



国防特色教材·化学工程与技术

计算化学

JISUAN HUAXUE

聂长明 廖力夫 主编

 **北京理工大学出版社**

BEIJING INSTITUTE OF TECHNOLOGY PRESS

北京航空航天大学出版社 哈尔滨工程大学出版社

哈尔滨工业大学出版社 西北工业大学出版社



国防特色教材·化学工程与技术

计算化学

主 编 聂长明 廖力夫

北京理工大学出版社

北京航空航天大学出版社 哈尔滨工程大学出版社
哈尔滨工业大学出版社 西北工业大学出版社

内容简介

计算化学是化学、计算方法、统计学和程序设计等多学科交叉融合的一门新兴学科。它运用数学、统计学与计算机程序设计的方法,进行化学、化工的理论计算,试验设计,数据与信息处理、分类、分析和预测。

本书主要包括:数理统计分析、一元回归分析、多元校正回归基础、主成分分析与多元校正、模式识别方法、神经网络在化学中的应用、构效关系研究和分子拓扑指数、分子模拟、实验设计与优化、MATLAB在化学中的应用等。

本书可用作化学类、化学工程与工艺、制药工程、高分子材料、环境科学与工程等专业本科生和研究生教材,也可作为对计算化学有兴趣的化学、化工等专业技术人员和青年教师的参考书。

图书在版编目(CIP)数据

计算化学/聂长明,廖力夫主编. —北京:北京理工大学出版社,2010.1
国防特色教材. 化学工程与技术
ISBN 978-7-5640-2817-6

I. 计… II. ①聂…②廖… III. 计算化学-高等学校-教材 IV. O6-04

中国版本图书馆 CIP 数据核字(2009)第 225628 号

计算化学

聂长明 廖力夫 主编
责任编辑 王玲玲

*

北京理工大学出版社出版发行
北京市海淀区中关村南大街5号(100081) 发行部电话:010-68944990 传真:010-68944450
<http://www.bitpress.com.cn>

北京圣瑞伦印刷厂印刷 全国各地新华书店经销

*

开本:787毫米×960毫米 1/16 印张:21.75 字数:445千字
2010年1月第1版 2010年1月第1次印刷 印数:1~3000册

ISBN 978-7-5640-2817-6 定价:55.00元

前 言

计算化学是化学、计算方法、统计学和程序设计等多学科知识融合的一门综合性新课程。它运用数学、统计学与计算机程序设计的方法,进行化学、化工的理论计算,试验设计,数据与信息处理、分类、分析和预测。

现代的化学研究中,计算机已经成为不可缺少的有力工具。学习使用计算机解决化学领域中遇到的各种数值计算,已成为化学中不可或缺的一部分。计算化学的研究领域非常丰富,本书主要内容包括数理统计分析、一元回归分析、多元校正回归基础、主成分分析与多元校正、模式识别方法、神经网络在化学中的应用、构效关系研究和分子拓扑指数、分子模拟、实验设计与优化、MATLAB在化学中的应用等。

本书可用作化学类、化学工程与工艺、制药工程、高分子材料、环境科学与工程等专业本科生和研究生教材,也可作为对计算化学有兴趣的化学、化工等专业技术人员和青年教师的参考书。

本书由聂长明、廖力夫任主编。参加本书编写工作的有林英武、彭国文、戴益民、武亚新、姜赛红、文松年、肖方竹、徐印堂、陈炫。在本书编写过程中,我们努力按照国防特色教材的要求进行编写、统稿和定稿。我们感谢国防科工局“十一五”规划教材基金对本书的资助,并感谢南华大学领导和北京理工大学出版社在成书过程中给予的大力支持和积极帮助。同时编写此书参考了不少书籍和期刊,本书的出版同这些图书以及有关论文的作者的辛勤工作是分不开的,在此一并表示衷心的感谢。由于篇幅有限,因此本书仅将主要的书籍和期刊列入参考文献中。

