残基与精氨酸、组氨酸和蛋氨酸组成的强 hub 节点与由疏水残基亮氨酸和异亮氨酸组成的弱 hub 节点，并发现这些节点在整合蛋白质二级结构元素中起着重要作用，从而有助于嗜热蛋白的稳定性。Guzel 等(2017)构建了核糖体的接触拓扑图，并在信号传导路径所经过的残基中发现高比例的抗生素靶点残基，该发现意味着抗生素通过干扰信号传导以阻断蛋白质合成过程。Ribeiro 和 Ortiz (2014)依据不同的标准计算拓扑图中残基距离，发现皮尔逊相关系数与广义相关系数均不能准确识别变构过程中的重要残基与信号传播路径，他们提出基于共价相互作用计算蛋白质网络，并成功应用于咪唑甘油磷酸合酶(Imidazole glycerol phosphate synthase)变构信号传播的识别，证明了该方法的可靠性。由此可见，基于动态残基相互作用网络研究蛋白质变构信号传导，以探索其对蛋白质耐热性的影响因素是卓有成效的。

最短路径计算常用于蛋白质相互作用网络的拓扑分析、关键蛋白质识别(嘉泽宁等，2011)以及蛋白质功能预测。计算图的最短路径算法常用的有 Dijkstra 算法、Floyd 算法与 Bellman-ford 算法。Dijkstra 算法(Lin and Kernighan, 1973)采用贪心策略进行逐层拓展，直到搜索到目标点。该算法的时间复杂度为 $O(v^2)$，空间复杂度为 $O(v)$，其中 v 为节点数，执行效率较高，但其缺点在于无法处理带有负权值的边。Floyd 算法(Lin, 1965)核心是一种动态规划思想，通过计算 $d_{ij}=\min\{d_{ik}+d_{kj}, d_{ij}\}$ 确定节点 e_i 与节点 e_j 之间的最短路径大小，其中 k 为节点 e_i 与节点 e_j 之间的穷举断点。该算法的优势在于可处理带有负权值的边，在稠密图中性能良好，但算法时间复杂度为 $O(v_3)$,空间复杂度为 $O(v_2)$，计算量较大。Bellman-ford 算法(Wang and Crowcroft, 1996)也利用了动态规划思想，对图 G 进行 $v-1$ 次松弛操作，每次松弛操作中对每条边 e_{ij} 计算 $d_j=\min\{d_i+e_{ij},d_j\}$，因此，该算法的时间复杂度为 $O(ev)$，其中 e 为边数，v 为节点数，则该算法在边数大于节点数的情况中复杂度较大。本文采用 Dijkstra 算法计算残基相互作用网络的最短路径。

频繁模式挖掘算法是一种挖掘事物集中项集与项集关联程度的算法，在计算物流网络路径、网站日志路径等有重要应用。Apriori 算法是基于关联规则的频繁项集挖掘算法。该算法通过多次扫描事物集以生成长度依次递增的频繁项集。由于传统 Apriori 算法计算冗余较大，人们对此进行深入研究与改进。FP-Growth 算法将事物映射为路径，以此生成 FP 树，从而在树中进行递归挖掘频繁项，该算法相比 Apriori 算法，仅需扫描两次事物集，计算速度更快。杨俊瑶等(2015)提出了基于物流网络拓扑信息的 PMWTI 算法，通过最小代价矩阵获取 K 序列，并对 $K-1$ 序列根据代价容忍度进行剪枝。该算法能够较好地应用于物流网络中生成路径序列的情况。此外，朱益立等(2017)提出了基于有向无环图的算法进行频繁模式挖掘，通过扫描二进制事务集建立有向无环图，再进行剪枝深度搜索出频繁项集。该方法采用了用节点表示频繁项，用边表示频繁项之间的关联程度的思想，简洁易懂且计算效率较高。

本研究以常温型木聚糖酶与耐热性木聚糖酶为研究对象，获取蛋白质在模拟时间内的结构信息并计算残基相互作用网络，利用 Dijkstra 算法获取最短路径集，并采用基于图的算法进行频繁模式挖掘，得到木聚糖酶动态网络的主要通信路径并分析了其与木聚糖酶耐热性的关系。

1 结果与分析

本研究构建了常温木聚糖酶(Xyna_Strli)与耐热木聚糖酶(Xyna_Theau)在 300 K 与 400 K 下各 300个时间帧的残基相互作用网络，利用 Dijkstra 算法计算每一帧中所有残基两两之间的最短路径。在所得路径集中，筛选出现 250 帧以上的频繁路径集，根据这些路径集重新构建图。对于所构建的图选取阈值为 7 进行剪枝，通过深度搜索获得主干路径。对以上结果统计了路径数目(图 1)、路径权重(图 2)和路径长度(图 3)，并对 Xyna_Strli 在 300 K 与 400 K 下路径进行可视化(图 4; 图 5)，对 Xyna_Theau 在 300 K 与 400K

下路径进行可视化(图 6; 图 7)。

Xyna_Strli 在 300 K 下有路径数目 53 条(图 1)，总体路径权重为 17 162 (图 2)，路径长度总体分布在 10~13 (图 3)，相比于 400 K 下有 8 条路径，路径总权重 312，路径长度总体在 1~3，可以看出 Xyna_Strli 在 300 K 下残基间通讯更加紧密。相似地，Xyna_Theau 在 300 K 下的路径数目与路径权重也远远大于 400 K 下的。因此可以做出以下推论：温度越高，残基相互作用网络中通讯路径规模越小，残基间的通讯越弱。

