

文章编号:1671-4229(2023)01-0025-08

蛋白质与蛋白质相互作用预测模型综述

钱冰, 马宝山*, 刘玉

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

摘要: 蛋白质作为生命活动的物质基础,通过彼此之间的相互作用来参与生物信号传递、能量和物质代谢及细胞周期调控等,因此,预测蛋白质与蛋白质相互作用(PPIs)有助于从系统角度理解生命过程,同时为细胞机制的研究奠定重要基础。目前,高通量测序技术的进步使研究人员获取了海量的蛋白质相互作用数据。面对大量具有潜在可用性的PPIs数据,国内外研究人员提出了多种PPIs预测模型。文章综述了现有的基于计算的PPIs预测方法,并进行分类,讨论了其各自优缺点。最后对PPIs预测方法进行总结和展望,提出了未来可以深入研究的方向。

关键词: 蛋白质相互作用; 预测方法; 计算模型

中图分类号: TP 3-05; TP 391 **文献标志码:** A

A survey of models for protein-protein interactions prediction

QIAN Bing, MA Bao-shan*, LIU Yu

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China)

Abstract: As the material basis of life activities, proteins participate in biological signal transmission, energy and substance metabolism and cell cycle regulation through their interactions with each other. Therefore, predicting protein-protein interactions (PPIs) helps to understand life processes from a systemic perspective, and provides an important foundation for the study of cellular mechanisms. At present, advances in high-throughput sequencing technologies have enabled researchers to acquire massive amounts of proteomics data. Faced with the large amount of potentially available data, researchers have proposed various computational models for PPIs prediction. In this paper, we review the existing computational methods for PPIs prediction, classify them, and discuss their respective advantages and disadvantages. Finally, we conclude with a summary of the methods for predicting PPIs, and suggest directions that can be studied in depth in the future.

Key words: protein-protein interaction; prediction methods; computational models

蛋白质是生命的物质基础,是生命活动的主要承担者。作为构成细胞的基本有机物,蛋白质对细胞功能具有至关重要的作用^[1]。然而绝大部分的蛋白质并不是单独发挥作用的,而是通过蛋

白质与蛋白质之间的相互作用来实现的,即PPIs。蛋白质大分子通过与其他蛋白质作用,参与DNA转录和复制、细胞信号传导以及生物体生长与繁殖等,来实现相应的生命活动^[2]。

收稿日期: 2022-07-28; 修回日期: 2022-09-06

基金项目: 大连市科技创新基金资助项目(2020JJ27SN066)

作者简介: 钱冰(1999—),女,硕士研究生. E-mail: qb990112@163.com

*通信作者. E-mail: mabaoshan@dlnu.edu.cn

引文格式: 钱冰, 马宝山, 刘玉. 蛋白质与蛋白质相互作用预测模型综述[J]. 广州大学学报(自然科学版), 2023, 22(1): 25-32.

随着人类基因组计划的完成,越来越多的研究人员开始关注蛋白质组学的研究。通过高通量检测、基于生物信息学的计算预测以及相关文献挖掘等手段,已产出海量蛋白质相互作用数据,且随着技术的发展以及生物信息学科的进步,PPIs 数据仍在不断增长。为了高效稳妥地存储、管理并应用这些数据,研究人员提出数百种不同类型的蛋白质相互作用数据库,同时提供相应的检索工具,方便人们利用这些已知的 PPIs 数据验证新的科学假设、探索未知 PPIs 和分析生物学规律等。其中需要注意的一点是,通过生物信息学方法预测的 PPIs 是研究人员基于一定科学依据提出的,

尚未经过生物实验验证,只能作为新的研究方向或者计算方法的线索。

蛋白质相互作用数据库已成为构建蛋白质相互作用网络的主要资源、研究蛋白质功能的有效平台。作者总结了几种常见的 PPI 数据库,如表 1 所示。同时还有用于相关研究的 PPI 服务器, PPIsearch^[3] (<http://gemdock.life.nctu.edu.tw/ppisearch/>) 是一个 web 服务器,它可以快速识别同源 PPI,并推断查询蛋白质对的相互作用域和功能的可转移性。通过整理数据库信息,了解到 PPI 信息的多种获取方式,有利于后续的 PPI 预测。

