中国科学: 化学 2013 年 第 43 卷 第 1 期: 18 ~ 34
SCIENTIA SINICA Chimica www.scichina.com chem.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS

评 述

化学基元组学与药物创新

徐峻[①*], 顾琼[①], 刘海波[②], 周家驹[①], 卜宪章[①], 黄志纾[①], 鲁桂[①], 李丁[①], 魏冬青[③], 王领[①], 古练权[①]

① 中山大学药学院药物分子设计研究中心, 中山大学人类病毒学研究所, 广州 510006
② 中国医学科学院药用植物研究所, 北京 100194
③ 上海交通大学药学生命科学技术学院生物信息生物统计系, 上海 200240
*通讯作者, E-mail: junxu@biochemomes.com

收稿日期: 2012-08-06; 接受日期: 2012-08-14; 网络版发表日期: 2012-10-15
doi: 10.1007/s11426-012-4761-0

摘要 化学基元组学(chemomics)是与化学信息学、生物信息学、合成化学等学科相关的交叉学科. 生物系统从内源性小分子(天然砌块)出发, 通过酶催化的化学反应序列制造天然产物. 生物系统通过化学反应和天然砌块向目标天然产物"砌入"一组原子, 这样的一组原子称为化学基元(chemoyl). 化学基元组(chemome)是生物组织中所含有的化学基元的全体. 化学基元组学研究各种化学基元的结构、组装与演化的基本规律. 在生存压力和繁衍需求的驱动下, 生物系统已经进化出有效手段来合成天然产物以应付环境的变化, 并产生了丰富多彩的生物和化学多样性. 近年来, 人们意识到药物创新的瓶颈之一是药物筛选资源的日益枯竭. 化学基元组学可以解决这个瓶颈问题, 它通过揭示生物系统制备化学多样性的规律, 发展仿生合成方法制备类天然化合物库(quasi natural product libraries)以供药物筛选. 本文综述了化学基元组学的主要研究内容及其在药物创新各领域中的潜在应用.

关键词
化学基元组
化学信息学
生物信息学
药物创新
仿生合成

1 引言

2003 年, 人类基因组计划宣告完成, 它标志着人类在揭示生命的奥秘方面前进了一大步. 后续工作主要是发现基因并将基因与各种生命现象关联起来[1]. 基因组学(genomics)涉及大量数据处理, 它带动了生物信息学(bioinformatics)的蓬勃兴起. 生物信息学主要用来解决生命科学中三类问题: 大分子结构问题、基因序列问题以及处理基因组实验数据问题[2]. 生物信息学关注大分子层面上生物学功能的问题. 它认为基因是遗传物质的最小功能单位, 是 DNA 分子上具有遗传信息的一组核苷酸序列; 蛋白质是执行生物功能的基本单位, 它的基本结构单位是氨基酸. 然而, 基因的性质最终是由核苷酸的化学性质决定的, 蛋白质的性质最终由氨基酸的化学性质决定, 而核苷酸和氨基酸的化学性质则是由他们的化学结构性质决定的.

化学结构与性质相关问题属于化学信息学(chemoinformatics)的问题. 简言之, 化学信息学研究分子的化学结构与性质的关系. 化学信息学在生命科学领域主要解决三类问题: 小分子化学结构与性质的关系问题、小分子与大分子相互作用问题以及小分子片段的组合问题.

随着生命科学的飞速发展, 生物信息学与化学信息学逐渐发生深度交叉. 在化学中, 最简单的结构单元是原子, 将原子组装成分子的基本规律因为元素周期表的发现而获得圆满解决. 在生命科学中, 生物的多样性可以最终归结到基因的多样性, 而基因

的基本物质砌块由为数不多的分子片段组成, 一些主要规律已经被发现. 于是人们提出这样的问题: 有没有“生物元素周期表”存在[3]? 这个表中的元素可能多达十万以上, 元素之间的关系也非用一张长方形表格就能列出那么简单[4]. 更重要的是, 这“生物元素周期表”中的元素不是单个原子, 应该是一组化学基元, 即组成天然产物(大分子或小分子)的基本化学单元(chemoyl). 这些化学基元是如何划分的呢? 这些问题需要化学信息学与生物信息学的共同努力才能逐渐解决. 化学基元组学就是化学信息学与生物信息学交叉的产物.

实际上, 自从上世纪 60 年代三联体遗传密码子被证实以来, 人们一直在为构建“生物周期表”而努力. 例如, 为了寻找密码子中核酸组合模式与氨基酸物化性质之间的关系, Biro 等[5]设计了一种以氨基酸为中心(生物元素)的长方形“生物元素周期表”, 发现位于中央的核酸控制氨基酸的药效团性质, 如果中央核酸密码子是嘌呤类, 则决定氨基酸侧链的亲水性; 如果中央核酸码子是嘧啶类, 则决定氨基酸侧链的酸碱性(正负电荷).

生物系统远比化学系统复杂, 因此, “生物元素周期表”应该比元素周期表复杂得多, 它应该是个广义的概念, 其核心问题是

(1) 核酸和氨基酸就是生物系统的基本单元了吗? 核酸和氨基酸可不可以分解成更加基本的化学基元?

(2) 这些化学基元之间的组装规律是什么?

(3) 由化学基元组装而产生的为生物体系服务的化学多样性是怎样演化的?

2　天然产物的基本砌块

天然产物大分子的基本砌块(单体, 如核酸、氨基酸)的化学结构比较简单, 它们代表了在这个星球上的生物存在的统一性. 然而, 天然产物小分子的化学结构十分复杂, 它们是自然界生物多样性的物质基础, 是生物体对环境变化应答的产物(本文后面所讨论的天然产物都指的是天然产物小分子, 不包括蛋白质等生物大分子). 天然产物是否具有基本砌块? 这个问题不是用简单的方法能解析出来的. 因此, 在过去的几十年来, 人们就这个问题提出了许多模型.

2.1　药效团模型

药效团(pharmacophore)概念最初是药物化学家的经验总结. 人们在研究药物研究中发现, 药物分子中的某些子结构对药效比其他结构部分更重要, 如果这些子结构缺失, 整个分子就失去药效. 因此称这些子结构为药效团. 随着计算机辅助药物设计技术的发展, 药效团概念已经发展成具有药效所必需的原子或者原子团之间的三维空间配置的模式(图 1)[6]. 因此药效团模型与天然产物的组装没有关系.

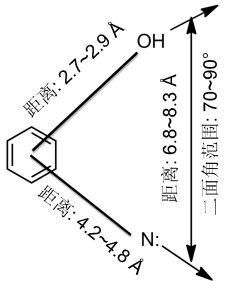


图 1　现代的三维药效团概念

2.2　优势结构模型

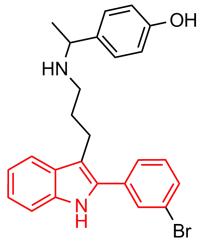
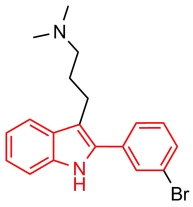
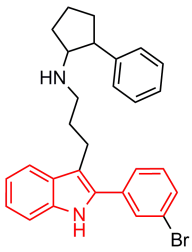
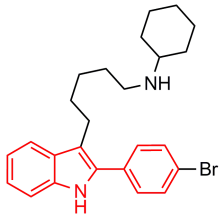
优势结构(privileged structure)是指那些能对多种不同药物靶标产生药理活性的结构片段, 由 Evans 和他的合作者于 1988 年首先提出[7]. 优势结构实际上是一种特定的化学结构骨架, 这种骨架可以针对某类靶标提供选择性的结合, 在一定范围容忍化学修饰, 并且有较好的成药(druggability)性质[8](表 1 列出了四个优势结构的实例[9], 优势结构用红色标出). 因此, 优势结构模型是一种化学结构骨架模型, 与具体生物靶标分子相关联, 而化学结构骨架并没有严谨的定义, 因此不能成为组成天然产物结构的化学基元.

如今, 优势结构的概念在药物化学研究中经常被用到, 药物化学家们主要利用这个模型作为一种加快新药发现的可能捷径, 但是没有对类药化合物的优势化学结构进行系统化的深入研究.

2.3　合成子模型

合成子(synthon)模型最初是由诺贝尔奖获得者 Corey 在 1967 年提出来的. Corey 通过严密的逻辑推理方法, 系统地分析一个目标化合物所含有的合成子. 对于一个目标化合物(一般是天然产物), 根据一

表 1 优势结构实例和对应的药物靶标

结构	靶标
	NPY5
	5HT-6
	CCR3, CCR5
	MCR4

组既定的反向合成规则(即合成反应的逆反应), 找到该目标化合物的前体, 这些前体作为子结构蕴含在目标化合物中, 叫做合成子. 如果前体的分子结构仍然不够简单或者在实验室中找不到, 就继续用相同的一组反向合成规则寻找前体的前体, 按此一直寻找下去, 直到所有的前体都可以在现实中找到, 或者所有的反向合成规则都已用尽. 这种对天然产物砌块的逻辑推理路线叫做反向合成树(图 2)[10].

