

· 中药作用机理辅助解析系统及其应用专题 ·

中药作用机理辅助解析系统的构建

张百霞, 骆四君, 颜静, 顾浩, 罗计, 张燕玲, 陶欧, 王耘*

(北京中医药大学 中药信息工程研究中心, 北京 100102)

[摘要] 针对目前基于网络药理学的中药作用机制解析存在对生物过程缺乏详细描述及网络模型构建效率比较低的问题, 构建“中药作用机理辅助解析系统”, 实现中药作用机制生物网络模型的自动建模。该研究以实体语法系统为理论框架, 融合中药方剂、单味药、化学成分、化学成分作用靶点、生化反应、信号通路及疾病相关靶点数据, 建立形式化模型, 编写推理引擎, 构建界面友好、操作简单的中药作用机制辅助解析平台, 可实现中药作用机制网络生物学模型的自动建模。“中药作用机理辅助解析系统”将为中药作用机制网络药理学模型提供自动高效的建模方法, 发挥其辅助解析中药作用机制的功能, 有助于对中药药性理论、配伍理论开展深入研究, 为中药新药的研究和开发提供参考。

[关键词] 实体语法系统; 中药; 作用机制; 辅助解析系统; 网络药理学

Construction of automatic elucidation platform for mechanism of traditional Chinese medicine

ZHANG Bai-xia, LUO Si-jun, YAN Jing, GU Hao, LUO Ji, ZHANG Yan-ling, TAO Ou, WANG Yun*

(Information Engineering Research Center for Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing 100102, China)

[Abstract] Aim at the two problems in the field of traditional Chinese medicine (TCM) mechanism elucidation, one is the lack of detailed biological processes information, next is the low efficient in constructing network models, we constructed an auxiliary elucidation system for the TCM mechanism and realize the automatic establishment of biological network model. This study used the Entity Grammar Systems (EGS) as the theoretical framework, integrated the data of formulae, herbs, chemical components, targets of component, biological reactions, signaling pathways and disease related proteins, established the formal models, wrote the reasoning engine, constructed the auxiliary elucidation system for the TCM mechanism elucidation. The platform provides an automatic modeling method for biological network model of TCM mechanism. It would be benefit to perform the in-depth research on TCM theory of natures and combination and provides the scientific references for R&D of TCM.

[Key words] entity grammar systems; traditional Chinese medicine; mechanism elucidation; auxiliary elucidation platform; network pharmacology

doi:10.4268/cjcm20151901

目前,多靶点、多环节、多途径的综合调节作用已成为中药学领域学者的基本共识^[1-2]。网络药理学是通过建立疾病-基因-靶点-药物的多层次生物网络,从生物网络的角度研究疾病的发生发展过程、认识药物与机体的相互作用^[3-4]。

网络药理学从分子层次认识药物与机体的相互作用,强调对信号通路的多途径调节,而非与疾病相关的个别基因或蛋白的局部关系,以高效解析药物作用机制,对药物进行重定位,进而指导药物的合理使用及二次开发^[5-6]。由于网络药理学

[收稿日期] 2015-06-11

[基金项目] 国家自然科学基金项目(81173568,81373985);教育部新世纪优秀人才支持项目(NECT-11-0605);北京中医药大学基本科研业务费自主选题项目(2015-JYB-XS054)

[通信作者] *王耘,教授,博士生导师, Tel: (010)84738620, E-mail: wangyun@bucm.edu.cn

具有整体、高效的特点,该方法被越来越广泛的应用到中药作用机制解析中^[4,7],而且被认为是颇具潜力的研究方向之一。

但是,目前的生物学网络多是针对蛋白质等单一生物分子类型建立的网络^[8],没有将代谢网络、蛋白质相互作用、细胞信号等信息进行有效整合,对药物作用的体内生物过程缺乏细致有效的描述;网络模型的建立多靠手工完成,无法实现批量处理,效率比较低,即缺乏一种高效、灵活的信息融合方法。本文融合中医药及生物学领域的的数据,以实体语法系统为理论框架,利用定性推理技术构建“中药作用机制辅助解析系统”平台并形成面向用户的平台,设计思路见图1,以实现中药作用机制的辅助解析。