表 1 PPIs 数据库、链接以及简介

Table 1 PPIs databases, links, and introductions

数据库	链 接	简 介
STRING	https://string-db.org/	整合了蛋白质之间所有已知和预测的关联,包括物理相互作用和功能关联 ^[4]
BioGRID	https://thebiogrid.org	管理和存储所有主要模式生物物种和人类的蛋白质遗传和化学的相互作用 ^[5]
BISC	http://bisc.cse.ucsc.edu	为用户提供蛋白质之间的物理直接相互作用的结构透视图和相关信息 ^[6]
PrePPI	http://bhapp.c2b2.columbia.edu/PrePPI	使用贝叶斯框架将预测和实验确定的 PPI 结合起来的数据库 ^[7]
HMNPPID	http://202.118.75.18:8082/HMNPPID.asp	人类恶性肿瘤的 PPI 数据库 ^[8]
BIND	http://download.baderlab.org/BINDTranslation/	存档了生物分子相互作用、复杂的路径信息 ^[9]
UniProt	https://www.uniprot.org/	为用户提供一套全面、高质量、可自由访问的蛋白质序列,并标注功能信息 ^[10]

关于蛋白质相互作用的预测,人们倾向于使用计算的方法,相较于传统的实验方法更具优势。传统的实验方法包括酵母双杂交技术^[11]、分裂泛素系统等^[12],优点是操作较为简单,价格便宜,但也存在许多弊端^[13],例如,酵母双杂交技术允许在活酵母中直接检测蛋白质之间是否存在相互作用^[14],但是存在一些蛋白质可单独转录启动,实现基因的表达,造成假阳性,并且该技术无法检测到微弱的瞬态相互作用,导致假阴性^[15]。同时还需要大量的人工和时间成本,效率低,无法满足大规模数据的检测^[16]。随着高通量实验技术的发展,越来越多蛋白质相互作用关系被确定,所构建的蛋白质相互作用网络规模日益扩大,并且复杂网络理论的成熟为研究蛋白质相互作用网络奠定了

一定基础。以研究蛋白质相互作用网络的形成机制为起点,结合网络拓扑参量与蛋白质生物信息数据^[17],利用数学和计算的方法来建模并分析^[18]。通过分析得到的网络特征,像网络的度分布、相关系数和聚类系数等能够直观体现复杂网络中各蛋白质之间的关联,为预测新链接提供强有力的论据^[17]。

本文主要总结了 6 类 PPIs 的预测模型,PPIs 预测模型及主要方法代码链接如表 2 所示。包括基于序列的模型、基于网络的模型、基于域的模型、基于结构的模型、基于深度学习的模型和基于多源数据的模型。通过总结现有的预测 PPIs 的模型,希望对今后 PPIs 预测模型的建立提供帮助。

表 2 PPIs 预测模型及部分方法代码链接

Table 2 PPIs prediction models and links to some algorithm codes

类 别	方法名	链 接
基于序列的计算模型	Pazos, et al	http://www.pdg.cnib.uam.es/mirrortree
	Sato, et al	http://timpani.genome.ad.jp/~parco/
	Hou, et al	http://www.ibi.vu.nl/downloads/RF_PPI/
基于网络的计算模型	NEOComplex	http://acolab.ie.nthu.edu.tw/bionetwork/NEOComplex
	GLIDE	https://bitbucket.org/kap_devkota/glide
基于域的计算模型	Zhang, et al	http://dx.doi.org/10.1016/j.jtbi.2016.02.026
	iLoops	http://sbi.imim.es/iLoops.php
	IntPred	www.bioinf.org.uk/intpred/
基于结构的计算模型	Sun, et al	repharma.pku.edu.cn/ppi
基于深度学习的计算模型	DPPI	https://github.com/hashemifar/DPPI/
基于多源数据的计算模型	idenPC-CAP	http://bliulab.net/idenPC-CAP/

1 PPIs 预测模型

1.1 基于序列的模型

研究表明,蛋白质序列由氨基酸构成,携带了大量有用信息,因此,人们尝试建立不同的基于序列的计算模型来研究蛋白质功能,构建蛋白质相互作用网络。大多数基于序列的模型通常考虑蛋白质对的序列相似性,从蛋白质序列中提取特征向量,然后利用机器学习的分类器预测 PPIs。

