

——第五届全国计算机化工应用学术年会(1995)论文集

# 计算机 在化工中的应用

中国化工学会 计算机化工应用专业委员会

● 杨友麒 温 浩 主编



化学工业出版社

# 计算机在化工中的应用

——第五届全国计算机化工应用学术年会（1995）论文集

中国化工学会 计算机化工应用专业委员会

扬友麒 温 浩 主编

化学工业出版社

• 北京 •

(京)新登字 039 号

### 内容提要

这本论文集是由中国化工学会计算机化工应用专业委员会为举办第五届全国计算机化工应用年会而征集编辑的,它从以下几个方面反映了我国近几年在这个领域的最新进展:计算机化工应用进展综述报告;流程模拟与优化;计算机辅助设计、批处理过程设计及过程综合与集成;过程支柱模拟、过程论断与监测;过程控制;管理信息系统 MIS 与集成制造系统 CIMS;计算机图象、数值等法与数据校正。

本书有助于化工、石化企业、设计院、研究院的广大从事计算机应用的工程技术人员知识更新,也可供化工高等院校的高等级学生、研究生和教师参考。

### 图书在版编目(CIP)数据

计算机在化工中的应用 杨友麒,温浩主编-北京:化学工业出版社,1995.10  
ISBN 7-5025-1615-8

I. 计… I. ①杨…②温… III. 计算机应用-化学工业 IV. TQ015.9

中国版本图书馆 CIP 数据核字(95)第 18276 号

---

出版发行:化学工业出版社(北京市朝阳区惠新里 3 号) 社长:俸培宗 总编辑:蔡剑秋  
经 销:新华书店北京发行所  
印 刷:中国化工信息中心印刷厂  
装 订:中国化工信息中心印刷厂  
版 次:1995 年 11 月第 1 版  
印 次:1995 年 11 月第 1 次印刷  
开 本:787×1092 1/16  
印 张:23  
字 数:58  
印 数:400 册  
定 价:72.00 元

## 序 言

中国化工学会计算机化工应用专业委员会已成立十周年了，这十年正是计算机事业大发展、大普及的十年。计算机在化工领域中的应用也不例外，获得了长足进展。这种进步表现在以下一些方面：首先随着微型 PC 机价格大幅度下降，性能/价格比显著提高，应用计算机的科技人员和管理干部愈来愈多。如果说十年前计算机应用还主要是少数专业队伍的事，那么今天计算机已成为广大化工科技和管理人员手中的得力工具；其次计算机在化工中的应用领域愈来愈广泛，如果十年前还主要用于科技计算和过程控制的话，那么如今在信息处理、管理信息系统 MIS、教育培训等领域应用愈来愈多，甚至已有些单位开始向计算机集成制造系统 CIMS 进军；第三计算机应用深度也不断提高，新概念、新技术层出不穷，如网络技术、专家系统、神经网络、模糊理论、图象技术、虚拟现实……等。它们的出现给化工领域中计算机的应用提供了美好的前景，一些过去认为难以想象的事，现在已可能做到了。当然，我国在计算机应用方面与国际先进水平相比尚有不小差距，尚不能满足快速发展的社会主义市场经济发展需要。例如我们电脑应用普及率尚不够高，微机利用率还较低；我们自己的开发成果商品化化工应用软件还很少，专业队伍不管从质量上或是数量上均不能满足发展要求。但我很高兴地看到计算机化工应用专业委员会一直在不懈地努力来减少这种差距，促进这一事业在我国不断发展，第五届全国计算机化工应用年会的召开和这本论文集由化学工业出版社正式出版就是一个很好证明。

这本论文集集中反映了广大化工、石化战线上的科技工作者在计算机应用方面的研究开发成果。希望这本集子的出版能在学术交流，知识更新上相互借鉴，进一步推动计算机化工应用工作向纵深发展起到积极作用。

中国化工学会理事长

1 康连生

# 目 录

## 一、综 述

1. 新药创制中的计算机辅助分子设计技术 ..... 周家驹 谢前 谢桂荣 孙红梅 冯军 (3)
2. 过程系统最优综合的混合整数规划法 ..... 尹洪超 袁一 (8)
3. 计算机辅助设计 (CAD) 在化工设计中的应用 ..... 赵世椿 (14)
4. 炼油厂的管一控一体化 (CIPS) ..... 李松年 (18)
5. 化工过程系统工程的现状和发展趋向 ..... 钱宇 (24)
6. 我国化工信息联机检索系统的现状及发展 ..... 余乃言 (29)
7. Internet 上的化学、化工资源 ..... 李晓霞 郭力 (33)
8. 故障诊断与人工智能 ..... 麻德贤 (42)