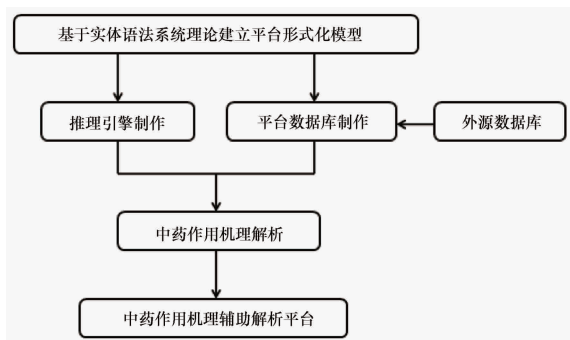


图1 中药作用机制辅助解析系统设计思路

Fig. 1 The design thought of auxiliary elucidation system for TCM mechanism

1 数据来源

对中药作用的体内生物过程进行细致的描述,需要整合中药、生化反应和疾病等多方面的信息。以下即为构建“中药作用机制辅助解析系统”平台所需数据的来源。

1.1 中药化学成分数据 中药化学成分数据来源于 TCMD 数据库^[9]、中药化学成分分析手册^[10]、TCMID (<http://www.megabionet.org/tcmid/>),综合上述数据库并通过查阅文献进行补充。

1.2 中药化学成分作用靶点数据 中药化学成分作用靶点数据来源于 Stitch 数据库^[11] (<http://stitch.embl.de/>)、文献报道的中药化学成分与蛋白质相互作用数据以及本实验室通过数据挖掘、虚拟筛选预测并得到验证的数据。

1.3 生物学反应及信号通路数据 生物学反应及信号通路数据来源于 Reactome^[12] 数据库 (<http://www.reactome.org/>)、hiPathDB^[13] 数据库 (<http://hiPathDB.kobic.re.kr>)、KEGG^[14] 数据库 (<http://www.genome.jp/kegg/>)及 STRING^[15] 数据库 (<http://string-db.org/>)。

1.4 疾病相关靶点数据库 疾病相关靶点数据来源于 TTD 数据库^[16-17] (<http://bidd.nus.edu.sg/group/cjttd/>)

TTD. asp.)、SciClips 数据库^[18] (<http://www.sciclips.com>)及 GeneCards 数据库^[19] (<http://www.genecards.org/>)。

2 方法原理

2.1 实体语法系统 实体语法系统^[20]是一种形式化语法系统,可用于形式化地描述复杂系统的组成单位、组织方式、变化规律。本研究引入实体语法系统,建立中药与人体组成的复杂系统的形式化模型。同时,利用实体语法系统作为中药作用机制辅助解析系统构建过程中各类信息融合的理论框架,在建立基于实体语法系统的形式化模型后,对各部分信息进行整理归类,统一到实体语法系统形式化模型中,结合各部分结构建立平台基本架构。

实体语法系统(entity grammar systems, EGS)是为生物复杂系统建模而提出的一种形式语法系统,源于 Chomsky (<http://www.chomsky.info/>)生成语法系统,是对 Chomsky 生成语法系统的进一步拓展。由于其高效灵活的特点,本课题组已将实体语法系统应用于中药作用机制解析^[21]及中药有效成分群辨识^[22]中。

实体语法系统用五元组 $G = (V_N, V_T, F, P, S)$ 表示,其中各项分别为非末端字符集、末端字符集、操作子集、规则集和初始字符。相比与 Chomsky 生成语法系统,实体语法系统增加了操作子集 F 。操作子集 F 中的每一个元素表示字符的一种组织方式,字符采用这些组织方式所组成的新的单元 F 和 V 统称为实体。实体既可以是具体的物体,也可以是抽象的概念。当不考虑或者不区分末端字符与非末端字符时,用四元组 $G = (V, F, P, S)$ 表示,其中 $V = V_N \cup V_T$ 。