Pazos 等^[19]通过比较序列间的进化距离来衡量蛋白质对的协同进化程度,程度越深,蛋白质对相互作用的可能性就越大,但这一方法会导致预测结果假阳性过多。Sato 等^[20]提出使用偏相关系数来衡量协同进化程度,有效减少了 PPIs 预测的假阳性。Hu 等^[21]提出名为 CoFex 的序列特征提取方法,通过考虑成对蛋白质序列间的协变部分来合成特征向量,再灵活运用分类器预测 PPIs,明显提高了预测精度。Hou 等^[22]利用蛋白质序列不同属性的特征,包括同源和异源特征,结合随机森林(RF)评估各种特征的重要性,从而预测 PPIs。结果表明,单一的序列特征无法准确判断 PPIs,结合不同属性确实可以提高预测准确性。

基于序列的模型(图 1)利用蛋白质序列信息来预测 PPIs,方法简单易懂。但由于蛋白质序列信息冗杂,在提取特征时,无法保证信息的完整性。同时,在利用机器学习分类算法时可能会因为信息冗余出现过拟合,造成预测的准确率较低。因此,后续使用序列特征时,要着重关注特征提取

方法,增强特征^[23],同时可以尝试多种分类器,选择预测效果更好的算法提高准确率。

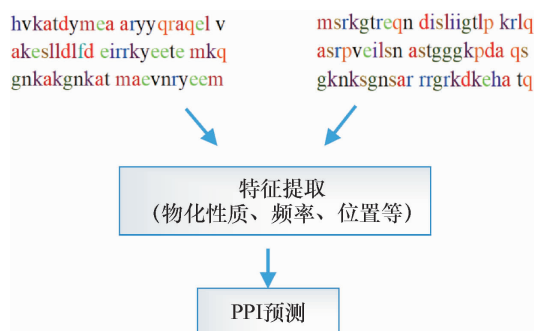


图 1 基于序列的模型

Fig. 1 Sequence-based model

1.2 基于网络的模型

基于网络的计算模型通过研究 PPIs 网络的拓扑结构,研究蛋白质对在 PPIs 网络中的连通性,并对其评分,从而确定蛋白质对是否具有相互作用。通常,用一个二元组 $G = \{V, E\}$ 来表示蛋白质相互作用网络,其中, $V = \{v_i\} (1 \leq i \leq N_v)$ 表示蛋白质集合,而 $E = [E_{ij}] (1 \leq i, j \leq N_v)$ 是 G 的一个 $N_v \times N_v$ 的邻接矩阵。当 $E = 1$ 时表示两个蛋白质相互作用, $E = 0$ 则相反^[24]。

Luo 等^[25]在子图上沿用度的概念以重新定义模块,并提出了一种新的聚类算法,结合 G-N 算法产生的边的相对顺序组装模块,更好地展现了蛋白质网络的拓扑特征。Ma 等^[26]提出名为 NEO-Complex 的方法,该算法综合了多网络对比(MNA)所获取的不同类型网络的结构信息,并且定义了边聚类系数 NECC,作为 PPIs 网络交互边

权重。该方法降低了缺少 PPIs 网络边缘的影响,有助于发现网络稀疏部分的蛋白质间相互作用。Kovács 等^[27]证明了若蛋白质 A 与蛋白质 B 中的一个相互作用蛋白质相似,则蛋白质 A、B 也相互作用,并在此基础上提出了 L3 模型,定义了一个依赖于长度为 3 的网络路径的度归一化分数,优于以往的链路预测方法。最新研究还将图、卷积矩阵用于链路预测,引发众多关注^[28]。Devkota 等^[29]提出了全局和局部集成扩散嵌入的预测方法,提高了 PPIs 网络非核心部分的研究价值。

基于网络的模型(图 2)通过挖掘蛋白质相互作用网络中的结构信息,如共同邻居、网络路径和几何嵌入等,来预测蛋白质节点间的链路(相互作用),结果较为可靠。但由于 PPIs 网络的复杂性,预测精度较低。因此,还需尽量完善 PPIs 网络,筛选出最相关的信息,同时也不可忽视真实生物网络的噪声等带来的影响。

