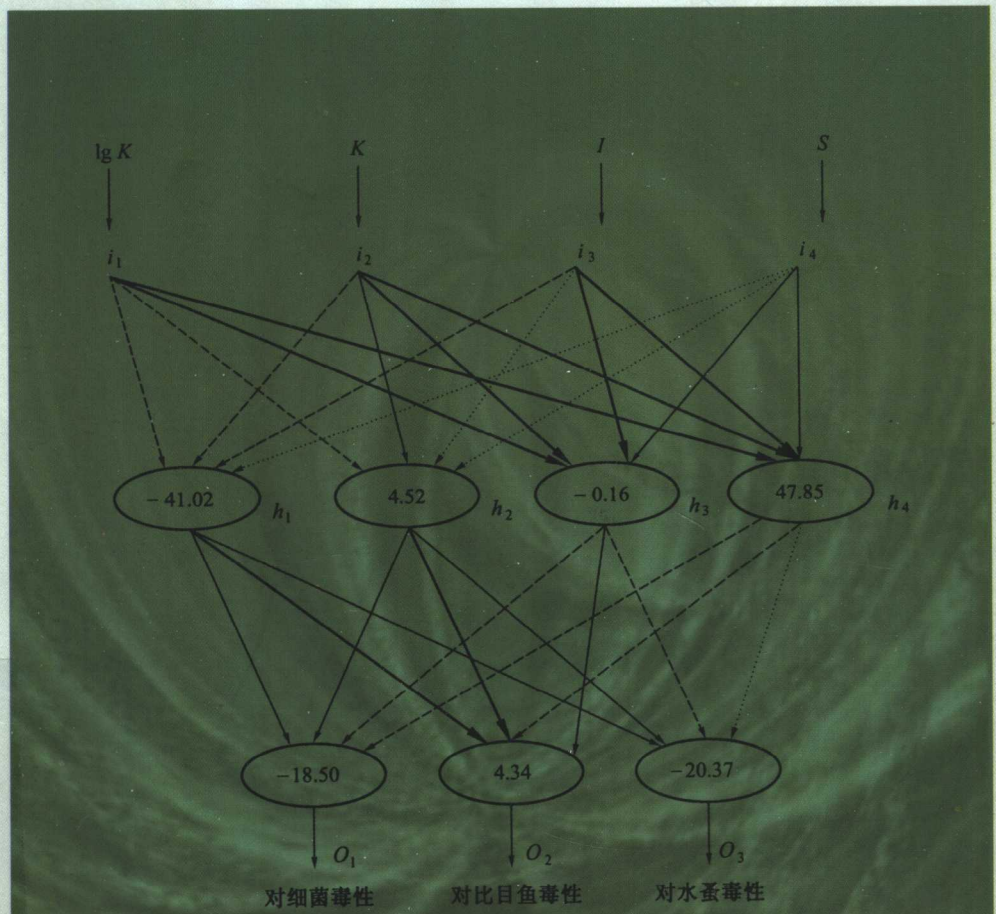
市政与环境工程系列研究生教材

定量构效关系及研究方法

王鹏 编著



哈尔滨工业大学出版社

市政与环境工程系列研究生教材

定量构效关系及研究方法

王　鹏　编著

哈尔滨工业大学出版社
哈尔滨

内 容 提 要

本书系统阐述了有机化合物定量构效关系及研究方法。全书由五章构成，分别介绍了定量构效关系的概念模式及环境科学领域应用研究的现状；定量构效关系研究中的分子结构数学表征方法，重点讨论了以分子连接性指数和自相关拓扑指数为代表的分于拓扑指数及研究方法；定量构效关系研究中的数学建模方法，重点讨论了回归分析和人工神经网络方法等。本书以定量构效关系研究方法分类成章，每章内容系统详尽，各章之间既相互联系，又相互独立自成体系；既体现了国内外在该领域的最新研究现状和前沿，又融入了作者本人的研究生课题工作成果，兼顾作为教材的系统性要求，具有较好的针对性、系统性和实用性以及较高的学术价值。

本书可作为高等学校环境科学与工程、化学、药学及其相关专业的教学用书，亦可作为相关领域的广大科技工作者的参考书和应用工具书。

图书在版编目(CIP)数据

定量构效关系及研究方法/王鹏编著.—哈尔滨：哈尔滨工业大学出版社，2004.9

(市政与环境工程系列研究生教材)