其次，观察同温度下两蛋白质的路径的数目、长度与权重，可以发现：Xyna_Strli 在 300 K 下与 400 K 的路径数目均大于 Xyna_Theau 的，这表明：同温度下，Xyna_Strli 的通讯相较于 Xyna_Theau 更为紧密。

最后，观察两温度下，Xyna_Strli 与 Xyna_Theau 路径数目减少的程度，可以发现：Xyna_Strli 路径数目减少程度较大，而 Xyna_Theau 减少程度较小。这表明：温度上升对于常温型木聚糖酶的影响大于耐热型木聚糖酶。

2 讨论

对于 Xyna_Strli，在 300 K 时通讯路径主要集中在 GLN205 所在的连通子图，并且该连通子图内颜色较深的路径构成了一些较为重要的通讯路径段(图 4)。温度升高后，一些路径段或是残基的消失导致路径之间彼此断开，路径规模减小，残基间的通讯变得零散(图 5)。对于这些变化，结合蛋白质二级结构与

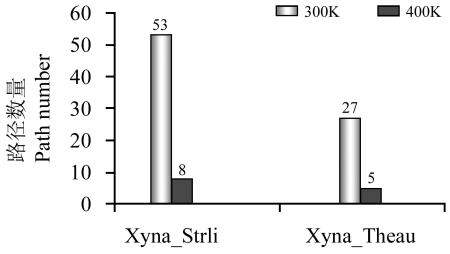

图 1 Xyna_Strli 与 Xyna_Theau 分别在 300 K 与 400 K 下路径数量

Figure 1 The number of paths at 300 K and 400 K for Xyna_Strli and Xyna_Theau

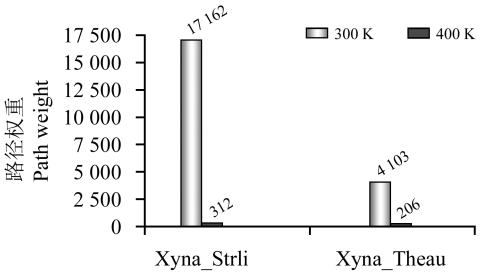

图 2 Xyna_Strli 与 Xyna_Theau 分别在 300 K 与 400 K 下路径总权重

Figure 2 The total path weights of Xyna_Strli and Xyna_Theau at 300 K and 400 K

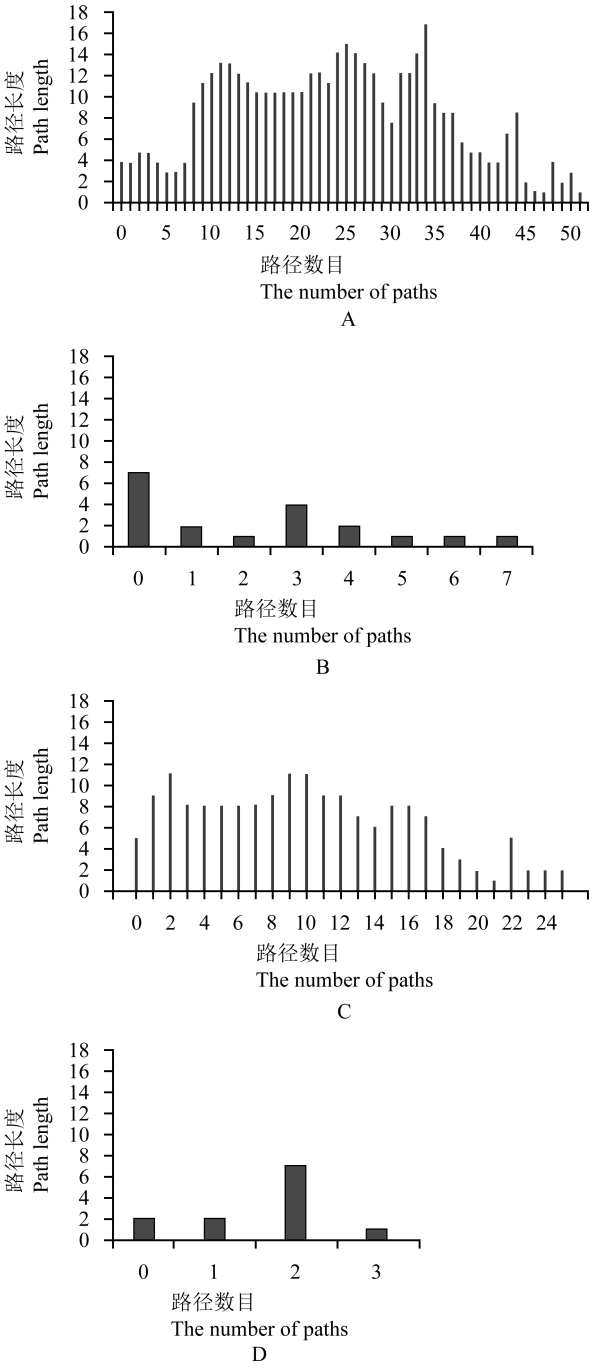

图 3 Xyna_Strli 与 Xyna_Theau 分别在 300 K 与 400 K 下路径长度分布

注：A: Xyna_Strli 在 300 K 下；B: Xyna_Strli 在 400 K 下；C: Xyna_Theau 在 300 K 下；D: Xyna_Theau 在 400 K 下