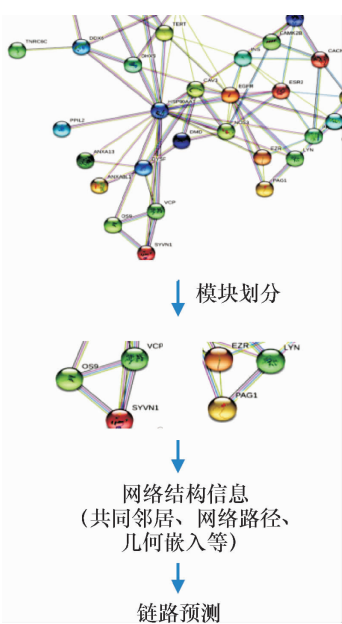


图 2 基于网络的模型

Fig. 2 Network-based model

1.3 基于域模型

PPIs 通常涉及域之间的结合,域是蛋白质折叠、进化和功能的基本单位^[30]。通过收集大量的蛋白质域信息对 PPIs 进行预测,蛋白质域与域相互作用信息可以从 3did、ArchDB、DOMINE 等公共数据库检索得到:3did 是一个三维相互作用域数据库,它收集了高分辨率三维结构的蛋白质中域与域相互作用^[31];ArchDB 是一个结构分类的数据

库^[32];DOMINE 是一个域与域交互的数据库^[33]。基于域的模型来预测 PPIs 存在两个相互作用的蛋白质 A 和 B,另外两个查询蛋白质为 C 和 D,如果蛋白质 C 中的域和蛋白质 D 中的域分别与蛋白质 A 和 B 中的域类似,那么 C 和 D 可能存在相互作用^[34]。

Singhal 等^[35]提出了 DomainGA,这是一种多参数优化方法,利用有关域-域相互作用的信息来预测 PPIs,DomainGA 算法预测的精度优于最大似然估计方法。Zhang 等^[36]利用 InterDom^[37]、3did 和 Pfam^[38]等数据库的信息建立了优化模型。它是一种基于域与域信息和蛋白质与结构域相互作用的 PPIs 预测新方法。与以往的 PPIs 预测方法相比,该方法在灵敏度、准确率等方面均取得了较好的结果,特别是降低了人体 PPIs 预测的假阴性率。Planas-Iglesias 等^[39]介绍了 iLoops,它是一种网络服务器,可以通过识别 SFs 来预测两种蛋白质是否具有相互作用,并能通过学习 PPIs 和 NIP 中的环或结构域模式来预测蛋白质对之间是否结合。

基于域的 PPIs 预测模型能够整合蛋白质域信息,较为准确地预测 PPIs。但是当蛋白质结构域信息不完整或者有错误时,PPIs 预测准确率会受到影响。未来应该通过多种途径来完善蛋白质结构域信息,从而使得蛋白质结构域信息在预测 PPIs 中发挥更大的作用。

1.4 基于结构的模型

蛋白质具有多种结构,包括一级结构^[40]、二级结构^[41]、三级结构以及四级结构,其中,三级结构对 PPIs 影响较为深刻^[42],可以利用这些结构信息预测 PPIs。基本原理是如果蛋白质 A 和蛋白质 B 分别与具有相互作用的蛋白质 C 和蛋白质 D 结构相似,那么蛋白质 A 和蛋白质 B 也可能具有相互作用^[43]。

Zhao 等^[44]假设界面结构相似的蛋白质可能共享相似的相互作用蛋白质,预测了 HIV-1 与人类蛋白质的相互作用,用 UniAlign 方法来对结构进行比对,该方法可以更好地捕捉保守残基的相似性,并得到了更好的预测精度。Northey 等^[45]提出了随机森林机器学习预测器 IntPred,它可以预测蛋白质-蛋白质界面位点。在进行集中测试时,IntPred 优于绝大部分现有的方法。Singh

等^[46]提出了一个 web 服务器 Struct2Net,它是一个基于结构信息预测 PPIs 的工具,它的优点在于可以通过不断集成最新的结构模板来改进其预测的精度。

基于结构的 PPIs 预测模型通过整合蛋白质的结构信息来预测 PPIs,该方法有助于相互作用以及位点的预测,但是蛋白质的结构数据有限,并且结构信息会随着时间、环境进行动态变化^[47],增加了预测 PPIs 的难度。

