



有机物分子电性距离矢量 表征及其应用

Structural Characterization of Organic Compounds by
the Molecular Electronegativity Distance Vector (MEDV)

刘树深



高等教育出版社

HIGHER EDUCATION PRESS



有机物分子电性距离矢量 表征及其应用

Structural Characterization of Organic Compounds by
the Molecular Electronegativity Distance Vector (MEDV)

刘树深



高等教育出版社

HIGHER EDUCATION PRESS

内容提要

本书为全国优秀博士学位论文的单行本。

以有机分子二维拓扑结构中各原子类型特征和电学性质为基础,构建了三种分子电性距离矢量描述子。即:①基于4种原子类型,应用各非氢原子相对电性与相对距离信息构建的分子电性距离矢量(MEDV-4)适用于描述分子结构较为简单、杂原子种类较少的分子体系;②基于13种原子类型,充分考察各种不同化学键的相对键长构造的分子全息距离矢量(MHDV)适用于描述分子中化学键类型较多,各化合物原子类型比较一致的分子体系;③基于13种原子类型,充分考虑非氢原子在分子中的电拓扑性质的分子电性距离矢量(MEDV-13)是一种通用型分子电性距离矢量,适用于各种有机化合物结构的表征。通过对大量有机分子体系分子结构的分子电性距离矢量表征与定量结构-性质/活性相关研究,表明该矢量描述子对各种化合物结构具有良好的分辨率,与分子的各种物理化学性质或生物活性具有良好的相关性。

图书在版编目(CIP)数据

有机物分子电性距离矢量表征及其应用 / 刘树深.
—北京:高等教育出版社,2005.8
ISBN 7-04-017611-4

I. 有... II. 刘... III. 有机化合物-分子电性-研究 IV. O621.14

中国版本图书馆 CIP 数据核字(2005)第 069834 号

出版发行	高等教育出版社	购书热线	010-58581118
社 址	北京市西城区德外大街4号	免费咨询	800-810-0598
邮政编码	100011	网 址	http://www.hep.edu.cn
总 机	010-58581000		http://www.hep.com.cn
		网上订购	http://www.landaco.com
			http://www.landaco.com.cn
经 销	北京蓝色畅想图书发行有限公司		
印 刷	北京中科印刷有限公司		
开 本	850×1168 1/32		
印 张	10.25	版 次	2005年8月第1版
字 数	260 000	印 次	2005年8月第1次印刷
插 页	2	定 价	21.80 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 17611-00

作者简介



刘树深教授,男,博士,1961年4月生,湖南新化人。1988年中国科学院理学硕士,2001年重庆大学工学博士,2003年6月南京大学环境科学与工程博士后流动站出站。1999年评聘为教授,2000年起为桂林工学院分析化学重点学科带头人,2004年评聘为南京大学环境科学与工程学科博士生导师。近年来主要从事化学计量学基础及其在生命与环境科学中的应用研究,已在化学计量学新方法、多元多维光谱分析、有机物分子结构有效表征、药物分子设计、定量构效关系研究方面取得一定进展。目前的工作主要集中在分子模拟方法及其软件开发与集成、有机污染混合物毒理学及其评价方法的基础与应用研究。目前主要承担和参与国家973项目、全国优秀博士学位论文作者基金、广西中青年学科带头人基金、国家自然科学基金以及广西自然科学基金等科学研究课题多项。曾获省级自然科学二等奖、省级优秀教材二等奖、全国优秀博士学位论文奖以及省自然科学优秀论文二等奖等多项。出版《基础化学计量学》一部,参译专著一部,参编研究生教材一部。在国内外学术刊物公开发表论文50余篇,其中37篇被SCI收录,已被SCI引用130余篇次。

导师简介



蔡绍哲教授,生物医学工程博士后流动站学术带头人。

主要从事生物力学与组织工程研究,已在细胞力学、血液和血管流变学理论、检测、治疗及药物研究方面取得一定进展。目前的工作主要集中于血液循环系统的组织工程,包括人工血管、人工血细胞、心肌组织修复的基础和应用研究。

蔡教授曾是重庆大学生物工程学院的首任院长、中国生物医学工程学会常务理事、中国生物物理学会理事、第11届国际生物流变学及临床血液流变学大会秘书长、重庆学术技术带头人(生物医学工程学科)、生物医学工程学报编委、生物物理学报编委、国外医学(生物医学分册)编委、中国血液流变学杂志编委、美国人工脏器杂志(ASAIO)评委以及教育部基因组药物工程中心(暨南大学)特聘研究员。

