

在[0,1]中取值,步长为 0.1,其他使用默认设置,并选择最优比对结果.

- (3) **MI-GRAAL 方法.**MI-GRAAL 方法整合了网络节点之间多种类型的相似性标准,并决定它们的权重,以此找到最优比对.为了保证算法比较的合理性,在实验中,剔除默认方法中基于蛋白质生物信息的相似性标准(BLAST-E 值),选择基于网络拓扑性的相似性标准的组合(graphlet 度特征、度、聚类系数、节点离心率和介数得分),使方法完全基于网络的拓扑相似性进行比对,并选择各相似性标准组合中的最优比对结果.
- (4) **CytoGEDEVO 方法.**GEDEVO 方法^[18]基于进化算法,使用图编辑距离作为优化模型来找到最佳比对.CytoGEDEVO 方法是基于 GEDEVO 方法的 CytoScape 软件上的扩展.为了保证算法比较的合理性,在实验中,设置迭代次数为 2 000,其他使用默认设置.

上述 9 组残基相互作用网络对比实验结果如表 2 和图 6 所示(图 6 中,横轴表示各组残基相互作用网络对,纵轴为 EC,每组柱状图从左到右分别表示 SI-MAGNA、GRAAL、MI-GRAAL、CytoGEDEVO 的比对结果的 EC).从实验中发现,SI-MAGNA 方法在各组网络对中均获得更优的比对结果,具有更高的 EC.这证实了 SI-MAGNA 方法在残基相互作用网络比对方面的表现更为出色,优于另外 3 种拓扑网络比对方法.

Table 2 Residue interaction network alignment results with different algorithms
表 2 不同方法下残基相互作用网络的比对结果

网络对	不同网络比对方法的结果 EC			
	SI-MAGNA	GRAAL	MI-GRAAL	CytoGEDEVO
1V8I&1MP2	0.685	0.251	0.603	0.574
9AME&1WVO	0.852	0.545	0.565	0.608
1TUX&1E0W	0.761	0.305	0.305	0.517
1XNB&1XND	0.894	0.486	0.617	0.524
1XXN&1M4W	0.849	0.516	0.622	0.534
1YNA&1PVX	0.970	0.796	0.922	0.907
3QMM&1I6W	0.937	0.682	0.734	0.591
3D2B&1I6W	0.945	0.735	0.735	0.681
3QMM&3D2B	0.964	0.804	0.734	0.765

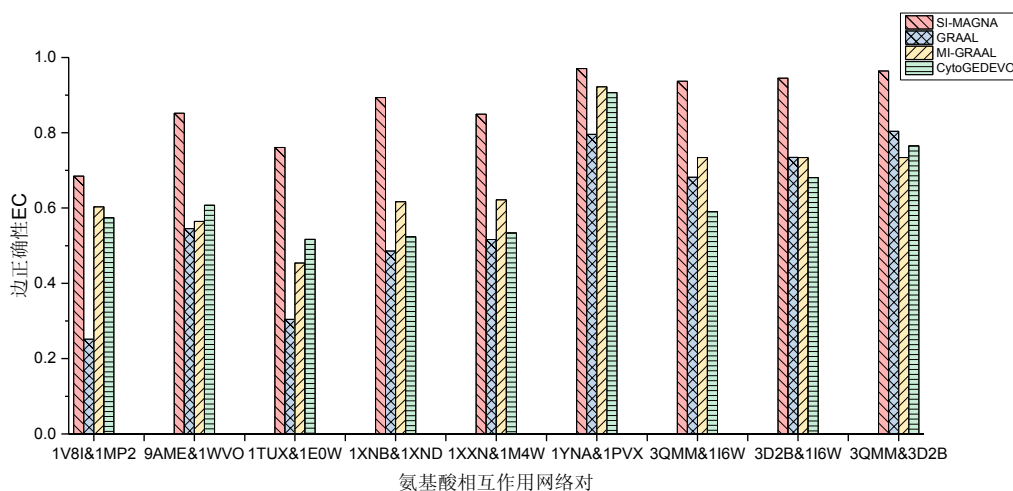


Fig.6 Residue interaction network alignment results (EC) with different algorithms