Corey 第一次把作为“艺术”的有机化学合成设计过程规范为可以严格地进行推理的逻辑体系. 此体系分为两部分: (1) 基本的有机合成砌块(合成子), 就像摆在棋盘上的棋子; (2) 对合成子所能进行各种组合操作(化学反应规则), 就像一组下棋的规则. 这种方法甚至可以用计算机程序来实现, 成为早期人工智能的经典作品.

本世纪初, Schreiber 等人在这个思路启发下, 提出了靶标导向、多样性导向[11]的合成方法, 试图解决药物筛选所需要的具有多样性和类药性化合物库的问题, 受到了业界的关注.

2.4 结构片段模型

20 多年前, 由于高通量药物筛选效率低下(高通量药物筛选的平均命中率远远小于 1%), 有人认为, 这与蛋白质靶标的活性口袋容纳不下尺寸较大的小分子化合物有关, 所以提出筛选小分子量的化合物(分子量 < 250 D). 由于小分子量化合物与靶标的结合作用有限, 活性相对较低, 人们将一些分子量小、活性较弱的化合物用化学合成法再组装成分子, 从而提高活性. 这种分子量小、活性较弱的化合物叫片段化合物(fragments), 而这种药物发现方法称为基于片段的药物发现(fragment-based drug discovery, FBDD)[12]. 表 2 给出两个基于片段的药物发现的实例[13]. 通过对片段的药物筛选, 得到活性的小分子片段, 然后用人工化学合成的方法将它们组装起来, 得到高活性的药物先导化合物.

这种方法对药物创新很有意义, 但是这类分子片段及其组装方法与天然产物的化学基元和自然界对这些化学基元的组装没有关系.

2.5 从合成子到化学基元

前文所述的各种关于小分子化合物组成的砌块模型, 大部分没有抓住自然界对天然产物的组装能力这个核心. 药效团模型反映了配体对接(docking)到受体所需要的空间配置和物理性质(如电荷、亲水性)的互补性与天然产物分子的基本组装单元没有关系; 优势骨架结构模型反映的是天然产物骨架对靶

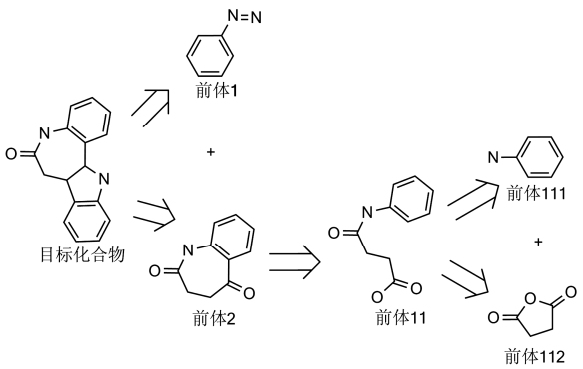
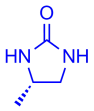
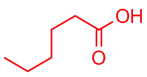
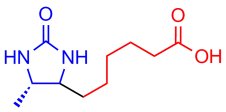
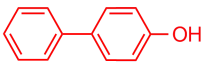
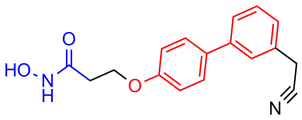


图 2 合成子与反向合成树

表 2　片段化合物和它们的组装

靶标	片段 1	片段 2	组装分子
抗菌蛋白	 K_i = 34 μmol/L	 K_i = 260 μmol/L	 K_i = 0.0004 nmol/L
基质溶素	 K_d = 17 mmol/L	 K_d = 280 mmol/L	 K_d = 15 nmol/L

标的选择性关系; 结构片段模型强调靶标分子活性位点对化学片段的选择. 只有合成子模型触摸到天然产物基本砌块和组装规律的核心. 它提供了把药物分子分解成基本组元的逻辑框架. 其缺陷是采用人类能够得到的化学试剂作为反向合成树生成的终结条件, 以人工化学合成反应作为规则集, 对目标分子进行系统化的划分. 这样, 思路虽然对了, 但只建成了基于人类化学合成能力的化学基元组装体系, 这个体系的合成能力不受内源性化合物种类的限制, 只受可用的化学合成试剂的限制, 可以用来组装天然的或非天然产物. 然而, 合成子模型启发了以破译天然产物组装奥秘为使命的化学基元组学的建立.

现在, 人们已经基本厘清了自然界从无机物(如二氧化碳、水、氮气)开始, 合成基本内源性化合物(如氨基酸)的原理[14]. 如图 3 所示, 自然界通过植物的光合反应与微生物的酶促氧化还原反应的耦合实现氨基酸的合成.

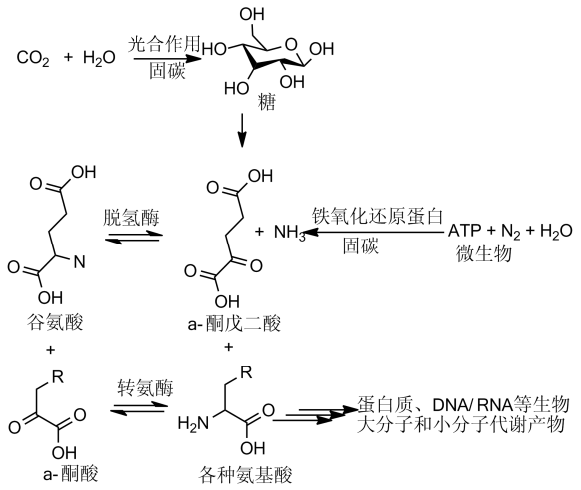


图 3　固碳化学反应与固氮化学反应网络的耦合实现了从无机到有机的飞跃

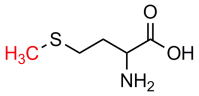
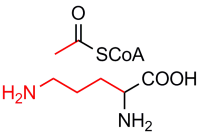
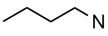
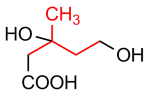
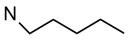
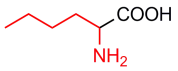
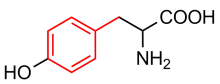
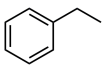
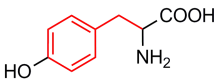
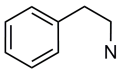
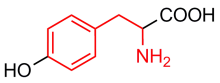
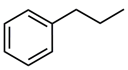
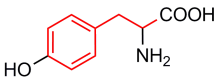
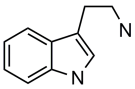
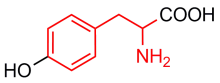
内源性分子并不是一个原子一个原子地"砌"成的, 而是以一组原子的形式"组装"到目标产物分子上去, 这一组原子就是化学基元. 但是化学基元在被"砌入"目标分子之前, 应该被承载在可以独立存在的分子中, 这样的分子被称为天然产物的砌块(building blocks)[15]. 砌块子的核心是化学基元, 它的其他部分一般是将砌块递送到目标产物的官能团, 有的是承载化学基元, 有的用来活化化学基元的特定原子以备连接到目标产物上, 有的向化学基元提供能量等.

表 3 中总结了十类天然产物的基本化学基元和它们对应的砌块. 如表 3 所示, 化学基元可看做是一种通式, 表示一组被贡献到目标天然产物的原子. 例如, C2 表示向目标天然产物分子"砌入"两个碳, 可以在环上或链上, 可以是饱和的, 也可以是不饱和的. 因此, 天然产物的化学基元不是独立的分子实体, "砌块"是它们的"载体", 可以在生物体中独立存在, 在酶的催化作用下, 向目标天然产物"递送"化学基元.

化学基元组(chemome)是化学基元的全体. 因此, 化学基元组学(chemomics)主要研究化学基元的结构、功能及合成产物. 其目标是揭示生物合成所用的基本砌块及其组装规律, 在此基础上通过仿生合成和化学合成的各种技术制备结构多样的类药化合物或类天然产物, 为新药或其他生物材料的创新提供新颖的物质资源和理论指导; 揭示生物小分子化学结构多样性的演化规律, 为天然产物的获取、改造和可持续利用提供理论依据与关键技术.

天然产物是生物体生存与繁衍的物质基础. 在漫长的生命演化史上, 对某种现存的物种而言, 应该

表 3 天然产物的化学基元和它们的砌块

化学基元*	符号	对应的砌块实例
	C1	
	C2	
	C4N	
	C5	
	C5N	
	C6C1	
	C6C2	
	C6C2N	
	C6C3	
	indole.C2N	

*化学基元是合成砌块对目标天然产物的原子团贡献. 例如, X–C 是一碳化学基元的通式, 表示向目标产物贡献一个碳; 其具体形式可以为饱和或不饱和碳, 在目标产物上键合的原子可以是 C, N 或 O. 其相应的砌块可以是蛋氨酸或其他内源性分子

还有许多天然产物由于对其本身的进化无意义或有害而被淘汰, 然而, 这些天然产物也许对于调控其他物种有用. 因此, 如果破解自然界合成生物分子的基本规律, 这些“灭绝”的天然产物可以被重新发现, 成为药物和生命科学研究的新资源.