蔡教授已培养博士后、博士、硕士30余人。他的博士生刘树深获2003年全国百篇优秀博士学位论文奖。

曾获国家自然科学三等奖,国家教委一、三等奖多项,发明专利一项、实用新型专利四项;国内外核心期刊文章90多篇(收录入SCI、EI文章20余篇),合著专著三部。

导师简介



李志良教授,男,1962年出生,汉族,博士。1989年理学博士,1992年留日博士后;1993年客座(留日非常勤)研究员,1995年教授(1994年中科院开放室客座研究员)。主攻学科为化学、应用化学、化学生物学;药学、生物药医学、新药创制。曾经完成及正在承担省、市级以上科研项目十多项。包括 a) 多维多元多模构效优化及分子设计, b) 多维多元计量化学及智能自然算法, c) 创新药物组合合成及分析表征, d) 衍生增效荧光与化学最优滤波, e) 神经网络在波谱解析中应用, f) 前后基因组分子结构表达与生物功能预测; g) 降血脂一类新药福希洛丁的研制; h) 仿生仿酶催化新技术与绿色化学制药工艺优化; i) 重庆地区有毒污染物构效关系及化学与综合治理等。先后获省第1、3届青年化学奖(1990, 1994), 部优秀科技青年(1992), 部科技进步二等奖(1995, 第1名), 省教委一等奖与国家教委三等奖(1996, 1997, 第1名), 霍英东青年教师(研究类)一等奖(1995)等。先后在 ACA, J. Chemom, Sci. China. B, TrAC, JCICS 等国内外核心学术刊物发表论文多篇, 数十篇被 SCI 及 EI 收录和上百次 SCI 引用。

前 言

2001 年作者在重庆大学生物工程学院攻读博士学位期间完成了博士论文《药物分子电性距离矢量及其应用》,2003 年经重庆大学推荐,重庆市学位委员会和国务院学位委员会组织同行专家评审,获得 2003 年度全国优秀博士学位论文。本书《有机物分子电性距离矢量表征及其应用》即是在博士论文基础上,经过精心修订与加工而成的。在这次修订中,主要根据高等教育出版社的出版要求,对论文格式和遣词造句方面作了全面修订,对所有数据进行了重新仔细核实,对所有图表按要求进行了重绘,对参考文献的格式作了统一修订并作了必要的调整。经过这次修订,力求使读者更容易了解与掌握本书介绍的方法原理,并能在实际研究工作中加以应用。

《有机物分子电性距离矢量表征及其应用》主要分为三个部分。

第一部分(第一章绪论)主要介绍矢量分子描述子的最近进展情况,使读者对分子结构表征或分子描述子的基本概念以及分子连接性、电拓扑状态、分子指纹、谱学描述子、分子比较场分析、分子相似性、WHIM 指数等一系列分子描述子有一个初步了解。

第二部分包括第二章和第三章,详细介绍作者提出的三种分子电性距离矢量的构建方法和计算实例,以及将分子电性距离矢

量应用于有机物定量结构-活性/性质相关(QSAR/QSPR)研究的技术方法。第二章在提出原子类型、原子属性、相对电性与相对距离等基本概念之后,详细地说明了基于4种原子类型和以相对电性与相对距离表征分子中各种原子类型之间相互作用而提出的基本分子电性距离矢量(MEDV-4)的方法理论;给出了基于13种原子类型,利用分子中各种化学键相对键长信息表征分子结构的分子全息距离矢量(MHDV)方法;发展了在13种原子类型和43种原子属性基础上表征复杂有机分子结构的更为通用的分子电性距离矢量方法(MEDV-13)。第三章主要说明矢量描述子研究中遇到的基本概念、模型构建与检验方法。主要内容包括矢量分子描述子的分辨率,以交互检验为目标函数的最佳子集回归、主成分回归以及偏最小二乘回归等模型化方法、模型质量评价等。