图 6 不同方法下的残基相互作用网络比对结果(EC)

2.3 SI-MAGNA方法的应用

为了探究蛋白质空间结构与热稳定性的联系^[19-22],本文构建了一系列的残基相互作用网络,利用 SI-MAGNA 方法进行比对和分析.通过残基相互作用网络比对,发现残基相互作用网络中节点和边的匹配,寻找相似的子结构(共同保守子图).残基相互作用网络对的基本信息和比对结果见表 3.

Table 3 Information and SI-MAGNA alignment results of different residue interaction networks**表 3** 不同网络的基本信息和比对结果

编号	网络对(PDB号)	节点数	边数	序列相似性(%)	比对结果	
					保守边数	EC
a	1V8I	150	517	39	354	0.685
	1MP2	187	625			
b	9AME	66	209	39	178	0.852
	1WVO	79	228			
c	1TUX	301	1 110	48	845	0.761
	1E0W	302	1 060			
d	1XNB	185	611	54	546	0.894
	1XND	190	631			
e	1XXN	185	643	65	546	0.849
	1M4W	197	633			
f	1YNA	193	642	88	623	0.970
	1PVX	194	633			
g	3QMM	358	1 291	93	1 210	0.937
	1I6W	359	1 270			
h	3D2B	358	1 300	97	1 229	0.945
	1I6W	359	1 270			
i	3QMM	358	1 291	97	1 245	0.964
	3D2B	358	1 300			

使用 CytoScape(<http://www.cytoscape.org/>)软件对 SI-MAGNA 方法产生的比对结果进行进一步处理分析,结果如图 7 所示,图中各组网络对分别为:(a) 1V8I(上左)、1MP2(上右)和两者比对的共同保守子图(下);(b) 9AME&1WVO(上右)和两者比对的共同保守子图(下);(c) 1TUX(上左)、1E0W(上右)和两者比对的共同保守子图(下);(d) 1XNB(上左)1XND(上右)和两者比对的共同保守子图(下);(e) 1XXN(上左)1M4W(上右)和两者比对的共同保守子图(下);(f) 1YNA(上左)1PVX(上右)和两者比对的共同保守子图(下);(g) 3QMM(上左)1I6W(上右)和两者比对的共同保守子图(下);(h) 3D2B(上左)1I6W(上右)和两者比对的共同保守子图(下);(i) 3QMM(上左)3D2B(上右)和两者比对的共同保守子图(下)。在图 7 每一组比对中,上方左右两个残基相互作用网络分别代表比对的残基相互作用网络对 G_1 和 G_2 ;下方一个残基相互作用网络为比对发现的子结构(共同保守子图),其中,节点表示为 $x_i=y_j, x_i \in V_1, y_j \in V_2$,这说明在比对中,节点 x_i 和 y_j 相匹配,边表示比对中的保守边。通过比较可以发现两个源网络与比对产生的子结构(共同保守子图)之间的异同,相似性高的部分通常对应于实现相似蛋白质功能的重要区域,而产生差异的部分则有很大的可能是蛋白质性质(如热稳定性)产生差异的原因。

