

发现·辨识·优化 ——中药新药设计的核心与关键*

□杨洪军** (中国中医科学院中药研究所 北京 100700)

雷 燕 (中国中医科学院 北京 100700)

唐仕欢 (中国中医科学院中药研究所 北京 100700)

陈建新 (北京中医药大学 北京 100029)

许海玉 李韶菁 陈 畅 黄璐琦** (中国中医科学院中药研究所 北京 100700)

摘 要:中药新药设计在中药新药研发中占有重要地位,对提升新药创制水平具有重大意义。本文围绕“组效关系”,提出了中药新药设计的概念、研究框架及技术方法,将处方发现、成分辨识和组方优化作为中药新药设计的核心与关键,以期对中药新药创制提供理论和技术支撑。

关键词:中药新药设计 组效关系 处方发现 成分辨识 组方优化 系统建模

doi: 10.3969/j.issn.1674-3849.2011.01.030

新药的创制是一项系统工程,无论中药还是西药,均包括研究与开发两个阶段。候选药物的确定是区分两个阶段的标志,即在确定候选药物之前为研究阶段,确定之后的工作为开发阶段。所谓候选药物是指拟进行系统的临床前试验并有可能进入临床研究的化学实体(西药)或中药提取物(中药)。研究阶段在新药创制流程中占有重要地位,能否找到合适的候选药物,则决定着该药后期开发的风险和最终

是否能成为上市新药。

就西药而言,为了寻求候选药物,提出了以“构效关系”为核心的药物设计(Drug Design),其研究的内容是药物发现的中心环节——先导物的发现途径(衍生与优化)以及所涉及的理论、技术和方法。计算机辅助药物设计(Computer-Aided Drug Design, CADD)在先导化合物的发现和优化方面,已取得了极大的成功^[1]。

单一化合物是基于“构效关系”进行设计的,与之相对应,多成分的复方中药除了“构效关系”外,更

收稿日期: 2011-01-24

修回日期: 2011-02-10

* 国家科技支撑计划项目(2008BAI51B03):基于有效成分群的中药组方设计技术研究,子课题负责人:杨洪军;国家重大新药创制专项(2009ZX09301-005):综合性中药新药研究开发技术大平台,负责人:曹洪欣;中国中医科学院基本科研业务费自主选题项目(Z02063):确有疗效中药、方剂的活性物质基础及其组效关系研究,负责人:杨洪军。

** 通讯作者:杨洪军,研究员,硕士研究生导师,主要研究方向:系统建模与中药组方优化研究,Tel:010-64014411-2948,E-mail:hongjun0420@vip.sina.com;黄璐琦,本刊编委,研究员,博士生导师,中国中医科学院副院长,中药研究所所长,主要研究方向:中药资源与分子生药学研究,Tel:010-64014411-2955,E-mail:huangluqi@263.net。

多侧重基于“组效关系”进行研究设计。所谓中药“组效关系”(Combination-Activity Relationship, CAR)是指在不同层次上的中药物质组合与药效活性之间的关联性,系统建模是分析“组效关系”的基本方法。为此,我们认为中药新药设计(Design of New Drug of Chinese Medicine, DNDCM)以“组效关系”为核心,围绕饮片层次上的中药新药处方发现、主要有效成分的辨识和提取物(组分或成分层次)的组方优化 3 个关键环节,以确定高效低毒候选药物为目标,中药新药设计流程,见图 1。

综上,就西药而言,研究阶段主要包括:靶标的确定、活性筛选发现先导化合物和先导化合物结构优化等 3 个环节,其中以上环节中先导化合物的发现和优化为新药设计,以“构效关系”为核心;而中药新药研究阶段则包括:中药处方发现、有效成分辨识和组方配比优化 3 个环节,我们将以上 3 个环节称为中药新药设计,以“组效关系”为核心。可见,“组效关系”与“构效关系”是中西药新药研究的主要区别。中药与西药新药研究流程对比,见图 2。

