

基于优化的纽曼随机游走模型挖掘甘草中的重要靶蛋白

姜强¹, 张颖², 于茜¹, 但汉字¹, 李佐静^{1*}

(沈阳药科大学 医疗器械学院, 辽宁 沈阳 110016; 2. 国家药品监督管理局, 北京 100037)

摘要: 目的 利用优化的纽曼随机游走模型, 挖掘甘草中常规拓扑分析方法容易忽视的重要靶蛋白。方法 模拟药物在体内吸收过程, 构建了多层随机游走网络。收集甘草的成分、靶蛋白、通路和治疗疾病等信息, 根据不同节点的特性, 设置了不同的转移概率, 并进行不同步长的游走。通过统计经过靶蛋白的游走次数作为衡量其重要性的依据, 筛选出前15%的靶蛋白, 并与归一化拓扑模型、MRWBC模型进行比较, 最终得到待验证的候选靶蛋白。结果 经过优化后的ARWBC模型, 在众多候选靶蛋白中筛选出了9个潜在的靶蛋白, 其中经过文献和分子对接验证了6个靶蛋白(CD40LG、IFNG、PRKCB、IGF2、E2F1、TGF1B)。结论 优化后的纽曼随机游走模型挖掘出了甘草中重要的靶蛋白, 这些候选靶蛋白的转录活性, 以期为甘草作为未来方剂的药理学研究提供有力的理论基础。

关键词: 随机游走; 甘草; 靶蛋白; 复杂网络

中图分类号: R 917 **文献标志码:** A

引用本文: 姜强, 张颖, 于茜, 等. 基于优化的纽曼随机游走模型挖掘甘草中的重要靶蛋白[J]. 沈阳药科大学学报, 2026, 43(4): 363-373.

Cite this article as: JIANG Qiang, ZHANG Ying, YU Qian, et al. Identifying the important target proteins in Licorice based on an optimized Newman random walk model[J]. Journal of Shenyang Pharmaceutical University, 2026, 43(4): 363-373.

甘草作为一种临床常用且不可或缺的中药材, 具有抗炎、抗病毒、抗菌、抗氧化、抗癌活性、免疫调节、保肝和保护心脏等药理作用, 主要用于治疗消化性溃疡^[1]与丙型肝炎^[2]等疾病。相关临床研究表明, 甘草富含三萜皂苷、黄酮类、异黄酮类^[3]和多糖类等生物活性成分, 其中甘草酸被认定为主要的生物活性成分^[4]。随着复杂网络在生物信息领域的广泛应用, 已有学者通过构建多层网络来挖掘中药材或方剂中的靶蛋白节点, 以系统阐释所筛选出的重要靶蛋白。

“随机游走”又称为“随机漫步”, 是指在给定一个网络和该网络中一个起点的前提下, 游走者从起点出发, 按照转移概率随机选择一个邻居节点并移动到该邻居节点, 然后将该邻居节点作为新的起点, 如此反复迭代, 直到达到停止条件(例如, 游走者到达目标节点或达到预先设定的游走次数)。在随机游走的过程中, 由于节点的重要性并不总是相同, 频繁出现在游走路径中的节点被认为是重要节点, 即所寻找的重要节点。介数中心性(betweenness centrality, BC)在网络分析中用于衡量节点在网络中的中介地位和影响力, 是

基于节点在网络中连接其他节点的路径数量和权重的度量。基于这两个理念, Newman^[5]运用介数中心性的概念探究网络中节点的重要性, 提出了一种通过多次随机游走来预估每个节点重要性的游走策略, 称为纽曼随机游走介数中心性(Newman random walk betweenness centrality, RWBC)。

近年来, 在生物网络方面的一些研究仍然局限于蛋白质相互作用网络的分析(protein-protein interaction, PPI), 而没有考虑到现实中药物的吸收利用往往受到多个因素影响的情况。本研究利用与目标网络相关的多维度信息, 模拟药物在人体内的吸收过程, 构建了一个包含成分、靶蛋白、通路和疾病等多个维度的复杂网络。将优化后的随机游走模型用于构建复杂网络并进行甘草潜在重要靶蛋白的预测。

1 材料与方法

1.1 构建复杂网络

1.1.1 数据来源

选择与甘草相关的生物网络信息进行网络的

收稿日期: 2023-11-03

作者简介: 姜强(1996-), 男, 硕士研究生, E-mail jiang_runrun@163.com; *通信作者: 李佐静(1977-), 女, 教授, 博士, 主要从事药物信息学和生物信息与计算机辅助药物设计以及药理学计算可视化研究, E-mail zuojing1006@hotmail.com。

构建并筛选重要靶蛋白,数据来源为 TCMSP 数据库、PubChem 数据库、STITCH 数据库、STRING 数据库、David 数据库和 KEGG 数据库,生物网络中的信息为节点,节点间的相互作用关系为边,具体节点数量与边的数量如表 1 所示。

首先根据吸收、分布、代谢、排泄 (ADME) 标准 ($OB \geq 30\%$, $DL \geq 0.18$) 从 TCMSP 数据库^[6]中筛选活性成分^[7],然后在 PubChem 数据库中收集用于分子对接验证的 SMILES 结构^[8],其中甘草的有效靶标来源为 DrugBank 和 SwissTarget 预

Table 1 Data summary for building complex networks

表 1 构建复杂网络数据汇总

Connection	Number of nodes	Number of edges	Data sources
I:Ingredients	84	—	TCMSP
T:Targets	227	—	PubChem
I-I:Ingredients-ingredients interacting relationship	84	68	STITCH
I-T:Ingredients-targets interacting relationship	84 227	1 489	STITCH
T-T:Targets-targets interacting relationship	227	3 808	STRING
T-P:Targets-pathway interacting relationship	227 179	2 923	David
P-D:Pathway-disease interacting relationship	179 64	64	KEGG

1.1.2 网络构建

邻接矩阵是常用表示图形结构的方阵,其中行和列分别对应于图中的节点坐标,矩阵中的元素表示节点之间的连接关系。利用邻接矩阵构建一个由节点和连接边组成的无向网络图 G ,进行节点间随机游走。根据数据长度 n 构建一个全零矩阵,见公式(1)。

$$Z = \begin{bmatrix} z_{i1} & z_{i2} & \cdots & z_{in} \\ z_{j1} & z_{j2} & \cdots & z_{jn} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nn} \end{bmatrix} \quad (1)$$