Figure 3 Path length distribution of Xyna_Strli and Xyna_Theau at 300 K and 400 K

Note: A: Xyna_Strli at 300 K; B: Xyna_Strli at 400 K; C: Xyna_Theau at 300 K; D: Xyna_Theau at 400 K

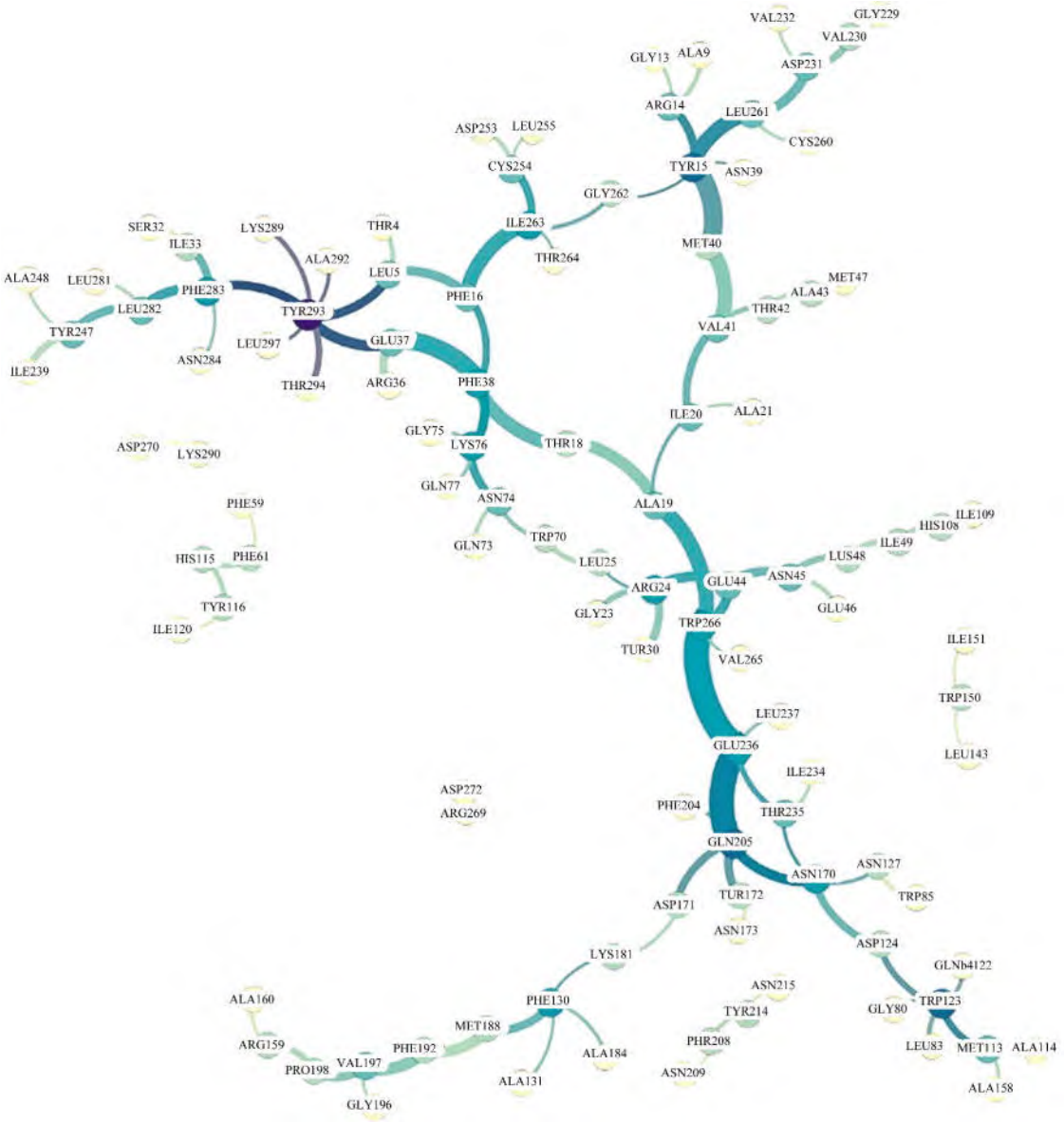

图 4 Xyna_Strli 在 300 K 下路径可视化

注: 图中节点之间的连线越粗, 颜色越深, 表示该路径段权重越大

Figure 4 The visualized picture of the path of Xyna_Strli at 300 K

Note: The thicker and the darker, the line between nodes in the figure, the greater the weight of the path segment