ISBN 7-5603-2073-2

Ⅰ.定… Ⅱ.王… Ⅲ.有机化合物-研究生-教教材 Ⅳ.0621

中国版本图书馆 CIP 数据核字(2004)第 084616 号

出版发行 哈尔滨工业大学出版社
社　　址 哈尔滨市南岗区教化街 21 号　邮编 150006
传　　真 0451-86414749
印　　刷 肇东粮食印刷厂
开　　本 787×1092　1/16　印张 13　字数 300 千字
版　　次 2004 年 9 月第 1 版　2004 年 9 月第 1 次印刷
书　　号 ISBN 7-5603-2073-2/X·17
印　　数 1～4 000
定　　价 20.00 元

序

多年来，化学家和药物学家对定量构效关系（QSAR）研究给予了足够的重视，环境科学家在该领域的研究是最近十几年的事。通过 QSAR 研究，使我们能基于污染物的化学结构来定量预测它们的生物毒性、化学活性和生物可降解性等环境行为。该研究属于多学科交叉的边缘学科，涉及到化学（如有机化学、结构化学）、数学（如线性代数、数理统计和拓扑学）、毒理学、环境科学、计算机科学等多门学科。该领域研究对于深入理解环境科学的基本原理、开发其在环境科学与工程领域的应用具有重要意义。

该书阐述了 QSAR 的基本原理，并对用于 QSAR 研究的化学、数学和人工智能方法进行了系统的描述。王鹏博士作为 Croucher 基金资助的访问学者曾在香港大学我的研究室工作，他严谨的科学研究作风，坚实的化学、数学和计算机应用基础给我留下了深刻的印象。我阅读了这本书的书稿，该书不仅编辑了大量的国内外最新研究文献，内容涵盖了诸如分子拓扑学、数理统计和人工智能等多门学科，而且令人信服地描述了 QSAR 在环境科学与工程领域的应用前景。

该书已被用做哈尔滨工业大学研究生的教材，它还可以作为在该领域从事科学研究人员的参考书。王鹏博士为环境科学及化学领域提供了一本有价值的著作，我认为他的努力必将会被大家所公认，并获得良好的评价。

香港大学教授　方汉平

序

化学一向被认为是一门实验科学，注重的是实验方法和结果。量子化学从微观的角度揭示了分子的组成和结构，但物质分子的性质与分子的结构特征之间又有什么样的关系，却一直未能有很好的解答。鉴于化学物质分子组成和结构的多样性，这一问题的解决显然不是一件容易的事，但这一问题的解决，特别是定量地（哪怕是部分地）弄清又确实会对科学和社会的发展，特别是化学、药学、生物学和环境毒理学的发展起重要的推动作用。因此，化合物定量构效关系（QSAR）的研究已引起了有关科学领域学者越来越多的重视，我国学者也不例外。鉴于 QSAR 的研究难度很大，又涉及到化学、数学和计算机科学等多门学科，目前的研究工作也还处于初创阶段，有关的研究队伍尚未真正形成，因此，有必要从相关学科的本科和研究生层次开设这方面的课程入手培养人才，以满足这方面的需求。王鹏老师的讲义是在数年教学实践基础上完成的，既反映了当前 QSAR 的研究成果和主要研究方向，又兼顾了相关学科学生的基础知识准备，是一本较好的具有承前启后作用的著作，值得予以正式出版。

吉林大学化学系教授　金钦汉

前　言

本书是在哈尔滨工业大学基础研究基金项目(9906731.050)、国家自然科学基金项目(50178022,50278023)研究的基础上,结合哈尔滨工业大学"环境化学"、"分子结构、性质与活性"、"分子拓扑学基础"等研究生课程的有关内容及作者近年来从事该领域研究的实践和体会,并参考了大量国内外相关文献编写而成的。

自从20世纪70年代中期,Hansch和Free－Wilson等借助计算机技术建立的结构－活性关系表达式,开创了定量构效关系(QSAR)研究的先河以来,QSAR研究在近几十年来得到迅猛的发展,并首先在定量药物设计领域取得了令人鼓舞的成功。研究和分析物质分子的基本结构特征与其从实验中表现出的一些性质及活性的关系(即所谓构效关系)已成为现代化学基础研究的重要内容,并得到了越来越多的化学、生物学、药学、环境科学研究者的重视。可以预计,对分子结构与性质或活性关系的充分阐释,将大大加速实现化学从经验科学向理论科学的过渡。作为一门新兴的交叉学科研究方向——定量构效关系及研究方法,在广泛的实际应用中显示出强大的生命力。近年来,QSAR被引入到环境科学与工程学科领域,在优化污染物筛选、化学品安全性评价、催化剂与功能材料计算机辅助分子设计、环境污染物迁移转化规律研究,水处理工程、清洁生产与绿色工艺等研究方向上取得了许多成功的范例。