实体语法系统包含结构数学中公理化方法的一般结构。公理体系由 3 部分构成:基本概念(基本对象及基本关系)、公理组、定理及证明。基本概念和公理组构成的公理系统是公理体系的基础部分。公理体系的这种构成与实体语法系统的构成具有一定的对应关系。实体语法系统 $G = (V, F, P, S)$ 中 V 对应公理体系中的基本对象, F 对应公理体系中的基本关系,两者合起来对应公理体系的基本概念。 P 则对应公理组,是公理体系中能够用于演绎的基本规则。 S 对应于利用公理体系进行推导和证明时的初始状态,它可以是作为出发点的基本公理,也可以是作为出发点的基本概念,或者是由基本概念衍生出的具体对象。利用规则 P 从 S 开始的推导或证明过程对应于公理体系中的证明和演绎过程,所得结果对应于经过证明的定理^[23]。

2.2 基于实体语法系统的形式化模型的建立 在本研究中 V 主要包含网络中各种类型的节点,如中药方剂,单味药,中药化学成分及其作用靶点,疾病及其相关蛋白,生物过程中涉及到的其他小分子、蛋白质和大分子复合物。 F 包含网络中的各种节点关系类型,如包含、促进、抑制等关系以及推理过程中形成的各类节点组织方式。 P 为基于上述实体推断中药与疾病相关蛋白作用关系的推理引擎。 S 表示系统初始状态,即推理的起点。用于中药作用机制解析的 V, F, P, S ,

其具体描述如下。

$$V = V_1 \cup V_2 \cup V_3 \cup V_4 \cup V_5 \cup V_6 \cup V_7$$

其中 V_1 代表中药方剂, V_2 代表单味药, V_3 代表单味药中的化学成分, V_4 代表化学成分作用的靶点, V_5 代表疾病, V_6 代表疾病相关蛋白, V_7 代表生物网络中涉及到的其他蛋白和分子。

$$F = \{ \text{start}(X, Y), \text{include}(X, Y), \text{midinclude}(X, Y), \text{link}(X, Y, Z), \text{draw}(X, Y, Z, W), \text{pict}(X, Y, Z, W), \text{distance}(X, Y, Z, W), \text{forward}(X, Y, Z, W), \text{backward}(X, Y, Z, W), \text{dis}(X, Y) \}$$

上述 F 中包含的基本关系类型为初始条件中包含的以及在推理过程中产生的全部节点关系。 $\text{start}(X, Y)$ 表示中药方剂 X 中包含中药 Y , $\text{midinclude}(X, Y)$ 表示中药 X 中包含化学成分 Y ; $\text{include}(X, Y)$ 表示中药方剂 X 中包含的化学成分 Y , $\text{link}(X, Y, Z)$ 表示生物分子 X 与生物分子 Y 间的作用关系为 Z (pos 为促进作用, neg 为抑制作用); $\text{draw}(X, Y, Z, W)$ 表示在推理过程中产生的节点 X 和 Y 间的关系; $\text{pict}(X, Y, Z, W)$ 表示推理的起点 X 到终点 Y 的作用关系和距离 (化学反应的步数); $\text{distance}(X, Y, Z, W)$ 表示网络中的节点 X 到终点 Y 的最短作用距离为 W ; $\text{dis}(X, Y)$ 表示与疾病 X 相关的蛋白为 Y 。在不同的节点关系中变量 X, Y, Z, W 可代表不同的含义, 但在相同的节点关系中若变量所在的位置相同, 虽然表示变量的字母不同, 但这些字母代表相同的含义。

$$P = P_1 \cup P_2 \cup P_3 \cup P_4 \cup P_5 \cup P_6 \cup P_7 \cup P_8 \cup P_9 \cup P_{10} \cup P_{11} \cup P_{12} \cup P_{13}$$

$$P_1 = \{ \text{include}(B, D) : \text{-start}(B, C), \text{midinclude}(C, D) \}$$

$$P_2 = \{ \text{draw}(B, C, R, 1) : \text{-include}(B, D), \text{link}(D, C, R) \}$$

$$P_3 = \{ \text{draw}(C, E, \text{pos}, M) : \text{-draw}(C, R, \text{pos}, N), \text{link}(R, E, \text{pos}), M = N + 1, N < 10 \}$$