一、中药新药处方发现

中药新药研发过程与化学药、生物技术药从实验室到临床的发现模式不同,以人体临床试验结果

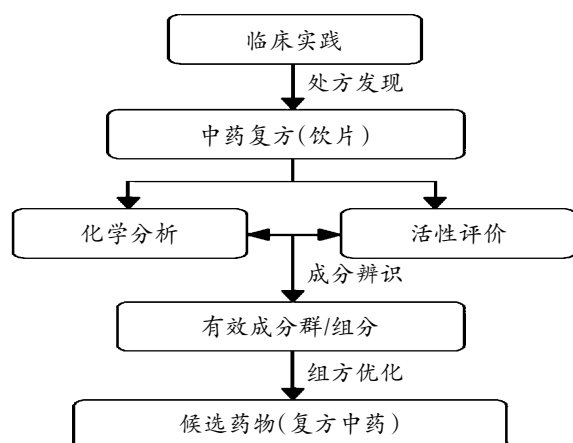


图 1 中药新药设计流程图

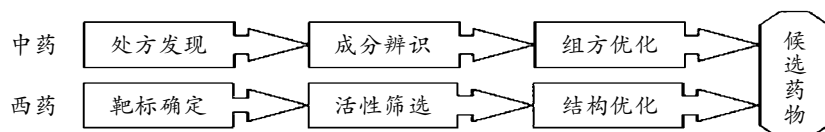


图 2 中药与西药研究对比流程图

为基础进行新药研发是中药显著特点。历代中医经过长期临床实践,积累了大量有效方剂和用药经验,以图书文献和在世的名医为载体得以保存。采用数据挖掘技术,对数据库中名老中医医案、验方、古代方剂、民族药等数据的挖掘,从数据中寻找组方规律,以数据挖掘和分析计算所确定的组方为基础,结合专家经验判断,再进行有针对性地处方筛选将提高命中率。这种新药处方发现方法,突破了以往完全依赖于实验评价和个人经验的局限,创新了现代中药处方发现方法。总之,处方发现是指以名医的个体经验和群体经验为基础,采用数据挖掘的方法,寻找用于新药研究的中药饮片处方。

数据挖掘方法有许多种,有监督解决的主要问题是自变量与因变量之间的关系,常用的有决策树、神经网络、支持向量机和贝叶斯等;无监督解决的主要问题是发现隐藏在自变量数据中的模式,常用的是聚类分析、主成份分析、熵聚类方法、隐变量、自组织神经网络映射法等。

处方发现的技术主要涉及到数据挖掘技术中的无监督技术,它们能在没有人为干预的情况下,根据变量的特点从处方数据中挖掘出隐含在其中的核心组合。鉴于方剂数据具有离散性强、因素多、样本少、高度非线性等特点,针对这些数据类型,目前在无监督方法中的复杂系统熵聚类、无监督随机神经网络、K 均值聚类 3 种技术能够高效、准确地从数据中提取出核心组合,实现处方发现。

1. 复杂系统熵聚类技术

复杂系统熵聚类的原理是把方剂数据看成是一个复杂系统,而构成此系统的因素就是所有的药物。这些因素(药物)都是变量,不同的方剂有不同的组合,而且在各个方剂中扮演的角色也不一样(如君、臣、佐、使)。在基于 Shannon 熵定义关联度的基础上,形成改进关联度系数法^[2],定量描述药物之间的关联度;为了有效进行发现方剂中的核心组合,提出了

基于“亲友团”的熵聚类算法^[3-4],可以快速形成强相关的多个类;在核心组合的基础上,进一步采用无监督的熵层次聚类^[5]实现新处方发现。基于复杂系统熵聚类技术,我们先后应用于流感、肺癆等疾病的处方分析^[6-7],经过专家判读,认为该方法可以作为分析中药组方规律的有

效辅助工具。

2. 无监督随机神经网络

无监督随机神经网络^[8]是把处方数据看成是相互作用的随机网络。各个药物之间的关联不是固定的,而是服从一定分布的。我们通过对处方数据编码,再输入到随机神经网络中,网络根据内部的运行,会得到每个变量下各个等级的激活概率,然后根据激活概率得到每个药物的强相关药物。在此基础上用聚类算法得到各个类。