其中 Z 表示邻接矩阵, z 表示节点,其下标表示坐标关系。

将不同的小网络模块链接在一起,根据节点之间的相互作用关系构建邻接矩阵。在节点 i 和节点 j 之间存在一条边的情况下,邻接矩阵中第 i 行第 j 列的元素为 1 反之为 0,由于是无向网络第 j 行第 i 列的元素也为 1 反之为 0。

对于同一个区域内节点,有连接关系的节点记作 1,无连接关系的记作 0,关系见公式(2)。

$$M^s[i,j] = \begin{cases} 1 & \text{if } (v_s^i, v_s^j) \in E_s \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

其中 M 表示连接关系, s 和 t 表示不同的层,

其次利用 UniProt 数据库 (<https://www.uniprot.org/>) 将收集的成分名称进行标准化为通用靶蛋白名称,在 David (<https://david.ncifcrf.gov/>) 数据库通过对靶蛋白进行富集分析得到对应通路信息。最后在 KEGG 数据库分别利用 STITCH (<http://stitch.embl.de/>)、STRING (<https://string-db.org/cgi/>)^[9] 数据库按照置信度水平 >0.4 进行相互作用关系筛选,并作为最终的映射,检索疾病与通路的关系从而获得相应疾病网络信息。

E 表示该区域节点的集合。

对于跨区域节点,有连接关系的节点记作 1,无连接关系的记作 0,关系见公式(3)。

$$M^s[i,j] = \begin{cases} 1 & \text{if } (v_s^i, v_t^j) \in E_{st} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

其中 M 表示连接关系, s 和 t 表示不同的层, E 表示该区域节点的集合。

1.2 网络分析

本研究利用与目标网络相关的生物信息,对节点进行分类构建节点转移概率分布的矩阵,在多层纽曼随机游走介数中心性模型^[10] (multilayer Newman random walk betweenness centrality, MR-WBC) 的基础上优化转移概率和游走策略,构建一个异步纽曼随机游走介数中心性模型 (asynchronous Newman random walk betweenness centrality, ARWBC),最后统计节点的游走次数以预测重要的靶蛋白。

1.2.1 随机游走过程中的转移概率

在复杂网络中,为了提高筛选重要靶蛋白的准确性,需要考虑与节点相关的邻居关系,度中心性常用来衡量一个节点在网络中与其它节点直接连接的数量。因此,将度中心性优化值 D 融入到转移概率中,见公式(4)。

$$D(i) = \frac{d_i}{N-1} \quad (4)$$

其中 d_i 表示与现有节点*i*相连的边数量, N 表示节点*i*所属网络的总数量。

根据复杂网络的特点,将节点集合划分为同一区域和不同区域两部分,并将节点间的连接关系分

为3种情况:一种情况是节点*i*的邻居节点*j*仅存在于同一区域内的连接关系(如图1A所示);另一种情况是节点*i*的同区域邻居节点*j*存在跨区域的邻居节点(如图1B所示);最后一种情况是节点*i*存在跨区域的邻居节点*j*(如图1C所示)。

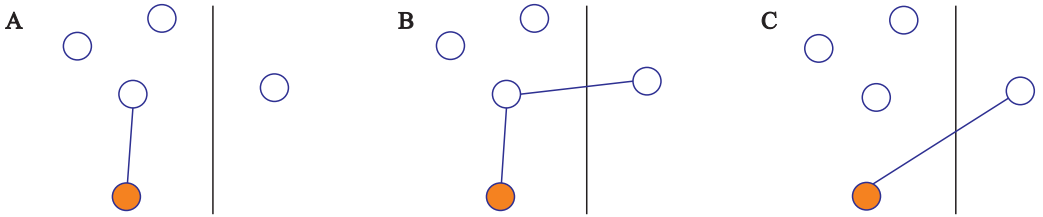


Fig. 1 The neighbor node is in the same area, and no other area is connected (A); The neighbor node is in the same area and is connected to other areas at the same time (B); The neighbor node is in another area (C)

图1 邻居节点在同一区域,且没有其他区域连接(A);邻居节点在同一区域,且同时连接到其他区域(B);邻居节点在另一区域(C)

为了选择进行深度游走和广度游走的情况,因此引入参数 λ 和 μ , D 表示该节点的度中心性优化值,参数 λ 代表节点以广度优先方式进行游走,参数 μ 表示节点将以深度优先方式进行游走。 $\sum_{j \in \theta} n_j$ 与 $\sum_{j \in \theta'} n_j$ 分别表示公式(5)同区域邻居节点*j*的总数和公式(6)(7)跨区域邻居节点*j*的总数。

不同类型节点转移概率,表示从该节点转移到邻居节点的概率,公式如下。

$$P_{ij} = D\lambda\mu \sum_{j \in \theta} n_j \quad (5)$$

$$P_{ij} = D\lambda(1-\mu) \sum_{j \in \theta'} n_j \quad (6)$$

$$P_{ij} = D(1-\lambda)(1-\mu) \sum_{j \in \theta'} n_j \quad (7)$$

1.2.2 随机游走过程中的游走策略选择

在常见的随机游走中,根据数据特点和研究目的可以选择不同的游走方法。最短路径中心性被认为是相对应力(stress)中心性的一种指标^[11]。Newman引入的随机游走方法^[12]表示从源节点开始到目标节点结束的随机游走经过所有包含的节点,并对所有节点间的路径进行均值处理,然后选取源节点到目标节点的最短路径。这种方法优化了随机游走时的最优路径选择。另外,BC常用于描述一个网络中通过节点的最短路径条数,具体见公式(8)。