$$P_4 = \{ \text{draw}(C, E, \text{pos}, M) : \text{-draw}(C, R, \text{neg}, N), \text{link}(R, E, \text{neg}), M = N + 1, N < 10 \}$$

$$P_5 = \{ \text{draw}(C, E, \text{neg}, M) : \text{-draw}(C, R, \text{pos}, N), \text{link}(R, E, \text{neg}), M = N + 1, N < 10 \}$$

$$P_6 = \{ \text{draw}(C, E, \text{neg}, M) : \text{-draw}(C, R, \text{neg}, N), \text{link}(R, E, \text{pos}), M = N + 1, N < 10 \}$$

$$P_7 = \{ \text{pict}(C, A, F, X) : \text{-draw}(C, E, F, X), \text{dis}(A, D), E = D \}$$

$$P_8 = \{ \text{distance}(A, B, D, Y) : \text{-\#min}\{X: \text{pict}(A, B, _, X)\} = Y, \text{pict}(A, B, D, Y), A! = B \}$$

$$P_9 = \{ \text{length}(Y) : \text{-distance}(C, B, X, Y, _) \}$$

$$P_{10} = \{ \text{backward}(A, B, E, 1) : \text{-link}(A, B, E), \text{distance}(_, B, _, _, _) \}$$

$$P_{11} = \{ \text{backward}(A, B, F, M) : \text{-link}(A, B, F), \text{backward}(B, C, E, N), M = N + 1, N < Y, \text{length}(Y) \}$$

$$P_{12} = \{ \text{forward}(C, D, X, 1) : \text{-distance}(C, _, _, _, _), \text{backward}(C, D, X, Y) \}$$

$$P_{13} = \{ \text{forward}(D, E, F, M) : \text{-forward}(C, D, X, N), \text{backward}(D, E, F, G), M = N + 1, N = Y - G, \text{length}(Y) \}$$

规则 P_1 用来推理中药方剂和化学成分的包含关系; 规则 $P_2 \sim P_7$ 用来推理中药化学成分到疾病相关蛋白的所有路径的距离, 其中 $N < 10$ 表示化学成分到疾病相关蛋白的距离小于 10, 在此可根据具体研究设置不同的数值; 规则 P_3 中 C 到 R 的作用方式为 pos , R 到 E 的作用方式为 pos , 则 C 到 E 的作用方式为 pos , 同理可了解 $P_4 \sim P_6$ 中包含的另外 3 种作用方式的推理; 一个化学成分作用于某个疾病相关蛋白可能经过多条路径, 规则 P_8 用来计算同一个化学成分到同一个疾病相关蛋白的最短距离; 规则 $P_9 \sim P_{13}$ 用来描述指定路径距离的化学成分到疾病相关蛋白经过的具体路径。

$$S = S_1 \cup S_2 \cup S_3$$

S_1 表示要研究的具体的中药方剂、中药或化学成分, 作为推理的起点; S_2 表示要研究的具体的疾病, 作为推理的终点; S_3 表示所有的生物分子形成的实体。

2.3 “中药作用机理辅助解析系统”的构建 通过对实体语法系统形式化模型的分析, “中药作用机理辅助解析系统”的构建需要以下 6 部分的内容: ① 中药化学成分数据; ② 中药化学成分作用靶点数据; ③ 生物分子相互作用数据; ④ 疾病及相关靶点数据; ⑤ 推理引擎; ⑥ 用户界面。“中药作用机理辅助解析系统”基本构架见图 2。为了保持系统的灵活性, 并未将方剂数据库包含在内, 用户可以根据方剂的具体药味组成, 分别进行解析, 然后通过整合结果, 实现对方剂的作用机制进行分析。为了以上各部分既紧密联系又保持相对独立, 可以单独更新, 为平台的升级和扩展提供方便。

3 结果

为使用户更方便地使用系统, 进一步建立了面向用户的可视化界面。在建立实体语法系统形式化模型的基础上, 将数据库格式与推理规则代码统一并优化, 让推理引擎高效执行。输入与输出界面则整合到用户界面上, 用户只需在前台输入中药的名称, 后台进行运算推理, 输出结果再反馈给用户, 形成一个简洁高效的平台, 解析结果保存为 CSV 格式的文件, 可用 Cytoscape, Gephi, GraphViz 等可视化软件进行更详细的分析, 系统界面见图 3。