1.5 基于深度学习的模型

深度学习作为机器学习的一个新的分支,它在很多机器学习应用的领域也取得了成功^[48],同时很多学者将深度学习也应用到 PPIs 预测中,运用深度学习来对 PPIs 进行预测的原理是从蛋白质中提取某些和相互作用相关的信息,并自动提取特征,而无需手工设计^[49],进而对 PPIs 进行预测。

Sun 等^[50]提出了使用深度学习中的堆栈式自动编码器对蛋白质的序列信息进行研究,预测蛋白质之间是否存在相互作用,结果显示该模型的平均准确率较高,并且与以往的方法比较,该模型在其他多套独立数据集上也取得了较好的准确率。这是深度学习首次通过蛋白质序列信息对 PPIs 进行预测,同时结果也显示了深度学习在该领域的优势。Hashemifar 等^[51]提出了一种新的深度学习框架 DPPI,该方法是根据序列信息对 PPIs 进行建模和预测。结果表明 DPPI 的精度 - 召回曲线下的面积优于其他几种基准方法,并且计算效率更高。Li 等^[52]提出了一种深度集成学习方法 EnAmDNN 来预测蛋白质相互作用,与之前只利用单一的方法对 PPIs 进行预测相比较,该方法综合了其他各种方法的优点和模型的优势,结果表明 EnAmDNN 比其他方法的预测性能更好。

深度学习在预测蛋白质相互作用时,既有它的可取之处,但也存在诸多不足。深度学习的优点在于它具有强大的学习能力,能够准确学习蛋白质各种信息的特征,从而生成更加准确的特征向量^[24]。其不足的地方在于深度学习存在过拟合和泛化^[53]。在以后的研究中可以不断改进模型,使深度学习在预测 PPIs 中发挥更大的作用。

1.6 基于多源数据的模型

基于多源数据的模型将多种生物相关信息,例如基因组表达数据、基因本体论、氨基酸频率等整合到 PPIs 网络中,重新构建资源更丰富、结构更

清晰的网络,并在该网络中预测 PPIs。

Chen 等^[54]提出了基于多源学习的蛋白质社区检测 (MLPCD) 算法,通过集成基因表达数据与 PPIs 网络信息,重建加权网络,并定义了社区模块化,利用基因本体标注评估蛋白质功能模块的丰富度,为后期预测提供了更多有力指标。实验结果证明了整合多种数据资源,可以识别更多可靠的相互作用蛋白质。Yu 等^[55]结合基因本体论、氨基酸频率等构造了不同类型的加权 PPIs 网络,综合考虑了密度、直径和余弦相似度等指标选择关键节点,对蛋白质进行分组或聚类,获得了较好的预测结果。Wu 等^[56]构建了由 RNA-RNA 相互作用、RNA-蛋白质相互作用以及 PPIs 组成的网络,提出 idenPC-CAP 算法,考虑了 RNA 与蛋白质之间的相互作用。实验结果表明该方法显著降低了预测时出现假阳性的比例。

基于多源数据的模型不仅可以研究 PPIs 网络,同时也关注蛋白质与其他生物分子间的互作网络,筛选更具有生物学意义的蛋白质群落,以此来提高 PPIs 预测性能。未来还需整合多源数据,加强对数据的质控和分类,使其更好的应用到 PPIs 网络的预测中。

2 展 望

蛋白质相互作用 (PPIs) 对探索生物过程至关重要,通过深入研究 PPIs,不仅有助于明确个体生命活动作用机制,还能用于探索疾病的发生机制^[57],改进现有的疾病治疗手段,寻找新的药物研发靶点^[58],为攻克许多至今仍困扰人类的疾病开辟道路。本文回顾了各种预测 PPIs 的计算方法,总结了前人的经验,希望为今后的 PPIs 预测提供指导和帮助。

本文将预测蛋白质相互作用的计算方法分为 6 类,实际上在已有文献中有其他的分类体系,不可否认这几类方法间有着密不可分的联系。有研究人员同时考虑两种或以上方法,综合其优点建立模型,往往会有更好效果。特别是深度学习作为近些年研究热点,在数据挖掘、模式识别、机器学习等方面取得了显著成果,广泛应用于计算机视觉、自然语言处理和生物信息学等。结合深度学习成为研究的一种趋势。Ye 等^[13]将多种已知的生物数据,例如,基因本体术语、蛋白质相互作