$$BC(i) = \frac{\sum_{p \neq q} I_{pq}^i}{I_{pq}} \quad (8)$$

其中, I_{pq} 表示源节点*p*到目标节点*q*最短路

径的数目, I_{pq}^i 表示从源节点*p*到目标节点*q*最短路径经过节点*i*的次数。

纽曼随机游走介数中心性考虑了随机游走时路径的数量,通过统计所有路径中该节点出现的次数来衡量节点的重要性,具体见公式(9)。

$$RWBC(i) = \frac{\sum_{p \neq q} I_{pq}^i}{n(n-1)} \quad (9)$$

其中 I_{pq}^i 表示从源节点*p*到目标节点*q*经过节点*i*的次数。

Islam等^[13]在研究协变量与马尔可夫模型之间的关系时发现协变量与转移概率存在依赖性,在高阶马尔可夫链中表现更为明显,且相对于高阶模型,对数据预测的捕捉更全面。本次研究规定,游走次数为10 000次,统计ARWBC模型重要节点。根据高阶模型和数据特点,设置参数 η 作为阈值,判断节点在游走时是否跳过当前节点到达其它节点,以减少节点之间距离过远导致错过重要节点,即在游走时考虑不同步长的游走策略。在获得某一时刻的转移概率矩阵后,在游走过程中判断节点下一步的游走是否为异步游走,具体见公式(10)。

$$Z_{ij} = \begin{cases} 1 & \text{if } \eta \leq 0.2 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

其中, η 表示异步游走概率的参数,将阈值设定为0.2。当该值小于阈值时进行异步游走,否则进行单步游走。单步游走时其转移概率矩阵如公式(11)(12)。

$$\tilde{P}_{ij} = P_{ij} = \begin{bmatrix} D\lambda\mu / \sum_{j \in \theta} n_j \\ D\lambda(1-\mu) \sum_{j \in \theta'} n_j \\ D(1-\lambda)(1-\mu) \sum_{j \in \theta''} n_j \end{bmatrix} \quad (11)$$

异步游走时其转移概率矩阵见(12)。

$$\tilde{P}_{ij}' = \eta P_{ij} \quad (12)$$

1.2.3 ARWBC 对甘草靶蛋白的预测

将数据库下载的甘草成分、靶蛋白、通路、疾病视为邻接矩阵中节点,它们之间相互作用关系作为连接关系构建邻接矩阵。其次根据不同类型节点的数量映射为邻接矩阵中对应的长度进行分层,包括成分层、靶蛋白层、通路层、疾病层。最后在随机游走开始前从第一层随机选择一个节点作为游走的起始节点,该节点在模型中根据邻居节点的转移概率矩阵进行游走。游走停止条件为到达疾病层模拟人体吸收,或者达到预先设定的游走步数为 10 000 步,即停止当前的游走,将此过程重复 10 000 次,靶蛋白层节点游走到的次数作为甘草节点重要性排名的依据。

1.3 ARWBC 模型验证

目前,酿酒酵母(*saccharomyces cerevisiae*)相关的蛋白相互作用及重要蛋白数据是所有物种中最完善且可靠性最高的数据之一^[14]。在本研究中,首先提取多个数据库中的酿酒酵母蛋白数据,建立重要酿酒酵母蛋白数据集和全部酵母蛋白数据集。然后,利用全部酵母蛋白数据集构建蛋白相互作用网络,并使用优化模型 ARWBC 对重要酿酒酵母蛋白进行预测。

基于本研究模拟人体药物吸收在体内不同层次节点相互影响的特点,将酿酒酵母蛋白随机分成 4 个部分,分别模拟实验设计中的成分、靶蛋白、通路和疾病。为了更好地评估优化模型 ARWBC,将其与接近度中心性模型(closeness centrality, CC)、介数中心性模型(betweenness centrality, BC)、度中心性模型(degree centrality, DC)以及优化前的模型 MRWBC 进行对比。考虑到重要蛋白的数量为 1 110 个,将模型预测排序靠前的酿酒酵母蛋白分别按照占比前 1%、2%、5%、10%、15%、20%、25% 的比例进行分组,并计算不同分组中重要蛋白的数量,最后在全部数据集中添加相应的统计学指标验证模型的可靠性。

1.4 拓扑模型对甘草靶蛋白的预测

传统运用网络药理学方法探究药材治疗疾病

表现出的多靶点、多通路作用机理时,会根据靶蛋白之间的相互作用强弱来判断相应靶蛋白在整个网络中的重要程度,从而筛选出重要靶蛋白。本次研究将甘草靶蛋白文件及其相互作用关系网络文件导入 Cytoscape 软件分别分析计算出这些靶蛋白的度中心性(degree centrality, DC)、介数中心性(betweenness centrality, BC)、接近中心性(closeness centrality, CC)和特征向量中心性(eigenvector centrality, EC)。虽然 DC 能评估节点重要性但是忽略了节点间路径的作用,而 BC 考虑了路径的重要性但在存在多条路径时节点重要性可能高估,同样 CC 考虑到了节点与其他节点之间的关系但会受到网络连通性的影响,EC 考虑了与其他节点的连接程度但在复杂拓扑网络中会受到较大的影响,这些指标侧重于不同的评估角度并且有各自局限性,在研究中对其预测指标进行了归一化处理,通过归一化后的 TI_i 值来评估筛选出的靶蛋白重要性。拓扑模型重要蛋白的评估见公式(13)。

$$TI_i = \frac{1}{4} \left(\frac{DC}{DC_{max}} + \frac{BC}{BC_{max}} + \frac{CC}{CC_{max}} + \frac{EC}{EC_{max}} \right) \quad (13)$$

其中 DC 代表靶蛋白的度中心性,BC 代表靶蛋白的介数中心性,CC 代表靶蛋白的接近中心性,EC 代表靶蛋白的特征向量中心性。

1.5 文献挖掘验证

在 PubMed 数据库中,使用关键词“liquorice”和“glycyrrhiza”检索了近 5 年的相关文献。然后,对筛选出的重要差异靶蛋白在 TTD 数据库中查找其与疾病的对应记录,再通过比对分析文献报道结果和数据库结果来判断候选靶蛋白的可靠性。

1.6 分子对接验证

分子对接技术广泛用于现代药物设计,不仅可以探索各种大分子靶标结合位点内采用的配体构象,还可通过评估分子间识别过程中涉及的关键现象来估计配-受体结合自由能,及配体结合模式和稳定配体-受体复合物的相应分子间相互作用。本次研究将筛选出的靶蛋白与甘草小分子进行分子对接,同时将对接结果与对照组进行对比来验证筛选出的靶蛋白。首先,对 PDB 数据库下载差异靶蛋白的晶体结构进行初步筛选,筛选条件:(1)蛋白晶体构象检测方法为 X-RAY;(2)晶体分辨率小于 3.0 Å。使用 Cytoscape 软件筛选出甘草中度中心性值排名前 5 的小分子成分