残基之间相互作用力讨论 Xyna_Strli 与 Xyna_Theau 残基网络的通讯特点以及温度对其影响。

路径 GLU37−PHE38−THR18−ALA19−TRP266−GLU236−GLN205 在 300 K 中是残基网络中权重最高的一条通讯路径，其连接着 β1、β2 与 loop8、β7 区域。而在 400 K 中，该路径依然是重要的通讯路径。相关研究表明：对于蛋白质中螺旋末端，turn 与 loop 区域等柔性区域，增加其刚性有利于结构稳定性，从而增强热稳定性(Russell et al., 1997; Vogt et al.,1997; Sani et al., 2018)。400 K 时，ALA19 与 TRP266，TRP266−GLU236 之间范德华作用力的数量大大增加，使得 β1 中的 ALA19 与处于 loop8 末端的 TRP266 连接更加紧密。同时，温度升高后，路径段 ARG269−ASP272、LYS290−ASP270、PHE283−LEU282−TYR247 使得 loop8 区域内部刚性增强，并且 ARG269−ASP272 处于 3_{10}− 螺旋两端。相关研究表明：3_{10}− 螺旋在高温下较为活跃，而收紧该区域可以使蛋白质获取更高的热稳定性(Kamal et al., 2011)。因此，400 K 时路径 ARG269−ASP272 使得 3_{10}− 螺旋收紧，其灵活性降低，热稳定性增加。

温度升高过程中，Xyna_Strli 中的热敏感区、C端以及 3_{10}− 螺旋等区域的使得路径断开，路径变短，网

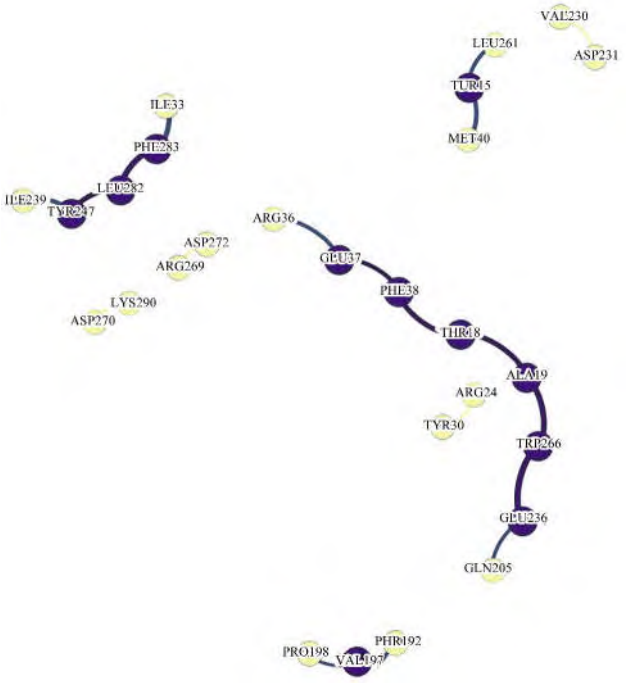

图 5 Xyna_Strli 在 400 K 下路径可视化

注: 图中节点之间的连线越粗, 颜色越深, 表示该路径段权重越大

Figure 5 The visualized picture of the path of Xyna_Strli at 400 K

Note: The thicker and the darker, the line between nodes in the figure, the greater the weight of the path segment

络通讯变得零散。有研究表明，常温木聚糖酶中，C端在高温中相较于 N 端更加活跃(Ding and Cai, 2013)。300 K 中残基 TYR293 通过路径TYR293−PHE283−LEU282 及其与相邻残基的通讯连接了 loop8、α8 与 α1。而 400 K 时，该残基在通讯网络中的消失不仅削弱了 loop8 与 β1 的通讯，并且 TYR293 与 LEU297、THR294、ALA292 之间的路径消失也证明了 C 端残基的活跃程度较高。常温木聚糖酶存在热敏感区，具体区域从 β3 至 loop4 (残基号 73~151)(Ding and Cai, 2013)。300 K 中处于热敏感区的残基 PHE130 连接着区域 α4 与 β5。温度升高后，PHE130 的活跃导致路径段 PHE192−VAL197−PRO198 与主干路径GLU37−PHE38−THR18−ALA19−TRP266−GLU236−GLN205 割裂，进而影响 α5 与 loop6、β7 的联系。此外，蛋白质中 3_{10}− 螺旋与 turn 在温度升高时容易发生变构从而影响结构稳定性。路径 LYS76−ASN74−TRP70−LEU25−ARG24−GLU44−A SN45−LYS48−ILE49，经过了 3_{10}− 螺旋与 turn 结构，连接 β3 与 α3，温度升高后，这些路径的消失表明这些结构的活跃影响了网络通讯。

总体而言，Xyna_Strli 的通讯路径主要集中在 β1、β2 与 loop8、β7 区域，增加这些区域的刚性，而温度升高，网络通讯总体减弱，由于热敏感区、C 端以及 turn、3_{10}− 螺旋的影响，其他区域的通讯减弱。

对于 Xyna_Theau，300 K 下路径集由两个大的连通子图构成，而 400 K 时，路径 PRO271−LYS291−LEU283− TYR248 所在的连通子图消失(图 6)，400 K 中重要的两条路径由 TRP126−ASP127 所在连通子图演化而来，并且保留了路径 ILE123−TRP126−ASP-127−ARG81−GLN42−TYR17 与TRP194−VAL199-PR-O200(图 7)。