由于编者水平有限和编写时间仓促,书中难免存在错误和疏漏,甚至谬误之处,敬请各位同仁和读者不吝赐教和指正。

编 者

目 录

第 0 章 绪论	1
0.1 什么是计算化学	1
0.2 计算机在化学中的应用	1
0.2.1 按化学体系分类	2
0.2.2 按计算机应用方法分类	4
0.3 计算化学的普及	5
0.4 计算化学未来的发展	6
0.5 结语	6
第 1 章 数理统计基础	8
1.1 误差	8
1.1.1 误差的定义	8
1.1.2 误差的来源	8
1.1.3 误差的类型	9
1.1.4 精密度和准确度	10
1.1.5 偶然误差的传递	11
1.1.6 系统误差的传递	12
1.2 基础统计学概念	13
1.3 区间估计	15
1.3.1 允许区间	15
1.3.2 总体均值的置信区间估计	16
1.4 结果的表示	18
1.4.1 有效数字的定义	18
1.4.2 有效数字与不确定度的关系	19
1.5 置信区间的其他应用	19
1.6 显著性检验	20
1.6.1 显著性水平	20
1.6.2 χ^2 检验	20
1.6.3 t 检验	21

1.6.4	F 检验	24
1.7	坏值的剔除	25
第2章	一元回归分析	28
2.1	一元线性回归	28
2.1.1	一元线性回归方程的求法	28
2.1.2	相关系数和显著性检验	31
2.1.3	一元线性回归的方差分析	33
2.1.4	斜率 b 和截距 a 的区间估计及斜率 b 的显著性检验	34
2.1.5	x 值和检测限的计算	36
2.1.6	标准加入法	38
2.1.7	借助回归线进行分析方法的比较	39
2.1.8	权重回归分析	41
2.2	一元非线性回归	43
第3章	多元校正分析基础	47
3.1	多元线性回归	47
3.1.1	多元线性回归的原理	47
3.1.2	多元线性回归模型的效果分析	51
3.1.3	多元线性回归的应用	56
3.2	经典最小二乘法	62
3.3	反推最小二乘法	64
第4章	主成分分析与多元校正	66
4.1	主成分分析	66
4.1.1	主成分分析的基本概念	66
4.1.2	主成分的计算原理	67
4.1.3	主成分的性质	68
4.1.4	主成分的计算方法	70
4.1.5	主成分数的判别	72
4.2	主成分回归	73
4.3	偏最小二乘回归法	74
4.4	目标因子分析	80
4.4.1	因子分析基本概念	80

4.4.2 投影矩阵	82
4.4.3 用投影矩阵进行目标因子分析	84
4.5 秩消因子分析	85
4.5.1 双线性数据矩阵	85
4.5.2 秩消因子分析的原理和步骤	87
4.5.3 广义秩消因子分析法	88
4.5.4 残差双线性分解因子分析法	89
第5章 模式识别方法	90
5.1 数据的表示及预处理	90
5.2 特征的提取和压缩	91
5.2.1 特征的提取	91
5.2.2 特征的压缩	92
5.3 相似系数和距离	93
5.4 模式识别方法	97
5.4.1 有管理的模式识别方法	97
5.4.2 无管理的模式识别方法	130
5.5 显示方法	136
5.5.1 线性映射	136
5.5.2 非线性投影	141
第6章 人工神经网络在化学中的应用	144
6.1 人工神经网络	144
6.1.1 人工神经网络的结构和功能	144
6.1.2 人工神经网络的学习方法	148
6.1.3 人工神经网络中的归一化问题	150
6.1.4 BP人工神经网络	151
6.2 人工神经网络信息流分析技术研究	154
6.2.1 ANN模型输入节点的筛选	154
6.2.2 人工神经网络的组织与运行	156
6.3 人工神经网络的应用	160
6.3.1 蛋白质的二级结构预测	160
6.3.2 谱图分析	163
6.3.3 定量构效关系	165