如果对天然产物酶促组装过程加以人工改造, 例如, 改变催化剂、加入化学试剂或让天然化学基元与人工化学基元杂聚, 应该产生大量的类天然产物, 这就是仿生合成(biomimetic synthesis) (图 4).

过去 60 多年来的药物研究经验表明, 天然产物和类天然产物都具有高的成药性, 它们对于药物创新和其他生物材料的创新都具有重要意义, 是战略性的物质资源[16].

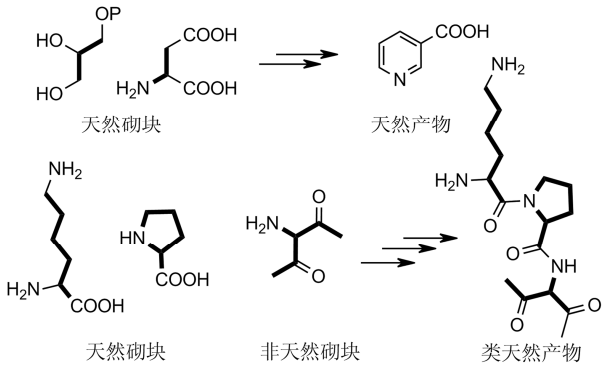


图 4 仿生合成制备类天然产物

3 化学基元组学的基本方法和技术

天然产物是自然界的生命体系根据对化学基元的需求以一组天然产物砌块为原料, 通过酶促合成组装而成的. 天然产物大分子(如蛋白质)的组装是在其他生物大分子的监控下依照程序在酶促条件下严格执行的, 尽量不出差错; 而且, 监控组装过程的细节被记录在以生物大分子为物质载体的基因里, 只能遗传, 不可“遗忘”, 这是保持物种统一性的基本条件. 大部分天然产物小分子的组装虽然也是在酶促条件下进行, 但是没有严格的监控, 组装完成之后, 有关天然产物小分子由哪些化学基元组成的化学印记也被“遗忘”. 如图 5 所示, 从脂肪酸的分子结构不能直接看出乙酰基是它的合成砌块, 不通过结合了生物合成知识的计算机算法解析, 直接从天然产物的结构是难以看出它的化学基元组成的.

3.1 天然产物化学基元的破译

天然产物的化学结构虽然复杂, 但是它的化学基元仍然是可以破译的. 以环氧聚微管素的化学基元破译为例(图 6): 运用前述的合成子分析方法可以推知, 环氧聚微管素是由三个 C2 砌块(基元族)、五个 C3 砌块和一个 C5C2 杂环砌块组装的. 由于化学基元砌块(相当于 Corey 所说的合成子)不能独立存在, 通过建立内源性化合物结构库, 可以确认每个化学基元砌块所对应的天然产物的砌块.

因此, 天然产物化学基元的破译需要下述条件:

(1) 一组天然产物合成的酶反应, 它的逆变换(即合成的反向操作)可以将天然产物分子结构解析

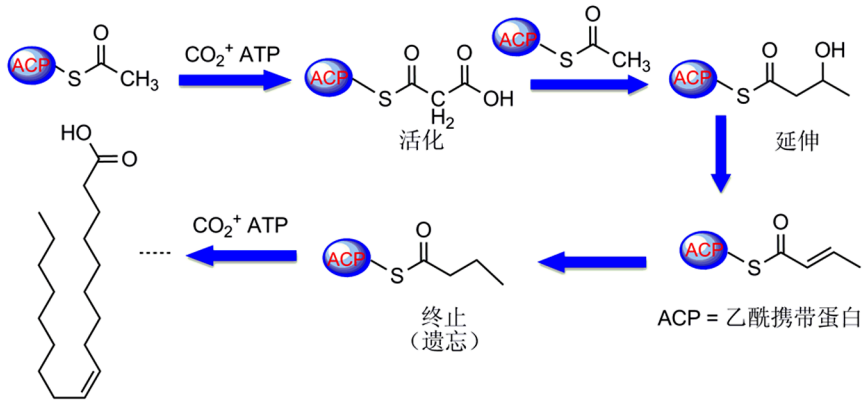


图 5　脂肪酸分子结构中的化学基元印记被“遗忘”

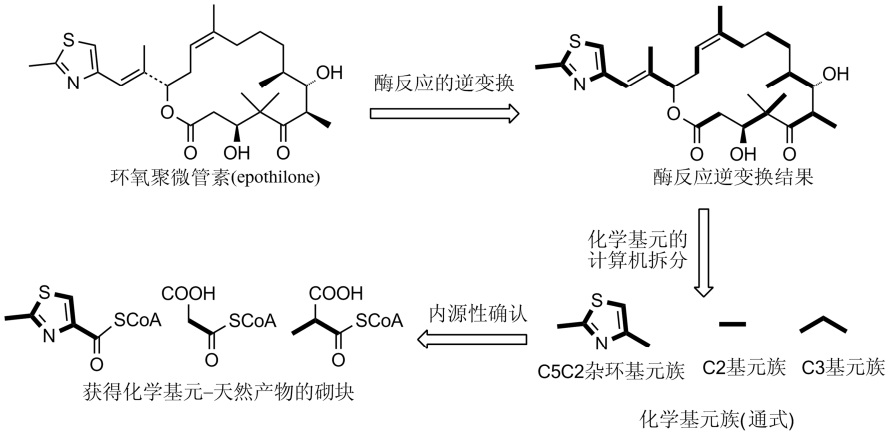


图 6　环氧聚微管素化学基元的破译

成对应的化学基元;

(2) 一个内源性化合物结构数据库, 库中每一个分子即天然产物砌块与酶相关联, 也与化学基元相对应;

(3) 天然产物砌块与化学基元之间的是多对多的关系路径图;

(4) 一组图论算法, 用于识别天然产物的化学结构, 利用酶反应数据库和内源性化合物结构库完成化学基元的拆分.

值得注意的是, 对给定的天然产物, 化学基元拆分结果有多种可能. 因为合成砌块与化学反应是多对多的关系, 合成砌块与化学基元也是多对多的关系, 它们之间构成极为复杂的关系网络. 说明组装天然产物有许多种途径.

化学基元的破译所需的酶反应网络数据已经有了丰富的积累. 人类、动植物的内源性化合物代谢反应途径已经被相对完备地总结在文献和数据库中, 著名的如 Michal 编纂的《生化代谢途径: 生物化学与分子生物学导航图[17]. 日本京都大学综合各方面的文献将生化代谢途径建成 KEGG 数据库[18] (近 7000 条内源性代谢反应); 德国爱尔兰根大学的 Johann Gasteiger 建成了可以进行生物化学反应检索的 BioPath 数据库[19] (近 4000 条内源性代谢反应). BioPath 数据库与 KEGG 数据库的不同在于 BioPath 数据库对反应中心进行严格的标引, 因而可以对代谢物的反应途径进行检索和重新构造.

3.2　化学基元组装的基本规律

在生物世界, 生物大分子合成和天然产物合成的共同特点是起始于共同的内源性砌块, 使用共同的组装装置(酶). 所使用的组装机制(酶反应)主要是水解、裂合、转移、异构、氧化还原和连接. 所不同

的是它们组装化学基元的目标和要求不同, 因而采用的组装平台不同.

生物大分子组装是程序性合成, 要求严格根据基因的信息准确地组装, 不犯错误. 当然世界上没有不犯错的体系, 虽然生物大分子组装过程经历严格的"质控", 仍然避免不了犯错误, 但其错误率仅在亿分之一到十亿分之一之间[20]. 低出错率保证了物种遗传的统一性. 然而, 低出错率也保证了物种应对环境的灵活性(变异适应的多样性). 有理由认为, 细菌的抗药性本质上是生物大分子组装出错率的积累造成的自然结果, 滥用抗菌素不过是加快了细菌的抗药性的速度. 假定细菌细胞繁殖出错率为 1% (变异), 每小时繁殖 2 次, 单个细菌细胞长成百万群落只要 10 h (2^{20} = 1048576), 即每一万细菌细胞至少积累一个变异体, 当携带变异体的菌群继续扩增, 抗药变异体出现的可能性就会大大增加[21]. 这也是有些细菌对尚未上市的药有抗药性的原因[22].

生物大分子蛋白质组装的基本规律总结如下:

起始材料: 氨基酸(基本代谢物);

组装平台 A: 核糖核蛋白体(在信使 RNA 模板上进行);

组装过程: 活化氨基酸→由 tRNA 带至与肽结合位置上→由氢键连在 mRNA 链的第(n+1)个密码子上→与肽链上当前的氨基酸形成肽键并释放 tRNA→mRNA 移位到第(n+2)个密码子处, 如此重复, 直到装配好最后一个氨基酸;

组装平台 B: 非核糖核蛋白体(在多模块酶模板上进行);

组装过程: 酶模板有 3 个主域, 对每个氨基酸执行 3 个操作: 选择氨基酸→活化氨基酸→与磷酸泛酰巯基乙胺(phosphopantetheine)共价键合→成肽键→后组装修剪[23].