以表 3 中 b 组比对 9AME 和 1WVO 为例,其与其余 8 组网络对相比较。尽管 9AME 和 1WVO 的序列相似性不高(39%),但通过残基相互作用网络比对发现,两者空间结构相似性很高,残基相互作用网络比对的 EC 达到 85.2%,且两者在各自生物体中实现相似功能。这也为“蛋白质的功能由序列信息和特定空间结构共同决定”^[12]提供依据。来自高纬度温带海域鱼类的 III 型抗冻蛋白质异构体 9AME 和来自人类的唾液酸合成酶 AFL 结构 1WVO 为两种生物的同源蛋白质,在各自生物体中实现非常相似的功能(抗冻蛋白)。9AME 的最适反应温度为 273K,1WVO 的最适反应温度为 293K~310K,两者在稳定性和活性方面具有不同的温度依赖性^[23]。为了了解两者网络结构的异同,对网络比对发现的子结构(共同保守子图)和二级结构相对应进行标记,如图 8 所示,其中,黄色标记的节点表示两源网络匹配上的 α -螺旋,橙色标记的节点表示两源网络匹配上的 β -折叠,紫色标记的节点表示两源网络匹配上的 3_{10} -螺旋,棕色标记的节点表示两源网络匹配上的 β -桥,蓝色标记的节点表示两源网络匹配上的弯曲,青色标记的节点表示两源网络匹配上的氢键转折,绿色标记的节点表示 9AME 中独有的 1 个 3_{10} -螺旋。通过残基相互作用网络比对发现,两者的结构均主要由 1 个 α -螺旋(9AME 中为残基 37-40,1WVO 中为残基 41-44)、2 个 3_{10} -螺旋(9AME 中为残基 19-21,57-59,1WVO 中为残基 23-25,61-63)和 2 个 β -折叠(9AME 中为残基 4-7,22-25,1WVO 中为残基 8-11,26-29)组成,此外,两者均包含一些 β -桥和弯曲结构。同时,9AME 中还独有 1 个 3_{10} -螺旋(残基 34-36)结构,而在 1WVO 中,无二级结构相同的结构与之匹配。根据残基相互作用网络比对结果,使用 PyMOL(<https://pymol.org/2/>)软件进一步对 9AME 和 1WVO 的序列和二级结构图进行标记,这能够更直观

地对照和分析两者在序列信息和空间结构上的异同,如图 9 所示,其中,黄色标记的序列和二级结构表示两源网络匹配上的 α -螺旋,橙色标记的序列和二级结构表示两源网络匹配上的 β -折叠,紫色标记的序列和二级结构表示两源网络匹配上的 3_{10} -螺旋,棕色标记的序列和二级结构表示两源网络匹配上的 β -桥,蓝色标记的序列和二级结构表示两源网络匹配上的弯曲,青色标记的序列和二级结构表示两源网络匹配上的氢键转折,绿色标记的序列和二级结构表示 9AME 中独有的 1 个 3_{10} -螺旋.