3. K 均值聚类技术^[9]

K 均值聚类技术是把处方数据看成是欧式空间中的点,它首先从 n 个数据对象任意选择 k 个对象作为初始聚类中心,而对于所剩下的其它对象,则根据它们与这些聚类中心的相似度,即距离,分别将它们要配与其最相似的(聚类中心所代表)聚类,然后再计算每个所获新聚类中心(该聚类中所在对象的均值),不断重复这一过程,直到标准测度函数开始收敛为止。

综上所述,3 种数据挖掘技术均适用于离散性强、因素多、样本少、高度非线性的方剂数据。三者之间各有优势,其中复杂系统熵聚类技术的优点在于:①不但能实现聚类,而且能够实现同一个药物在不同的类里面出现,符合处方发现的特点;②快速、高效地聚类,不但能从大量的数据中聚类,而且对于小样本也可以适用;③适合于非线性、多因素和高度离散的数据。与之相比,无监督随机神经网络更适用于高度非线性的数据。K 均值聚类技术则能快速、高效地把变量空间分成最优的几类,并可以进行层次聚类。在实际应用过程中,根据数据特点,可以选择以上相应的方法进行分析,实现核心组合的提取和新处方的发现。

二、中药有效成分辨识

中药复方的药效物质基础是一个比较笼统的概念,从“宏观到微观”或者“自上而下”,可以分为 4 个层次:有效药材(饮片)、有效部位、有效组分(群)、有效成分(群)。虽然各个层次之间不能完全等同,但是宏观层次的药效活性,必然是微观层次各种化学成分交互作用“涌现”的结果,因此,从根本上讲,饮片层次上的临床疗效必然是有效成分综合作用的结果。

面对通过临床实践及数据挖掘技术所确定中药

新药研发的处方,如何辨识有效成分是中药新药研究不可避免的关键科学问题之一。就目前状况而言,从中药中发现高活性单体化合物的难度越来越大,而且这种研究方法与中医理论严重脱节,无法回归到中医。中医历来强调药物之间的配伍,“七情合和”是中医组方的基本原则,认为“药有个性之特长,方有合群之妙用”。即使是单味药,治疗某一病症时,也往往多成分综合效应。如银杏叶国际通用标准提取物(EGb761)的成功开发等大量的中药研究实践告诉我们,片面追求中药及其复方中单一化学成分的研究之路,越来越窄;重视多种成分的综合效应,探讨多成分组合的药理效应规律,是符合中药自身作用特点的研究方法。

同时,现代中药复方新药研发的经验告诉我们,复方中的成分虽多,但是真正的药效成分却有限,而且往往是药材中含量比较高的主成分。当然,我们不排除含量过低成分的药效活性和在复方中的作用,但是在现阶段完全了解清楚复方中每个成分及其组成比例与活性的量效关系,既不现实也太理想化。所以在现实情况下,可以考虑选取有限的具有生物活性的多种单体成分,基于主要药效指标探讨主成分(或活性明确的部分微量成分)的组合规律,确定关键药效成分和辅助药效成分,在成分组合效应清晰的前提下,采用成分直接组方进行新药研发,或遵循成分最优配比,在组分层次上进行新药研发,这不失为一种可行的创制现代中药的研究模式。

各种新兴技术的发展为中药复方药效物质基础的研究提供了有力工具,如膜分离、血清药理学、血清药物化学、药动学、生物膜色谱、Caco-2 细胞、指纹图谱、计算机学和高通量筛选等技术的发展以及天然组合化学库和中药多靶点作用机制、系统生物学、化学物质组学等理论体系的发展,使药效物质基础研究取得了长足进步^[10],为中药有效成分辨识提供了技术支撑。

基于以上研究成果,我们提出了“基于系统建模和实验研究相结合的中药有效成分辨识”策略。该策略的核心思想“中药化学成分-入血成分-活性评价-建模分析-有效成分”,是充分分析中药的化学成分(中药物质组)和入血成分(血清药理学化学),然后进行多指标组合的药效学研究,再通过系统建模建立中药多成分组合与药效活性的关联性,辨识有效成分。该研究思路在川芎挥发油舒张血管有效成分辨