第三部分包括第四、第五和第六章,是分子电性距离矢量方法在各种 QSAR/QSPR 研究中的具体应用部分。第四章重点说明 MEDV-4 矢量分子描述子对有机分子结构的表征及其在 QSPR 研究中的应用。应用 MEDV-4 分子描述子与最佳子集回归模型化方法,建立了烷烃、烯烃、炔烃、多环芳烃等多种有机化合物的沸点、密度、临界压力、色谱保留指数等十多种物理化学性质与热力学函数的 QSPR 模型。第五章将 MHDV 和 MEDV-13 矢量分子描述子应用于几类典型生物活性物质的活性估计与预测。介绍了 MHDV 矢量描述子在两组二肽药物分子生物活性及一组取代苯化合物的生物可降解性模型中的应用,说明了 MHDV 方法的有效性。给出了 MEDV-13 方法在建立“标准”甾族化合物的亲和能模型与肠促胰酶肽-A 受体拮抗剂的生物活性的 QSAR 模型。第六章将 MEDV-13 矢量分子描述子应用于几类选择性抑制环加氧酶-2(cyclooxygenase-2, COX-2)非甾类抗炎药物分子结构的表征及其抑制活性的 QSAR 研究。

在本书完成之际,作者特别感谢导师重庆大学蔡绍哲教授和李志良教授在几年的研究学习与工作中给予的精心指导与悉心关

怀; 特别感谢中国科技大学施蕴渝院士在新型抗炎药物分子设计和定量构效关系研究方面给予的悉心指导; 特别感谢南京大学王连生教授在作者博士后研究与工作期间给予的精心指导、呵护关怀与热情鼓励; 特别感谢我的家人, 尤其是我亲爱的妻子刘海玲教授多年来始终如一的理解、支持、关心和无限的爱。正是他们多年来一如既往的指导与关怀, 才得以能够顺利完成本书。特别感谢高等教育出版社编辑的精心编辑与加工, 国务院学位委员会、教育部学位与研究生教育发展中心、国务院学位办评审专家以及高等教育出版社的大力支持使本书能够顺利出版。

由于作者水平有限, 书中缺点和错误在所难免, 恳请读者批评指正。

作 者

2005 年 3 月于南京大学科学楼

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其为人将承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人给予严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话：(010) 58581897/58581896/58581879

传 真：(010) 82086060

E - mail：dd@hep.com.cn

通信地址：北京市西城区德外大街 4 号

高等教育出版社打击盗版办公室

邮 编：100011

购书请拨打电话：(010)58581118

策划编辑 林 琳

责任编辑 应丽贞

封面设计 张 楠

责任绘图 杜晓丹

版式设计 史新薇

责任校对 尤 静

责任印制 宋克学

目 录

前言	I
第 1 章 绪论	1
1.1 引言	1
1.2 矢量分子描述子研究的某些进展	3
1.2.1 分子连接性指数	3
1.2.2 电拓扑状态指数	3
1.2.3 分子指纹	4
1.2.4 谱学描述子	6
1.2.5 CoMFA 与 GRID 描述子	7
1.2.6 与 CoMFA 相关的描述子	8
1.2.7 分子相似性描述子	9
1.2.8 WHIM 描述子	10
1.2.9 小结	11
1.3 本文的研究内容	12
参考文献	14
第 2 章 建立分子电性距离矢量	23
2.1 基于四原子类型的分子电性距离矢量(MEDV-4)	24
2.1.1 MEDV-4 中的几个基本概念	24
2.1.2 MEDV-4 的构建	29
2.1.3 MEDV-4 的计算机程序	33
2.2 分子全息距离矢量(MHDV)	35

2.2.1	原子类型、原子属性、相对键长	36
2.2.2	MHDV 的构建	37
2.3	基于 13 种原子类型的 MEDV	39
2.3.1	原子类型与原子属性	39
2.3.2	相对电性量度:修正的 E- 状态指数	42
2.3.3	MEDV-13 的构建	43
	参考文献	47
第 3 章	矢量描述子研究方法	49
3.1	矢量描述子对分子结构的分辨率	50
3.2	变量选择	53
3.2.1	描述子变量的初步选择	55
3.2.2	基于遗传算法的变量选择	56
3.2.3	基于最佳子集回归的变量选择	57
3.3	多元线性回归(MLR)	59
3.3.1	模型回归系数	60
3.3.2	MLR 中几个重要统计量	62
3.3.3	标准回归系数和置信区间	63
3.4	主成分回归(PCR)	63
3.4.1	PCR 基本原理	64
3.4.2	从 PCR 模型到 MLR 模型	65
3.5	偏最小二乘回归(PLSR)	66
3.5.1	PLSR 基本原理	66
3.5.2	从 PLSR 模型到 MLR 模型	68
3.5.3	原始描述子变量的重要性分析	69
3.6	模型质量评价	70
3.6.1	模型估计能力	70
3.6.2	模型稳定性	71
3.6.3	模型预测能力	72
3.7	应用程序 VSMP 简介	73
3.7.1	VSMP 1.0 主要模块与功能	73
3.7.2	应用程序 VSMP 的工作流程	75
	参考文献	76