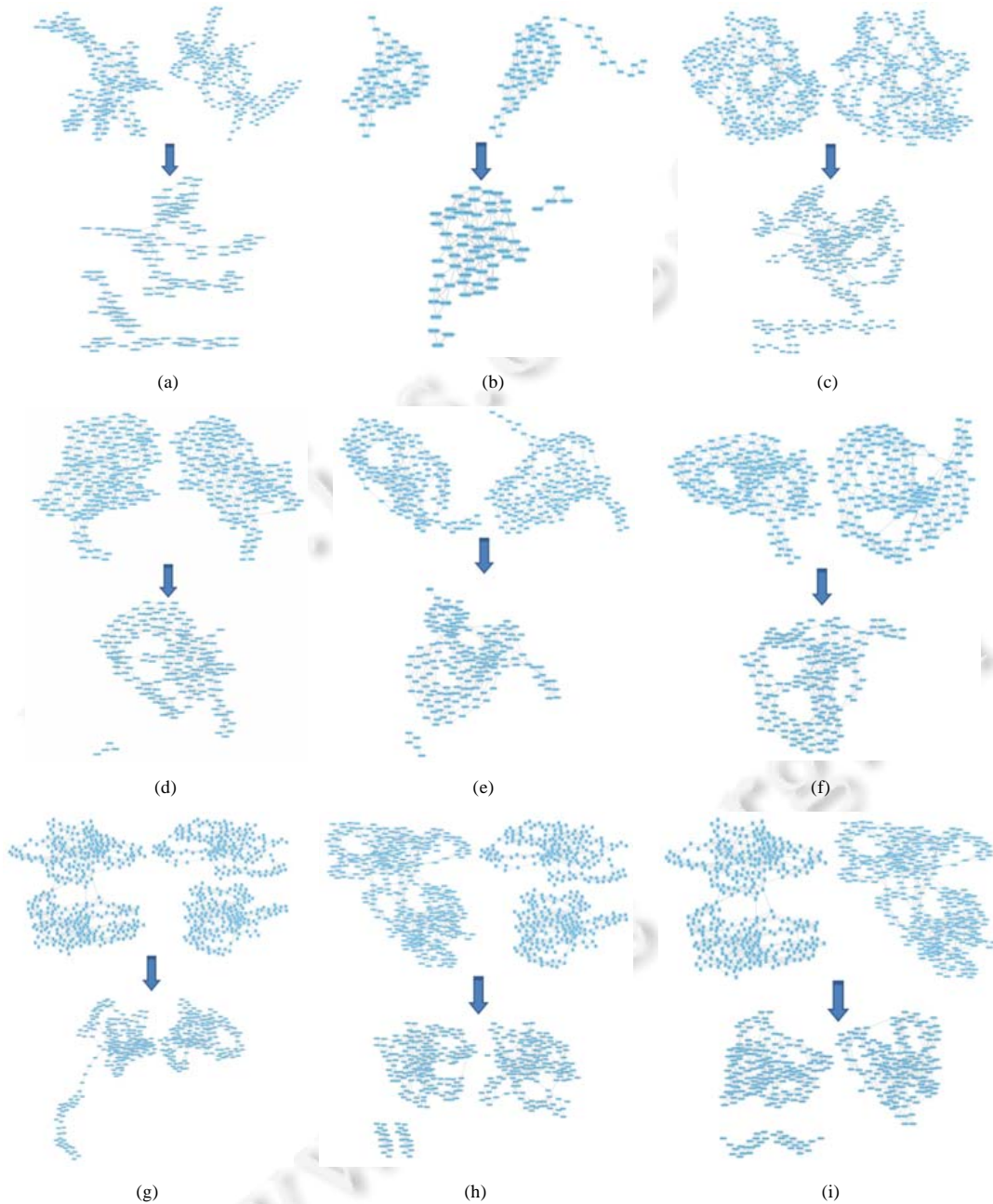


Fig.7 Common conserved subgraph by SI-MAGNA alignment algorithm

图 7 SI-MAGNA 网络比对方法产生的共同保守子图

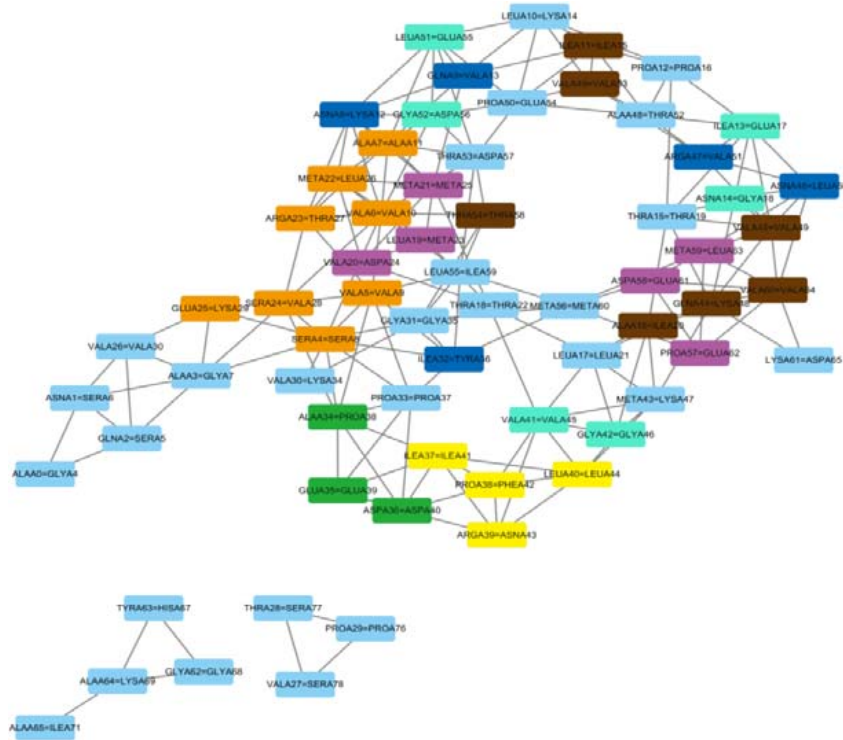


Fig.8 Common conserved subgraph of the 9AME&1WVO by SI-MAGNA alignment algorithm
图 8 SI-MAGNA 方法发现 9AME&1WVO 的共同保守子图

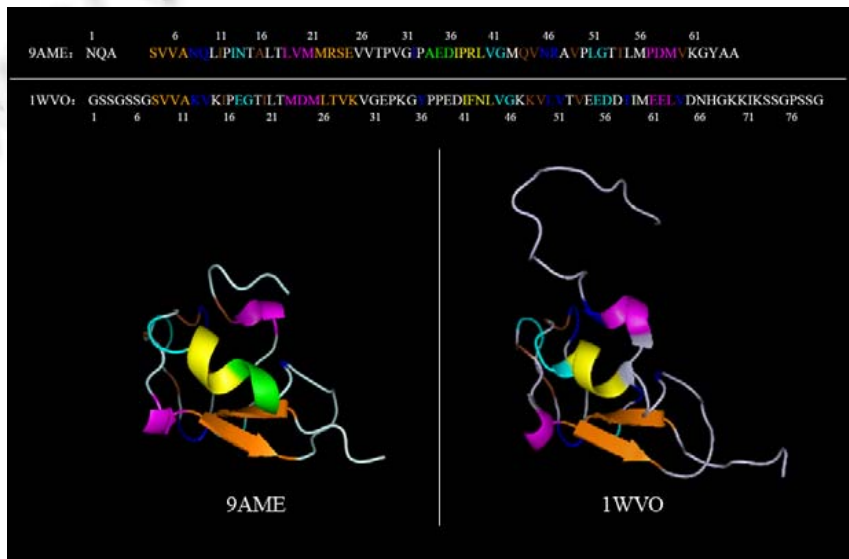


Fig.9 Illustration of sequence and secondary structure of the 9AME&1WVO
图 9 9AME&1WVO 的序列和二级结构对照图

从图 9 中可以发现,9AME 和 1WVO 构成相同二级结构的序列相似性很低,但两者主要的二级结构却非常相似.因此可以推断,尽管形成这些特定结构的残基有所不同,但两者共同包含的重要结构(1 个 α -螺旋、2 个 3_{10} -