(表2),在 PubChem 数据库中获取它们对应的 smile 二维结构式,将小分子二维结构在 Discovery Studio 软件中加氢变成 3D 结构,然后在该软件中对靶蛋白晶体结构进行去水加氢处理。选择 3 种

模型预测出排名靠前同时被现有研究普遍认为是甘草中重要靶蛋白的 IL6 作为分子对接结果对照。在版本为 v2.1 的 IGDOCK 软件中选择“general”模式,其余参数设置为默认进行分子对接。

Table 2 Information on the top five active ingredients of licorice degree centrality

表 2 甘草度中心性前五的活性成分信息

Component	Lipinski rules			B/%	DL
	$M_w \leq 500$	H-bond acceptors ≤ 10	H-bond donors ≤ 5		
Naringenin	272.27	5	3	59.29	0.21
Quercetin	302.25	7	5	46.43	0.28
Formononetin	268.28	4	1	69.67	0.21
Kaempferol	286.25	6	4	41.88	0.24
7-Methoxy 2-methyl isoflavone	266.31	3	0	42.56	0.20

2 结果与验证

2.1 优化后模型可靠性分析

在 DEG 数据库 (database of essential genes)^[15] 中检索到了 1 110 个重要酿酒酵母蛋白,在 DIP 数据库 (database of interacting proteins)^[16] 和 UniProt 数据库中检索酿酒酵母蛋白,并经过交集等数据处理,得到了包含 5 085 个酿酒酵母蛋白的完整数据集。利用 STRING 数据库检索了这些蛋白之间的相互作用关系,共得到了 22 798 条蛋白相互作用边。

在预测前 25% 酵母蛋白重要蛋白时,前 2% (101) 酵母蛋白由于模型随机游走加上异步游走的策略影响重要蛋白预测数量低于其他模型,从 5% (253) 开始模型预测数量超过其他模型,并且

随着酵母蛋白数量增加,同百分位酵母蛋白的重要蛋白预测数量稳定增加同时处于领先地位(图 2)。为了进一步评价 ARWBC 模型,用几个常用的统计指标来评估模型(表 3)。酿酒酵母蛋白共有 5 085 个,包括 1 110 个关键蛋白和 3 975 个非关键蛋白,所有的蛋白按照不同的算法结果进行降序排序。真阳性 (true positive, TP) 与假阳性 (false positive, FP) 分别表示前 1 110 个蛋白中关键蛋白的数目与非关键蛋白的数目,假阴性 (false negative, FN),与真阴性 (true negative, TN) 分别表示排序靠后 3 975 个蛋白中关键蛋白的数目与非关键蛋白的数目。根据 TP、FP、FN、TN 分别计算了几种算法模型结果的敏感度 (sensitivity, SN)、特异性 (specificity, SP)、阳性预测值 (positive predictive value, PPV)、阴性预测值 (negative

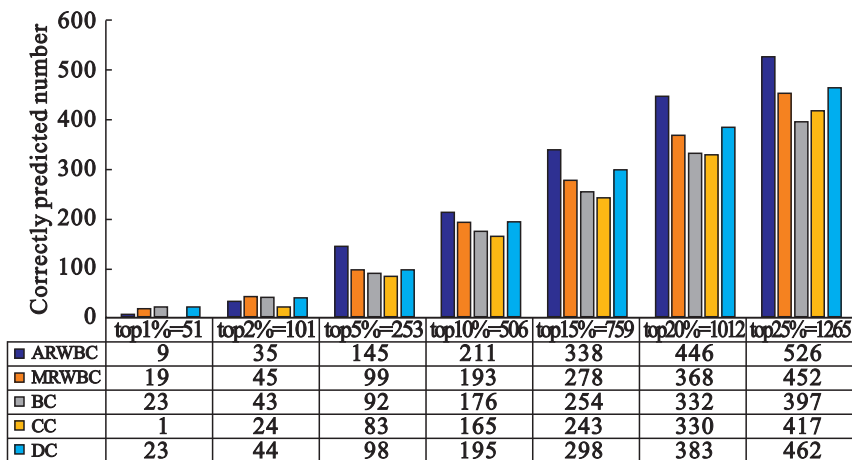


Fig. 2 Number of predictions of key target proteins in different model of saccharomyces cerevisiae

图 2 不同模型中酿酒酵母重要靶蛋白预测数量

predictive value, NPV)、F- 指标(F-measure, F) 和准确率(accuracy, ACC), 其结果如表4所示。根据统计结果优化后的模型(ARWBC)无论是比优化之前的模型(MRWBC)还是传统的拓扑模型在

统计学指标上都有一定程度的提升, 并且其准确率达到75.6%, 相较于优化前提升了3.4%, 比差异最大的传统模型CC准确率提升了4.8%, 这些结果说明优化后的ARWBC模型具有可靠性。

Table 3 Common statistical terms

表3 常用的统计学术语

Measure	Formula
SN	$SN = TP / (TP + FN)$
SP	$SP = TN / (TN + FP)$
PPV	$PPV = TP / (TP + FP)$
NPV	$NPV = TN / (TN + FN)$
F	$F = (2 \times SN \times PPV) / (SN + PPV)$
ACC	$ACC = (TP + TN) / (P + N)$

Table 4 Statistical results of different statistical algorithm models

表4 不同统计学算法模型统计学结果

Model	SN	SP	PPV	NPV	F	ACC
ARWBC	0.441	0.844	0.441	0.844	0.441	0.756
MRWBC	0.364	0.822	0.364	0.822	0.364	0.722
DC	0.374	0.825	0.374	0.825	0.374	0.727
BC	0.326	0.812	0.326	0.812	0.326	0.706
CC	0.332	0.813	0.332	0.813	0.332	0.708