天然产物组装的基本规律总结如下:

起始材料: 基本或次级代谢物(化学砌块);

辅助材料: ATP(提供活化所需的能量), 辅酶(化学结构砌块的载体);

组装平台: 无;

组装过程: 活化(activation)→延伸(elongation)→终止(termination)(参见图 5);

合成条件: 按照一定次序组合的一组酶(代谢网络).

为了保持生物系统的稳定性, 天然产物的化学基元通常是稳定的, 只有需要拿它建造新分子时才将其活化. 活化是通过加入新官能团实现的. 如图 7 所示, C2 基元以乙酰基形式存在于 SCoA 砌块分子中. 由二氧化碳加成导致甲基上碳原子活化, 指示该碳原子已经准备好被定点组装到目标化合物上, 载体的腺苷部分则负责提供将 C2 砌块添加到目标化合物上所需要的能量. 这样的分子像一颗"导弹", 有 3 个功能部件: 战斗部(红色)、导航识别部(蓝色)和推进部(黑色).

由于不像生物大分子蛋白质组装那样由组装平台来严格控制组装过程, 天然产物的组装给予生物体在应对环境时有较大的灵活性. 如图 8 所示, 砌块 A 和 B 携带化学基元 C2, 砌块 C 携带化学基元 C3. 通过调控酶的种类、砌块的种类以及浓度, 可以组装

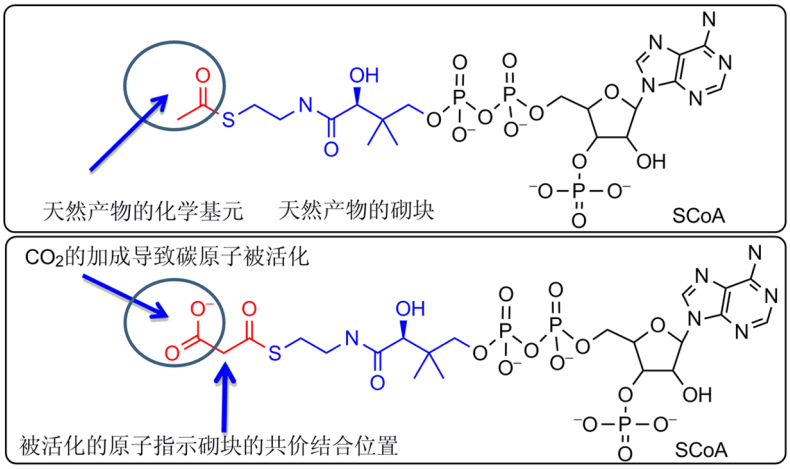


图 7 天然产物砌块的活化

出复杂度和种类看起来完全不同的天然产物.

虽然砌块 A 和 B 都携带 C2 化学基元, 骨架相同, 但是, A 的活化基位于一端, 为羰基加 SCoA(巯基辅酶 A); 而 B 的活化基位于两端, 分别为羧基和羰基加 SCoA(巯基辅酶 A). 两种砌块在基元组装线上的作用效果不同, A 在一端与其他砌块装配, 合成线性分子. 砌块 B 和 C 携带的化学基元(粗线表示的骨架)两端都可以与其他砌块实现扩充组装, 合成线性、树状或环状分子. 在生物化学合成酶的催化下, 脂肪酸就是由 A 开始组装, 以 A 反复自我扩增, 最后以 A 终止, 组装成线性分子. 砌块 A、B、C 的组合组装成环状分子; 砌块 A 和 C 组装成树状分子. 分子尺寸和结构类型由酶的种类、底物浓度控制.

生物大分子组装由多种生物大分子协同控制, 严格按照编码组装. 而天然产物组装的灵活性体现在为了应付环境变化(抗逆)的靶向组装. 在图 9 中, 同样是两种氨基酸, 在编码组装平台上, 用一般的氨羧缩合反应, 组装成肽; 而在靶向组装中, 生物体系为了抗逆的需求, 根据物种竞争的需要, 合成β-内酰胺骨架(青霉素), 阻止其他菌类肽聚糖上两个 D-丙氨酸的缩合.

由此可见, 即使在化学基元种类不变的情况下, 生物系统对化学结构多样性的生产能力往往出人意料, 也为人工化学合成设计带来新的思路.

3.3　化学多样性在自然界中的演化

生物个体都要经历诞生、成熟和衰老的过程. 各个过程中, 生产的内源性物质的品种和相对含量都不相同, 其所含的内源小分子(含有化学基元的生物化学砌块)的种类和含量都经历了不断的演化, 它们的演化谱对指导中药方剂研究和中药药材采收条件有重要意义.

如民谚所说: “三月茵陈四月蒿, 过了四月当柴烧”. 茵陈的药用成分是在其植物尚幼时丰度较高,

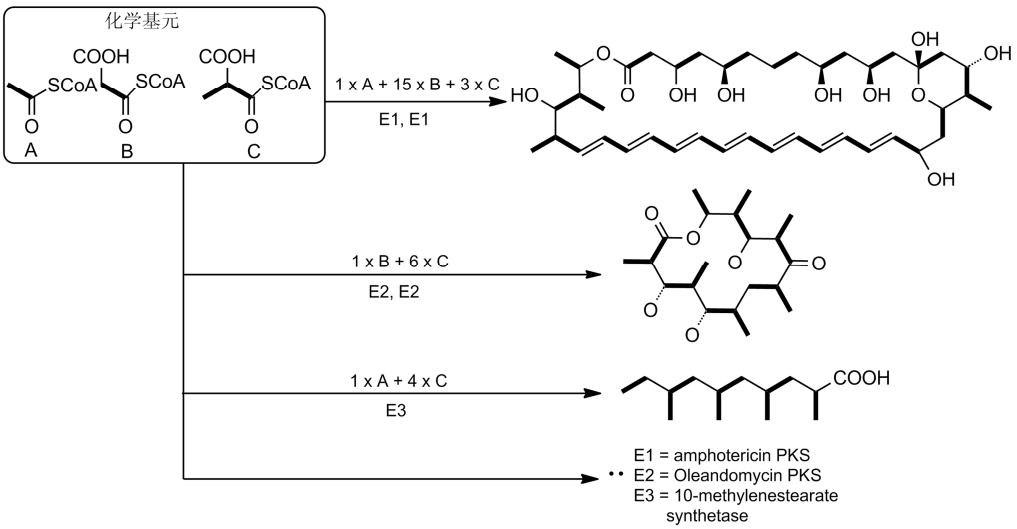


图 8　从化学基元组装天然产物

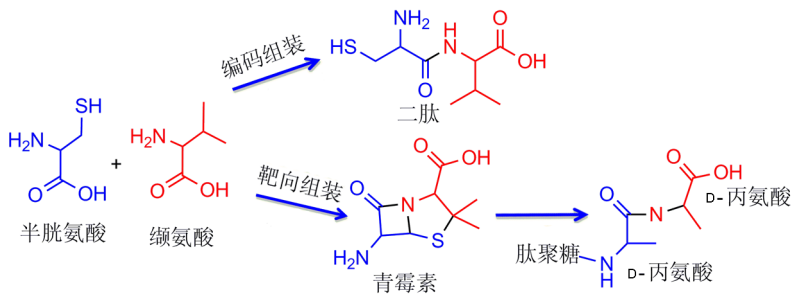


图 9　编码组装和靶向组装

以后逐渐减少以至于失去药用价值.

又如达菲的前体莽草酸(shikimic acid), 在植物和微生物中广泛分布, 但是只在少数植物(如八角科植物果实)中大量存在. 分析莽草酸的代谢网络(图 10)可知, 它是植物制造许多其他重要内源性物质的中间体, 不是生物化学合成的最终产物. 因此, 为了高效地从自然界捕获莽草酸, 需要知道植物中莽草酸的演化规律. 通过了解莽草酸的生物合成途径, 还可以根据需要收获其他重要的生物合成砌块.

上世纪 80 年代, 人们已经注意到天然产物的产生与物种个体发育有关. 例如, 喹诺里西啶碱、毒芹碱、类固醇碱分别在羽扇豆、铁杉和土豆的嫩芽中合成[24]. 这些生物化学多样性与个体发育过程相关性的规律应该是中药自然资源的采集和利用的科学基础.

然而传统上中药的采集与炮制主要依赖于经验. 为了充分利用和保护中药药用资源, 应该系统地研究药用植物所含的内源性砌块与个体发育(ontogenesis)、代谢网络的关系. 根据药用植物的化学砌块生成的时间和代谢网络途径来确定采集时间、种植和提取技术以及生物多样性保护方式才是合理的.

另外, 生物体系中化学砌块多样性的演化也与物种从低级向高级进化相关. 其规律应该是中药组方中的同性味取代的科学基础, 也是对于濒危物种的保护的基本依据. 以咖啡酸为例, 它在植物中广泛分布, 随着植物种群从低等向高等的演化, 通过自聚、杂聚所产生的代谢物的分子结构呈现多样化趋势(图 11). 这些代谢物的合成要求更专一的酶, 对细胞的功能进行更精细的调控. 弄清它们相应酶的演化图谱和咖啡酸代谢物的演化图谱之间的关系以及演化的动因, 不仅可以揭示自然演化之谜, 更为药物研发提供广阔而新颖的分子结构多样性空间.

