

doi: 10. 3969/j. issn. 1004 - 275X. 2020. 09. 004

基于序列的蛋白质相互作用预测方法研究进展

孙晶京

(山西农业大学 文理学院, 山西 太谷 030801)

摘 要: 蛋白质相互作用 (PPI) 在所有生物体的细胞功能和生物学过程中都起着至关重要的作用, 可以帮助人们更好地了解感染机理, 开发药物和优化治疗方法。主要回顾和比较了基于氨基酸序列特性进行 PPI 预测的方法, 并讨论了该领域的技术挑战。

关键词: 蛋白质相互作用; 蛋白质序列; 计算方法

中图分类号: Q51 文献标识码: A 文章编号: 1004 - 275X (2020) 09 - 013 - 03

Research progress of sequence - based protein - protein interaction prediction methods

Sun Jingjing

(College of Arts and Sciences , Shanxi Agricultural University , Shanxi Taigu 030801)

Abstract: Protein - protein interaction (PPI) plays a vital role in the cellular functions and biological processes of all organisms , and can help people better study the molecular mechanism of diseases , develop drugs and optimize treatment methods. This article mainly reviews and compares methods for predicting PPI based on amino acid sequence characteristics , and discusses technical challenges.

Key words: protein - protein interaction; protein sequence; calculation method

蛋白质是生物大分子, 可以与其他各种分子相互作用, 如 DNA、RNA、小的配体和其他蛋白质。这些相互作用是促进细胞中大多数生物学过程的主力, 包括基因表达, 细胞生长、增殖, 营养吸收, 形态, 运动性, 细胞间通讯以及细胞凋亡。在分子水平上了解这些相互作用对于开发新的治疗方案, 注释蛋白质功能, 研究疾病的分子机制以及描述蛋白质相互作用网络非常重要。因此蛋白质的相互作用 (PPI) 预测得到了研究人员的广泛关注, 并提出了很多方法, 包括物理化学实验方法和计算方法。物理化学实验技术可以识别蛋白质之间的物理化学相互作用, 进而可以预测蛋白质之间的功能关系。这些技术包括基于酵母双杂交的方法、免疫共沉淀、串联亲和纯化、蛋白质芯片和混合方法等。尽管这些技术已经成功地鉴定了几种物种的几种重要的相互作用蛋白, 如酵母和拟南芥等, 但是, 实验方法在检测 PPI 方面存在费时费力、检测效率低、难以识别弱相互作用等缺点。因此, 人们开始寻求一些新的计算方法来验证实验结果并加速对未知蛋白质相互作用的预测。

1 计算方法

近年来, 已提出许多用于预测蛋白质相互作用的计算方法。这些方法可以分为两类: 基于结构的预测和基于序列的预测。基于结构的预测方法受到由实验确定的蛋白质结构相对较少的约束, 相反,

基于序列的预测方法只需要蛋白质序列即可预测蛋白质相互作用。因此, 本文主要介绍基于序列的 PPI 预测方法的研究进展。

基于序列的 PPI 预测方法主要是利用氨基酸 (AA) 特性, 比如疏水性, 理化特性, 进化谱, AA 组成, AA 平均值或滑动窗口上的加权平均值等。可以分为统计方法和机器学习的方法。

1.1 基于统计的预测方法

1.1.1 镜像树

Pazos 和 Valencia^[1]提出的镜像树是一种基于统计的方法, 基于相关蛋白质家族序列之间进化距离的比较和系统树的拓扑相似性来预测 PPI。其距离来自 McLachlan 氨基酸同源性矩阵的残基相似性的平均值。将树之间的相似度用于构建树的距离矩阵之间的相关性。镜像树方法不需要创建系统树, 而仅分析基础距离矩阵, 因此, 此方法独立于树的构建方法。尽管镜像树方法不需要完全测序的基因组, 但要求在考虑的所有物种中都存在直系同源蛋白。所以, 当更多的物种基因组可用时, 可以应用的蛋白质就更少了。此外, 由于距离矩阵不是系统树的最佳表示, 因此基于对距离矩阵的比较可能引入一些误差。