用数据、基因百科全书等作为深度学习模型的输入,合并多维特征向量,降低网络的维度,获取了更有效的预测。

目前大多数预测模型都用部分数据先对模型进行训练,数据集的质量与预测精度息息相关。因此,对数据样本进行预处理,建立具有可解释性

的特征集,是评估模型性能的关键。此外,作者还注意到,仅在有限的生物中验证了大规模的 PPIs。鉴于蛋白质的进化守恒^[59],一些蛋白质在不同物种间存在相似的表达,利用现有数据可以预测其他物种的 PPIs。同时,积累多种生物的蛋白质组学数据有利于进一步推进 PPIs 预测研究^[60]。

参考文献:

- [1] You Z H, Zhong M, Huang H Y, et al. A novel method to predict protein-protein interactions based on the information of protein sequence[C]//IEEE International Conference on Control System, Computing and Engineering. Piscataway: IEEE, 2012, 210-215.
- [2] Ma D C, Diao Y B, Guo Y Z, et al. A novel method to predict protein-protein interactions based on the information of protein-protein interaction networks and protein sequence[J]. Protein and Peptide Letters, 2011, 18(9): 906-911.
- [3] Chen C C, Lin C Y, Lo Y S, et al. PPIsearch: A web server for searching homologous protein-protein interactions across multiple species[J]. Nucleic Acids Research, 2009, 37(S2): W369-W375.
- [4] Szklarczyk D, Gable A L, Nastou K C, et al. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets[J]. Nucleic Acids Research, 2021, 49(D1): 605-612.
- [5] Oughtred R, Stark C, Breitkreutz B J, et al. The BioGRID interaction database: 2019 update[J]. Nucleic Acids Research, 2019, 47(D1): 529-541.
- [6] Juettemann T, Gerloff D L. BISC: Binary subcomplexes in proteins database[J]. Nucleic Acids Research, 2010, 39(S1): D705-D711.
- [7] Zhang Q C, Petrey D, Garzón J I, et al. PrePPI: A structure-informed database of protein-protein interactions[J]. Nucleic Acids Research, 2012, 41(D1): D828-D833.
- [8] Li Q, Yang Z, Zhao Z, et al. HMNPPID-human malignant neoplasm protein-protein interaction database[J]. Human Genomics, 2019, 13(1): 1-10.
- [9] Bader G D, Betel D, Hogue C W. BIND: The biomolecular interaction network database[J]. Nucleic Acids Research, 2003, 31(1): 248-250.
- [10] UniProt Consortium. UniProt: The universal protein knowledgebase in 2021[J]. Nucleic Acids Research, 2021, 49(D1): 480-489.
- [11] 林存刚, 徐慧, 陈玉成. 酵母双杂交系统研究及其应用进展[J]. 微生物学杂志, 2005(6): 85-89.
- [12] Xing S, Wallmeroth N, Berendzen K W, et al. Techniques for the analysis of protein-protein interactions in Vivo[J]. Plant Physiology, 2016, 171(2): 727-758.
- [13] Ye J, Wang S, Yang X, et al. Gene prediction of aging-related diseases based on DNN and mashup[J]. BMC Bioinformatics, 2021, 22(1): 597.
- [14] Brückner A, Polge C, Lentze N, et al. Yeast two-hybrid, a powerful tool for systems biology[J]. International Journal of Molecular Sciences, 2009, 10(6): 2763-2788.
- [15] Farooq Q U A, Shaikat Z, Aiman S, et al. Protein-protein interactions: Methods, databases, and applications in virus-host study[J]. World Journal of Virology, 2021, 10(6): 288-300.
- [16] Khatun M S, Shoombuatong W, Hasan M M, et al. Evolution of sequence-based bioinformatics tools for protein-protein interaction prediction[J]. Current Genomics, 2020, 21(6): 454-463.
- [17] 王荣全. 基于蛋白质相互作用网络的蛋白质复合物识别算法研究[D]. 长春: 吉林大学, 2020.
- [18] 骆嘉伟, 梁成, 宋丹, 等. 蛋白质相互作用网络演化模型研究进展[J]. 计算机应用, 2013, 33(3): 816-820.
- [19] Pazos F, Valencia A. Similarity of phylogenetic trees as indicator of protein-protein interaction[J]. Protein Engineering, 2001, 14(9): 609-614.
- [20] Sato T, Yamanishi Y, Horimoto K, et al. Partial correlation coefficient between distance matrices as a new indicator of pro-