## 二、化工过程模拟与优化

9. 低压脱甲烷系统操作模拟及优化分析 ..... 王弘轼 周沛 宋维端 (49)
10. 计算机程序在苯加氢生产环己烷工艺中的应用 ..... 张美琴 郑丽凤 (56)
11. 芳烃联合装置优化工艺计算软件 ..... 张绍华 (60)
12. PET 生产过程酯化段多指标多自由度操作优化 ..... 徐顺喜 张素贞 叶心宇 (64)
13. 石油化工生产装置在线模拟与优化技术应用探讨 ..... 庄芹仙 (69)
14. 高压煤气化流程模拟和产品气中甲烷含量预测 ..... 党梦海 (76)
15. 氯丙醇减压皂化过程的数学模拟 ..... 刘家祺 孙广智 刘禾 (84)
16. VPDB—蒸气压数据库的设计与实现 ..... 李平 梁英华 马沛生 徐明 陈鹏 (90)
17. 干法水泥厂窑尾改造工况操作参数的模拟预测研究 ..... 刘浙 (93)
18. 正丁醇与顺丁烯二酸酐反应精馏的模拟 ..... 王秀珍 李俊台 赵松钧 (98)
19. 用基团对应状态法 (CSGC) 估算不同温度下液体粘度 ..... 李平 马沛生 郭西峰 (103)
20. 过程模拟中程度设计的内外双层法 ..... 王维 张吕正 陆小华 王延儒 时钧 (108)
21. AB5 型稀土储氢化合物性质预测 ..... 沈剑韵 任洪岩 文振环 李国勋 (114)

## 三、化工过程动态仿真与计算机监测

22. 定性仿真在加氢装置仿真培训系统的应用 ..... 陶振麟 汪婕 宾伟雄 祝如松 (123)
23. 重油加氢反应系统动态仿真数学模型的建立 ..... 李镭 沈静珠 (131)
24. 重油加氢装置动态仿真系统的集成 ..... 赵春华 阎全忠 沈静珠 (138)
25. 基于 Windows 平台的 DCS 仿真培训系统开发技术 ..... 王静 俞蒙槐 (144)
26. 自校准的多路测温系统 ..... 张春远 何丕廉 (148)
27. 计算机网络化机器在线监测技术 ..... 汪家铭 (152)
28. 环保大气自动监测显示控制系统 ..... 刘天龙 索玉和 (157)
29. 模拟移动床 Parex 过程的建模与仿真 ..... 刘建国 袁立嵩 顾幸生 蒋慰孙 (163)

#### 四、CAD, 间歇过程及过程综合

30. 氧化锌脱硫剂制备参数的优化 ..... 韩寿祖 黄自兴 (171)
31. 多产品间歇化工过程最优设计—混合模拟退火方法  
..... 王春峰 权红印 许锡恩 (177)
32. 萃取精馏溶剂计算机辅助分子设计的研究 ..... 吴锦元 杨振生 (186)
33. 基于 MILP 模型的多时段非清晰精馏网络合成方法 ..... 袁希钢 刘春江 (195)
34. 低温过程系统的能量优化综合 ..... 李会泉 王世广 姚平经 (205)
35. 用夹点法合成换热网络 ..... 封瑞江 李晓鸥 萨恒福 (211)

#### 五、计算机过程诊断与控制

36. 硫酸焙烧炉映象理解及其控制策略 ..... 钱泰华 汪其华 陈文兵 程乃毅 (219)
37. 微机控制集气管系统 ..... 钟海洋 姚英海 乔鸿鹄 (223)
38. 变压吸附制氢装置的程序控制 ..... 陈永江 黄河 (228)
39. FIA-NH-01 流动注射分析仪智能控制器软件的实现 ..... 方寿海 (233)
40. 工控机湿法磷肥生产过程控制系统 ..... 钱江 黄西孟 殷向东 (238)
41. 酚醛树脂自动定量包装机的单片微机控制器 ..... 殷向东 钱江 黄西孟 (242)
42. STD 总线计算机过程监控系统和 KMM 程序控制系统的设计与应用  
..... 邹志云 林健 (250)
43. 石油炼制生产装置集散控制系统上的模式识别在线开环优化  
..... 张俊 源立基 马林年 杨彬颜 王环滔 (257)
44. 用 Turbo C 实现炉温的实时控制 ..... 殷向东 (263)
45. 芳胺氮原子上烷基化过程微机控制系统 ..... 戴晋林 张炳宏 刘子岩 (274)
46. 大型机组在线监测与诊断系统的建立与应用 ..... 杨蜀生 辜学明 (279)

#### 六、计算机信息管理系统 MIS 及集成制造系统 CIMS

47. 江苏江山药厂 JSMIS 的开发与应用 ..... 包士海 刘莹玉 (287)
48. 在市场经济体制下建立化工企业生产、经营决策系统的思考  
..... 沈祖志 谢敦礼 李浩 叶福根 吴红梅 (292)
49. 大庆化肥厂生产经营综合管理信息系统的设计 ..... 彭松涛 (297)
50. 计算机集成制造 (CIM) 与严格在线模拟 (ROM) 技术 ..... 张志橐 (304)
51. 微机局域网在日用化工中的应用 ..... 邓淑萍 (314)
52. 谈 MIS 开发过程中的问题与策略 ..... 赵吉伟 (319)
53. 过程工业 CIMS 框架及实现技术探讨 ..... 杨爱东 郭壮 袁一 (322)

#### 七、计算机图象, 数值计算方法和数据校正

54. 过程工业测量数据中的过失误差的侦破与校正 ..... 杨友麒 滕荣波 (333)
55. 过程信息流与过程计算 ..... 程乃毅 杨小健 蔡瑞英 (342)
56. 大型过程优化问题 SQP 算法的研究进展及其发展动向  
..... 方寿海 李振夏 程乃毅 (347)
57. 过程映象及其理解 ..... 程乃毅 (355)