螺旋和 2 个 β -折叠)以及这些结构间的重要相互作用与两者在各自生物体中实现类似的功能具有极大的关系^[24-25],可以从这部分的重要残基及其与周围结构间的重要相互作用入手,进一步探讨实现抗冻蛋白功能的结构.9AME 中独有的 1 个 3_{10} -螺旋结构以及它与周围结构间的重要相互作用,可能是导致两者在稳定性和活性方面具有不同温度依赖性的原因^[26],因此可以从这部分的重要残基切入,从结构的角度进一步研究热稳定性产生差异的根源.

3 结果与讨论

针对残基相互作用网络比对,对蛋白质-蛋白质相互作用网络比对算法 MAGNA 进行了改进,在其优化函数中引入了蛋白质的序列信息(即残基匹配度),并探讨拓扑信息和序列信息对网络比对的影响程度,使改进的 SI-MAGNA 算法更加适用于残基相互作用网络的比对.与此同时,通过实验证实,SI-MAGNA 方法在残基相互作用网络比对方的表现是优秀且稳定的,并且优于现有的基于网络拓扑的经典网络比对方法.此外,构建了同源蛋白质的残基相互作用网络,使用 SI-MAGNA 方法进行网络比对和分析,探索蛋白质结构对蛋白质性质、功能的影响.