- tein-protein interactions[J]. *Bioinformatics*, 2006, 22(20): 2488-2492.
- [21] Hu L, Chan K C. Extracting coevolutionary features from protein sequences for predicting protein-protein interactions[J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2016, 14(1): 155-166.
- [22] Hou Q Z, De Geest P, Vranken W F, et al. Seeing the trees through the forest: Sequence-based homo- and heteromeric protein-protein interaction sites prediction using random forest[J]. *Bioinformatics*, 2017, 33(10): 1479-1487.
- [23] Agrawal S, Sisodia D S, Nagwani N K. Augmented sequence features and subcellular localization for functional characterization of unknown protein sequences[J]. *Medical Biological Engineering & Computing*, 2021, 59(11/12): 2297-2310.
- [24] Hu L, Wang X, Huang Y A, et al. A survey on computational models for predicting protein-protein interactions[J]. *Briefings in Bioinformatics*, 2021, 22(5): bbab036.
- [25] Luo F, Yang Y, Chen C F, et al. Modular organization of protein interaction networks[J]. *Bioinformatics*, 2007, 23(2): 207-214.
- [26] Ma C Y, Chen Y P, Berger B, et al. Identification of protein complexes by integrating multiple alignment of protein interaction networks[J]. *Bioinformatics*, 2017, 33(11): 1681-1688.
- [27] Kovács I A, Luck K, Spirohn K, et al. Network-based prediction of protein interactions[J]. *Nature Communications*, 2019, 10(1): 1-8.
- [28] Coşkun M, Koyutürk M. Node similarity based graph convolution for link prediction in biological networks[J]. *Bioinformatics*, 2021, 37(23): 4501-4508.
- [29] Devkota K, Murphy J M, Cowen L J. GLIDE: Combining local methods and diffusion state embeddings to predict missing interactions in biological networks[J]. *Bioinformatics*, 2020, 36(S1): 464-473.
- [30] Ta H X, Holm L. Evaluation of different domain-based methods in protein interaction prediction[J]. *Biochemical and Biophysical Research Communications*, 2009, 390(3): 357-362.
- [31] Mosca R, Céol A, Stein A, et al. 3did: A catalog of domain-based interactions of known three-dimensional structure[J]. *Nucleic Acids Research*, 2014, 42(D1): D374-D379.
- [32] Espadaler J, Fernandez-Fuentes N, Hermoso A, et al. ArchDB: Automated protein loop classification as a tool for structural genomics[J]. *Nucleic Acids Research*, 2004, 32(S1): D185-D188.
- [33] Raghavachari B, Tasneem A, Przytycka T M, et al. DOMINE: A database of protein domain interactions[J]. *Nucleic Acids Research*, 2008, 36(S1): D656-D661.
- [34] Dong S, Provart N J. Analyses of protein interaction networks using computational tools[M]. *Two-Hybrid Systems*. New York, NY: Humana Press, 2018: 97-117.
- [35] Singhal M, Resat H. A domain-based approach to predict protein-protein interactions[J]. *BMC Bioinformatics*, 2007, 8(1): 1-19.
- [36] Zhang X, Jiao X, Song J, et al. Prediction of human protein-protein interaction by a domain-based approach[J]. *Journal of Theoretical Biology*, 2016, 396: 144-153.
- [37] Ng S K, Zhang Z, Tan S H, et al. InterDom: A database of putative interacting protein domains for validating predicted protein interactions and complexes[J]. *Nucleic Acids Research*, 2003, 31(1): 251-254.
- [38] Finn R D, Bateman A, Clements J, et al. Pfam: The protein families database[J]. *Nucleic Acids Research*, 2014, 42(D1): D222-D230.
- [39] Planas-Iglesias J, Marin-Lopez M A, Bonet J, et al. iLoops: A protein-protein interaction prediction server based on structural features[J]. *Bioinformatics*, 2013, 29(18): 2360-2362.
- [40] Bock J R, Gough D A. Predicting protein-protein interactions from primary structure[J]. *Bioinformatics*, 2001, 17(5): 455-460.
- [41] Wardah W, Khan M, Sharma A, et al. Protein secondary structure prediction using neural networks and deep learning: A review[J]. *Computational Biology and Chemistry*, 2019, 81: 1-8.
- [42] Kotlyar M, Rossos A E M, Jurisica I. Prediction of protein-protein interactions[J]. *Current Protocols in Bioinformatics*, 2017, 60: 1-14.
- [43] Lewis A C, Saeed R, Deane C M. Predicting protein-protein interactions in the context of protein evolution[J]. *Molecular BioSystems*, 2010, 6(1): 55-64.