识中,获得成功应用。选取不同产地的川芎,提取挥发油,运用 GC-MS 进行化学成分分析,同时,检测舒张血管活性,运用 LARS 和 LASSO 回归算法的数据挖掘方法,构建川芎挥发油化学成分与血管活性之间的数学模型,在此基础上辨识出 13 个与血管活性相关的化学成分^[11]。

三、中药新药组方优化

国家重点基础研究发展计划(“973”计划)项目《方剂关键科学问题的基础研究》提出了“组分配伍”研制现代中药的新模式,证实从饮片配伍过渡到组分配伍是完全可行的^[12]。组分中药具有物质基础和作用机理基本清楚,质量可控性好,疗效确切,便于服用与携带等特点,是中药研发的重要方向。在有效成分辨识清楚的基础上,通过中药资源及药物经济学的综合评判,既可采取成分直接组方优化,亦可采用所辨识的有效成分为质量控制指标的有效组分进行组方优化。

所谓的组方优化是以有效成分或标准组分为对象,通过系统建模,分析中药物质(成分或组分)组合与药效活性的关联性,在成分或组分的类别及配比上,寻求最优组合。可见,组方优化仍然以“组效关系”为核心。

在优化效能上取决于试验设计,在超出原有试验设计范围,进行最优指标值和条件的预测计算缺乏进一步优化能力,因此难以求得最优。另外,对于中药面对“多指标、非线性、离散性和小样本”的数据分析问题,其精度差,难以准确地得出各因素与目标的关系。因此,迫切需要适合中药数据特点的数据处理方法。

1. 多指标综合特征提取技术

在新药研发过程中,无论药物的适应症是单病种,还是多病种以及和证候的结合,在进行药效评价时,无一例外均需要选择多种动物模型,而且观测指标大部分不是单一的。针对这种情况,每一种模型优化的组分配比未必一致,若同时对这些指标进行优化,很有可能得到最优配比是空集。

为此,我们提出用主成分分析和基于熵的广义指标方法分别对多模型的药效指标进行综合特征指标的提取,提取出代表每一种模型的综合特征指标,再用系统建模的方法对这些目标进行优化,得到最优配比。

2. 基于系统建模的多目标综合优化技术

根据“多指标、非线性、离散性和小样本”的数据特点,我们认为基于熵的神经网络多目标优化技术、基于模型选择的回归优化技术和进化计算多目标优化技术,是较为可行的优化技术。

回归方法优化技术是一种模型选择算法,它通过均匀设计得到的实验结果建立成分组合与多个药效指标之间确切的方程,然后通过约束求解得到多个药效指标的最大值,最大值对应的成分组合之间的配比即为最优配比。我们改进了 LARS 和 LASSO 方法,使得其能刻画成分组合与药效之间的非线性关联,并使用相关的优化方法提高药效。以川芎的 3 个主要化学成分(藁本内酯、丁基苯酞、阿魏酸)为范例,依据均匀设计,通过实验得到各个配比对应的血管舒张活性,在改进的两种方法下建立成分与药效之间的非线性关联,并使用枚举法提供活性,再通过实验验证的方法证实了这种关联,使得活性达到了 90% 以上,远远超过均匀设计过程中最大的活性(60%)^[5]。

神经网络优化技术通过网络多个神经元权重的线性组合刻画有效成分群与多个药效活性指标的高度非线性关系,并在此基础上通过对目标函数进行优化得到最优配比。

进行组方优化,建立成分组合与药效活性相关联的精确数学模型是主要的技术难点。为此,我们提出“均匀设计-药效试验-系统建模-药效验证-完善模型”的研究策略,尤其采用基于复杂系统的多目标优化技术,可以充分保证模型的精确性,进而有效揭示有效成分群与多个药效指标的关联规律。