有机物定量构效关系(QSAR)研究涉及数学(如图论、拓扑学、数理统计、线性代数等)、化学(如有机化学、生物化学、结构化学等)和计算机科学等多门学科。目前,系统阐述定量构效关系的书很少,特别是适合高等学校学生学习理解的有关定量构效关系的参考书和教科书就更少,而专门系统讨论QSAR研究方法的书籍尚未见出版。但愿本书的出版能起到抛砖引玉的作用。

全书比较系统地介绍了定量构效关系研究的基本原理、研究方法及在环境科学中的应用。全书共分5章。第1章对定量构效关系研究的基本情况作概括性描述,重点介绍了QSAR研究中的分子结构参数;第2章分别从污染物化学和生物降解过程的定量构效关系、环境毒理学中的定量构效关系以及生物毒性快速检测技术研究等四个方面阐述了定量构效关系在环境科学研究中应用的基本原理、相关方法和应用实例;在第3章中结合我们在该领域开展的科研工作,较为详细地介绍了分子拓扑指数及研究方法;第4章介绍了定量构效关系研究中的数学方法,包括回归分析和其他多元统计分析方法;在第5章中结合我们的研究工作介绍了目前研究比较活跃的人工神经网络方法及在QSAR研究中的应用。在本书的最后列出了相关参考文献100余篇,供读者深入学习时参考。全书自成体系,力求反映QSAR研究的前沿水平,突出其学科交叉特色,注重相关基础理论探讨,注重相关计算机程序设计与开发,并适当介绍了作者在该领域的研究工作。

参加本书编写工作的有很多是我的同事和学生。他们参加了本书中相关文献的检索、收集和整理,以及提供相关研究数据和参与部分章节的编写,他们是:陈春云(第1章,

第 4 章),高大文(第 2 章,第 5 章),龙明策(第 3 章),杨蕾、郑彤、苏建成、郭晓燕、陈传品、蔡臻超、周鑫、范志云、甄卫东、林益池、蒋益林等也参与了本书编写工作及从事大量的相关工作。本书是我们研究小组师生共同努力的结晶。本书得以出版要感谢韦永德教授、徐崇泉教授、黄君礼教授等对本研究工作的指导、关心和帮助;感谢于秀娟副教授,孟宪林副研究员对本研究工作的协助;感谢吉林大学金钦汉教授、中国科学院长春应用化学研究所汪炳武研究员、日本东京工业大学阿部光雄教授、哈尔滨工业大学周定教授、香港大学方汉平教授等对作者的培养、关心和帮助;特别感谢方汉平教授和金钦汉教授在百忙当中为本书作序。

在编写此书时参考了不少书籍和期刊,本书的出版同这些图书及有关论文的作者的辛勤工作是分不开的,在此也向他们致谢。因篇幅有限,仅择主要书刊录入参考文献中。

由于编者水平所限,书中的疏漏及不妥之处在所难免,欢迎读者批评指正。

作 者

2004 年 2 月

目　录

第 1 章　有机物定量构效关系

1.1　定量构效关系及研究现状

1.1.1　定量构效关系

有机化合物结构与活性定量相关(定量构效关系,Quantitative Structure-Activity Relationship,QSAR)的研究,最初作为定量药物设计的一个研究分支,是为了适应合理设计生物活性分子的需要而发展起来的。它对于设计和筛选生物活性显著的药物,以及阐明药物的作用机理等均具有指导作用。特别是近二三十年来,由于计算机技术的发展和应用,使 QSAR 研究提高到了一个新的水平,QSAR 的研究日益成熟,其应用范围也正在迅速扩大。目前,QSAR 不仅已成为定量药物设计的一种重要方法,而且在环境化学、环境毒理学等领域中也得到了广泛的应用。许多环境科学研究者通过各种污染物结构 - 毒性定量关系的研究,建立了多种具有毒性预测能力的环境模型,对已进入环境的污染物及尚未投放市场的新化合物的生物活性、毒性乃至环境行为进行了成功的预测、评价和筛选,这些都说明 QSAR 在环境领域中已显示出极其广阔的应用前景。

所谓定量构效关系,就是定量地描述和研究有机物的结构与活性之间相互关系。定量构效关系分析是指利用理论计算和统计分析工具来研究系列化合物结构(包括二维分子结构、三维分子结构和电子结构)与其效应(如药物的药效学性质、药物代谢动力学参数、遗传毒性和生物活性等)之间的定量关系,即采用数字模型,借助理化参数或结构参数来描述有机小分子化合物(药物、底物、抑制剂等)与有机大分子化合物(酶、辅酶或有机大分子)或组织(受体、细胞、动物等)之间的相互作用关系。