第 4 章 MEDV-4 用于 QSPR 研究	79
4.1 烷烃分子结构表征与沸点 QSPR 模型	79
4.1.1 烷烃分子的 MEDV-4	80
4.1.2 150 个烷烃沸点 QSPR 模型	97
4.1.3 沸点模型的比较分析	102
4.2 烷烃其他性质 QSPR 模型	105
4.2.1 密度与折光指数	105
4.2.2 临界温度与临界压力	109
4.2.3 表面张力	112
4.2.4 摩尔体积、摩尔折光率与蒸发热	114
4.2.5 热容与热焓	117
4.2.6 液态与气态生成焓	122
4.2.7 色谱保留指数	124
4.3 链烃及醇结构表征与沸点相关	126
4.3.1 63 个烯烃	127
4.3.2 147 个链烃	133
4.3.3 106 个醇	134
4.4 芳烃化合物结构表征与性质相关	158
4.4.1 66 个不含杂原子芳烃与沸点	158
4.4.2 209 个多环芳烃与色谱保留指数	165
参考文献	181
第 5 章 生物活性物质 QSAR 研究	184
5.1 MHDV 用于两组二肽分子 QSAR 研究	185
5.1.1 二肽分子 MHDV 描述子	186
5.1.2 最佳主成分数的选择	214
5.1.3 主成分回归模型与似回归 (MLR-like) 模型	218
5.1.4 偏最小二乘回归模型与似回归模型	222
5.1.5 结果比较	225
5.2 MHDV 构建取代苯生物降解模型	226
5.2.1 取代苯分子结构表征	227
5.2.2 最佳主成分数的确定	227
5.2.3 似回归模型与检验	230

5.3 MEDV - 13 研究 31 个甾族化合物的生物活性	239
5.3.1 31 个甾族化合物的结构表征	239
5.3.2 最佳主成分与 PLS 模型	243
5.3.3 变量选择与 MLR 模型	244
5.3.4 结果比较分析	248
5.4 MEDV - 13 研究 CCK 的生物活性	254
5.4.1 CCK 化合物的结构表征	254
5.4.2 变量选择与最佳子集	255
5.4.3 多元线性回归模型	257
参考文献	261
第 6 章 几类 COX - 2 抑制剂 QSAR 研究	264
6.1 吡啶美辛及其酰胺与酯	266
6.1.1 吡啶美辛衍生物的结构表征	266
6.1.2 最佳主成分与重要描述子选择	267
6.1.3 PLSR 模型对应的似回归模型	273
6.2 二芳基环戊烯酮类(DAP)抑制剂	275
6.2.1 DAP 分子的结构表征	275
6.2.2 最佳主成分选择	276
6.2.3 QSAR 模型	279
6.3 噻唑酮与噻唑酮类(TOS)抑制剂	288
6.3.1 TOS 分子的结构表征	288
6.3.2 最佳主成分选择与 PLSR 分析	290
6.3.3 最佳子集回归与 MEDV 变量分析	296
6.4 3,4-二芳基噻唑酮(DAA)类抑制剂	302
6.4.1 DAA 分子结构表征	302
6.4.2 主成分选择与 PLSR 模型	305
参考文献	310
附录 发表与博士论文有关的学术论文	315
1 被 SCI 收录的学术论文	315
2 其他论文	317

第1章 绪 论

1.1 引 言

分子结构表征 (molecular structural characterization, 简称 MSC) 是当代药物分子设计与药效评价中不可缺少的重要部分。MSC 是药物化学、有机化学等多个学科领域中一个古老而又年轻的课题, 其手段随科学的发展与技术的进步而不断地更新、不断地完善。由于新分子的不断发现、不断创造, 特别是组合化学技术的出现, 使得在短时间内合成成千上万个分子成为可能。然而, 分子结构, 尤其是三维分子结构的实验测试技术进展缓慢, 已越来越难以跟上新分子爆炸式增长的速度, 数以万计的分子结构等待测试。近年基因组学、蛋白质组学与糖组学的相继兴起, 大量核酸、蛋白质和糖等生物大分子的结构特别是高级结构(三维)的测试难度更大, 测试的数量就更少。据报道目前已经测试过的生物大分子的独立三维结构不过 2000 余种^[1]。随着分子力学与分子动力学技术的兴起, 特别是多拷贝随机分子动力学在药物创制中的应用, 大量模拟分子的生物活性急需评价。这不仅需要相互作用能量的估