1.1.2 共同进化差异

Liu 等^[2]提出了一种共同进化差异 (CD) 的方法来预测人类蛋白质的相互作用。该方法基于两个假设, 一个是 PPI 对可能具有相似的替代率, 另一

个是, PPI 更可能在相关物种之间保存。CD 定义为两个蛋白质之间取代率差异的绝对值。CD 方法结合了来自许多物种的相互作用蛋白对的共同进化信息。该方法未使用多元比对, 因此比其他比对方法(如镜像树) 花费时间更少。该方法不要求所研究物种具有直系同源性的蛋白质, 但增加物种数量有利于提高方法的准确性。尽管此方法可以确定给定蛋白质对相互作用的可能性, 但它不能推断相互作用的特定特征, 如界面中的相互作用残基。

1.2 基于机器学习的预测方法

1.2.1 自协方差

Guo 等^[3]提出了一种使用自协方差(AC) 和支持向量机(SVM) 来预测 PPI 的方法。其中用 AC 来表征序列中相距一定距离的残基之间的相互作用。这样蛋白质序列则由一系列 AC 表征, 然后使用 AC 向量作为输入, 构建 SVM 模型。实验表明, 一个 AA 残基与其 30 个邻位 AA 的相互作用将有助于表征 PPI 信息。该方法的优点是 AC 包含了 AA 残基的远程相互作用信息, 这在 PPI 识别中很重要。

1.2.2 相似比对

Zaki 等^[4]提出了一种基于蛋白质一级结构成对相似性比较的 PPI 预测方法。每个蛋白质序列通过滑动窗口创建的 AA 子序列的相似值来表示。然后使用这些值作为输入, 构建 SVM 模型。因此, 蛋白质序列的相似性暗示了其同源性, 也揭示了结构和功能的相似性。

1.2.3 氨基酸组分

Roy 等^[5]研究了氨基酸组分(AAC) 在 PPI 预测中的作用及其常见特征(例如域、元组特征和标识产物特征) 的性能。每个蛋白质对用 AAC 和域特征来表示。AAC 以单体和二聚体特征来表示。单体特征是由单个 AA 特征组成, 而二聚体特征是由连续 AA 对的特征组成。然后将这些特征离散化为二进制特征。域特征是由离散化的域名标识来表示。为了将 AAC 与其他基于非域序列的特征进行比较, 创建了元组特征和标识产物特征。并使用三个机器学习分类器(逻辑回归, SVM 和朴素贝叶斯) 对提出的方法进行了评估。结果表明, AAC 在不同数据集和分类器中所做的贡献几乎等同, 这表明 AAC 包含了识别 PPI 的重要信息。AAC 适用于任何蛋白质序列, 尤其在缺乏域信息时非常有用。AAC 也可以结合其他特征来提高 PPI 预测的能力。

1.2.4 蛋白质相互作用的通用计算机模拟预测器(UNISPPI)

Valente 等^[6]提出了 UNISPPI 方法。该方法使

用 20 种 AA 理化性质的频率和组成, 用于训练决策树分类器。频率特征集包括 20 个 AA 在蛋白质序列中的百分比。每个 AA 归入与理化性质相关的三个不同组之一, 并计算每个特征的每个组的百分比, 最终得到共 21 个组成特征, 从而构建了组成特征集。结果表明, 仅使用 AA 频率就足以预测 PPI。另外, 天冬酰胺, 半胱氨酸和异亮氨酸的 AA 频率是区分相互作用和非相互作用蛋白对的重要特征。UNISPPI 的主要优点是简单和计算成本低, 因为少量特征用于训练决策树分类器。但是, 决策树分类器通常会出现过度拟合的问题。

1.2.5 ETB - Viterbi

Kern^[7]提出的 ETB - Viterbi 是 ipHMMs 中的一种具有早期追溯机制的解码算法(“交互特征隐马尔可夫模型”), 该算法旨在最佳地将输入序列中相互作用的 AA 残基之间的长距离相关性纳入其中。它能够捕获长距离相关性以提高预测精度, 且不受序列方向的影响。但是, Viterbi 算法在内存和计算时间上是非常昂贵的。因为隐马尔可夫模型训练涉及 Viterbi 算法的重复迭代, 且可能无法收敛到给定训练集的真正最佳参数集, 也可能出现过度拟合现象。