4 化学基元组学与药物创新

药物创新始于对活性化合物的筛选, 而药物筛选的成功取决于筛选技术的可靠性、用于筛选的物质资源的成药性和化学结构的多样性.

今天, 超高通量药物筛选、高含量药物筛选和各种组学已经成为药物工业界的常规技术, 而原创药物产出率仍然不高, 原因之一就是可供药物筛选的化学结构多样性依然没有大的突破.

用于药物筛选的化合物库大多数来自化学合成或者组合化学合成, 此类方法受化学合成试剂的品种和化学合成手段的限制. 另一个来源是天然产物, 但天然产物含量少、提取成本高, 不利于资源的可持

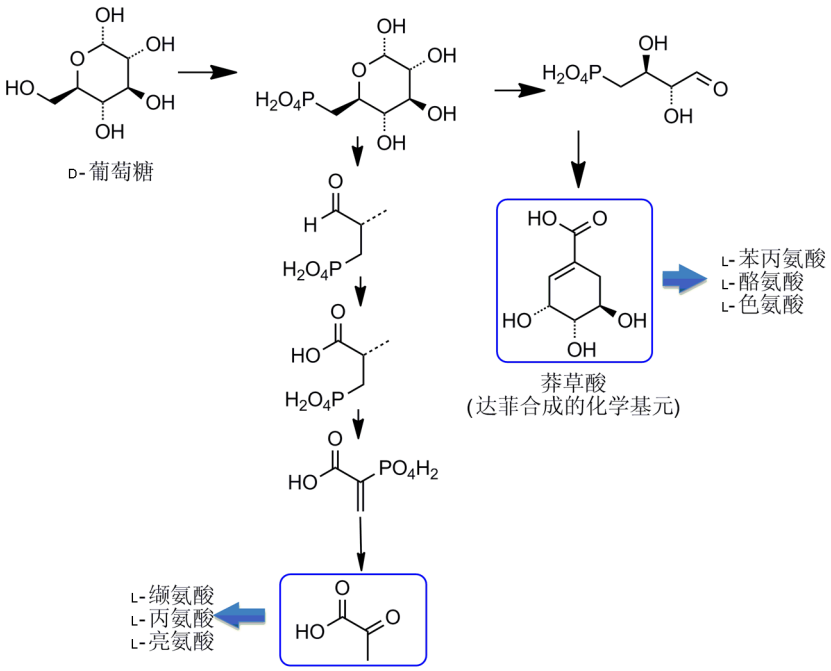


图 10 莽草酸作为许多基本代谢物的中间体

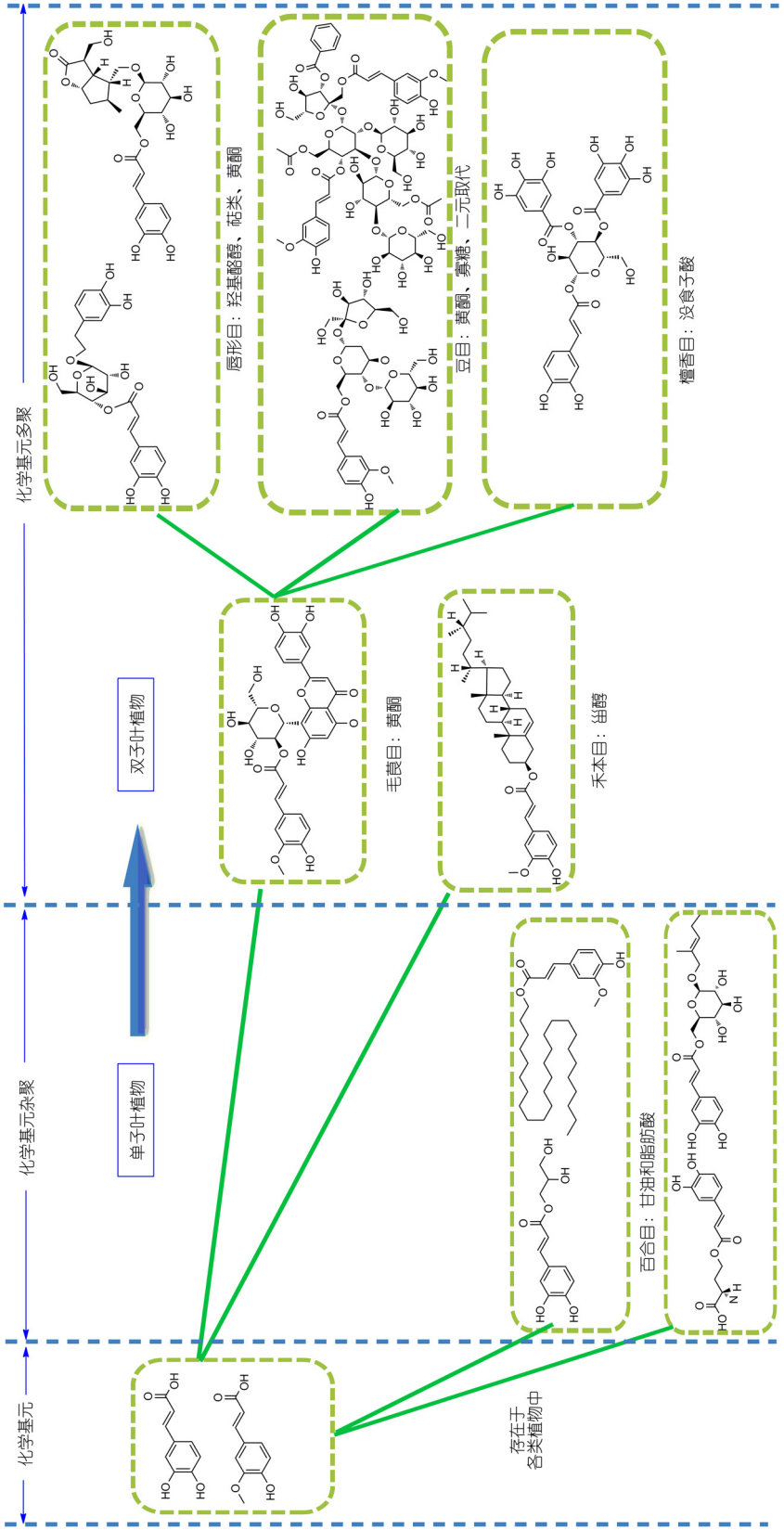


图 11　咖啡酸衍生物的多样性与物种演化

续利用. 为了解决药物筛选的资源问题, 人们提出类天然化合物的策略[25]. 化学基元组学研究的目的之一就是用化学合成技术仿真自然界的生物合成规律, 以相对较低的成本、快速地合成大量的类天然产物, 为药物筛选提供可持续发展的物质资源.

4.1 天然产物的仿生合成

自然界合成天然产物是在常规条件下(非高温高压等极端条件)进行的, 从简单生物小分子开始, 如氨基酸、肌酸、二氧化碳等, 通过酶和辅酶的催化以及催化反应网络的耦合提高效率并保证产物的化学结构的多样性. 天然产物的结构多样性常在人类意料之外, 纯粹用化学合成方法往往难以模拟.

化学合成技术与自然界合成能力相比虽然看起来相形见绌, 但也有独特的优势: 化学合成可以在非常规条件下进行, 如无氧、无水、高温高压、使用微波加热等; 通过在孤立体系中进行, 使环境因素的影响大为简化, 得到纯度高的产物; 可以进行在自然条件下不能完成的化学反应.

仿生合成就是要结合天然合成和人工合成的优点, 以较低的代价, 高效地制备类天然产物, 以满足药物筛选的需求. 表 4 列出 3 种合成策略的特点. 如图12所示, 天然合成采用蛋白质(此处是乙酰携带蛋白, ACP)为组装平台. 人工的组合化学合成采用高聚物微球代替蛋白质作为组装平台. 而仿生合成可以结合两者特点, 即仿照天然合成的合成规划, 但采用高聚物微球取代昂贵而敏感的蛋白质作为合成平台. 用高聚物微球代替蛋白质不仅可以降低成本, 对微球上的基团还可以进行化学修饰, 使之兼具载体和活化功能, 提高仿生合成的灵活性和效能.

自然合成的规律会给化学合成方法带来很多启示. 在弄清机理之后, 通过对自然合成过程的改造, 仿生合成可以大大提高合成效率. 如图 13 所示, 短杆菌肽是一种天然合成的环肽. 天然合成需要以含巯基的肽携带蛋白为载体, 在短杆菌肽合成酶催化下合成.