# 一、综 述



# 新药创制中的计算机辅助分子设计技术

周家驹 谢 前 谢桂荣 孙红梅 冯 军

(中国科学院化工冶金所计算机化学开放实验室 (LCC))

## 1. 挑战

近年来,由于我国专利法进行了重大修改,改为对化合物进行保护,加之知识产权问题提上议事日程,与“化合物”产权问题密切相关的一系列在国民经济中举足轻重的产业,特别是医药、农药、精细化学品产业都面临重大的转变。逐步建立我国自己的新药研究、创制、开发的完整体系是一项紧迫的战略任务。

与工业发达国家相比,在建立这一体系中(至少)有两个关键的薄弱环节。一是缺乏先进的生物活性测定方法,即先进的药物筛选方法;二是缺乏计算机辅助分子设计软件工具。这就给我国计算机化学工作者提出了一个极具挑战性的课题。

本文主要结合近几年来 LCC 实验室开展计算机辅助分子设计(CAMD)研究的进展和体会,论述如何在工作中从国家的急需出发,进行必要的应用基础研究和充分的应用研究,尽快向广大新药研制者提供实用型软件工具。关于药物分子设计的一般性原理、方法和技术,只做概念性的扼要介绍。

## 2. 概述

顾名思义,“分子设计”意指在化学合成之前,按照人们的意愿设计出具有某种指定性能的化合物;或是从天然产物出发进行结构改造,设计性能更优越的化合物。在这里,“指定性能”是指特定的生物活性,如杀虫活性、杀菌活性、除草活性、植物生长调节活性、某种酶的抑制活性等。目前,分子设计方法是在药物设计方面形成研究热点,其实就其方法论而言,也完全适用于精细化工新产品的设计,同样有很大的潜力。

分子设计的核心问题是在分子结构的变化上做文章。按照分子结构变化的程度不同,可分为两类问题。一类是在原有化合物结构骨架不变的基础上进行修改,例如在多数情况下,只是改变一些取代基团,这种设计称为“结构优化(structure optimization)”;另一类是改变原有化合物的结构骨架,产生新的骨架,而其生物活性或其它性能指标(如对人或环境的毒性)又要优于原来的化合物,这种结构上大的变动称为“结构衍化”或“结构产生(structure generation)”。在研制新药的不同阶段,这两类问题的解决各有各的用途,均不可忽视。同时,在“结构优化”和“结构衍化”之间并无明显的界限。例如,在环上进行杂原子的替换,或用杂原子代替碳原子,既可看成是结构优化,也可看成是结构衍化。但叫什么并不特别重要,重要的是环上原子的替换常常能突破他人的专利。

应该指出,就新药创制而言,人们对结构衍化的问题感到更大的兴趣。这一方面是因为结构优化的方法相对而言较为成熟,而结构衍化的方法只是在近年来才有突破性的进展,在文献中称为“de novo”方法,即“全新”药物设计,成为研究的热点。另一方面原因是,从商业上市场竞争角度,以及专利保护的角度来看,只有在分子骨架上有大的变动,或是发现全新的结构才能突破他人的专利,发明新的专利,才具有市场竞争力。

### 3. 三类基础研究

如果把分子设计作为一种高技术（笔者赞同这一观点），那么，什么是它的研究基础呢？大体上，下列三个方面的研究构成计算机辅助分子设计的基础，即结构信息处理、结构性质关系研究及分子模型化方法。

#### 3.1. 结构信息处理

用计算机对大量的分子结构信息进行表达、存储、检索、分析等，统称为“结构信息处理”，是计算机化学基础研究的重要内容。

化合物的结构信息可分为三个层次。二维的结构信息是指分子中的原子及其连接关系，即通常所说的分子结构式。在计算机内部，二维结构信息可用二维连接表或线性代码表示，这是结构信息的基础层次。三维的结构信息是指分子中各组成原子在三维空间中的坐标。有了原子坐标，也就有了键长、键角、二面角、分子的空间构象等详尽丰富的立体化学信息，这是结构信息的核心层次，在分子设计中非常重要。第三个结构层次我们叫做电子能级层次，指由分子力学、量子化学计算产生的分子轨道，电子密度分布等更深刻地反映分子性质的信息集合。在寻找化学规律中，电子能级层次的结构信息潜力很大，应予以特别注意。

#### 3.2. 结构性质关系研究

基于实验结果，寻找结构与性质（含生物活性）的关系，也是计算机化学基础研究的内容。实际上，结构性质关系研究是化学研究的核心问题。

60年代，美国的 Hansch 教授和日本的 Fujita 教授提出了“定量结构—活性关系”（Quantitative Structure-Activity Relationship, QSAR）方法。其核心思想是用“立体”、“电子”、“疏水”这三类物理化学参数把结构信息和生物活性相关联，并应用一些具体的参数实现了建立定量方程，预报那些属于同一系列，但尚未进行合成的，当然其生物活性也是未知的化合物的生物活性，用来进行药物分子设计。至今，Hansch 方法做为 QSAR 方法中最常见的一种，仍广泛地应用于医药、农药、环境等领域。当然，30年来，QSAR 方法有了很大的发展，特别是量子化学信息的引入和 3D-QSAR 方法的提出（以“可比较的分子场分析—CoMFA 方法”为代表）。但是，从“立体”、“电子”、“疏水”三个方面来关联分子结构和生物活性，发现规律，进而根据发现的规律来设计目标分子的思想仍然具有重要的指导意义。扼要地说，立体因素就是指药物分子和受体口袋之间的空间适配性；电子因素与生化反应的机制密切相关；而疏水性则既和传输过程有关，又和药物分子和受体之间能否有效契合有关。我们认为，对 Hansch 方法不应狭义地理解为就是用传统的参数建立一个方程，而应理解为进行分子设计的一般化指导思想，这一思想是把化学（结构信息）和生物学（生物效应）连接起来的一座桥梁。