2.2 不同模型对甘草靶蛋白的挖掘

将拓扑模型的预测结果与优化前后的模型预测得到的重要靶蛋白进行了对比, 分别收集3种模型预测的前15%的靶蛋白(表5), 然后将这些靶蛋白进行重复度对比(图3), 蓝色表示归一化拓扑模型, 其预测的相较于其余两个模型差异靶蛋白数量为15.4%(8个), 红色为优化前模型(MRWBC)预测, 其预测的差异靶蛋白数量为11.5%(6个), 而绿色为优化后模型(ARWBC)预测, 其预测的差异靶蛋白数量为17.3%

(9个), 接近35%(18个)靶蛋白在3种模型中都有出现。根据前文在酿酒酵母蛋白数据集上ARWBC模型预测结果统计学指标的提升和前25%酿酒酵母重要靶蛋白预测数量的差异, 说明本次优化后模型(ARWBC)能力较强, 将ARWBC得到区别于归一化拓扑模型和MRWBC模型的靶蛋白作为甘草候选的潜在重要靶蛋白(表6), 这部分靶蛋白将作为本研究进行验证和分析甘草治疗疾病时潜在重要靶蛋白对象。

Table 5 Top 15% target proteins of different models

表5 不同模型前15%靶蛋白

Topology	MRWBC	ARWBC	Topology	MRWBC	ARWBC
AKT1	PTEN	AKT1	CCND1	IGFBP3	CCND1
IL6	MYC	MAPK1	ERBB2	ERBB2	RAF1
TNF	IL1B	TNF	NOS3	CCL2	NFKBIA
TP53	TP53	TP53	SIRT1	RAF1	IKBKB
EGFR	PTGS2	PTGS2	CCL2	GSK3B	PRKACA
VEGFA	IL6	MAPK14	CXCL8	SPP1	EGF
MYC	AKT1	IL6	PPARA	EGF	AR
IL1B	CHUK	PPARG	MMP2	CD40LG	PTEN
CASP3	CXCL8	RELA	HMOX1	MAPK1	CASP8
PPARG	MAPK14	CASP3	MAPK1	NFKBIA	HIF1A

(to be continued)

Continued table 5

Topology	MRWBC	ARWBC	Topology	MRWBC	ARWBC
FOS	IGF2	MAPK8	NFKBIA	E2F1	CXCL8
MMP9	HIF1A	EGFR	RELA	PRKCB	CDK2
PTGS2	PPARG	FOS	AR	TGFB1	MMP9
CAT	VEGFA	GSK3B	CDKN2A	HSP90AB1	PRKCA
HIF1A	EGFR	IL1B	CASP8	CASP8	HSP90AB1
PTEN	IFNG	MYC	CAV1	CCND1	NOS2
EGF	FOS	VEGFA			

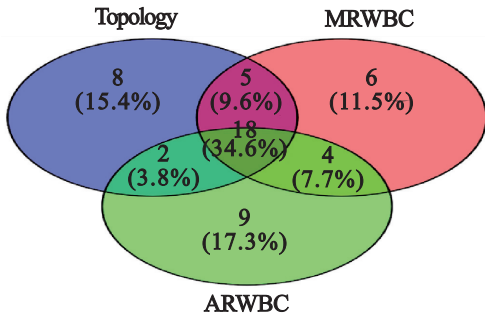


Fig. 3 Intersection analysis of key target proteins in different models

图 3 不同模型中重要靶蛋白交集

2.3 结果验证

2.3.1 文献挖掘验证

通过 TTD 数据库检索发现 ARWBC 筛选出的 9 个差异靶蛋白,其中有 8 个在 TTD 数据库得到验证(表 6)。考虑到后续分子对接可行性,由于构象关系仅留下了 6 个靶蛋白进行文献验证。文献指出(表 7)这些靶蛋白治疗的疾病主要为癌症、免疫疾病、炎症和糖尿病等。其中部分靶蛋白在甘草中有协同作用报道,例如, Yoshida 等^[17]通

过小鼠实验发现甘草甜素(glycyrrhizin, GL)在治疗肝纤维化过程中起着重要作用,其主要机理是通过显著增加抗纤维化细胞因子干扰素(interferon-gamma, IFNG),利用 IFNG 对 CD40LG 的刺激来增强免疫调节,从而实现对肝纤维化的治疗作用。相关文献报道和靶蛋白在数据库记载的统一性验证了筛选出的靶蛋。

Table 6 ARWBC differential target protein

表 6 ARWBC 差异靶蛋白

Entry	Gene symbol	TTD ID
O15111	CHUK	NA
P01344	IGF2	T90572
P01579	IFNG	T94307
P17936	IGFBP3	T33455
P10451	SPP1	T00032
P29965	CD40LG	T14755
Q01094	E2F1	T57059
P05771	PRKCB	T40276
P01137	TGFB1	T97257

Table 7 Literature verification of ARWBC differential target protein

表 7 ARWBC 差异靶蛋白的文献验证

Target	Gene symbol	Treatment of disease	Reference
T90572	IGF2	Cancer	[18 – 20]
T94307	IFNG	COVID-19	[21 – 23]
T33455	E2F1	Dementia	[24]
T14755	CD40LG	Diabetes	[25 – 26]
		Systemic lupus erythematosus	[27 – 28]
		Rheumatoid arthritis	[29 – 30]
T40276	PRKCB	Rheumatoid arthritis	[31]
		Human immunodeficiency	[31]
		central nervous system	[31]
T97257	TGFB1	Rheumatoid arthritis	[32]
		Liver cancer	[32 – 33]

2.3.2 分子对接验证

ARWBC 模型筛选出的重要差异靶蛋白分子信息见表 8, 使用 IGMDOCK 软件对部分符合分子对接条件的靶蛋白 (CD40LG、IFNG、PRKCB、IGF2、E2F1、TGFB1) 进行了分子对接, 并计算了对接能量。对接能量越小表示对接效果越好, 而对接能量小于 $-80 \text{ kcal}\cdot\text{mol}^{-1}$ 则表示具有结合效

应。结果显示, 进行分子对接验证的重要靶蛋白它们的对接结合能主要集中在 $-95 \text{ kcal}\cdot\text{mol}^{-1}$ 附近, 对照组 IL6 结合能量同样集中在 $-95 \text{ kcal}\cdot\text{mol}^{-1}$ 附近, 具体的分子对接结果如表 9 所示, 并通过 R 语言将结合能量进行了可视化分析, 结果见图 4。