四、结 语

“单靶点、高亲和”的现代创新药物研究模式已经表现出局限性,“多靶点、低亲和”的药物研究方法日趋受到重视。国际上,提出了网络药理学的概念,在重大疾病治疗上,提出“鸡尾酒”疗法、固定剂量复合制剂等,这些都是新药研发模式的重大转变,与中药的用药理念日趋一致。但是,鉴于中西医药毕竟是两个不同的理论体系,如何提高中药新药的创制能力是中医药发展的重大科学问题,为此,我们提出了中药新药设计的概念、研究框架及技术方法,以期对中药新药创制提供支撑。

就目前来看,“组效关系”与“构效关系”是中西

药新药设计的核心区别,但是,随着中西药研发理念的日趋一致,我们有理由相信,“组效关系”与“构效关系”相结合的新药设计模式,是未来产生中西药统一的新医药学的切入点。

参考文献

- 1 吴吉安, 嵇汝运. QSAR 在新药设计中的成功应用. 国外医学. 药学分册, 1989, 16(1):8~16.
- 2 杨洪军, 赵亚丽, 唐仕欢, 等. 基于熵方法分析中风病方剂中药物之间的关联度. 中国中医基础医学杂志, 2005, 11(9):706~709.
- 3 陈建新, 陈静, 王伟, 等. 复杂系统熵聚堆算法及其在中风病临床的应用. 系统仿真学报, 2008, 20(15):4155~4158.
- 4 唐仕欢, 陈建新, 杨洪军, 等. 基于复杂系统熵聚类方法的中药新药处方发现研究思路. 世界科学技术-中医药现代化, 2009, 11(2):225~228.
- 5 Hongjun Yang, Jianxin Chen, Shihuan Tang, et al. New drug R&D of Traditional Chinese medicine-Role of data mining approaches. Journal of Biological Systems, 2009, 17(3):329~347.
- 6 TANG Shi-huan, CHEN Jian-xin, LI Geng, et al. Research on Component Law of Chinese Patent Medicine for Anti-influenza and Development of New Recipes for Anti-influenza by Unsupervised Data Mining Methods. Journal of Traditional Chinese Medicine, 2010, 30(4):288~293.
- 7 唐仕欢, 陈建新, 杨洪军, 等. 基于复杂系统熵方法的抗肺癆方剂组方规律研究. 中国中药杂志, 2010, 35(24):3368~3370.
- 8 陈建新. 中医证候的复杂系统建模及其与疾病的相关性研究. 中国科学院研究生院, 博士论文, 2008:12.
- 9 Selim, Z., Ismail M. A. K-mean type algorithm. IEEE T PATTERN ANAL, 1994, 6(1):81~87.
- 10 彭苗苗, 方芸. 中药复方药效物质基础研究进展. 中国药房, 2010, 21(7):659~661.
- 11 陈畅. 基于数据挖掘的川芎挥发油有效成分辨识研究. 西南交通大学, 硕士论文, 2009:11.
- 12 张伯礼, 王永炎. 方剂关键科学问题的基础研究——以组分配伍研制现代中药. 中国天然药物, 2005, 3(5):258~261.

Discovery, Identification and Optimization of New Drug of Chinese Medicine - The Core and Key Design

Yang Hongjun¹, Lei Yan², Tang Shihuan¹, Chen Jianxin³, Xu Haiyu¹, Li Shaojing¹, Chen Chang¹, Huang Luqi¹

(1. Institute of Chinese Material Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China;

2. China Academy of Chinese Medical Sciences, Beijing 100700, China;

3. Beijing University of Chinese Medicine, Beijing, 100029, China)

Abstract: Design of New Drug of Chinese Medicine (DNDCM) is very important in the research and development of new drug in Chinese medicine. DNDCM will improve the level of new drug research and development in Chinese medicine. In this paper, we put forward the conception, research frameworks and technical routes of DNDCM centering on Combination-Activity Relationship (CAR). In addition, prescription discovery, component identification and prescription optimization are the core and key of DNDCM, which will be expected to offer the theoretical and technological support for R&D of new drug in Chinese medicine.

Keywords: Design of New Drug of Chinese Medicine, Combination-Activity Relationship, Prescription discovery, Component identification, Prescription optimization, System modeling

(责任编辑: 李沙沙 张志华, 责任译审: 王 晶)