在药物和环境研究领域中,QSAR 分析具有如下两方面的功能:

(1)根据所阐明的构效关系的结果,为设计、筛选或预测任意生物活性的化合物指明方向。

(2)根据已有的化学反应知识,探求生理活性物质与生物体系的相互作用规律,从而推论生物活性所呈现的机制。

QSAR 的要点是从化合物的结构出发来建造某种数学模型,然后运用这种模型去预测化合物的活性或性质,从而为新分子的设计、评价提供理论依据。目前,几乎所有探索化合物结构 - 活性关系的分析方法都是以统计学为基础的。最常用的方法为 Hansch 分析法和 Free-Wilson 分析法,此外,模式识别、人工智能及其他数理统计法也已得到了广泛应用。

20 世纪 60 年代,Hansch 和 Free-Wilson 分别用数理统计方法并借助计算机技术建立的结构 - 活性关系表达式,标志着 QSAR 时代的开始。在他们开创性的研究工作之后,许多新方法相继不断涌现,目前已有 20 多种方法。尽管这些方法形式多样,但都符合相同

的原理,它们的应用都是以下面的前提为基础的:

(1)假定化合物的结构和生物活性之间存在一定的关系。也就是说,结构 S 和活性 A 之间存在函数关系 $F(S,A)=0$。

(2)根据已知化合物结构 - 活性数据建立的函数 $F(S,A)=0$,可以外推至新的化合物。

(3)化合物的结构可用适当的结构描述符来表示。

1.1.2 定量构效关系研究的发展历程

结构 - 活性关系研究可以追溯到科学发展的初期,其发展历史大致可分为对结构 - 活性关系的朴素认识、对结构 - 活性定性关系研究(SAR)和对结构 - 活性定量关系研究(QSAR)三个阶段。先后出现了许多研究方法,其中有些已经在实践中得到了很好的应用。

定量构效关系的发展经历了如下几个阶段:

(1)早期朴素认识:很早以前,人们就已认识到物质的反应性与其结构之间存在着一定的关系。由于当时对物质认识水平的肤浅,这种对结构 - 活性关系的认识是朴素的和原始的。

(2)定性阶段:就在 1869 年门捷列夫提出元素周期表的几乎同一时期,Crum-Brown 和 Frazer 开创了 SAR 研究的先河,他们认为,化合物的生物活性与其结构之间存在着某种函数关系,即

$$\Psi = f(C)$$

其中,Ψ 是化合物活性的某种度量,C 代表化合物的结构特征。

SAR 研究的系统开展始于 19 世纪末 20 世纪初 Richet、Meyer 和 Overton 等人的研究。Richet 的研究发现醇和酯在水中的溶解度越大,其毒性越小;Meyer 和 Overton 等人发现,简单的中性有机物(醇、酮、酯等)对生物的麻醉效力与它的油 - 水分配系数有关。

(3)定量阶段:1964 年,Hansch 等人从研究取代基与活性的关系出发,建立了线性自由能关系模型(LFER),从而使构效关系的研究从 SAR 转向 QSAR。与此同时,Free 和 Wilson 提出了 QSAR 的取代基贡献模型。近年来,随着对分子结构的深入认识,以及数理统计方法的引入,QSAR 的研究正向三维发展,先后提出位穴模型、比较分子场方法等,不仅取得了令人欣慰的成果,而且开辟了更为广阔的应用前景。

QSAR 的研究同时也促进了分子结构的研究,先后引出了很多新的结构参数。如拓扑结构指数,包括 Hosoya 的 Z 指数(1971)、Kier 分子连接性指数 χ(1976)、Balaban 的 J 指数(1979),Simon、Crippen(1980 年)等人又引入了一系列三维结构参数,这些结构参数大大丰富和促进了 QSAR 的发展。

鉴于环境污染物的多样性和复杂性,1977 年在 Win Olsor 大湖水质国际会议和加拿大在"大湖水质协议"中也要求发展和应用 QSAR 方法。1978 年美国采用 QSAR 方法估计化学品的热力学性质及毒性分类。1983 年 8 月,在加拿大 Mc-Master 大学召开了"QSAR 在环境毒理学中的应用"研讨会,并出版了论文集。1986 年美国 EPA 出版的"Research outlook"中提出应该利用 QSAR 方法预测环境化学物质的特性及其活性。近几年来,随着平衡分配法在有机污染物环境行为研究中的突出应用,环境化学和毒理学领域 QSAR 的研究非