6.3.4	模式识别	167
6.3.5	化学反应产物及产率的预测	169
第7章	构效关系研究和分子拓扑指数	170
7.1	图论的基本概念	170
7.2	化学结构和图论	171
7.2.1	化学图	171
7.2.2	分子的矩阵表示	173
7.3	化学结构的二维连接表	176
7.4	主要的拓扑指数及其应用	176
7.4.1	距离矩阵指数	178
7.4.2	邻接矩阵指数	191
7.4.3	量子拓扑指数 a_N 及广义 a_N 指数	212
第8章	分子模拟	225
8.1	分子模拟基本方法	225
8.1.1	量子力学	225
8.1.2	分子力学	230
8.1.3	QM/MM 法	235
8.1.4	分子动力学	236
8.1.5	分子蒙特卡洛法	239
8.2	分子模拟软件简介	240
8.2.1	ChemOffice 简介	241
8.2.2	HyperChem 简介	249
8.2.3	Gaussian 简介	253
第9章	实验设计与优化	259
9.1	化学实验设计基础	259
9.1.1	试验指标	260
9.1.2	因素和水平	260
9.1.3	同时试验和序贯试验	260
9.1.4	试验最优化和解析最优化	261
9.1.5	有效实验存在的条件	261
9.1.6	实验设计的基本原理	261

9.1.7 实验设计的步骤	262
9.2 正交实验设计	263
9.2.1 正交表	263
9.2.2 正交实验设计的一般步骤	264
9.3 析因实验设计	266
9.3.1 析因设计表	266
9.3.2 析因实验设计的一般步骤	267
9.4 均匀实验设计	270
9.4.1 均匀设计表	270
9.4.2 均匀实验设计	270
9.5 响应曲面法	273
9.5.1 形状已知的响应界面法	273
9.5.2 形状未知的响应界面法	274
9.6 序贯优化:单纯形法	276
9.6.1 固定步长单纯形法	277
9.6.2 可变步长单纯形法	278
第 10 章 MATLAB 在化学中的应用	280
10.1 MATLAB 编程基础	281
10.1.1 MATLAB 的基本操作	281
10.1.2 MATLAB 的变量、常数和数据类型	282
10.1.3 数据的输入和输出	283
10.1.4 图形可视化	284
10.2 MATLAB 在化学、化工中的应用	285
10.2.1 数据插值	285
10.2.2 方程(组)求解	286
10.2.3 微分方程(组)求解	289
10.2.4 回归分析	292
附录	296
附录 1 常用概率分布表(附表 1.1~附表 1.5)	296
附录 2 常用正交设计表和均匀设计表(附表 2.1~附表 2.24)	306
附录 3 矩阵代数初阶	326
参考文献	330

第 0 章 绪 论

0.1 什么是计算化学

计算化学(Computational Chemistry)是在计算机的帮助下求解化学及相关领域问题的交叉学科,是最近十年中发展的最快的化学研究领域之一。然而,究竟什么是计算化学呢?由于目前其在各种化学研究中广泛的应用,我们并不容易给它一个很明确的定义。简单地说,计算化学是根据基本的物理化学理论(通常是量子化学),以大量的数值运算方式来探讨化学系统的性质,并依据化学理论编制出计算机程序用于计算化学分子结构和性质,模拟生物大分子的变化和状态,模拟超分子材料、纳米材料的变化过程和结构性质。计算化学主要包括基于量子理论的量子化学(Quantum Chemistry),用于研究分子反应过程机理、基于分子力学理论的分子动力学(Molecular Dynamics),基于统计物理理论的在大分子体系模拟中得到广泛应用的分子模拟(Molecular Simulation)技术及分子建模(Molecular Modeling)等领域。

在计算化学中,最常见的例子是以量子化学计算来解释实验上各种化学现象,帮助化学家以较具体的概念来了解、分析观察到的结果。除此之外,对于未知或不易观测的化学系统,计算化学还常扮演着预测的角色,提供进一步研究的方向。另外,计算化学也常被用来验证、测试、修正或发展较高层次的化学理论。

计算化学是有着悠久历史的研究领域,自 1920 年量子力学理论建立以来,许多科学家曾尝试以各种数值计算方法来深入了解原子与分子的各种化学性质。然而,在计算机广泛使用之前,此类的计算由于其复杂性而只能应用在简单的系统与高度简化的理论模型之中。因此,计算化学仍是一门须具有高度量子力学与数值分析素养的人从事的研究,而且由于其庞大的计算量,绝大部分的计算工作需依靠昂贵的大型计算机主机或高阶工作站来进行。