路径ILE123−TRP126−ASP127−ARG81−GLN42−TYR17 在 300 K 时作为通讯网络的重要路径，连接 β1、β2 与 β3 区域，在 400 K 时该路径在网络中仍然扮演重要角色。并且，相比于 Xyna_Strli 的重要路径分散于 loop 区域与 β 区域，Xyna_Theau 的路径更多集中在 β 区域。众所周知，蛋白质二级结构中，loop 区域与 turn 容易受到高温影响，其次是 α 螺旋结构，而相比之下，β 折叠结构更为稳定(Mamonova et al., 2013; Rathi et al., 2016)。因此，可以推测通讯路径在 β 区域的集中使得网络通讯受温度影响较小。

另外，值得注意的是，路径 TRP194−VAL199−PRO200 在 400 K 中仍然存在，然而其在 300 K 中并非权重较高路径，由此可以看出该路径在温度升高过程中较为稳定。并且，在 Xyna_Strli 中，路径PHE-192−VAL197−PRO198 有着相似的表现。从结构上而言，在 Xyna_Theau 中，PRO200、VAL199 处于 α5 与 β6 之间的无规则卷曲区，与之相连的 TRP194 处于 α5 螺旋，该路径降低了经过的 turn 结构的灵活性。相似地，在 Xyna_Strli 中 PRO198、VAL197 位于 α5 与 β6 中的无规则卷曲区域，TRP192 则处于 α5 螺旋区域中，并且该路径位于 Xyna_Strli 中热不敏感区域(Ding and Cai, 2013)，这一点与该路径在温度上升过程中的稳定表现有所呼应。从组成上来看，两条路径都由相同的氨基酸序列组成，并且脯氨酸在不稳定区域中降低二级结构的灵活性(Kamal et al.,2012)，由此推断该路径的构成成分也使得此路径在高温下显示出一定的稳定性。

总体而言，Xyna_Theau 的通讯路径集中在 β1、β2、β3 区域，相比于 Xyna_Strli，路径集中在 β 折叠结构带来一定的稳定性。

此外，对比相同温度下 Xyna_Theau 与 Xyna_Strli 的通讯路径网络，根据序列对比可以发现部分路径有一定相似性，意味着两蛋白质部分结构间通讯方式存在一定的相似性。Xyna_Theau 与 Xyna_Strli 分别在 300 K 与 400 K 中相同的路径以及路径所连

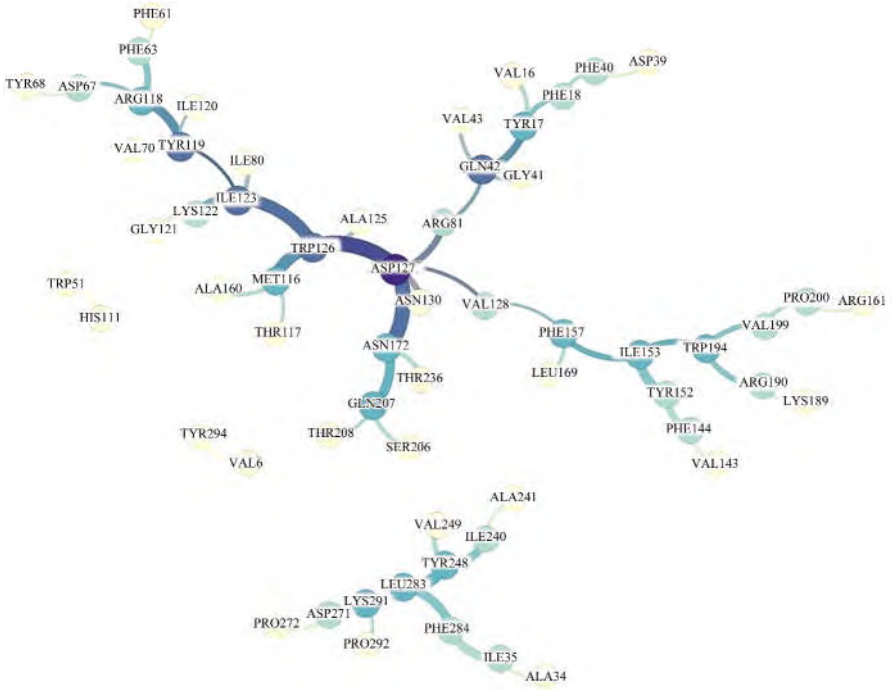

图 6 Xyna_Theau 在 300 K 下路径进行可视化

注: 图中节点之间的连线越粗, 颜色越深, 表示该路径段权重越大

Figure 6 The visualized picture of the path of Xyna_Theau under 300 K

Note: The thickerand the darker the line between nodes in the figure, the greater the weight of the path segment

接区域(表 1)。对比 400 K 下的相同路径，Xyna_Theau 保留了路径 ASP127−TRP126−MET116 并且存在重要路径段 LYS122−ILE123 和 ARG81−ASP127，这些路径段在高温下使得 β4、α3 之间紧密相连。而在 Xyna_Strli中消失的对应路径存在于 Xyna_Strli 的热敏感区域。因此可以做出推论：在高温情况下 Xyna_Strli 在热敏感区域的通讯受到较大影响。而 Xyna_Theau 中则相反，多条重要路径使得 β4、α3 等区域之间联系紧密。