- [44] Zhao C, Zang Y, Wei Q, et al. HIV1-human protein-protein interaction prediction based on interface architecture similarity [C]//IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Piscataway: IEEE, 2017: 97-100.
- [45] Northey T C, Barešić A, Martin A. IntPred: A structure-based predictor of protein-protein interaction sites[J]. *Bioinformatics*, 2018, 34(2): 223-229.
- [46] Singh R, Park D, Xu J, et al. Struct2Net: A web service to predict protein-protein interactions using a structure-based approach[J]. *Nucleic Acids Research*, 2010, 38(S2): W508-W515.
- [47] 于建涛, 郭茂祖, 蔡禄. 蛋白质相互作用及其网络预测方法研究进展[J]. *电子学报*, 2007, 35(B12): 1-7.
- [48] Chan H P, Samala R K, Hadjiiski L M, et al. Deep learning in medical image analysis[J]. *Advances in Experimental Medicine and Biology*, 2020, 1213: 3-21.
- [49] Shi C, Chen J, Kang X, et al. Deep learning in the study of protein-related interactions[J]. *Protein and Peptide Letters*, 2020, 27(5): 359-369.
- [50] Sun T, Zhou B, Lai L, et al. Sequence-based prediction of protein protein interaction using a deep-learning algorithm[J]. *BMC Bioinformatics*, 2017, 18(1): 1-8.
- [51] Hashemifar S, Neyshabur B, Khan A A, et al. Predicting protein-protein interactions through sequence-based deep learning [J]. *Bioinformatics*, 2018, 34(17): 802-810.
- [52] Li F, Zhu F, Ling X, et al. Protein interaction network reconstruction through ensemble deep learning with attention mechanism[J]. *Frontiers in Bioengineering and Biotechnology*, 2020, 8: 390.
- [53] Li H, Gong X J, Yu H, et al. Deep neural network based predictions of protein interactions using primary sequences[J]. *Molecules*, 2018, 23(8): 1923.
- [54] Chen J, Li K, Bilal K, Metwally A A, et al. Parallel protein community detection in large-scale PPI networks based on multi-source learning[J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2018. doi:10.1109/TCBB.2018.2868088.
- [55] Yu Y, Zheng Z. Protein complex identification based on weighted PPI network with multi-source information[J]. *Journal of Theoretical Biology*, 2019, 477: 77-83.
- [56] Wu Z, Liao Q, Fan S, et al. idenPC-CAP: Identify protein complexes from weighted RNA-protein heterogeneous interaction networks using co-assemble partner relation[J]. *Briefings in Bioinformatics*, 2021, 22(4): bbaa372.
- [57] Rabbani G, Baig M H, Ahmad K, et al. Protein-protein interactions and their role in various diseases and their prediction techniques[J]. *Current Protein & Peptide Science*, 2018, 19(10): 948-957.
- [58] Murakami Y, Tripathi L P, Prathipati P, et al. Network analysis and in silico prediction of protein-protein interactions with applications in drug discovery[J]. *Current Opinion in Structural Biology*, 2017, 44: 134-142.
- [59] Hu J, Li J, Chen N, et al. Conservation of hot regions in protein-protein interaction in evolution[J]. *Methods*, 2016, 110: 73-80.
- [60] Ding Z, Kihara D. Computational methods for predicting protein-protein interactions using various protein features[J]. *Current Protocols in Protein Science*, 2018, 93(1): e62.

【责任编辑:周全】