Table 8 Molecular information of ARWBC differential target proteins

表 8 ARWBC 差异靶蛋白分子信息

Gene symbol	PDB ID	Determination method	Resolution/ \AA
IGF2	5IEI	X-RAY	2.8
IFNG	7W7Q	X-RAY	1.96
E2F1	7TOA	X-RAY	1.41
CD40LG	1ALY	X-RAY	2.0
PRKCB	2I0E	X-RAY	2.6
TGFB1	5VQF	X-RAY	2.9
CHUK	5EBZ	X-RAY	4.5
SPP1	4AN5	ELECTRON MICROSCOPY	8.8
IGFBP3	7WRQ	ELECTRON MICROSCOPY	3.6

Table 9 Molecular docking energy of ARWBC differential target protein ($\text{kcal}\cdot\text{mol}^{-1}$)

表 9 ARWBC 差异靶蛋白分子对接能量 ($\text{kcal}\cdot\text{mol}^{-1}$)

Component	IL6	CD40LG	IFNG	PRKCB	IGF2	E2F1	TGFB1
Naringenin	-94.03	-97.64	-96.82	-97.76	-100.6	-86.49	-109.6
Quercetin	-99.59	-107.1	-99.28	-112.4	-109.9	-103.5	-116.3
Formononetin	-90.47	-91.38	-104.4	-94.24	-97.46	-93.30	-85.70
Kaempferol	-96.34	-94.40	-93.79	-105.4	-96.86	-100.9	-116.8
7-Methoxy 2-methyl isoflavone	-100.31	-97.93	-82.48	-87.36	-77.88	-82.70	-97.79

Molecular docking heatmap

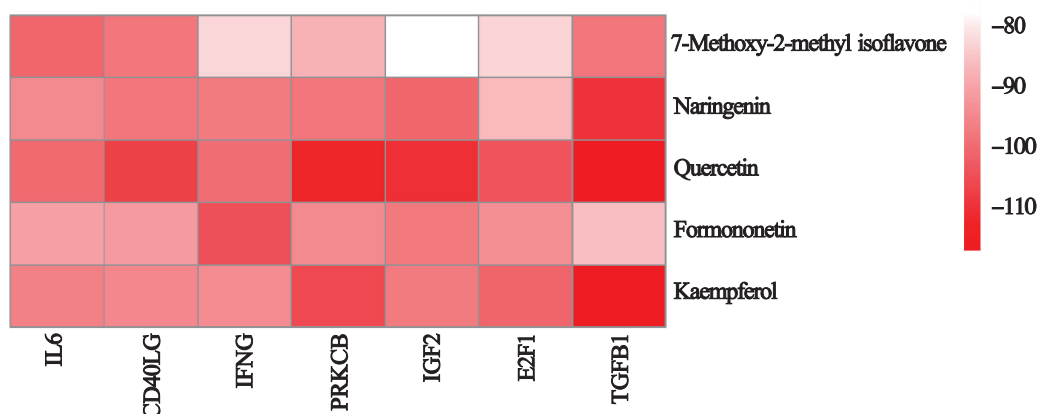


Fig. 4 Heatmap of ARWBC differential target proteins docking with small molecules in the top 5 of the licorice degree value

图 4 ARWBC 差异靶蛋白与甘草度值前 5 的小分子对接热图

3 讨论

在本研究中,通过优化多层随机游走模型,提高了模型对重要靶蛋白的预测能力。相比传统的拓扑模型,ARWBC模型通过对目标网络相关的多维度信息引入更加贴合现实中躯体对药物的吸收受到多个因素影响,同时更大程度的利用了筛选重要靶蛋白时相关异质网络信息,而异步游走策略的选择让模型在筛选重要靶蛋白时发掘出常规拓扑模型易忽略的靶蛋白。

在对甘草重要蛋白进行挖掘验证时,使用了拓扑模型、MRWBC模型和ARWBC模型对预测出的甘草前15%重要靶蛋白进行验证,通过交集处理,筛选出了9个差异靶蛋白。文献挖掘验证结果显示,在TTD数据库中有6个差异靶蛋白与某些疾病相关联。分子对接结果显示,ARWBC筛选出的重要差异靶蛋白和对照靶蛋白IL6对接结合能均在 $-95 \text{ kcal} \cdot \text{mol}^{-1}$ 附近,它们表现出了较好的对接活性,说明这些靶蛋白与甘草小分子成分之间存在较紧密的结合关系,进一步证明优化模型挖掘甘草的重要靶蛋白的可靠性。

模型筛选出的差异靶蛋白为甘草治疗相关疾病提供了证据。其中E2F1结合甘草次酸阻断细胞周期来治疗非小细胞肺癌^[34],同时它还在复合中药PHY906中治疗结肠癌^[35]。CD40LG靶蛋白,其主要作用是刺激T细胞增殖和细胞因子产生,实验发现CD40LG缺乏的个体常表现出中性粒细胞减少、口腔溃疡与直肠溃疡;在“补阳还五汤”中CD40LG具有治疗心肌缺血的作用^[36]。IFNG靶蛋白在一些知名方剂如“黄芪建中汤”和“桂黄清热颗粒”中分别具有治疗口腔溃疡^[37]和鼻粘膜炎症^[38]的作用。同样PRKCB在治疗一些疾病中体现出了它的重要作用,抑制PRKCB的表达可以提高NF- κ B信号通路的活性,调节B淋巴瘤细胞的死亡,从而治疗系统性红斑狼疮^[39]。IGF2作为胰岛素相关靶蛋白,其相关基因片段的缺失被视为结直肠癌的潜在风险标志,其形成的复合体(IMP2/IGF2BP2)能影响治疗迟发糖尿病进程^[40]。挖掘得到的差异靶蛋白在方剂治疗疾病时所展现的机理,部分在甘草治疗疾病中有所体现,如甘草次酸结合E2F1然后阻断细胞周期达到对非小细胞肺癌的治疗。这些差异靶蛋白所体现出的作用对解释甘草治疗结肠癌、直肠溃疡、炎症等相关疾病时涉及到的分子水平复杂作