3 结语

本文对基于序列的蛋白质相互作用预测的计算技术进行了回顾, 总结了几种相关的现有方法, 并对它们进行了分类和比较。显然, 为了达到合理的预测精度, PPI 预测仍然需要大量的研究工作。在现有 PPI 预测方法中的特征更多关注 AA 理化性质, 而忽略了蛋白质的生物学特性。在未来工作中, 应更加关注将生物学知识纳入预测方法中。

参考文献:

- [1] Pazos F, Valencia A. Similarity of phylogenetic trees as an indicator of protein - protein interaction [J]. Protein engineering, 2001, 14(9): 609 - 614.
- [2] Liu C H, Li K C, Yuan S. Human protein - protein interaction prediction by a novel sequence - based co - evolution method: co - evolutionary divergence [J]. Bioinformatics, 2013, 29(1): 92 - 98.
- [3] Guo Y, Yu L, Wen Z, et al. Using support vector machine combined with auto covariance to predict protein - protein interactions from protein sequences [J]. Nucleic acids research, 2008, 36(9): 3025 - 3030.
- [4] Zaki N, Lazarova - Molnar S, El - Hajj W, et al. Protein - protein interaction based on pairwise similarity [J]. BMC bioinformatics, 2009, 10(1): 150.
- [5] Roy S, Martinez D, Platero H, et al, Werner - Washburne M. Exploiting amino acid composition for predicting protein

(下转第 17 页)

1. 87 mmol/g。

综上所述,复合金属氧化物较单一金属氧化物表现出更高的催化活性,主要原因是,催化剂比表面积、酸碱位浓度提高;引入第三种金属氧化物、磷酸、磷钨酸等对复合氧化物改性,可大幅度提高催化剂的结构性能,提高催化剂活性。

3 结语

CO₂ 是一种储量巨大,且非常廉价的碳源,由 CO₂ 和甲醇直接合成 DMC 是一种绿色环保的资源化路线。复合金属氧化物比表面积大,酸碱性位浓度高,适合作为合成反应催化剂。比表面积和酸碱性位浓度是合成 DMC 的关键因素,因此,对金属氧化物改性,如磷酸改性、磷钨酸改性、引入另一种氧化物,提高金属氧化物的比表面积和酸碱性位浓度是提高催化剂活性的发展方向。

参考文献:

- [1] Friedlingstei P, Jones M W, O'Sullivan M, et al. Global Carbon Budget 2019[J]. Earth Syst. Sci. Data, 2019, 11(4): 1783–1838.
- [2] Wu X L, Xiao M, Meng Y Z, et al. Direct synthesis of dimethyl carbonate on H₃PO₄ modified V₂O₅ [J]. J. Mol. Catal. A – Chem., 2005, 238(1–2): 158–162.
- [3] Aymes D, Ballivet – Tkatchenko D, Jeyalakshmi K, et al. A comparative study of methanol carbonation on unsupported SnO₂ and ZrO₂ [J]. Catal. Today, 2009, 147(2): 62–67.
- [4] Kang K H, Joe W, Lee C H, et al. Direct Synthesis of Dimethyl Carbonate from Methanol and Carbon Dioxide Over CeO₂(X) – ZnO (1–X) Nano – Catalysts [J]. J. Nanosci. Nanotechnol., 2013, 13(12): 8116–8120.
- [5] 陈红萍. CO₂ 和甲醇直接合成碳酸二甲酯非均相催化体系的研究[D]. 天津: 河北工业大学, 2014.
- [6] Tomishige K, Sakaihorii T, Ikeda Y, et al. A novel method of direct synthesis of dimethyl carbonate from methanol and carbon dioxide catalyzed by zirconia [J]. Catal. Lett., 1999, 58(4): 225–229.
- [7] Chen L, Wang S P, Zhou J J, et al. Dimethyl carbonate synthesis from carbon dioxide and methanol over CeO₂ versus over ZrO₂: comparison of mechanisms [J]. RSC Adv., 2014, 4(59): 30968–30975.
- [8] Jung K T, Bell A T. An in situ Infrared Study of Dimethyl Carbonate Synthesis from Carbon Dioxide and Methanol over Zirconia [J]. J. Catal., 2001, 204(2): 339–347.