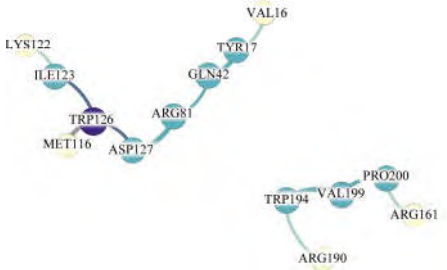

图 7 Xyna_Theau 在 400 K 下路径进行可视化

注: 图中节点之间的连线越粗, 颜色越深, 表示该路径段权重越大

Figure 7 The visualized picture of the path of Xyna_Theau under 400 K

Note: The thicker and the darker the line between nodes in the figure, the greater the weight of the path segment

本研究使用 Dijkstra 算法与频繁模式挖掘算法对常温型木聚糖酶与耐热型木聚糖酶进行频繁路径挖掘，以此探究分子通讯与耐热性机制之间的关系。结果发现：同温度下，Xyna_Strli 的频繁路径数目、权重与长度均高于 Xyna_Theau 的，表明 Xyna_Strli 通讯较 Xyna_Theau 更为活跃，且对比温度升高频繁路径的变化，可发现 Xyna_Strli 受到温度升高的影响更大；结合蛋白质二级结构发现 Xyna_Strli 的频繁路径集中在 β1、β2 与 loop8、β7 区域,而 Xyna_ Theau 的频繁路径则紧密联系了 β1、β2、β3 区域，并且通过序列对比确定两蛋白质一些相同的路径，表明两蛋白质部分区域间通讯路径相同；此外，对比 300 K 与 400 K 时的路径，可以看出 Xyna_Strli 的通信受到 C 端、turn 转折、热敏感区域影响较大。综上所述，通过频繁路径挖掘了解蛋白质内部通讯情况，可进一步阐述蛋白质的耐热机制。该项研究为使用定点突变、引入二硫键等等方法增强木聚糖酶的耐热性提供借鉴，从而提升食品加工、饲料制造等工业的生产效率

3 材料与方法

3.1 木聚糖酶无向加权网络的构建

本研究针对来自嗜热子囊菌的耐热型木聚糖酶(*Thermoascus aurantiacus xylanase*, Xyna_Theau)和来

表 1 300 K 下两蛋白质相同路径对比

Table 1 Comparison of the same pathway of the two proteins at 300 K

序号 No.	Xyna_Strli	Xyna_Theau	连接区域 Connection area
1	GLN205−ASN170−ASN127; THR235−ASN170−ASP124−TRP123−MET113	MET116−TRP126−ASP127−ASN172−GLN207	β5, loop6, β3
2	ALA248−TYR247; ILE239−TYR247−LEU292−PHE283−ILE33−SER32	ALA241−ILE240−TYR248−LEU283−PHE284−ILE35−ALA34	α7, loop8, α1
3	ARG159−PRO196−VAL197−PHE192	TRP194−VAL199−PRO200−ARG161	α4, α5
4	PHE16−PHE38; TYR15−MET40	GLN42−TYR17−PHE16−PHE40	β1, β2

表 2 400 K 下两蛋白质相同路径对比

Table 2 Comparison of the same pathway of the two proteins at 400 K

序号 No.	Xyna_Strli	Xyna_Theau	连接区域 Connection area
1		MET116−TRP126−ASP127	β5,loop6,β3
2	ILE239−TYR247−LEU292−PHE283−ILE33		α7,loop8,α1
3	PRO196−VAL197−PHE192	TRP194−VAL199−PRO200−ARG161	α4, α5
4	TYR15−MET40	GLN42−TYR17−PHE16	β1,β2

自变铅青链霉菌的常温型木聚糖酶 (*Streptomyces lividans xylanase*, Xyna_Strli) 进行网络构建及频繁路径挖掘。相关研究表明，基于欧式距离构建的网络不能准确识别信号传导的重要残基和路径，Andre 等提出基于相互作用力的方法来构建蛋白质网络，该方法已在实践中被证明可识别已知的酶的传播特点(Ribeiro and Ortiz, 2014)。本研究采用基于相互作用力的方法构建残基网络，具体做法如下：

步骤一：以 RSCB 库中获取的 Xyna_Theau 与 Xyna_Strli 两个蛋白质的三维结构数据为基础，将蛋白质分别放入包含 18 269 个和 17 415 个 TIP3P 水分子的矩形水盒子中，并分别加入 2 个和 5 个 Na+ 进行系统中和。采用 Amber 软件进行分子动力学模拟，系统在 NPT 系综下最小化和平衡，模拟在 NVT 系综下进行，力场为 Amber ff99SBildn 力场。模拟温度为 300 K 和 400 K，模拟时间为 300 ns，每个时间步长为 2 *fs*，每 5000 步保存 1 次构象。这样可以获取不同温度下模拟时间内的分子运动轨迹文件；并在模拟时间内截取 *N* 帧(*N*=300)快照。

步骤二：考虑到残基之间通信强度差异，在步骤一的基础上，以残基作为节点，残基之间作用力数目之和的倒数作为边权，构建 *N* 个无向加权网络{G_1, G_2,⋯G_n}。