用机理提供了参考。

由于本研究数据来源为多个在线公开数据库,不同数据库的数据来源、格式和数据更新与同步的延迟等原因,可能会导致数据一致性和样本偏倚的风险。其次用酿酒酵母蛋白验证模型根据研究特点模拟了数据分布,分布特点可能会有更精细化的参数调优选择。最后本研究的靶蛋白是根据算法筛选然后文献验证和分子对接验证,在数据库挖掘中没有获得本研究配体-受体亲和力实验的数据报道,可能需要对这些靶蛋白进行进一步实验验证确保其药理活性。

综上所述,通过优化多层随机游走模型,本研究筛选出了6个重要的甘草靶蛋白,这些候选靶蛋白的转录活性为未来甘草作为方剂的药理学研究提供了有力的理论基础。

参考文献:

- [1] JALILZADEH-AMIN G, NAJARNEZHAD V, ANASSORI E, et al. Antiulcer properties of *Glycyrrhiza glabra* L. extract on experimental models of gastric ulcer in mice [J]. *Iran J Pharm Res*, 2015, 14(4): 1163.
- [2] ADIANTI M, AOKI C, KOMOTO M, et al. Anti-hepatitis C virus compounds obtained from *Glycyrrhiza uralensis* and other *Glycyrrhiza* species [J]. *Microbiology*, 2014, 58(3): 180-187.
- [3] WANG G, HAN Z W. The protective action of *Glycyrrhiza* flavonoids against carbon tetrachloride hepatotoxicity in mice [J]. *Acta Pharmaceutica Sinica (药学报)*, 1993, 28(8): 572-576.
- [4] BATIHA G E S, BESHBIHSHY A M, EL-MLEEHA A, et al. Traditional uses, bioactive chemical constituents, and pharmacological and toxicological activities of *Glycyrrhiza glabra* L. (Fabaceae) [J]. *Biomolecules*, 2020, 10(3): 352.
- [5] NEWMAN M E. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality [J]. *Phys Rev E*, 2001, 64(1): 016132.
- [6] RU J, LI P, WANG J, et al. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines [J]. *J Cheminform*, 2014, 6: 1-6.
- [7] WANG S, WANG H, LU Y. Tianfoshen oral liquid: a CFDA approved clinical traditional Chinese medicine, normalizes major cellular pathways disordered during colorectal carcinogenesis [J]. *Oncotarget*, 2017, 8(9): 14549.
- [8] DAINA A, MICHELIN O, ZOETE V J N A R. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules [J]. *Nucleic Acids Res*, 2019, 47(W1): W357-W364.
- [9] SZKLARCZYK D, FRANCESCHINI A, KUHN M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored [J]. *Nucleic Acids Res*, 2010, 39(suppl_1): D561-D568.
- [10] ZHANG A J, ZHANG Y, QI G P, et al. Study on network

- toxicology of *Pinelliae Rhizoma* based on multilayer network random walk [J]. *Journal of Shenyang Pharmaceutical University*(沈阳药科大学学报),2025,42(12):1120–1127.
- [11] FREEMAN L C. A set of measures of centrality based on betweenness [J]. *Sociometry*,1977,40(1):35–41.
- [12] NEWMAN M E A. A measure of betweenness centrality based on random walks [J]. *Soc Networks*,2005,27(1):39–54.
- [13] ISLAM M A, CHOWDHURY R I. A higher order Markov model for analyzing covariate dependence [J]. *Appl Math Model*,2006,30(6):477–488.
- [14] GIAEVER G, CHU A M, NI L, et al. Functional profiling of the *saccharomyces cerevisiae* genome [J]. *Nature*,2002,418(6896):387–391.
- [15] ZHANG R, LIN Y J N A R. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes [J]. 2009, 37 (suppl_1):D455–D458.
- [16] XENARIOS I, RICE D W, SALWINSKI L, et al. DIP; the database of interacting proteins [J]. *Nucleic Acids Res*,2000,28(1):289–291.
- [17] YOSHIDA T, ABE K, IKEDA T, et al. Inhibitory effect of glycyrrhizin on lipopolysaccharide and d-galactosamine-induced mouse liver injury [J]. *Eur J pharmacol*,2007,576(1/3):136–142.
- [18] WANG K L, YU Y C, HSIA S M. Perspectives on the role of isoliquiritigenin in cancer [J]. *Cancers*,2021,13(1):115.
- [19] WANG Z Y, NIXON D W. Licorice and cancer [J]. *Nutr Cancer*,2001,39(1):1–11.
- [20] KIM M, PARK S C, LEE D Y. Glycyrrhizin as a nitric oxide regulator in cancer chemotherapy [J]. *Cancers*, 2021, 13 (22):5762.
- [21] GAJEWSKI A, KOŚMIDER A, NOWACKA A, et al. Potential of herbal products in prevention and treatment of COVID-19. Literature review [J]. *Biomed Pharmacother*,2021,143:112150.
- [22] BOOZARI M, HOSSEINZADEH H. Natural products for COVID-19 prevention and treatment regarding to previous coronavirus infections and novel studies [J]. *Phytother Res*,2021,35(2):864–876.
- [23] AL-KAMEL H, GRUNDMANN O. Glycyrrhizin as a potential treatment for the novel coronavirus (COVID-19) [J]. *Mini-Rev Med Chem*,2021,21(16):2204–2208.
- [24] ZULFUGAROVA P, ZIVARI-GHADER T, MAHARRAMOVA S, et al. A mechanistic review of pharmacological activities of homeopathic medicine licorice against neural diseases [J]. *Front Neurosci*,2023,17:1148258.
- [25] YANG L, JIANG Y, ZHANG Z, et al. The anti-diabetic activity of licorice, a widely used Chinese herb [J]. *J Ethnopharmacol*,2020,263:113216.
- [26] TAN D, TSENG H H L, ZHONG Z, et al. Glycyrrhizic acid and its derivatives; promising candidates for the management of type 2 diabetes mellitus and its complications [J]. *Int J Mol Sci*,2022,23(19):10988.
- [27] HOFFMANN J, GENDRISCH F, SCHEMPP C M, et al. New herbal biomedicines for the topical treatment of dermatological disorders [J]. *Biomedicines*,2020,8(2):27.
- [28] KOWALSKA A, KALINOWSKA-LIS U. 18 β -Glycyrrhetic acid; its core biological properties and dermatological applications [J]. *Int J Cosm Sci*,2019,41(4):325–331.
- [29] SHARMA T, SHARMA P, CHANDEL P, et al. Circumstantial insights into the potential of traditional Chinese medicinal plants as a therapeutic approach in rheumatoid arthritis [J]. *Curr Pharm Design*,2022,28(26):2140–2149.
- [30] HAN R, REN H C, ZHOU S, et al. Conventional disease-modifying anti-rheumatic drugs combined with Chinese herbal medicines for rheumatoid arthritis: a systematic review and meta-analysis [J]. *J Tradit*,2022,12(5):437–446.
- [31] YANG G, YU Y. Immunopotentiating effect of traditional Chinese drugs-ginsenoside and glycyrrhiza polysaccharide [J]. *Proc Chin Acad Med Sci Peking Union Med Coll*,1990,5(4):188–193.
- [32] WU F, LI X, JIANG B, et al. Glycyrrhetic acid functionalized nanoparticles for drug delivery to liver cancer [J]. *J Biomed Nanotechnol*,2018,14(11):1837–1852.
- [33] STECANELLA L A, BITENCOURT A P, VAZ G R, et al. Glycyrrhizic acid and its hydrolyzed metabolite 18 β -glycyrrhetic acid as specific ligands for targeting nanosystems in the treatment of liver cancer [J]. *Pharmaceutics*,2021,13(11):1792.
- [34] ZHU J, CHEN M, CHEN N, et al. Glycyrrhetic acid induces G1-phase cell cycle arrest in human non-small cell lung cancer cells through endoplasmic reticulum stress pathway [J]. *Int J Oncol*,2015,46(3):981–988.
- [35] SU Z, ZHOU C, QIN S, et al. The significant pathways and genes underlying the colon cancer treatment by the traditional Chinese medicine PHY906 [J]. *Evid-Based Complement*, 2017,2017(1):1–7.
- [36] ZHANG H, WANG W R, LIN R, et al. Buyang Huanwu decoction ameliorates coronary heart disease with Qi deficiency and blood stasis syndrome by reducing CRP and CD40 in rats [J]. *J Ethnopharmacol*,2010,130(1):98–102.
- [37] JIANG K W, MU K Q, QU J M, et al. Analysis of the antitussive mechanism of blood compounds of Qingfei Decoction based on the technology of HPLC-Q/TOF MS and network pharmacology with molecular docking [J]. *Journal of Shenyang Pharmaceutical University*(沈阳药科大学学报),2024,41(12):1546–1561.
- [38] KONG Y, HAO M, CHEN A, et al. SymMap database and TMNP algorithm reveal Huanggui Tongqiao granules for Allergic rhinitis through IFN-mediated neuroimmuno-modulation [J]. *Pharmacolo Res*,2022,185:106483.
- [39] SHENG Y J, GAO J P, LI J, et al. Follow-up study identifies two novel susceptibility loci PRKCB and 8p11.21 for systemic lupus erythematosus [J]. *Rheumatology*,2011,50(4):682–688.
- [40] CHRISTIANSEN J, KOLTE A M, NIELSEN F C. IGF2 mRNA-binding protein 2; biological function and putative role in type 2 diabetes [J]. *J Mol Endocrinol*,2009,43(5):187–195.