残基相互作用网络比对方法将会成为研究蛋白质的空间结构、性质和功能的重要工具.通过残基相互作用网络比对方法进行比对分析,解释蛋白质结构的形成机理、探究蛋白质结构-功能关系,并可以将其运用于分子设计、分子筛选、药物设计等诸多领域.

References:

- [1] Meng L, Striegel A, Milenkovic T. Local versus global biological network alignment. *Bioinformatics*, 2016,32(20):3155–3164.
- [2] Chang S, Jiao X, Wang MH, Tian XH. Progress in amino acid networks of proteins. *Progress in Modern Biomedicine*, 2011,11(1): 190–193 (in Chinese with English abstract).
- [3] Kelley BP, *et al.* PathBLAST: A tool for alignment of protein interaction networks. *Nucleic Acids Research*, 2004,32(2):83–88.
- [4] Koyuturk M, Kim Y, Topkara U, Subramaniam S, Szpankowski W, Grama A. Pairwise alignment of protein interaction networks. *Journal of Computational Biology*, 2006,13(2):182–199.
- [5] Mina M, Guzzi PH. AlignMCL: Comparative analysis of protein interaction networks through Markov clustering. In: *Proc. of the 2012 IEEE Int'l Conf. on Bioinformatics and Biomedicine Workshops (BIBMW)*. 2012. 174–181.
- [6] Singh R, Xu JB, Berger B. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. of the National Academy of Science*, 2008,105(35):12763–12768.
- [7] Oleksii K, Milenkovic T, Vesna M, Wayne H, Natasa P. Topological network alignment uncovers biological function and phylogeny. *Journal of the Royal Society Interface*, 2010,7(50):1341–1354.
- [8] Kuchaiev O, Przulj N. Integrative network alignment reveals large regions of global network similarity in yeast and human. *Bioinformatics*, 2011,27(10):1390–1396.
- [9] Hashemifar S, Xu J. HubAlign: An accurate and efficient method for global alignment of protein-protein interaction networks. *Bioinformatics*, 2014,30(17):i438–i444.
- [10] Saraph V, Milenkovic T. MAGNA: Maximizing accuracy in global network alignment. *Bioinformatics*, 2014,30(20):2931–2940.
- [11] Hashemifar S, *et al.* ModuleAlign: Module-based global alignment of protein-protein interaction networks. *Bioinformatics*, 2016, 32(17):i658–i664.
- [12] Sun X, Lu ZH, Xie JM. *Bioinformatics Foundation*. Beijing: Tsinghua University Press, 2005 (in Chinese).
- [13] Malek M, Ibragimov R, Albrecht M, Baumbach J. CytoGEDEVO—Global alignment of biological networks with cytoscape. *Bioinformatics*, 2016,32(8):1259–1261.
- [14] Greene LH, Higman VA. Uncovering network systems within protein structures. *Molecular Biology Reports*, 2003,334(4):781–791.
- [15] Bode C, Kovács IA, Szalay MS, Palotai R, Korcsmáros T, Csermely P. Network analysis of protein dynamics. *FEBS Letters*, 2007, 581(15):2776–2782.
- [16] Estrada E. Universality in protein residue networks. *Biophys*, 2010,98(5):890–900.