随着计算机硬件的飞速发展和带动,计算化学快速发展,计算化学手段和成果被广泛地应用于与国民经济相关的领域,如化工生产、药物设计、新药研制以及环境污染物监测与控制等。许多化学现象在计算的帮助下获得了成功解释,甚至实验科学中的某些错误认识也得到了纠正。同时,科学家依据计算结果获得了具有预期功能的新型材料和高效药物,在科学研究和社会效益两方面均取得了巨大成功。

0.2 计算机在化学中的应用

计算机是一种多功能的设备,可用于计算、拟合模拟、制表、绘图、选择、判别、存贮、检索、

统计、管理、自动控制、人工智能、专家系统等方面。计算机在化学中的应用可从不同角度分类：按化学体系，可从解决化学各分支学科的问题分类；按应用方法，则是从计算机的功能应用来分类。

0.2.1 按化学体系分类

1. 计算机在分析化学中的应用(简称计算分析)

(1) 数据处理

利用一元统计,可对同一项目的若干次测量数据进行统计处理,计算置信区间、标准误差、变动系数等。利用二元统计,可以计算浓度与滴定体积或浓度与吸光度之间的直线方程线性回归法)。

(2) 条件预测

根据溶液平衡原理,考虑副反应系数校正,形成精确的数学模型,可对化学分析条件进行预测,例如显色反应最合适的 pH 的预测、离子交换色谱法中淋洗液浓度和用量的预测等。在较复杂的情况下,可以利用计算数学方法。假设有 10 种金属离子与 10 种络合剂共存,它们之间的竞争反应可用迭代法预测,计算机对每种络合物用迭代法处理,获得收敛结果的给出答案,迭代 999 次仍不收敛者弃去,总共不多于 10 万个数据的计算。如果按常规方法以每个数据平均费时 6 min 计,一个人要三年半才能算完,用计算机处理不到 1 h 便可得出答案。同时也为化学分析中确定哪种离子参加反应、哪些离子被掩蔽等条件,获得可靠的预测效果。

(3) 提高选择性

准确测定指定的组分,消除干扰。一般可概括为下列两种模型。

① 平衡模型,以各种平衡常数为依据,把共存的每种平衡都写成一个方程式,形成一组方程。在测得某些未知量之后,就可把被测物质的共存干扰物质的质量分数一起计算出来。这种模型适于处理化学分析问题,但受到平衡常数的精密度和高浓度溶液中活度校正的准确度的限制。

② 当量模型,以广义的当量关系,即测定信息与被测物质量分数的关系为依据。

这些测定信息可以是滴定体积、沉淀质量、吸收、发射、电流、电压、波峰的高度或面积等。将它们组成方程组,可把多种组分的质量分数一起计算出来。这种模型适用于化学分析和各种仪器分析,准确度高于平衡模型,但也受到某些限制。

(4) 提高灵敏度

改善信噪比、提高分辨率,常采用数学方法,使原来测量不出来的量能被测出。其方法有累加平均法、导数光谱法、傅里叶变换法、信号相关法和卷积法等。

(5) 实现仪器自动化和智能化

仪器自动化发展迅速,内容包括数据采集(将仪器测得的模拟量通过模数转换电路转换为数字,以便计算机处理)、数据处理(自动记录、换算、校正、平滑)、自动控制(用程序控制进样、加压、升温、调节等操作)以及屏幕指导(操作人员不用带纸笔和操作规程,一切工作都由屏幕提示,人机对话,操作过程和结果都由机器打印记录)等。

仪器智能化是一个新的课题,是仪器自动化并配备专家系统的产物,其低级阶段是配备小型数据库,能选择实验条件,存贮、调用谱图等;其高级阶段是用专家系统指导人们工作,检查仪器,对操作人员辅导、答疑等。

2. 计算机在有机化学中的应用(简称计算有机)