Identifying the important target proteins in Licorice based on an optimized Newman random walk model

JIANG Qiang¹, ZHANG Ying², YU Qian¹, DAN Hanyu¹, LI Zuoqing^{1*}

(1. School of Medical Devices, Shenyang Pharmaceutical University, Shenyang 110016, China; 2. National Medical Products Administration, Beijing 100037, China)

Abstract: Objective To identify important target proteins in licorice that are often overlooked by conventional topological analysis methods using an optimized random walk model based on Newman random walk. **Methods** A multi-layer random walk network was constructed to simulate the drug absorption process. By collecting information on licorice components, targets, pathways, and diseases, different transfer probabilities were assigned to different nodes, and asynchronous long walks were conducted within a specified number of steps. The frequency of traversing target proteins was used as a measure of their importance. Top 15% target proteins were identified and compared with the normalized topology and MRWBC models to select the final candidate proteins for validation. **Results** The optimized ARWBC model yielded 9 candidate proteins, of which 6 proteins (CD40LG, IFNG, PRKCB, IGF2, E2F1, and TGF1B) were confirmed through literature review and molecular docking. **Conclusion** The optimized Newman's random walk model effectively uncovers important target proteins in licorice. These candidate proteins with transcriptional activity could serve as a theoretical foundation for future pharmacological studies on licorice formulations.

Key words: random walk; Licorice; target protein; complex network

(上接第 308 页)

The impact of psychoneural factors on atopic dermatitis and the interventional role of traditional Chinese medicine

LI Yaxing¹, HE Mengna¹, ZHANG Huiyuan¹, ZHU Liqin^{2*}, CHEN Yunlong^{1*}

(1. School of Traditional Chinese Pharmacy, Tianjin University of Traditional Chinese Medicine, Tianjin 301600, China; 2. Department of Pharmacy, Tianjin First Central Hospital, Tianjin 300100, China)

Abstract: Atopic dermatitis (atopic dermatitis, AD) is one of the most common chronic inflammatory skin diseases, characterized by skin barrier dysfunction, combining the interaction of multiple links such as microecological imbalance and immune system disorders. In the theory of traditional Chinese medicine, "by observing external symptoms, one can understand internal organ conditions and identify the underlying disease". It indicates that the occurrence of skin diseases is a further manifestation of the dysfunction of internal organs on the skin. AD can extend to the gastrointestinal tract and respiratory system, and itching and pain can lead to sleep disturbances, circadian rhythm changes, and cardiovascular and cerebrovascular complications. At present, AD-related treatment drugs and methods are becoming more and more abundant, but there is still a lack of considering the occurrence and development of the disease from a holistic perspective. This review will systematically summarize psychoneural problems and gut microenvironment as the starting point, and then explore the research progress of traditional Chinese medicine in the treatment of AD.

Key words: atopic dermatitis; psychoneural; gut microenvironment; alternative therapies; traditional Chinese medicine