3.2 计算最短路径

为获取 *N* 帧下的路径信息，需要获取每一帧下节点 S_s 与 S_t(s,t∈{1,2,3,⋯,n})之间的最短路径。本文利用 Dijkstra 算法挖掘任意两残基间的最短路径。具体步骤如下：

步骤一：选取节点 *Ss* 为源点，节点 S_t 为终点，其中(s,t∈{1,2,3,⋯,n})。定义集合 *D* 为已标记的节点集合，*S* 为未标记节点集合，数组 *P* 记录最短路径中对应残基先驱，d_i 为到节点 S_i 的路径距离。

步骤二：初始化 *D* 集合，仅包含源点 S_s，*S* 集合包含其余为标记的所有节点。

步骤三：遍历 *S* 集合中所有节点。对于每一个未标记节点 S_i∈*S*，遍历 *D* 集合中所有已标记节点 S_j，令 d_i=min{d_i,d_j+e_{ij}}，其中 e_{ij} 为节点 i 到节点 j 的距离。如果修改 d_i=d_j+e_{ij}，则 p_i=*j*。

步骤四：在所有未标记节点 *S* 中，选取 d_i 最小的节点放入集合 *S* 中。

步骤五：判断如果集合 *D* 中出现终点 S_t，则结束。否则返回步骤三。

3.3 频繁路径挖掘

对于从动态网络中挖掘到的路径集合，需要根据路径段出现频率挖掘出频繁路径。朱益立等(2017)提出了基于图的频繁模式挖掘算法，该算法的思想是：以节点表示各个项，以节点与节点之间的边的权值表示项与项之间的支持度，边权越大，所连接的项集之间支持度越高。根据该思想，对于 Dijkstra算法挖掘出的最短路径集，重新构建图并进行频繁路径挖掘，具体方法如下：

步骤一：统计 300 帧中所有路径中出现的帧数，并选定阈值 F，筛选出帧数大于 F 的路径集 M。

步骤二：建立图 G。遍历路径集 M 中每一条路径 $E=\{v_1,v_1\cdots v_n\}$，如果节点 v_i 与节点 v_j 在路径 E 中相邻，则判断如果边 e_{ij} 与边 e_{ji} 已存在，则将边权加 1；如果边 e_{ij} 与边 e_{ji} 不存在，则将边加入图 G 中，置边权为 1.

步骤三：对所得图 G 进行剪枝。选取阈值 K，将边权小于 K 的值减去，删去非频繁子序列。

步骤四：基于深度搜索遍历图 G，同时标记已经过的路径，防止由于环导致的死循环。

剪枝步骤保证了基于图的频繁模式挖掘符合先验原理，即：频繁序列的子序列也一定是频繁序列。相比于 Apriori 算法，其需要对事务集进行多次扫描，计算量大且耗费时间长。而本研究所采用基于图的频繁模式挖掘算法只需扫描一遍事务集，计算量较小。

作者贡献

王馨颖、丁彦蕊是本研究的实验设计者和实验研究的执行人；王馨颖完成数据分析、论文初稿的写作；冯一诚，胡凯惠以及李心羽参与实验设计、试验结果分析；丁彦蕊是项目的构思者及负责人，指导实验设计、数据分析、论文写作与修改。全体作者都阅读并同意最终的文本。

致谢

本研究由国家自然科学基金项目(21541006)资助。

参考文献

Amitai G., Shemesh A., Sitbon E., Shklar M., Netanely D., Venger I., and Pietrokovski S., 2004, Network analysis of protein structures identifies functional residues, J. Mol. Biol., 344(4): 1135-1146.

Bai W.Q., Zhou C., Xue Y.F., Huang C.H., Guo R.T., and Ma Y. H., 2014, Three-dimensional structure of an alkaline xylanase Xyn11A-LC from alkalophilic *Bacillus* sp. SN5 and improvement of its ther mal performance by introducing arginines substitutions, Biotechnology Letters, 36(7):1495-1501.

Brinda K.V., and Vishveshwara S., 2005, A network representation of protein structures: implications for protein stability, Biophys. J., 89(6): 4159-4170.

Collins T., Gerday C., and Feller G., 2010, Xylanases,xylanase families and extremophilic xylanases, FEMS Microbiol. Rev., 29(1): 3-23.

Ding Y.R., and Cai Y.J., 2013, Conformational dynamics of xylanase a from *Streptomyces lividans*: Implications for TIM-barrel enzyme thermostability, Biopolymers, 99 (9): 594-604.

Georis J., de Lemos Esteves F., Lamotte-Brasseur J., Bougnet V., Devreese B., Giannotta F., Granier B., and Frère J.M., 2000, An additional aromatic interaction improves the thermostability and thermophilicity of a mesophilic family 11 xylanase: structural basis and molecular study, Protein Sci., 9 (3): 466-475.

Guzel P., and Kurkcuoglu O., 2017, Identification of potential allosteric communication pathways between functional sites of the bacterial ribosome by graph and elastic network models, Biochim. Biophys Acta Gen. Subj., 1861(12): 3131-3141.

Jeong M.Y., Kim S., Yun C.W., Choi Y.J., and Cho S.G, 2007, Engineering a de novo internal disulfide bridge to improve the thermal stability of xylanase from *Bacillus stearothermophilus* No. 236, J. Biotechnol., 127(2): 300-309.