然而, 通过厘清它的合成机理, 化学家采用简单的 1-乙酰氨基, 2-巯基乙烷代替肽携带蛋白, 用氨水做催化剂即可高效率地合成[26]. 通过分子动力学模拟发现, 具有下述化学基元组合通式(D-X)-L-P-(L-X)$_n$ 的肽序列在溶液中易于形成环肽. D 和 L 为氨基酸的立体构型, X 代表任意氨基酸, n 为大于 1 的整数.

表 4 天然合成、人工合成与仿生合成的特点

	合成砌块	合成条件	合成规划
天然合成	内源分子	自然条件 复杂体系	导入活化起始物→ 与引物结合延伸→ 载体释放得到产物
化学合成	化学试剂	可控条件 孤立体系	分析合成子→ 前体试剂(活化)→ 执行合成(手性控制+延伸+释放)
仿生合成	内源分子 化学试剂	可控条件 简化的复杂体系	导入活化起始物→ 与引物结合延伸→ 载体释放得到产物

4.2 类天然化合物库的构建

为了构建类天然化合物库, 首先需要选择化合物库的骨架. 骨架可以是与生物功能有明确的相关关系的化学基元或者化学基元的组合. 图 14 列出一个生物功能与化学基元关系网络的实例.

通过对周家驹等人建立的中药化学结构数据库中收集的天然产物的化学结构的研究, 我们发现[27], 有些天然产物化学基元或者结构骨架与许多生物活性相关, 而有些化学基元或者结构骨架与生物活性相关不多. 肽类分子除外, 以α-氨基酸为骨架的天然产物种类很少. 然而苯丙氨酰类(C6C3 基元)在天然产物中广泛分布, 而且与很多生物活性相关. 从苯环上单碳取代衍生出来的化学基元对天然产物的骨架结构多样性演化具有重要意义. 如图15所示, 苯环单碳取代化学基元在中药药用植物中有广泛分布, 而且具有多样化的生物活性, 在构建类天然化合物(类药性化合物)库的研究中具有重要意义. 以苯并吡喃类为例, 它们的主要生物作用依次为: 细胞毒、抗氧化、抗菌/抗病毒、抗炎、抗代谢性疾病、抗肿瘤. 其中具有抗氧化活性的苯并吡喃类化合物主要来自甘草、淫羊藿、黄芩、鸢尾、白果、鱼藤、黄花蒿、橘皮、葛根、桑白皮、吴茱萸、刺桐、鸡血藤、沉香、苦参、茵陈、射干、穿心莲、九里香、丹参、侧柏等.

进一步分析发现, 作为抗氧剂的苯并吡喃类化合物, 它们的结构特征通式如图 16 所示, 该通式所定义的化学空间就是一个类天然化合物的库. 通过计算机模拟技术, 从图 16 的 Markush 结构出发, 枚举出全部的靶向的虚拟类天然化合物库. 再通过化学基元解析技术, 利用仿生合成设计, 提出虚拟类天然化合物库的制备方案, 经过基于结构和基于配体的虚拟筛选, 可以鉴定新一代的靶向药物先导化合物.

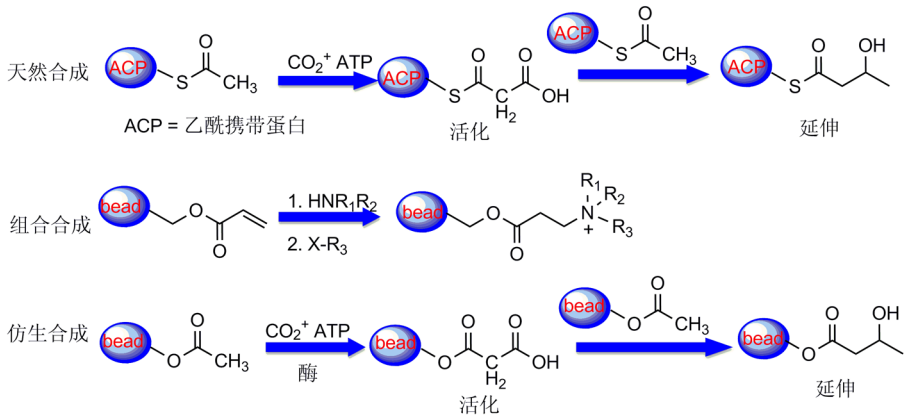


图 12　仿生合成与天然合成及组合合成的实例

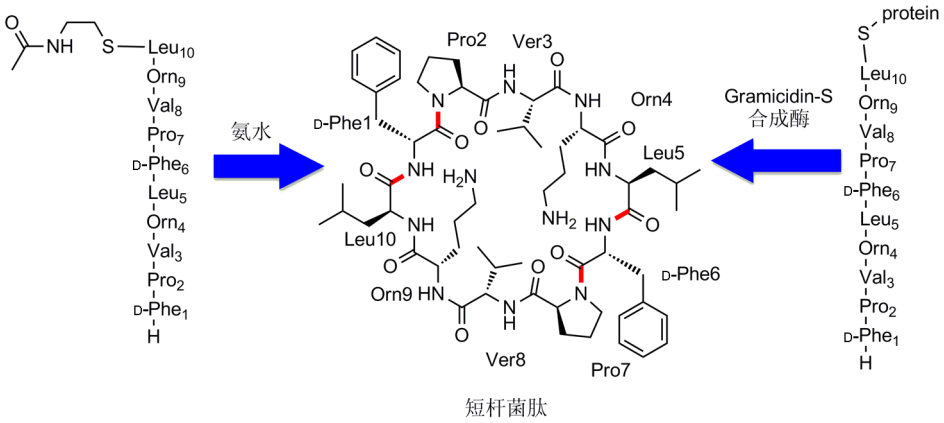


图 13　短杆菌肽的仿生合成与天然合成

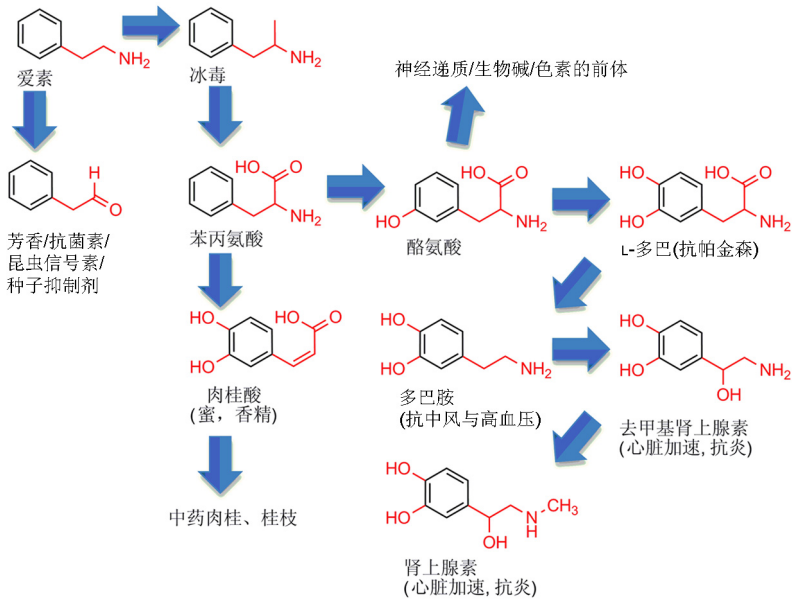


图 14　生物功能与化学基元的关系网络

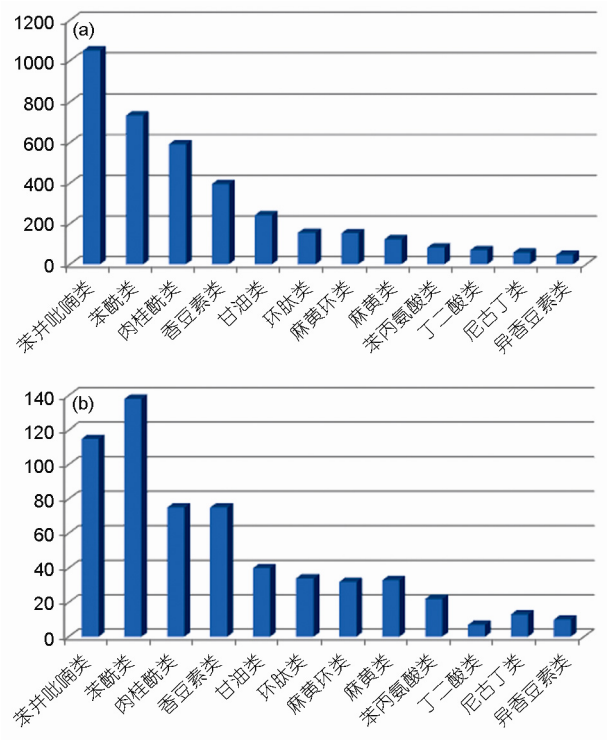
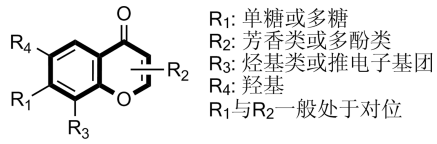


图 15 (a) 化学基元与中药药用植物种数的关系; (b) 化学基元与生物活性数目之间的关系



61种来自中药的天然产物

图 16 作为抗氧剂的类天然化合物库

4.3 基于化学基元的药效网络

传统的药物创新模式以寻找高选择性的单靶标调控为目标. 然而, 越来越多的证据显示大多数配体同时作用于多个靶标, 而靶标之间可能存在着复杂的信号转导网络关系, 很难设计或发现药效作用强又没有任何副作用的化合物.