常活跃，Mackay 及其同事依此建立了颇有影响的泛逸度模型，QSAR 被有效地应用于沉积物质基准的研究中。

1.1.3　定量构效关系研究现状及分析

选择和设计合适的分子结构描述参数、研究采用合适的技术和方法建立 QSAR 模型以及开发快速生物活性检测体系和技术是目前 QSAR 研究的三大热点。采用有效的算法建立 QSAR 模型是 QSAR 研究的核心步骤。自 Hansch 于 1964 年构建了定量的线性自由能关系模型形成 QSAR 学科以来，经过许多研究者的努力，目前已经有多种 QSAR 模型，不同的 QSAR 模型的效果是有区别的，而且适用于不同的情况。

1.1.3.1　传统的数值模型研究

在过去的 QSAR 研究中，人们首先想到的是利用回归的方法来达到建立结构与活性的关系模型，能达到这一目的的回归方法有三种：

(1)传统的多元线性回归分析法(Multivariate Linear Regression，简称 MLR)；

(2)主成分回归分析(Principal Component Regression，简称 PCR)；

(3)偏微分最小二乘法(Partial Least Squares Regression，简称 PLS)。

QSAR 研究在刚刚起步时应用的是多元线性回归法，由于多元线性回归法能给出明确的数学表达式，因此，它在定量构效关系中的应用非常广泛。

何艺兵等人应用一级反应动力学模型研究水生生物中毒机理，推导出毒性与结构的相关方程，把该方程应用到取代芳烃化合物定量结构与活性关系的建立，取得了很好的结果。王连生等人采用多元逐步回归方法，在芳烃类有机物结构与活性相关的模式参数研究中，通过参数相关分析，从多个信息参数中筛选出 7 种典型分子表征参数，从理论上表述了有机物生物活性效应取决于有机物与生物靶分子的结合量和反应过程中靶分子含量。郑红等人系统地综述了电子参数 σ 在构造多元线性回归 QSAR 模型中的应用，把电子参数引入到定量构效关系中，使得 QSAR 方程相关系数精确度明显提高。靳立军等人研究的取代苯甲醛衍生物对大型蚤的 48 h 急性毒性，采用多元线性回归方法建立了 QSAR 模型，得出取代苯甲醛衍生物对大型蚤的急性毒性是一种反应性毒性机制，毒性大小主要取决于苯环上取代基的 Hammett 电效应常数的大小的结论。

虽然多元线性回归分析方法在 QSAR 模型构建中贡献很大，并对结构与毒性间的毒理学解释提供了方便。但它构建 QSAR 模型时也存在一些问题，其主要缺点是受结构变量集的维数限制。分子结构参数很多，况且还在不断增加，选用不同的结构变量集，则可以建立多种不同的构效关系，这给科学的解释回归结果带来相当大的困难，也给最终的构效关系表达式带来混乱。针对多元线性回归法的问题，研究工作者采用主成分回归分析法和偏微分最小二乘法来克服维数限制。这两种方法都先将变量数目经计算机彻底简化，两者相比，主成分分析法得到的解更有普遍性；偏微分最小二乘法运算较快，在实际应用中更受欢迎。

总的来说，这些 QSAR 模型预报能力较差。Benigni 等人于 1989 年建立的致变活性 QSAR 模型(REPAD)能以 90% 的正确率将致癌剂和非致癌剂区别开，但用它预测另一组化合物的正确率只有 60%。Hileman 等人对以回归为基础的不同 QSAR 模型进行比较研究表明，它们预测啮齿类动物致癌活性的最高正确率仅为 60% ~ 65%。

传统的数值分析法建立的 QSAR 模型缺乏预测能力的原因主要有两点：

(1)某些对化合物活性有明显激活或抑活效应的特殊子结构或分子片段很难用数值表达，在上述方法中只能忽略；

(2)化合物的构效关系一般是非线性的，而且有些结构变量彼此相关。

1.1.3.2 人工神经网络模型研究

以第二代专家系统著称的人工神经网络于 1990 年给 QSAR 研究带来定量模型化思想上的重大变革，1990 年数学家证明了带有 S 形变换的多层前馈神经网络能够相当近似多维空间的任何实型连续函数，也就是说，它有较强的模拟多元非线性体系的能力。人工神经网络大多是通过例子学习，不断修正连接权值，产生判别函数，利用判别函数对学习集进行分类和预测。由于人工神经网络的模拟和预测能力都很强，因此，人工神经网络算法更适宜构造结构与活性之间的关系，近年来，它在定量构效关系的研究中得到了广泛应用。