70年代，英国剑桥大学研制有机晶体结构数据库（CSD 系统）的学者们提出了“结构信息系统分析”的概念。当时的具体目标是对 CSD 系统中的十多万种有机晶体的三维结构信息进行系统的研究，以便在理论上修正一些化学发展早期形成的概念（例如球形原子的概念）。所谓“系统分析”，就是基于数据库中的大量实验数据，用统计或其它方法来发现隐含于其中的，尚未发现的结构性质关系的规律。他们认为，这种研究方法将导致化学研究方法上的大变革。事实上，剑桥学者们提出的这个概念对于新药创制同样非常重要。具体说，在建立 QSAR 模型时，不仅要充分应用本实验室的实验结果（包括合成，生测的结果），同时也要尽可能地应用他人的实验结果（特别是数据库及专利文献中的大量实验结果）。很明显，建立在

大数据集合基础上的模型要比建立在小数据集合基础上模型能提供更多的有用信息。

### 3.3. 分子模型化方法

分子模型化是一个含意相当宽的概念,泛指在计算机上描述、模拟分子的形象、行为及相互作用等,是计算机化学中十分活跃的研究领域。例如在三维图形工作站上用“搭建模型”的方法对从有机小分子到蛋白质大分子的三维形象给以直观的表达(包括分子的静态描述和动态描述)。事实上,仅此一点,就给化学家带来了极大的方便和重要的启发。长期以来,人们似乎已经习惯于用纸面上二维的分子结构式代表一个分子,但分子的形象本来是三维的。现在,在计算机屏幕上可以方便地显示分子在三维空间中的本来面目,包括蛋白质、核酸、甚至病毒。当旋转分子中任意一个 $\sigma$ 键,可得到一系列能量不同,生物活性不同的构象,计算其能量,观察不同的构象,研究何者是所谓的“药效构象”,也已经相当方便。上述搭建分子模型是分子模型化中最简单的一种。进一步,可以模拟分子的行为及相互作用,例如用分子动力学的方法来模拟药物小分子和受体大分子以及环境中的溶剂分子的相互作用。

## 4. 两类分子设计方法

就新药创制而言,主要有两类药物分子设计的方法,看所研究的具体体系的背景知识而定。如果所研究的体系在分子生物学水平上有很好的背景知识,例如已经知道某类除草剂是某种酶的抑制剂,甚至已知该靶标酶,或酶和底物的复合物的三维结构(通常首先制备该酶的单晶样本,然后用X-射线衍射法决定其三维结构,即蛋白质大分子中所有原子的三维坐标),药物设计就可以从靶标酶或复合物的三维结构出发。这被称为“基于受体的分子设计(receptor-based molecular design)”,有人用“量体裁衣”来比喻。总的来看,这种情况比较少,特别对农药而言就更少。如果虽然有一些靶标酶的知识,但酶的三维结构尚属未知,这时的药物设计可称为“不基于受体的分子设计”,有人用“量衣裁衣”来比喻。这种情况比较多,特别对农药而言就更多。由于蛋白质提纯及三维结构测定的工作通常不是在短期内能轻易完成的,而我们近年的主要任务是在农药方面,所以把主要的研究方向定在第二类方法上,以求能把所发展的药物设计方法不失时机地投入应用。与此同时,当体系的背景知识充分时,也采用基于受体结构信息的第一类方法。

## 5. 进展介绍

根据“立足基础,面向应用,面向国家急需”的指导思想,以及上述对学科和任务等方面的分析,我们把研究的重点定为:发展一般化的分子设计方法,开发普适性的软件工具。下面从几个方面交流有关的研究进展。

### 5.1. 辅助筛选的 CASAC 软件

CASAC 软件是一个能在微机上运行的“计算机辅助筛选生物活性化合物”软件。其核心功能是对一组(有共同分子骨架的)化合物分子的结构信息进行计算机自动处理,形成数值化结构因子矩阵;随后结合其活性数据自动进行多元统计和人工智能处理,在统计或模型的意义下,找出结构与活性的关系,预测未知活性化合物的活性。为有效地启动下一轮合成,设计出最可能具有高活性,高选择性的化合物名单。

使用 CASAC 软件的起点是一组实验数据,它可以来自实验室,也可以来自数据库。预测(未知化合物活性)能力的强弱取决于软件当前开发深度以及该组实验数据的质量和数量两方面,因此对实验数据的质量和数量有一定要求。