利用该系统得到的大黄降血脂作用机制网络在 Cytoscape 中的可视化结果见图 4。图中红色节点为大黄的代谢产物, 玫红色节点为高血脂相关靶点, 紫色为小分子, 橙色为蛋白或蛋白复合物, 绿色节点为生化反应的名称。从图中可以分析每个化学成分的直接作用蛋白及其作用效应的生物学传递路径, 由此可以分析大黄可能影响到的靶点与疾病, 可以用于药物生物效应及其潜在副作用的分析。

4 讨论

以网络药理学和中医药理论为指导, 针对目前基于网络药理学的中药作用机制解析存在对生物过程缺乏详细描述

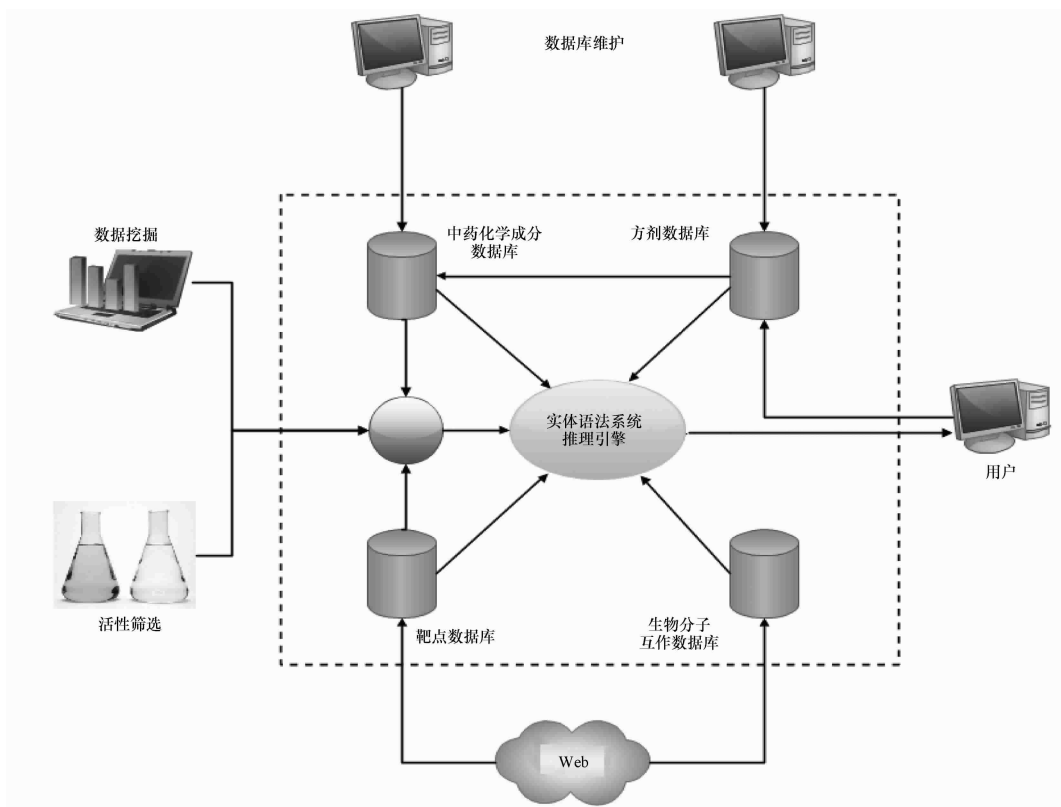


图2 中药作用机理辅助解析平台基本架构

Fig. 2 The fundamental framework of auxiliary elucidation platform for the action mechanism of TCM