实际上, 没有绝对单靶标作用的药物, 只有尚未知晓更多作用靶标的药物. 对由复杂原因引起的疾病尤为如此[28]. 例如, β-内酰胺的抗病毒机理其实是通过 2 个以上蛋白质诱导细胞死亡; 典型的靶向药 Gleevec 也是通过多个信号转导机制而抗癌的.

近年来, 网络药理学(network pharmacology)受到国内外业界的关注. 网络药理学结合系统生物学方法, 运用网络分析技术研究基因多态性、连接性和冗余性. 有人认为, 它不过是 Paul Janssen 针对药物作用和副作用做广泛的 QSAR 研究的新提法而已[29]. 也有人认为网络药理学就是系统药理学的网络研究方法[30]. 网络药理学所研究的网络主要有:

(1) 蛋白质-蛋白质作用网络(PPI). 其节点是蛋白质分子, 边代表蛋白质之间的结构相似度或有无调控关系;

(2) 药物-药物作用网络(DDI). 其节点是药物分子, 边代表药物分子之间是否有共同的靶标(drug targets)(图 17), 或共同的治疗效果(medical indications), 即药效网络;

网络药理学主要解决下述问题:

(1) 确证药物靶标. 其主要标准是: 具有可筛选性的酶, 此酶是导致疾病的主要因素, 对疾病起着核心驱动作用, 在疾病发展途径上起决速作用并且是唯一的; 靶标与临床观察事实相关; 活性位点与配体有亲和力且分子量较低, 有三维结构或模型, 最好不在人体系统中; 该靶标的抑制或活化不会干扰与疾病过程无关的其他蛋白或基因.

(2) 确证调控药物靶标的先导化合物. 化合物应该主要作用于药物靶标, 不作用于反靶标(anti-targets). 如果疾病是由多个靶标协同作用引起的, 该化合物最好同时调控相关的靶标(多重靶向药), 或者设计若干化合物, 分别作用于各自的靶标, 进行协同治疗(复方药).

然而, 药物-药物作用网络还有另外一种形式. 如果以天然产物的化学基元作为网络的节点, 两个化学

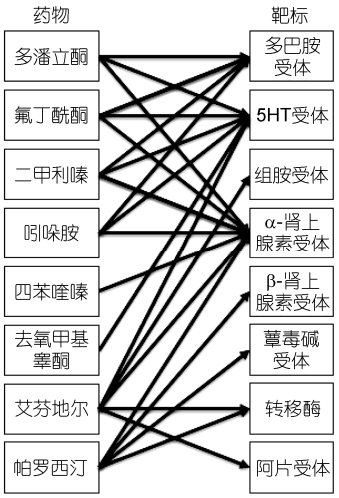


图 17 药物-靶标关系网

基元形成天然产物的关系作为边, 而此边与生物活性相关, 这样就做成了基于化学基元的药效网络.

通过对天然产物数据库(dictionary of natural products, DNP, http://www.chemnetbase.com/)的研究, 我们发现化学基元之间配对形成天然产物有一些有趣的规律(详情另文发表). 图 18(a)显示 20 种天然氨基酸为主的化学基元的配对产生天然产物而形成的网络. 因为甘氨酸没有侧链, 从网络中移去之后, 得到图 18(b). 有些化学基元之间配对次数远远高于其他基元之间的配对. 如络氨酸和苯丙氨酸配对所产生的天然产物数目超过 43. 而脯氨酸与苯丙氨酸、色氨酸与苯丙氨酸配对产生的天然产物只发现一种. 这可能是由于这类产物对物种进化的意义不大而未能保留下来, 或者由于其他原因.

基于化学基元的药效网络还需要深入地研究, 例如, 已知的化学基元组合所产生的天然产物与靶标的关系如何? 那些不常见的化学基元组合所形成的类天然产物有哪些? 它们为什么未能在进化史上被保留下来? 它们潜在的靶标有哪些?

5　结语

化学基元是构成生物大分子和小分子的基本化学单元, 是探究自然界从简单分子砌块合成复杂天然产物的钥匙. 每个化学基元都对应于若干个可以独立存在的内源小分子, 这些内源小分子是天然产物的砌块. 所以化学基元与天然产物的砌块之间是多对多的关系, 需要认真研究才能解开天然产物的自然合成的奥秘.

分子的成药性的正确预测对于药物创新至关重要. 虽然我们早就知道分子的成药性与结构相关, 但是我们以前对分子的成药性与组成药物分子的化学基元的认识不足, 也缺少研究.

随着对化学基元组研究的不断深入, 天然产物的化学基元组装规律被逐渐揭示, 不仅使我们能够发现新的途径制备类天然化合物库以供新药筛选, 还对什么样的化学基元组合能够产生成药性好的化合物有新的认识.

研究化学基元的组装方法, 必然要研究化学反应机理. 这个机理的核心就是底物分子在化学反应过程中过渡态的构象变化. 很多药物的构象是因为与酶的底物的过渡态高度相似, 并与酶有更高的亲和力而起作用的, 因此化学基元的组装方法的研究结果能够帮助人们设计仿过渡态的化合物库. 也许在不久的将来还会出现专门研究酶反应底物过渡态的构象组学, 为药物的虚拟筛选增加新的手段.

化学基元及其对应的天然产物砌块随着药用植物个体生命周期的时间坐标变化的谱图、化学基元及

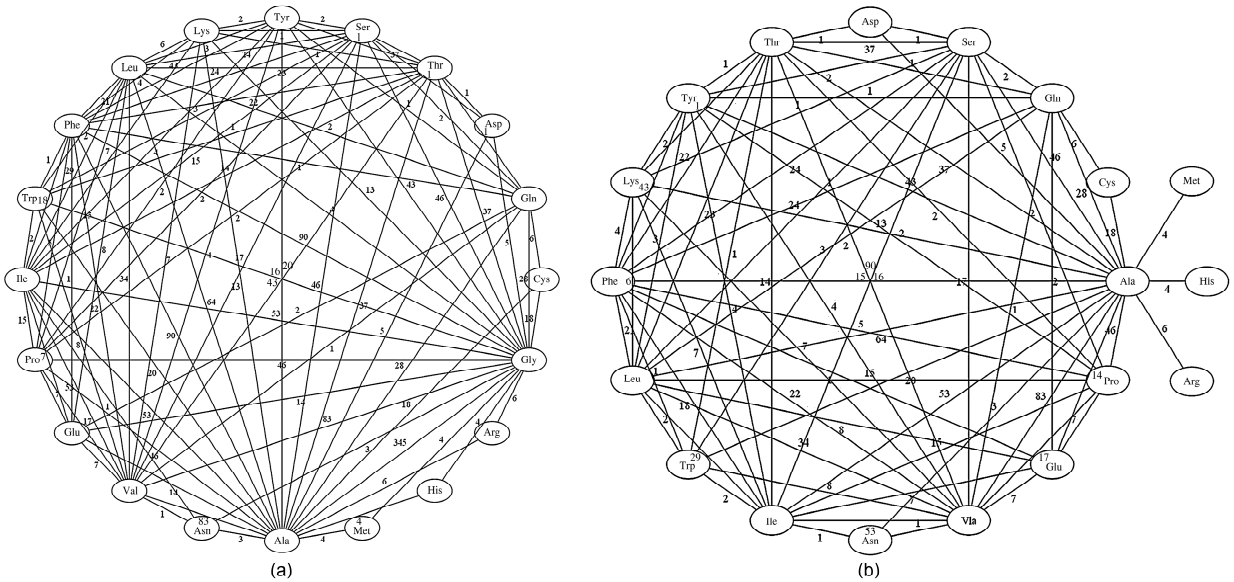


图 18　基于化学基元的药效网络. (a) 20 种天然氨基酸之间配对网络; (b) 去掉甘氨酸后形成的网络

其对应的天然产物砌块在不同物种之间的分布和演化规律, 化学基元的组合产物与生物活性的相关网络将是未来对天然产物可持续利用的指南, 也是中国传统医药学的科学发展的平台.

近年来兴起的合成生物学(synthetic biology)[31]的核心思想是从核酸开始合成细胞的基本部件, 通过人工干预的方式迫使生物体(如细菌)合成氨基酸或其他指定的天然产物.

化学基元组学与合成生物学的相似之处在于两者都追求在自然合成能力的启发下, 制备人类所需要的功能分子. 两者的不同在于化学基元组学侧重对生物合成系统的化学机理的揭示, 以化学反应的方法来模拟和简化天然产物的合成过程, 而合成生物学的终极目标是能够合成生命.