计,而且还需要定量结构与活性相关规律的建立。因此,从分子二维(2D)拓扑结构产生 2D 分子结构描述子,或先通过分子模拟技术产生三维(3D)结构,再产生 3D 分子结构描述子,进而研究定量分子结构-活性/性质相关(quantitative structure-activity/property relationships,简称 QSAR/QSPR)、完成目标靶分子与配体识别过程以及设计新型配体分子进而创制新型分子就显得特别重要和有实用意义^[2]。

大量事实表明,一个化合物的物理、化学或生物性质取决于分子中各个原子的三维构型。QSAR/QSPR 研究近年来发展迅速,并已或正在对生物分子模拟、医农药、合成化学、药物创制、环境毒性评价等与生命科学、化学科学和环境科学等多个研究领域产生越来越深刻的影响。QSAR/QSPR 技术已经成为研究物质理化性质与生物活性以寻求分子解释的所有学科领域中不可或缺的强有力工具^[3,4]。要建立 QSAR/QSPR 模型,必须首先对分子结构进行表征从而产生分子描述子(molecular descriptor,简称 MD)。目前流行的 MD 包括二维拓扑描述子、能量描述子、量子力学描述子和三维分子场描述子等^[5]。另外还有用来描述基于分子在大量生物化学实验分析中的行为^[6]或基于目标分子对大量不同的蛋白质位点^[7]而产生的分子描述子。从描述分子结构的层次考虑,这些 MD 又可分为基于二维结构的所谓 2D 分子描述子和基于三维结构的 3D 分子描述子。从分子描述子个数考虑,MD 又可分为单参数的标量描述子和多参数的“矢量”型描述子。

“矢量分子描述子”是本论文概括总结的一个概念,用来特指那些由多个等价的标量描述子共同表征一个分子结构的描述子集合。这里使用的“等价”是特别重要的,那些从不同方法产生或从不同参数库选择的编码不同结构特征信息的分子描述子集合不是矢量分子描述子。比如由编码不同结构信息的各种拓扑指数与量子力学描述子集合起来表征或描述一个分子的结构,这些 MD 集合就不是矢量分子描述子。我们提出的基于四原子类型的分子电

性距离矢量^[8,9]中包含的 10 个分子描述子,全部是基于二维分子结构的距离矩阵,以同样的计算公式从非氢原子的相对电性和原子之间的相对距离而得到的,是矢量分子描述子。又如各阶分子连接性指数的集合也是矢量分子描述子。由于分子描述子种类繁多,加上旧的方法不断改进,新的方法不断提出,要对所有这些分子描述子进行全面的分析讨论是不可能的,也是不可取的。本章仅以作者的见识和观点对属于“矢量”型的分子描述子作一些简单的讨论与分析。

1.2 矢量分子描述子研究的某些进展

1.2.1 分子连接性指数

各阶分子连接性指数系列(${}^m\chi$)集合,即零级、一级、二级、三级等分子连接性指数(${}^0\chi, {}^1\chi, {}^2\chi, {}^3\chi, \dots$)是一个典型的矢量分子描述子。分子连接性指数由 Randic^[10]提出并由 Kier 及其合作者扩展^[11~13]。业已表明分子连接性指数与多种物理化学性质存在良好的相关性,如分子极化度^[11,14]、水溶性与沸点^[14]、分配系数^[15]、范德华体积^[16]等。分子连接性指数也能够解释多种分子结构类型物质的生物化学性质包括大量环境领域中的多种应用^[17,18]。大量研究与应用表明,分子连接性是一个基于分子二维拓扑结构的最重要指数之一,可用于处理含环、多重键和含杂原子的分子体系,已成为应用最广泛的分子描述子之一^[19~22]。近年来,国内已有多本专著介绍分子连接性方法及其应用^[23,24],在此不再详述。

1.2.2 电拓扑状态指数

一个富含信息的分子描述子必须是计算快速、操作容易和易于解释的。正如在最近的几部著作^[25,26]和多篇研究论文^[27~31]中