使用 CASAC 软件时,输入简单(只需一个输入文件),输出丰富(同时可有多组输出),应用面广,工作效率高。只要有共同分子骨架,软件中的 SFM 模型就可以处理结构上的多种变化。定量及定性两类活性数据均可处理。特别是无需人为地选择和试算重要结构因子,也无需选择和查找各种物理化学参数。软件中的参数库提供了 340 种结构因子的三类物化参数,包括文献中的各类传统参数,还包括我们提出的基于自洽迭代概念的基团电负性参数(EN)和基于疏水亲水基团贡献加合法的分子疏水参数。这些新的计算参数有力地支持了软件的智能化和自动化。

该软件目前包括多种模型:有用于结构信息处理的结构框架模型(SFM)、主成分分析;有用于处理定性活性数据的逐步判别分析法、星座图判别法、簇类分析法、非线性映照法、偏最小二乘法、人工神经网络法;有用于处理定量活性数据的逐步回归法(即 Hansch 方法)、多因变量回归法、人工神经网络法等。几年来,CASAC 有力地支持了不少单位创制新农药,新医药的工作。

### 5.2. 改进的神经网络方法

新近的一个发展是人工神经网络法的改进。神经网络反传算法近年来在各个领域得到广泛应用。但它有两个缺点。一是网络信息无从解释,二是有时出现过拟合现象,使至关重要的模型预报能力降低。针对这两个缺点我们进行了研究。一是用交叉验证法提出了合理的网络收敛判据,防止过拟合,保证了模型的预报能力。二是把常规的回归或判别方法和人工神经网络联合使用,使 CASAC 能更有效地处理非线性体系。首先用常规方法在诸多结构因子中确定进入模型的重要结构因子,进而用人工神经网络建立非线性模型。两种方法互补,常规法为网络提供变量及其物理解释,网络法用来建立包括高度非线性关系的精确模型。这一方法组合经多次验证,都得到良好的效果,有希望成为一种普适的研究工具。

### 5.3. 3D-QSAR 方法研究

在近年国际上提出的各种 3D-QSAR 方法中,CoMFA(比较分子场分析)是最典型的一种。该方法主要是应用探针原子进行立体场和静电场的场分析,得出三维等值线图,从而得到三维构效关系。其基本假设是药物分子与受体的作用主要是非键作用,用活性区域立体场和静电场分布的不同来解释药物分子不同的生物活性。

该法的基本步骤主要有:(1)构建分子,并进行分子力学及半经验量子力学优化;(2)输入定量的生物活性信息;(3)根据一定的重叠规则进行分子叠合;(4)按分子的大小确定一个长方体区域,在该区域内设定三维的网格,选择合适的探针按一定的步长(通常为  $2\text{\AA}$ )移动,每移动一步,计算一次探针原子和各分子的立体排斥能和静电能;(5)用偏最小二乘(PLS)法作交叉验证,建立 3D-QSAR 模型;(6)产生三维等值线图。模型建立后,就可以在工作站上根据三维等值线图的指引,随时设计感兴趣的分子,计算机则随时预报该分子的活性。实现了在合成之前,可以先在计算机上设计活性高,又易于合成的分子。我们应用 CoMFA 于几个体系。实验结果表明,当数据质量较高,结构和活性的分步较好且重叠合理时,一般能得到预报能力较强的模型。

在运用该方法时,难点在于构象的选取和重叠规则的确定。但对于刚性强的共同分子骨架,通常没有大的困难。另一点要注意的是,该法原则上适用于体外的定量活性数据。

### 5.4. 全新(de novo)分子设计

如何利用计算机化学的成就解决“结构产生”的问题,即根据已有的结构和活性信息,在计算机上设计出全新结构的先导化合物,这是最具吸引力的研究方向。正是在这一方向

上,近年来有了重大的突破,称为全新(de novo)分子设计。我们针对除草剂开发中最重要的乙酰乳酸合成酶抑制剂体系,研究了商业软件 Apex-3D 专家系统和构象搜索相结合的,发现全新结构先导化合物的分子设计实用技术,近期获得了可喜的进展。

Apex-3D 是一个能自动识别药效团的专家系统。“药效团”是具有生物活性的分子中的特定的空间结构,该特定结构被认为可能是与受体作用的活性位点。药效团是一个三维的概念,由分子中特定的中心及它们之间的距离等元素所组成。这些中心是一些具有特殊性质的原子或虚原子,虚原子可以是环中心、疏水区域、或氢键位置。该专家系统的具体实施过程如下:

- (1) 根据化合物活性将其分类(例如分为两类)。
- (2) 建立分子的三维结构并算出用于药效团识别的各种指数,如 HOMO、LUMO、原子电荷、 $\pi$  电荷密度、原子疏水性、孤对电子数等。
- (3) 识别出属于同一活性级别的所有化合物的共同结构模式(fr)。这些共同结构模式中有些是药效团,有些则不是。
- (4) 统计共同结构模式 fr 在各活性级别化合物中的出现频率。由这些频率值计算出两个用于确定药效团的统计参数:可能性(Probability)和可靠性(Reliability)。
- (5) 根据训练过程中自动优化得出的可能性阈值  $t_p$  和可靠性阈值  $t_R$  决定结构模式 fr 是否为药效团。
- (6) 把拥有同一药效团的分子叠加并进行优化,建立 3D-QSAR 模型,使活性与分子性质和活性点联系起来。
- (7) 用 3D-QSAR 模型预测尚未合成化合物的生物活性。