Villemin 等人运用误差反向传播(BP)算法的多层人工神经网络构造多环芳烃化合物结构与致癌性的关系模型，该模型把此类化合物分为两大类，即活性和非活性；模型的总预测精度达 86%。Vracko 等人利用与几何的和电子的结构有关的描述符作为结构参数来构建结构与致癌能力人工神经网络 QSAR 模型，去掉异常值后，获得预测相关系数 $R = 0.83$。Gini等人改进了含氮芳香化合物致癌性预测的 BP 算法的人工神经网络模型，输入参数是选择不同类型的分子描述符，输出参数是 TD_{50}，即给出表达致癌性的连续数字参数，依据主成分分析减少输入参数的个数，构建人工神经网络模型。在研究中使用了 104 个分子，获得相关系数 $R^2 = 0.69$，剔除 12 个异常值后，$R^2 = 0.82$。Gini 等人在混合系统内耦合专家系统和人工神经网络，该方法能够利用每个方法的优点。在构建 QSAR 模型中，除应用 BP 算法人工神经网络外，近年来 RBF 等其他算法的人工神经网络也在 QSAR 中得到了很好的应用。

在 QSAR 构建中，应用较多的结构参数是分子描述符。在构建网络前必须首先计算所要预测化合物的分子描述符，为了克服这一问题，Igor 研究了一个神经装置以表达有机化合物结构与活性间的关系，这个神经装置构建成类似生物视觉系统，并有软件支持。该方法事先没有分子描述符的计算，它的解释和预测能力相当甚至超过使用分子描述符的 QSAR 模型。

通过前面的讨论，我们知道，利用人工神经网络构建 QSAR 模型更能反应结构与活性之间的非线性关系。但采用人工神经网络 QSAR 模型比采用多元线性回归 QSAR 模型究竟好多少？关于这个问题，国内外仍有很多学者在进行研究。

王桂莲等人应用人工神经元网络进行了对多氯酚的定量构效关系研究，为了研究多氯酚结构－毒性关系，作者归纳出全部 19 种多氯酚的 3 个活性参数：对细菌(TL81)毒性(Y1)，对比目鱼毒性(Y2)，对大型水蚤毒性(Y3)；选用的 4 个结构参数为：辛醇－水分配系数($\lg K_{OW}$)，离解常数(K_a)，一阶分子连接性指数(I)，分子自由表面(S)；先对其中 12 种多氯酚的结构－活性数据进行神经网络非线性关联，再用所得到的神经网络模型预测其余 7 种多氯酚的毒性。为了比较，作者还采用多元线性回归法建立多氯酚的结构－毒性关系方程式，并进行毒性预测。经对计算值与实验值的比较表明，人工神经网络法的相关系数约为 0.99，多元线性回归法的相关系数约为 0.92，前者的百分误差也明显小于后

者。可见，人工神经网络模型在模拟和预测多氯酚的结构－毒性关系上都优于多元线性回归分析。

Tabak 等人应用 BP 算法研究有机物的结构与降解性能关系，在“学习集”中计算结果与实验结果符合得很好，正确率超过 90%，预测集中正确率也超过 90%。Aoyama 等人研究了 16 个解裂霉素抗癌药物的构效关系，人工神经网络算法的分类与预测结果均优于自适应最小二乘法(ALS)；另对 29 个芳基丙烯酰哌嗪类抗高血压活性化合物分类，正确率为 90%(优于 ALS 的 62%～76%)，经随机抽样训练神经网络得到的分类正确率为 90%，预测正确率为 75%。

石乐明等人采用 BP 人工神经网络对 97 种磺酰脲类、SUH－除草剂的两类(活与非)生物活性进行分类，发现并剔除奇异样本，分类正确率为 100%和预测正确率为 82%。孙立贤等人运用基于误差反向传播的三层人工神经元网络来研究酚类化合物的结构－活性关系，所得结果优于逐步回归法，运用全部 8 个变量的人工神经元网络所得的正确率为 100%，而用逐步回归法选得重要变量组合为($^0\chi^v$，$^4\chi^v$，$^5\chi^v$，$\lg K_{OW}$)，由此建立的相关方程表达式，其正确率只有 83.87%。沈洲等人运用人工神经网络研究含硫芳香族化合物对发光菌的毒性构效关系，并与多元线性回归方法相比较，得出多元线性回归方法的学习训练均方差为 0.012 1，预测均方差为 0.016 8；而人工神经网络算法的训练均方差为 0.002 1，预测均方差为 0.009 2；结果表明人工神经网络明显优于多元线性回归方法。