- [17] Schlicker A, Domingues FS, Rahnenführer J, *et al.* A new measure for functional similarity of gene products based on gene ontology. *BMC Bioinformatics*, 2006,7:Article No.302.
- [18] Ibragimov R, *et al.* GEDEVO: An evolutionary graph edit distance algorithm for biological network alignment. In: Proc. of the German Conf. on Bioinformatics 2013 (GCB 2013). Gottingen: Schloss Dagstuhl-Leibniz-Zentrum Fuer Informatik, 2013. 68–79.
- [19] Faisal FE, Zhao H, Milenkovic T. Global network alignment in the context of aging. *IEEE/ACM Trans. on Computational Biology and Bioinformatics*, 2015,12(1):40–52.
- [20] Yan LC, Su JG, Chen WZ, Wang CX. Study on the characters of different types of amino-acid networks and their relations with protein folding. *Progress in Biochemistry and Biophysics*, 2010,37(7):762–768 (in Chinese with English abstract).
- [21] Wang XQ, Ding YR, Mu ZL, Cai YJ. Research on the relationship between iron superoxide dismutase amino acid networks and thermostability. *Acta Biophysica Sinica*, 2014,30(2):146–156 (in Chinese with English abstract).
- [22] Tan ZB, Li JF, Wu MC, Yin X, Hu D, Dong YH. Research advance on engineering thermostability of lipase. *Journal of Food Science and Biotechnology*, 2014,33(7):673–681 (in Chinese with English abstract).
- [23] Sangeeta K, Debjani R. Comparative structural studies of psychrophilic and mesophilic protein homologues by molecular dynamics simulation. *Journal of Molecular Graphics and Modelling*, 2009,27(8):871–880.
- [24] Guo XL, Gao L, Chen X. Models and algorithms for alignment of biological networks. *Ruan Jian Xue Bao/Journal of Software*, 2010,21(9):2089–2106 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3860.htm> [doi: 10.3724/SP.J.1001.2010.03860]
- [25] Yang JL, Li J, Dong LH, Grunewald S. A heuristic algorithm to align protein interaction networks. *Journal of Biomathematics*, 2011,26(3):569–575 (in Chinese with English abstract).
- [26] Tao SH, Ding YR. Research on the relationship between thermostability and structure of xylanase based on residue interaction network alignment. *Chinese Journal of Biochemistry and Molecular Biology*, 2018,34(7):760–768 (in Chinese with English abstract).

附中文参考文献:

- [2] 常珊, 焦雄, 王美华, 田绪红. 蛋白质氨基酸网络研究进展. *现代生物医学进展*, 2011, 11(1): 190–193.
- [12] 孙啸, 陆祖宏, 谢建明. *生物信息学基础*. 北京: 清华大学出版社, 2005.
- [20] 严立成, 苏计国, 陈慰祖, 王存新. 不同类型氨基酸网络参量与蛋白质折叠的关系. *生物化学与生物物理进展*, 2010, 37(7): 762–768.
- [21] 王雪芹, 丁彦蕊, 牟兆琳, 蔡宇杰. 超氧化物歧化酶氨基酸网络与耐热性的关系研究. *生物物理学报*, 2014, 30(2): 146–156.
- [22] 谭中标, 李剑芳, 郭敏辰, 殷欣, 胡蝶, 董运海. 脂肪酶热稳定性改造研究进展. *食品与生物技术学报*, 2014, 33(7): 673–681.
- [24] 郭杏莉, 高琳, 陈新. 生物网络比对的模型与算法. *软件学报*, 2010, 21(9): 2089–2106. <http://www.jos.org.cn/1000-9825/3860.htm> [doi: 10.3724/SP.J.1001.2010.03860]
- [25] 杨家亮, 李军, 董骝焕, Grunewald S. 一个生物网络比对的启发式算法. *生物数学学报*, 2011, 26(3): 569–575.
- [26] 陶斯涵, 丁彦蕊. 基于残基相互作用网络比对的木聚糖酶热稳定性研究. *中国生物化学与分子生物学报*, 2018, 34(7): 760–768.



陶斯涵(1993–), 女, 湖南株洲人, 硕士, 主要研究领域为计算智能, 生物信息学.



丁彦蕊(1976–), 女, 博士, 教授, CCF 专业会员, 主要研究领域为计算智能, 生物信息学.