实施过程中一个重要的问题是多重构象的处理。如果不考虑计算机的问题,当然可对每一分子中的每一个 $\sigma$ 键进行旋转,再进行能量优化筛选出有限个数的低能构象。但实际上处理的构象数目太多是不现实的,必需采用适当的策略。

我们与南开大学合作,用合成的 31 个化合物为训练集,用该法建立的初步模型来预测文献上结构各异的 12 个已知 ALS 酶抑制剂,结果有 11 个预报为高活性。全新结构的先导化合物正在设计中。

应该说明,全新(de novo)分子设计方法刚刚提出不久,有待实践的检验。而且,在建立有效模型的方法上,思路肯定还有许多工作要做,也值得去做。

# 过程系统最优综合的混合整数规划法

尹洪超 袁一

(大连理工大学化工学院)

## 摘 要

本文扼要分析了过程系统综合的基本方法,简介了超结构混合整数规划法及其特点,综述了混合整数非线性规划算法的研究进展,评述了过程综合的混合整数规划法的发展现状与趋势。

**关键词:**过程系统 优化综合 混合整数规划法

## 1. 前言

过程系统综合作为化工系统工程学科的核心内容,是过程设计的关键步骤,它是在系统设计的初始阶段,根据不完备的信息产生各种可能的流程方案,并选择最优的系统结构<sup>[1]</sup>。这实际上是一个从众多可能方案中选择最优流程的复杂的多目标最优组合问题。长期以来主要靠工程师的经验,自 60 年代后期才开始提出了一些有关系统综合的理论和方法<sup>[2]</sup>。系统综合的基本方法可分为经验规则法、热力学目标法和数学规划法<sup>[3,4]</sup>。

(1)经验规则法(heuristic method)。它是应用一些经验积累下来的直观推断规则和调优规则,剔除一些不可能和不合理的方案,得到一个较优的可行解<sup>[5]</sup>。这种直观推断调优法是在工程实际中应用最多和最为成熟的方法,可充分发挥设计者的实践经验。随着计算机人工智能技术的发展,这些经验规则可形成一个知识库,由此建立起基于知识的专家系统,在一定的推理策略指导下来进行过程系统综合<sup>[6]</sup>。由于对不同的系统对象经验规则完全不同,因而这种方法通用性较差,亦难于进行不同子系统之间的同步协调综合。

(2)热力学目标法(thermodynamic target approaches)。这类方法是基于热力学原理和分析方法,将过程系统的能耗最少或能量利用效率最大作为系统综合的一个重要目标。其中有代表性的方法有 Goggioli 等人的有效能分析法<sup>[7]</sup>,华贲等人的热经济优化综合法<sup>[8]</sup>,梅田等人的有效能图分组综合法<sup>[9]</sup>,Linnhoff 等人的夹点技术与能量集成方法<sup>[10,11]</sup>。特别是有效能分析法和夹点技术法已在实用中取得了巨大成就,可用于不同系统的优化综合与能量集成。但这类方法主要追求系统的热力学目标,不能同时考虑能耗与费用之间的同步优化权衡,所得到的系统流程可能出现明显的不合理方案,需要进一步进行能量松弛和调优。

(3)数学规划法(optimization techniques)。数学规划法是把系统综合问题归结为一个纯数学的有约束多变量最优化模型,然后采用适当的算法得到问题的解。对这种方法有人持某些怀疑态度<sup>[3]</sup>,主要争论在于数学规划法需要掌握娴熟的建模技巧和必要的优化算法;对大规模系统综合问题其求解效率低;对严重非凸非线性混合离散模型的求解算法尚不成熟,往往只能得到局部最优解;而且运算过程不透明,无助于其物理实质的理解。但 Grossmann 等人认为<sup>[3,12,13]</sup>,经验规则法和热力学目标法虽已取得了巨大成就,但它们即不能提供一个求解不

同类型问题的通用结构化方法,亦不能保证最优;同时也不能明确考虑总体系统综合时各子系统之间的相互影响和同步优化。采用数学规划法,特别是混合整数线性规划(MILP)法,可以回避这些局限性,因而这种方法的研究引起了许多学者的极大兴趣。

## 2. 超结构混合整数规划法

利用数学规划法进行过程系统设计所采用的优化算法主要有线性规划(LP),非线性规划(NLP),混合整数线性规划(MILP)与混合整数非线性规划(MINLP)。其中前两种方法主要用于流程结构基本确定的系统进行设计参数或操作条件的有约束连续变量优化;后两种方法统称为混合整数规划(MIP)法,它不仅包含表示设计参数(如设备尺寸)和操作条件(如压力、温度、流量等)的连续性变量,而且还包含表示流程结构和单元选择的离散变量(整数变量或0—1变量),因而它可以同时进行流程结构优化和操作条件的优化。