张爱茜等人研究采用误差反向传递人工神经网络预测有机化合物生物降解性能，并同运用多元线性回归预测结果相比较，结果表明，人工神经网络对这类复杂问题有极高的求解能力，预测的均方误差为 0.001 02，远低于多元线性回归方法模拟的预测误差 0.015 91。孙晞等人运用三层误差反向传播网络对 51 种胺类有机物进行了结构－毒性关系的研究，结果表明，神经网络对急性毒性 LD_{50}具有良好预测效果，大大优于多元线性回归分析和判别分析。郭明等人直接应用化合物的分子结构式产生的结构描述参量，研究了 45 个酚类化合物的麻醉毒性和分子结构之间的相关性，用多元线性回归分析和神经网络法建立了相应的数学模型，并用其预测了 5 个酚类化合物的麻醉毒性。结果表明，用神经网络所得的结果优于多元线性回归分析结果。

虽然人工神经网络模型具有非线性交换、自适应能力、自组织特性、较好的容错性、外推性等优点，并且在各个领域已经得到了广泛的应用，但目前仍然存在如下一些问题。

1.收敛速度问题

目标函数下降速度很慢，通常需数千步或更多次迭代。其原因很多，如常用的传递函数——Sigmoid 函数本身存在无穷多次导数，而多次情况下只用了一次导数，致使收敛速度很慢。另外，网络的隐含层及隐含层节点数目的选择尚无理论上的指导，仅凭经验选取，众多研究仅采用三层网络。

2.局部最优解问题

网络在学习过程中各梯度分量值趋小，停留在某一“平台”上，目标函数不再下降，达不到预定的值，学习无法继续下去。

3.学习、预测效果问题

有时网络学习效果不理想，有时学习效果理想而预测效果不理想。其影响因素很多，其中网络结构参数与学习参数的选择、样本选择及其数量为主要因素。

除以上问题外，人工神经网络模型不像传统的数值算法那样给出明确的构效关系表达式，目前的人工神经网络模型仍属于黑箱系统，即输入输出关系不明确；并且它与传统的数值算法一样，不便考虑难以数值化的化合物特殊子结构。

1.1.3.3 分子结构参数研究

分子结构参数的选择与确定，是 QSAR 研究中非常重要的环节。目前，主要有三种结构参数，即理化参数、拓扑指数和量子化学参数。

经典的 QSAR 研究主要采用理化参数来表达分子的结构信息，以分子式为基础，根据实验测得的经验参数与相应的性质如药效、污染物的生态毒性等建立定量关系式。例如，以 Hansch 方法为代表的线性自由能关系法就属于这一种，该方法是用一些取代基的理化参数如分配系数 $\lg K_{OW}$、Hammett 的 σ 电子参数、摩尔折射 MR 或立体参数 E_s 与分子的生物活性进行回归分析建立 QSAR 模型，这种方法在实际应用中已有较大的进展，也确实解决了一些实际问题。但该方法的缺点是所用的参数大多是由实验测定的，一方面是过程比较繁杂，另一方面由于分子本身的复杂性和周围环境的影响，使实验值存在一定的误差，因此，采用该方法所作的预测的可靠性还有待提高。

采用量子力学的方法对分子进行精确计算，以了解分子的全部信息，这是了解分子活性本质的好方法。但该方法计算繁琐复杂，并非具有普通基础的人能掌握的，而薛定谔(Schrodinger)方程的近似计算又会失去许多信息，使该方法受到很大的限制，因此迄今为止，该法尚未获得广泛推广。

分子连接性方法是由 Kier 和 Hall 等人根据拓扑理论，在 Randic 的分子分枝指数基础上提出和发展起来的一种新方法。该方法能根据分子结构式的直观概念对分子结构作定量描述，使分子间的结构差异实现定量化。例如，正丁烷和异丁烷在结构上存在着差异，但正丁烷与异丁烷的差异程度比起正戊烷和异戊烷的差异是大还是小呢？仅根据化学键的直观概念不能回答这个问题，而借助分子连接性指数就能解决这个问题，以此为基础就能建立分子结构和相应性质的定量关系式。

分子连接性方法不需测定所研究分子的实验参数，也不需要解复杂的薛定谔方程，只需直接根据分子的拓扑结构，就能把理化性质或生物学性质(活性)的加和性和构成性以分子连接性函数的方式译制出来。利用这种函数式，一方面可预测一些分子的未知性质；另一方面，可根据需要设计具有一定性质或活性的分子。前者可用来在化学或环境科学研究领域中评价和预测化合物的反应性和污染物的生态毒性，后者则可在药物设计或合成方面具有指导作用。