Jia Z.N., Yang G., and Zheng W.P., 2011, Shortest path-based identification of essential proteins, Guangxi Daxue Xuebao (Journal of Guangxi University) (Nature Science Edition), 36 (5): 764-769. (嘉泽宁, 杨贵, 郑文萍, 2011, 基于最短路径的关键蛋白质识别研究, 广西大学学报 (自然科学版), 36 (5): 764-769.)

Kamal M.Z., Ahmad S., Molugu T.R., Vijayalakshmi A., Deshmukh M.V., Sankaranarayanan R., and Rao N.M., 2011, *In vitro* evolved non-aggregating and thermostable lipase: structural and thermodynamic investigation, J. Mol. Biol., 413(3): 726-741.

Kamal M.Z., Mohammad T.A., Krishnamoorthy G., and Rao N. M., 2012, Role of active site rigidity in activity: MD simulation and fluorescence study on a lipase mutant, PLoS ONE, 7(4): e35188.

Lin S., 1965, Computer solutions of the traveling salesman problem, Bell Labs Technical Journal, 44(10): 2245-2269.

Lin S., and Kernighan B.W., 1973, An effective heuristic algorithm for the traveling salesman problem, Annals of Ops. Res., 21(2): 498-516.

Ling Z.M., Kang Z., Liu Y.., Liu S, Chen J., and Du G.C., 2014, Improvement of catalytic efficiency and thermostability of recombinant *Streptomyces griseus* trypsin by introducing artificial peptide, World J. Microbiol. Biotechnol., 30 (6): 1819-1827.

Mamonova T.B., Glyakina A.V., Galzitskaya O.V., and Kurnikova M.G., 2013, Stability and rigidity/flexibility-two sides of the same coin? Biochimica et Biophysicaacta, 1834 (5): 854-866.

Rathi P.C., Fulton A., Jaeger K.E., and Gohlke H., 2016, Appli-

cation of rigidity theory to the thermostabilization of lipase A from *Bacillus subtilis*, PLoS Comput. Biol., 12(3): e1004754.

Ribeiro A.A., and Ortiz V., 2014, Determination of signaling pathways in proteins through network theory: importance of the topology, J. Chem. Theory Comput., 10(4): 1762-1769.

Russell R. J., Ferguson J. M., Hough D. W., Danson M. J., and Taylor G. L., 1997, The crystal structure of citrate synthase from the hyperthermophilic archaeon pyrococcusfuriosus at 1.9 A resolution, Biochemistry, 36(33): 9983-9994.

Sani H.A., Shariff F.M., Rahman R., Leow T.C., and Salleh A. B., 2018, The effects of one amino acid substitutions at the c-terminal region of thermostable l2 lipase by computational and experimental approach, Mol. Biotechnol., 60(1): 1-11.

Vogt G., Woell S., and Argos P., 1997, Protein thermal stability, hydrogen bonds, and ion pairs., J. Mol. Biol., 269 (4): 631-643.

Wang Y.W., Fu Z., Huang H.Q., Zhang H.S., Yao B., Xiong H.R., and Turunen O., 2012, Improved thermal performance of thermomyceslanuginosus GH11 xylanase by engineering of an N-terminal disulfide bridge, Bioresour Technol., 112: 275-279.

Wang Z., and Crowcroft J., 1996, Quality-of-service routing for supporting multimedia applications, IEEE J. Sel. Areas Comm., 14(7): 1228-1234.

Yang H.M., Yao B., Luo H.Y., Zhang W.Z., Wang Y.R., Yuan T. Z., Bai Y.G., Wu N.Y., and Fan Y.L., 2005, Hydrophobic interaction between β−sheet b1 and b2 in xylanase XYNB influencing the enzyme thermostability, Shengwu Gongcheng Xuebao (Chinese Journal of Biotechnology), 21(5): 414-419. (杨浩萌, 姚斌, 罗会颖, 张王照, 王亚茹, 袁铁铮, 柏映国, 伍宁丰, 范云六, 2005, 木聚糖酶 XYNB 分子中折叠股 b1 和 b2 间的疏水作用对酶热稳定性的影响, 生物工程学报, 21(5): 414-419.)

Yang J.Y., Meng Z.Q., and Jiang L., 2015, Logistics frequent path sequence mining algorithm based oil topological information, Jisuanji Kexue (Computer Science), 42(4): 258-262. (杨俊瑶, 蒙祖强, 蒋亮, 2015, 一种基于拓扑信息的物流频繁路径挖掘算法, 计算机科学, 42(4): 258-262.)

Zhao Z.X., Hou S.L., Lan D.M., Wang X.M., Liu J.S., Khan F.I., and Wang Y.H., 2017, Crystal structure of a lipase from *Streptomyces* sp. strain W007−implications for thermostability and regiospecificity, FEBS J., 284(20): 3506-3519.

Zhu Y.L., Deng Z.R., and Xie P., 2017, Mining frequent itemsets algorithm based on directed acycline graph, Jisuanji Gongcheng yu Sheji (Computer Engineer and Design), 38 (5): 1237-1241. (朱益立, 邓珍荣, 谢攀, 2017, 基于有向无环图的频繁模式挖掘算法, 计算机工程与设计, 38 (5): 1237-1241.)

(责任编辑 蹇慧)