图3 平台主页

Fig. 3 The homepage of platform

及网络模型构建效率比较低的问题,本研究以实体语法系统为理论框架,建立形式化模型并构建了“中药作用机理辅助

解析系统”平台,实现了中药作用机制网络生物学模型的辅助构建。本研究所建立的此平台是对中药作用机制解析研

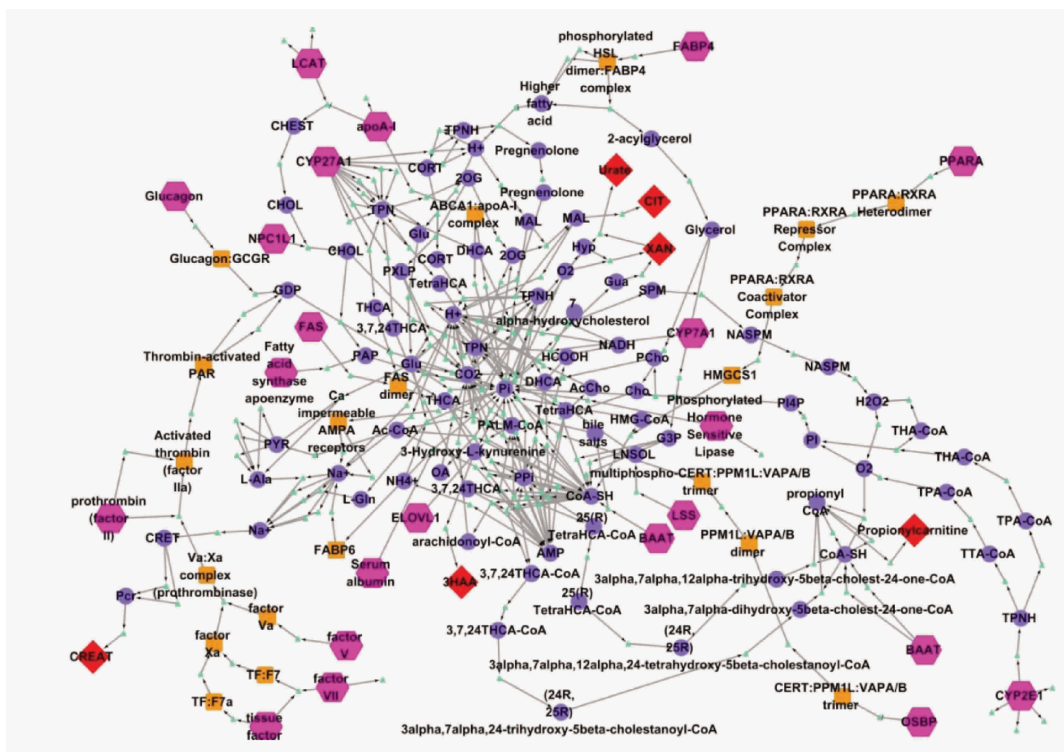


图4 药用大黄代谢产物降血脂作用机理的网络模型

Fig. 4 The network model of antihyperlipidemic mechanism of rhubarb metabolite

究模式的尝试与探索。在中药作用机制研究中,引入实体语法系统,实现对中药方剂、单味药、中药化学成分、疾病靶点、生物分子网络的统一描述,使中药作用机制网络生物学模型的自动建模成为可能。以数据库和逻辑推理引擎为组件,搭建高度集成的系统,探索性地开展了以系统为核心的中药作用机制解析模式,高效输出中药作用机制网络生物学模型。

该系统不仅可以对中药方剂、单味药、单个化学成分进行作用机制解析,也可以对西药或西药复方进行作用机制的解析,还可以按用户需求对多味中药或多个成分自行组方和组分配伍进行作用机制解析,构建新复方药物的分子作用网络模型,为新复方药物的研究和开发提供依据。将平台输出的结果绘制成生物网络,使中药药性、配伍等理论的研究转换为对生物网络的研究,从而进行更深层次的挖掘,在分子层面和生物网络层面对中医药理论进行探讨。

由于不同数据库收录完整度不同,存在部分成分无靶点输出、部分靶点无生物网络输出的现象。相信随着中医药及生物学领域数据库的不断充实,对推理引擎的不断完善,中药作用机制辅助解析平台能更好地应用于中药网络药理学研究。

【参考文献】

[1] 郝海平,郑超满,王广基.多组分、多靶点中药整体药代动力学研究的思考与探索[J].药学学报,2009,44(3):270.