分子连接性方法由于具有方便、简单、所用指数不依赖于实验等优点，同时用分子连接性函数预测的某些理化性质其误差接近于实验误差，因此，在创建后的十多年时间内，已在多种研究领域中得到广泛的应用，大量的研究成果也反过来进一步验证了分子连接性方法的应用价值和预测能力；同时，作为一种新方法，也得到了发展和完善，指数的含意也越来越明确，特别是在用电子数和轨道数方面定义连接性指数，使从分子拓扑和电子信息角度上解释分子连接性指数成为可能，也丰富了连接性指数的结构意义，为在非统计学角度上分析和解释模型函数打下了基础。Kier 和 Hall 作为该方法的创始人，在应用推广方面也做了大量的工作，为该方法的完善做出了重大的贡献。以分子连接性指数为代表的分子拓扑指数的引入，为 QSAR 研究注入了新的活力，成为 QSAR 结构信息参数的研究

热点。

1.1.3.4　生物毒性测定技术研究

由于对化学品需求的不断增加，大量有毒化学品被释放到环境中，使自然生态环境面临巨大的威胁。迅速而简便地检测和筛选环境中众多的外来化学品，尤其是检测有毒化学品的环境毒理效应显得越来越重要。有毒化学品对生物的毒理作用主要取决于有毒化学品的毒性程度和暴露水平。目前，化学检测的手段虽然已能精确地测定化学物质，甚至是痕量的浓度，但化学物质对生物的毒性作用只能通过生物测试的手段来获得。传统的毒性试验通常采用单一物种进行实验。但进入环境的化学品数量越来越多，这种方法不能满足快速检测的需要，发展快速、简便、灵敏和低廉的微生物检测技术无疑具有重要意义。因此，生物毒性测定技术越来越受到人们的重视。

微生物接触有毒化学品后，可造成细胞内蛋白质变性、遗传物质破坏或细胞膜破裂导致胞内物质外漏，从而对微生物造成毒性危害。用适当的指标把这些危害效应反映出来，就可以对有毒化学品的毒性程度和浓度大小做出评价。根据微生物毒性试验测定的指标和在环境监测中的应用，毒性检测一般可以分为如下三大类型：细菌发光检测；细菌生长抑制、呼吸代谢速率或菌落数检测；生态效应检测。目前应用较多的生物毒性试验是前两种类型。

1.有机物对发光细菌毒性的检测

早在 1889 年就有人证实毒物能降低发光菌的发光强度，1966 年发光细菌首次被用于检测空气样品中的毒物。1978 年美国 Backman 仪器公司研制成功一种生物发光光度计（或称生物毒性测定仪），商标名称为“Microtox”，所用菌剂为明亮发光杆菌（Photobacterium Phosphoreum）NRRLB－1177 菌株的冻干粉。仪器与冻干粉均轻便可携，检测费用低廉，方法简便快速，因此，该仪器的问世推动了各国环境工作者利用发光菌进行毒性检测这一领域的研究。

在我国，自从 1981 年 Tchan 来华介绍发光细菌生物测定技术以来，对该项技术的研究和应用有了长足进展。先后研制出第一代 GDJ－2 型（中科院南京土壤所研制）和第二代 LB 系统（类似 Microtox，华东师大和南京无线电仪器厂合制）生物发光计（或称毒性测定仪），中科院南京土壤所和华东师大均研制出明亮发光杆菌的冻干粉，用于测定。二代仪器的制造和应用研究先后于 1984 年和 1986 年通过鉴定，并已应用于水、土壤（LB 系统还能用于大气）环境生物毒性的监测。顾宗濂等利用国产 GDJ－2 型生物发光计测定了 6 种重金属离子毒性和 8 种重金属离子混合液毒性，还测定了各类排污厂排放废水的毒性，证明了发光菌发光强度同重金属离子浓度呈显著负相关。Kenneth 等人研究发光菌对水和土壤的酸或非酸提取液中生物体有害物质的检测，研究表明，不用酸处理的金属比用酸处理的金属有更强的结合力。

对有机化学品进行发光菌毒性检测的目的之一是建立 QSAR 模型，进而预测有机化学品的毒性。由于目前评价有机化学品的毒性经常采用对鱼的毒性（LC_{50}）测定方法，所以有必要建立有机化学品对发光菌的毒性（EC_{50}）和对鱼的毒性（LC_{50}）的相关性（这方面工作国内外均有报道）。赵元慧等人应用 Free-wilson 法和分子连接性法研究了 46 种取代芳烃对发光菌的毒性 EC_{50}，建立结构和活性相关方程，并讨论了 EC_{50} 和 LC_{50} 的相关性。袁东星等人应用发光菌测定了 13 种氯代芳烃和 27 种硝基芳烃的毒性，比较了 17 种硝基