化学基元组方法的特色是从自然的角度来求解天然产物或药物分子的基本组成砌块问题. 通过揭示生物化学合成规律, 发展药物创新所需的关键性物质制备与筛选技术. 传统的天然产物化学以天然代谢物为起点, 以天然产物的化学修饰物为终点. 其缺点是产物的化学多样性受限于已知的骨架结构和现有的合成试剂与化学合成手段.

基于化学基元组学的类药化合物的设计与制备技术, 始于对生物化学合成基本砌块的解析, 终于生物化学合成基本砌块与化学合成砌块的重组和仿生合成与化学合成手段的整合, 以拓展药物筛选的化学多样性空间.

化学基元组学是交叉学科, 需要化学、生物学、信息学、和数学领域的科技工作者的共同努力.

致谢 本工作得到科技部药物创新重大专项(2010ZX09102-305)、广东省创新人才团队引进专项(2009010058)、国家高技术研究发展计划(863 计划, 2012AA020307)、国家自然科学基金(81173470)资助, 特此致谢.

参考文献

1 Collins FS, Morgan M, Patrinos A. The human genome project: lessons from large-scale biology. *Science*, 2003, 300: 286–290
2 Luscombe NM, Greenbaum D, Gerstein M. What is Bioinformatics? A proposed definition and overview of the field. *Methods Inf Med*, 2001, 40: 346–358
3 Dhar PK. The next step in biology: A periodic table. *J Biosci*, 2007, 32: 1005–1008
4 Lander ES. The new genomics: Global views of biology. *Science*, 1996, 274: 536–539
5 Biro JC, Benyó B, Sansom C, Szlávecz Á, Fördös G, Micsik T. A common periodic table of codons and amino acids. *Biochem Biophys Res Commun*, 2003, 306: 408–415
6 Van Drie JH. Pharmacophore discovery–Lessons learned. *Curr Pharm Design*, 2003, 9: 1649–1664
7 Evans BE, Rittle KE, Bock MG, Dipardo RM, Freidinger RM, Whitter WL. Methods for drug discovery: Development of potent, selective, orally effective cholecystokinin antagonists. *J Med Chem*, 1988, 31: 2235–2246
8 DeSimone RW, Currie KS, Mitchell SA, Darrow JW, Pippin DA. Privileged structures: Applications in drug discovery. *Comb Chem & High T Scr*, 2004, 7: 473–493
9 Lewell XQ, Judd DB, Watson SP, Hann MM. RECAP——Retrosynthetic combinatorial analysis procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J Chem Inf Comput Sci*, 1998, 38: 511–522
10 Corey EJ. General methods for the construction of complex molecules. *Pure Appl Chem*, 1967, 14: 19–38
11 Burke MD, Schreiber SL. A planning strategy for diversity-oriented synthesis. *Angew Chem Int Ed*, 2004, 43: 46–58
12 Böhm HJ. Site-directed structure generation by fragment-joining. *Perspect Drug Discovery Des*, 1995, 3: 21–33
13 Rees DC, Congreve M, Murray CW, Carr R. Fragment-based lead discovery. *Nat Rev Drug Discov*, 2004, 3: 660–672
14 Miflin BJ, Lea PJ. Amino acid metabolism. *Ann Rev Plant Physiol*, 1977, 28: 299–329
15 Dewick PM. *Medicinal Natural Products: A Biosynthetic Approach*. Wiley, 2009
16 Munos B. Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov*, 2009, 8: 959–968
17 Michal G. *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*. Wiley, 1999
18 Kotera M, Hirakawa M, Tokimatsu T, Goto S, Kanehisa M. The KEGG databases and tools facilitating omics analysis: Latest developments involving human diseases and pharmaceuticals, next generation microarray bioinformatics. *Methods Mol Biol*, 2012, 802: 19–39
19 Reitz M, Sacher O, Tarkhov A, Trümbach D, Gasteiger J. Enabling the exploration of biochemical pathways. *Org Biomol Chem*, 2004, 2: 3226–3237

20 Johnson RE, Washington MT, Prakash S, Prakash L. Fidelity of human DNA polymerase η. *J Biol Chem*, 2000, 275: 7447–7450
21 Pray LA. DNA replication and causes of mutation. *Nat Educ*, 2008, 1: 1
22 Brown JR, Gentry D, Becker JA, Ingraham K, Holmes DJ, Stanhope MJ. Horizontal transfer of drug-resistant, aminoacyl-transfer-RNA synthetases of anthrax and Gram-positive pathogens. *EMBO reports*, 2003, 4: 692–698
23 Strieker M, TanovićA, Marahiel MA. Nonribosomal peptide synthetases: Structures and dynamics. *Curr Opin Struct Biol*, 2010, 20: 234–240
24 Wiermann R. Secondary plant products and cell and tissue differentiation. In: Conn EE, Ed. *The Biochemistry of Plants*. New York: Academic Press, 1980. 85–137
25 Nicolaou KC, Pfefferkorn JA, Roecker AJ, Cao GQ, Barluenga S, Mitchell HJ. Natural product-like combinatorial libraries based on privileged structures. 1. General principles and solid-phase synthesis of benzopyrans. *J Am Chem Soc*, 2000, 122: 9939–9953
26 Bu XZ, Wu XM, Ng NLJ, Mak CK, Qin CG, Guo ZH. Synthesis of gramicidin S and its analogues via an on-resin macrolactamization assisted by a predisposed conformation of the linear precursors. *J Org Chem*, 2004, 69: 2681–2685
27 Zhou J, Xie G, Yan X. *Encyclopedia of Traditional Chinese Medicines—Molecular Structures, Pharmacological Activities, Natural Sources and Applications*. Berlin: Springer-Verlag, 2011
28 Janga SC, Tzakos A. Structure and organization of drug-target networks: insights from genomic approaches for drug discovery. *Mol Bio Syst*, 2009, 5: 1536–1548
29 Hopkins AL. Network pharmacology: The next paradigm in drug discovery. *Nat Chem Biol*, 2008, 4: 682–690
30 Berger SI, Iyengar R. Network analyses in systems pharmacology. *Bioinformatics*, 2009, 25: 2466–2472
31 Purnick PEM, Weiss R. The second wave of synthetic biology: From modules to systems. *Nat Rev Mol Cell Biol*, 2009, 10: 410–422

Chemomics and drug innovation

XU Jun[1*], GU Qiong[1], LIU HaiBo[2], ZHOU JiaJu[1], BU XianZhang[1], HUANG ZhiShu[1], LU Gui[1], LI Ding[1], WEI DongQing[3], WANG Ling[1], GU LianQuan[1]

1 Research Center for Drug Discovery (RCDD), School of Pharmaceutical Sciences and Institute of Human Virology, Sun Yat-Sen University, Guangzhou 510006, China
2 Institute of Medicinal Plant Development, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100194, China
3 Biostatistics College of Life Sciences and Biotechnology Shanghai Jiaotong University, Shanghai 200240, China
*Corresponding author (email: junxu@biochemomes.com)

Abstract: Chemomics is an interdisciplinary study using approaches from chemoinformatics, bioinformatics, synthetic chemistry, and other related disciplines. Biological systems make natural products from endogenous small molecules (natural product building blocks) through a sequence of enzyme catalytic reactions. For each reaction, the natural product building blocks may contribute a group of atoms to the target natural product. We describe this group of atoms as a chemoyl. A chemome is the complete set of chemoyls in an organism. Chemomics studies chemomes and the principles of natural product syntheses and evolutions. Driven by survival and reproductive demands, biological systems have developed effective protocols to synthesize natural products in order to respond to environmental changes; this results in biological and chemical diversity. In recent years, it has been realized that one of the bottlenecks in drug discovery is the lack of chemical resources for drug screening. Chemomics may solve this problem by revealing the rules governing the creation of chemical diversity in biological systems, and by developing biomimetic synthesis approaches to make quasi natural product libraries for drug screening. This treatise introduces chemomics and outlines its contents and potential applications in the fields of drug innovation.

Keywords: chemomics, chemoinformatics, bioinformatics, drug innovation, biomimetic synthesis



徐峻, 1958 年出生, 中山大学药学院教授. 1989 年毕业于中国科技大学近代化学系, 获理学博士学位. 1990~1992 年, 先后在澳大利亚国立大学(ANU)和加拿大麦基尔大学(McGill)做博士后研究. 1993~2008 年, 在美国药物工业界历任资深科学家直至研发总监, 曾任职于 TRIPOS、伯乐(BIO-RAD)、德国勃林格殷格翰制药公司(Boehringer Ingelheim Pharmaceuticals, Inc.)等. 2009 年 3 月起任中山大学药学院教授, 博士生导师、药物分子设计中心主任. 专长于研究计算机算法及其在化学、药学研究中的应用. 主要研究领域为药物化学、药物设计、化学/生物信息学、图论算法、基于科学原理的中药创新. 从事抗病毒药物、抗菌素、抗代谢类疾病的药物和抗神经退行性疾病的药物研发. 与美国、日本、德国、荷兰、澳大利亚工业界和学术界均有课题合作, 是美国洛杉矶华人作家协会会员, 著有诗歌散文集《云外心绪》(2006 年, 安徽人民出版社).