- [9] Yoshida Y, Arai Y, Kado S, et al. Direct synthesis of organic carbonates from the reaction of CO₂ with methanol and ethanol over CeO₂ catalysts [J]. Catal. Today, 2006, 115(1–4): 95–101.
- [10] Liu B, Li C M, Zhang G Q, et al. Oxygen Vacancy Promoting Dimethyl Carbonate Synthesis from CO₂ and Methanol over Zr – doped CeO₂ Nanorods [J]. ACS Catal., 2018, 8(11): 10446–10456.
- [11] Tamboli A H, Chaugule A A, Gosavi S W, et al. Ce_xZr_{1-x}O₂ solid solutions for catalytic synthesis of dimethyl carbonate from CO₂: Reaction mechanism and the effect of catalyst morphology on catalytic activity [J]. Fuel, 2018, 216: 245–254.
- [12] Zhang Z F, Liu Z W, Lu J, et al. Synthesis of Dimethyl Carbonate from Carbon Dioxide and Methanol over Ce_xZr_{1-x}O₂ and [EMIM]Br/Ce_{0.5}Zr_{0.5}O₂ [J]. Ind. Eng. Chem. Res., 2011, 50(4): 1981–1988.
- [13] Kang K H, Lee C H, Kim D B, et al. NiO/CeO₂ – ZnO Nano – Catalysts for Direct Synthesis of Dimethyl Carbonate from Methanol and Carbon Dioxide [J]. J. Nanosci. Nanotechnol., 2014, 14(11): 8693–8698.
- [14] Liu B, Li C M, Zhang G Q, et al. Direct Synthesis of Dimethyl Carbonate from CO₂ and Methanol over CaO – CeO₂ Catalysts: The Role of Acidic – basic Properties and Surface Oxygen Vacancies [J]. New J. Chem., 2017, 41(20): 12231–12240.
- [15] Eta V, Mäki – Arvela P, Wärnå J, et al. Kinetics of dimethyl carbonate synthesis from methanol and carbon dioxide over ZrO₂ – MgO catalyst in the presence of butylene oxide as additive [J]. Appl. Catal. A – Gen., 2011, 404(1–2): 39–46.
- [16] Fu Z W, Zhong Y Y, Yu Y H, et al. TiO₂ Doped CeO₂ Nanorod Catalyst for Direct Conversion of CO₂ and CH₃OH to Dimethyl Carbonate: Catalytic Performance and Kinetic study [J]. ACS Omega, 2018, 3(1): 198–207.
- [17] La K W, Jung J C, Kim H, et al. Effect of acid – base properties of H₃PW₁₂O₄₀/Ce_xTi_{1-x}O₂ catalysts on the direct synthesis of dimethyl carbonate from methanol and carbon dioxide: A TPD study of H₃PW₁₂O₄₀/Ce_xTi_{1-x}O₂ catalysts [J]. J. Mol. Catal. A – Chem., 2007, 269(1–2): 41–45.

收稿日期: 2020 – 08 – 10

基金项目: 湖南省自然科学基金(No. 2019JJ50566)、湖南省教育厅科学研究项目(No. 18CJ017)、湖南省大学生创新创业训练计划资助项目和湘南学院校级科学研究项目(2018XJ21)。

作者简介: 曾妮(1986 –), 女, 湖北恩施人, 硕士, 助教, 研究方向: 催化剂设计及开发、CO₂ 资源化利用。

通信作者: 时米东(1986 –), 男, 山东菏泽人, 博士, 讲师, 研究方向: 催化剂设计及开发、CO₂ 资源化利用、有机合成。

(上接第 14 页)

- protein interactions [J]. PloS one, 2009, 4(11): e7813.
- [6] Valente G T, Acencio M L, Martins C, et al. The development of a universal in silico predictor of protein – protein interactions [J]. PloS one, 2013, 8(5): 65587.
- [7] Kern C, Gonzalez A J, Liao L, et al. Vijay – Shanker. Predicting interacting residues using long – distance information and novel decoding in hidden markov models [J]. Nanoscience, IEEE Transactionson, 2013, 12(13): 158–164.

收稿日期: 2020 – 07 – 15

基金项目: 国家自然科学基金(31871571); 国家自然科学基金青年项目(61803245); 山西省科技攻关计划(201903D211002; 201603D221037–3); 中国博士后基金(2017M621105); 山西省优秀博士来晋工作奖励资金科研项目(SXYBKY2018040); 山西农业大学科技创新基金(201308)

作者简介: 孙晶京(1982 –), 女, 讲师, 研究方向: 生物信息学。