利用混合整数规划法解决过程系统综合问题的最大困难之一在于过程流程可能的设计方案数目太多,有可能出现组合方案爆炸的问题,使模型表述和搜索求解难以实现<sup>[4]</sup>。因此需要首先建立一个包含多种可能方案的超结构(superstructure),即从一组确定的流程方案中选择最优解<sup>[3,4]</sup>。为克服数学规划法所固有的缺点,基于经验规则法和热力学目标法的优点,Grossmann提出了将三种方法结合起来应用,发挥各自的优点的策略<sup>[3,13]</sup>,即在建立包含许多可能方案的超结构中,可以采用一些经验规则或热力学方法排除那些不可能或不合理方案,同时还可用于产生一个较好的初始可行解。经过这样的初步筛选所得到的系统超结构,不仅包含那些可行的较好方案,而且搜索空间大大缩小,问题规模可以控制在合理求解的范围内。因此,数学规划法作为前两种方法的有力补充,并不排除经验规则法和热力学目标法的应用,而是与之相结合成为一种有效通用的结构化方法。在这种计算环境下,设计师仍可置身于决策过程中,发挥他们的经验特长。这无疑有助于最优解的获得,回避了经验规则法和简单分解法经常遗漏较优解的缺点集成。

利用混合整数规划法进行过程系统综合,需要在假设的超结构基础上,正确地设置连续变量 $x$ 和离散变量 $y$ ,建立描述超结构的混合整数非线性规划模型<sup>[14,15]</sup>:

$$\begin{aligned} \min \quad & c=f(x,y) \\ \text{s. t.} \quad & h(x,y)=0 \\ & g(x,y)\leq 0 \\ & x\in X=\{x|x\in R^n, x^L\leq x\leq x^U\} \\ & y\in Y=\{y\in\{0,1\}^m\} \end{aligned} \quad (1)$$

其中 $x$ 为 $n$ 维连续变量,表示系统的操作参数或设计参数,其上、下界为 $x^L, x^U$ ;  $y$ 为 $m$ 维0—1变量,表示系统中可能存在的过程单元取舍的结构参数;目标函数 $c=f(x,y)$ 一般为变量 $x, y$ 的非线性函数,表示所追求的目标,通常取系统的总费用或总收益;等式和不等式约束方程,表示超结构流程必须满足的物料平衡、能量平衡、物理性能约束、设计规定以及逻辑条件等约束。通过求解式(1)的模型,可以从超结构流程中选取在规定目标条件下的最优流程方案和操作条件。

利用混合整数规划法进行过程系统综合具有如下几个特点:

(1)可以同时进行系统结构和参数优化。由于两类变量即相互联系又相互影响,采用混合整数规划法在流程方案决策的同时也实现了连续变量的决策。

(2)可以处理用0—1变量表示的离散约束和逻辑约束。这种模型允许设计者设置一些有

关方案选择的结构约束条件,可以克服连续型优化结果中单元数太多等缺点,以获得更为满意的流程结构。

(3)混合整数规划法提供了一个系统综合的通用结构化方法。采用相同的数学工具,可以表达和求解完全不同的系统综合问题,而且可以对不同子系统之间进行同步联合优化和能量集成。

(4)在费用目标函数中采用以 0-1 变量项表示的固定费用函数,既可以反映投资费用的规模经济性,又改变了模型的非凸性。

### 3. 混合整数非线性规划算法综述

混合整数规划法是过程系统综合的一个重要工具。但对于一般形式的 MINLP 问题,曾一度认为不可求解或难于求解<sup>[13]</sup>。因此,开始人们是先将其简化为 LP 或 NLP 问题,其主要局限性在于它们一般只适用于流程拓扑结构固定的参数优化,当用于超结构系统综合时由于不能恰当地反映费用目标的规模经济性而产生非凸的局部解,而且结果中可能包含太多的单元或明显不合理的结果。

数学规划法系统综合的进一步发展,是将问题模化为一个 MILP 模型<sup>[3,13]</sup>,即式(1)中的目标函数和约束函数均为线性形式。其求解方法主要有分支定界法,隐数枚举法,割平面法等。其中比较成熟和实用的方法是分支定界法,现有多种 MILP 算法程序<sup>[4,13]</sup>,这些算法的健壮性都很好,求解效率有些差别,处理问题的规模可达几千个变量和约束,但目前对整形(0-1)变量要求不超 200 个。

MILP 方法仍是对过程系统综合的一种线性近似,它对一般的非线性系统模型无能为力。近十年来,人们开始对 MINLP 算法这一富有挑战性的研究领域予以重视。虽然人们传统地认为 MINLP 问题求解极为困难,或求解费用很高,但最新的研究进展表明对大规模 MINLP 问题的求解在技术上是可行的<sup>[16]</sup>。文献认为,求解 MINLP 的有效算法只有对模型中的 0-1 变量是与连续变量分开的线性项的情况下才是可解的。因而大多数文献都是考虑下列 MINLP 模型的求解<sup>[14,16,17]</sup>:

$$\begin{aligned} \min \quad & c = C^T y + f(x) \\ \text{s. t.} \quad & Ay + g(x) \leq 0 \\ & x \in R^n = \{x | x^L \leq x \leq x^U\} \\ & y \in \{0,1\} \end{aligned} \quad (2)$$

其中  $f(x)$ ,  $g(x)$  是只与连续变量有关的非线性凸函数(或线性函数),而 0-1 变量  $y$  在目标函数和约束中均以线性形式出现。这种假设改善了函数的非凸性,有利于最优解的获得,同时也减少了求解的难度。

对上述 MINLP 问题的算法,主要有以下几类:

(1)分支定界法(Branch and Bound Method)<sup>[13,16]</sup>。这种方法与 MILP 里的分支定界法相似,第一步先对整数条件进行松弛,求解松弛后的连续型 NLP 问题;若得到的解满足整数条件,即为原问题最优解;否则松弛的 NLP 问题的解提供了原问题最优解的一个下界(对极小化问题而言),并连续执行一个树枚举,其中在每一个节点上对应一个 NLP 问题,其中有些整变量取固定值,而其他整变量松弛为在<sup>[0,1]</sup>上取值的连续变量,这样在每个节点上对应的 NLP 问题的解提供原 MINLP 问题目标值的一个上界,目标值超过该上界的所有节点予以删除,并进一步搜索下一分支,直至得到最优解。分支定界法的最优策略在对所有节点的树枚举中只搜索

其中的一部分节点即可得到最优解。但对大规模问题所要枚举的节点数目很大,每一节点都涉及一个大规模 NLP 子问题的求解,而这个子问题不能象 MILP 中 LP 子问题那样有效地改进。因此非线性分支定界法一般计算费用很高。

(2) 广义 Benders 分解法 (Generalized Benders Decomposition Method, 简称 GBD 法)<sup>[19,14]</sup>。这种方法是通过划分变量将问题分解。Benders 分解法用于求解 MILP 问题时,是将连续变量和整形变量分开,利用线性规划可行解表示定理,把原问题转化分解为若干子问题,通过求解一系列子问题导出原问题的解。广义 Benders 分解法求解 MINLP 时也是将变量划分为若干组复杂变量和非复杂变量(MINLP 模型中的 0-1 变量一般认为是复杂变量),将 MINLP 问题分解为交替求解的 NLP 子问题和 MILP 主导问题。NLP 子问题是原 MINLP 问题中所有 0-1 变量取一组固定值时对应的连续形优化问题,其最优解为原问题的解提供一个上界;MILP 主导问题是在简约的复杂变量空间中由原问题投影得到,它为后续的 NLP 子问题给出一组新的整数变量值。当原问题为凸规划时,主导问题为原问题的解提供一个下界,该下界对每个主循环而言是单调增加的。当上、下界等于(或大于)当前上界时,过程收敛,最优解由这个界给出。这种方法所需的迭代次数一般很大,亦需求解大型 NLP 子问题。

(3) 外逼近法 (Outer-approximation Algorithm, 简称 OA 法)<sup>[20,14]</sup>。这种方法是由 Durant 和 Grossmann 提出的,在算法上与 GBD 法基本相同,即将 MINLP 问题分解为交替求解的 NLP 子问题和 MILP 主导问题,分别提供原问题解的一个上、下界。与 GBD 法不同之处在于 MILP 主导问题的给出是基于连续可行域的外部近似,这些外部近似是通过在(0-1)变量取固定值的 NLP 子问题的解点上的函数线性近似获得的。相对于 GBD 法而言,外逼近主导问题为原问题提供一个较好的下界,因而迭代次数明显减少,求解效率有所提高。

另外,Mawengkan 和 Murtagh<sup>[21]</sup>还提出一个可行性方法,其主要思想是将松弛的 NLP 的解至少圆整到一个局部下降的解,这种方法不能保证全局最优,但取得了非常好的结果。上面提到的几种 MINLP 算法,都能推广到含有非线性等式约束  $h(x)=0$  的情况。分枝定界法简单地将等式约束添加到松弛的 NLP 子问题中即可;对 GBD 法主问题不变;而 OA 法处理等式约束是通过 Kocis 与 Grossmann<sup>[14]</sup>所提出的等式松弛(OA/ER)策略,即在 NLP 子问题  $k$  次解处,线性化等式增加到主问题中,通过拉格朗日乘子的符号松弛来实现。

近年来对上述算法又作了一些推广。对于 GBD 法 Floudas 等人<sup>[22]</sup>建议划分连续变量为引进到主问题中的复杂变量和在 NLP 子问题中优化的非复杂变量。在 0-1 变量为线性,连续变量为双线性的情况下,MINLP 问题可分解成 LP 子问题(0-1 变量固定)和 MILP 主问题,LP 子问题与 MILP 主问题均可获得全局最优解,但 MINLP 全局收敛性不一定由此而得到。因为主问题的下界由于非凸性引进的拉格朗日函数不一定合理。但是 Floudas 等人显示的计算结果成功率较高。对于 OA/ER 算法,Kocis 及 Grossmann<sup>[14]</sup>提出一个两阶段策略的推广:第一阶段通过全局及局部的非凸性检验,识别非凸性;第二阶段对于证明为非凸的约束线性化,通过引进到主问题中罚函数的松弛变量进行松弛化。多个例题验证表明这个策略有 80% 的算例为全局最优。Viswanathan 及 Grossmann<sup>[16]</sup>在 OA/ER 算法的基础上,提出了在主问题中采用一个扩张惩罚函数的 AP/OA/ER 算法。这个算法通过一个松弛 NLP 问题来提供一个 0-1 变量的初始值。如果得到的是整数解,算法停止;否则,构成一个 MILP 主问题,类似于两阶段法的主问题,直到 NLP 的解不再下降,AP/OA/ER 过程停止。而当凸性条件满足时,除了初始点的选取不同外,AP/OA/ER 算法与 OA/ER 算法一致。计算经验表明这种算法对非凸性问题有较高的健壮性。