- [2] 杨义芳,萧伟.基于多成分多靶点的中药药动/药效相关性研究解读与策略[J].中草药,2013,44(12):1521.
- [3] Andrew H. Network pharmacology: the next paradigm in drug discovery[J]. Nat Chem Biol, 2008, 4: 682.
- [4] Li X, Wu L, Jin Y, et al. A network pharmacology study of Chinese medicine QiShenYiQi to reveal its underlying multi-compound, multi-target, multi-pathway mode of action [J]. PLoS ONE, 2014, 9(5): e95004.
- [5] Yu S, Guo Z, Guan Y, et al. Combining ZHENG theory and high-throughput expression data to predict new effects of Chinese herbal formulae [J]. Evid Based Complement Alternat Med, 2012,doi:10.1155/2012/986427.
- [6] Tao W, Xu X, Wang X, et al., Network pharmacology-based prediction of the active ingredients and potential targets of Chinese herbal Radix Curcumae formula for application to cardiovascular disease[J]. J Ethnopharmacol, 2013, 145(1): 1.
- [7] 吴红钉,丘小惠,朱伟,等.网络药理学方法探讨清热中药治疗冠心病作用机制[J].中华中医药杂志,2011,26(5):1004.
- [8] Nandy S K, Jouhten P, Nielsen J. Reconstruction of the yeast protein-protein interaction network involved in nutrient sensing and global metabolic regulation[J]. BMC Syst Biol, 2009, 18: 3.
- [9] Yan X J, Xie G R. Molecular structure database of traditional Chinese drugs and primary applications [J]. Comput Appl Chem, 1998,15(5):277.

- [10] 常新全,丁丽霞. 中药化学成分分析手册[M]. 北京:学苑出版社, 2002:16.
- [11] Michael K, Christian von M, Monica C, et al. STITCH: interaction networks of chemicals and proteins[J]. *Nucleic Acids Res*, 2008, 36: 684.
- [12] Vastrik I, DEustachio P, Schmidt E, et al. Reactome: a knowledge base of biologic pathways and processes[J]. *Genome Biol*, 2007, 8(3): 39.
- [13] Yu N, Seo J, Rho K, et al. hiPathDB: a human-integrated pathway database with facile visualization[J]. *Nucleic Acids Res*, 2012, 40: 797.
- [14] Minoru K, Susumu Goto. KEGG: kyoto encyclopedia of genes and genomes[J]. *Nucleic Acids Res*, 2000, 28(1): 27.
- [15] Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life[J]. *Nucleic Acids Res*, 2015, 43: 447.
- [16] Chen X, Ji Z L, Chen Y Z. TTD: therapeutic target database [J]. *Nucleic Acids Res*, 2002, 30(1): 412.
- [17] Zhu F, Han B, Kumar P, et al. Update of TTD: therapeutic target database[J]. *Nucleic Acids Res*, 2010, 38: 787.
- [18] Payne T. New comprehensive oncology biomarker database launched by SciClips[J]. *Epigenomics*, 2012,4 (1): 10.
- [19] Safran M, Dalah I, Alexander J, et al. GeneCards Version 3: the human gene integrator[J]. *Database (Oxford)*, 2010, doi: 10.1093/database/baq020.
- [20] Yun W. Entity grammar system: a grammatical tool for studying the hierarchal structures of biological systems[J]. *Bull Mathem Biol*, 2004, 66(3): 447.
- [21] Jing Y, Yun W, Sijun L, et al. TCM grammar systems: an approach to aid the interpretation of the molecular interactions in Chinese herbal medicine[J]. *J Ethnopharmacol*, 2011, 137(1): 77.
- [22] Luo J, Ren Y L, Gu H, et al. dTGS: method for effective components identification from traditional Chinese medicine formula and mechanism analysis [J]. *Evid Based Complement Alternat Med*, 2013,doi:10.1155/2013/840427.
- [23] 王耘,乔延江. 实体语法系统与中医药理论现代化[J]. *世界科学技术——中医药现代化*,2004, 6(5):6.

[责任编辑 马超一]