(1) 谱图检索

物质的不同结构引起谱图上的不同特征。因此,谱图的检索就成为有机分析的重要手段,常用的有红外、核磁、质谱等谱图。例如,由实验测出未知物的红外谱图,把它和标准谱图对照,参照质谱数据求得相对分子质量,就可求得未知物的组成和结构。但是,标准谱图数量太大,如果有18万张标准谱图,每2s翻阅1张,一个人要半个月才能翻完一遍,还谈不上思考和比较。若将谱图信息数字化,用计算机进行检索,就可以迅速指出实测谱图与哪一张标准谱图相同,或与哪几张标准谱图相似程度最大,这将为分析者提供解决问题的线索。

(2) 差谱技术

实测谱图的可靠性通常存在一些问题,如溶剂、基体的影响,共存物质的干扰等。一般试样本身就是未知物,欲将它提纯为纯化合物测谱是困难的,这就产生了差谱技术,即用差减的方法产生相应于纯化合物的谱图。

传统的差谱是用光学方法,如利用参比溶液、双光束补偿等方法,对于识别未知浓度的干扰物质有困难。利用计算机执行差谱程序,可将干扰物质的标准谱图通过换算,与试样的谱图进行差减,达到扣除基体、数据平滑、多组分逐级差谱等效果,为有机物的成分、结构分析提供新的手段。

(3) 结构解析

结构解析方法是利用已有的光谱、波谱数据,由人工归纳出结构单元与谱图性质关系的“知识规则”,存入计算机,作为逻辑判断的标准。试样数据输入时,计算机推理判断,指出试样结构的若干种可能方案。这种方法模拟了化学专家的智能,属于“化学专家系统”的研究。

结构解析的目标是结构自动分析,将未知物在红外光谱仪、核磁共振谱仪等几台仪器上同时测谱,所得数据联机送入计算机进行实时处理。在屏幕上显示出平面或立体结构图形,不过这种工作仅在小范围内实现,要处理天然有机化合物等复杂问题为时尚早。

(4) 合成路线设计

文献中已有大量有机合成路线,这是进行新物质合成的基础,但是人们难以全部掌握如此多的合成方法。利用数据库方法把已有合成路线存入计算机中,可从不同途径加以利用:

① 逆向追溯,提出欲合成某种目标物质时,机器从已有合成路线追溯,知道该物质可由 A、B 两物质在什么条件下合成;进一步追溯 A 可由 C 和 D 合成,B 可由 E 和 F 合成,如此直到找到一些廉价易得的物质作为合成原料;

② 顺向预测,已有大批原料,让计算机判断用这些原料能合成什么有用物质;

③ 途径选择,机器找出一批合成路线后,让机器从中选出最符合要求(如成本最低、产率最高、方法最简、污染最少等)的合成路线。

0.2.2 按计算机应用方法分类

1. 数值计算

主要是利用计算数学方法,对化学各专业的数学模型进行数值计算求解。例如量子化学、结构化学中的一些演绎性的计算,分析化学中的条件预测,化工中的各种应用计算等。

2. 化学模拟

模拟是计算机应用的重要方面,化学模拟主要有:

① 数值模拟,例如,欲从工作曲线测量数据归纳成数学公式,可用曲线拟合法,这是较简单的模拟。有时用一种数值计算方法就能完成任务。

② 过程模拟,欲总结某一复杂过程的测试数据,形成整套的规律和数学模型时,可能涉及许多种数值模拟工作。过程模拟能预测反应效果,在生产中起重要指导作用。

③ 实验模拟,例如,为了弄清几种参数(反应物浓度、温度、压力)对产量的影响,可在建立数学模型后,逐个改变参数,让机器回答其产量。这样,若干小时或若干天才能完成的实验,在计算机上用若干分钟就能得出结果。

化学模拟的另一种形式是在屏幕上显示反应设备和反应现象的实体图形,或反应条件(数据)与反应结果(数据)的坐标图形。将一种操作方法或条件输入,屏幕上即显示相应的实验效果,通常用于计算机辅助教学。

3. 模式识别

在化学中应用较广的是统计模式识别法。这是一种统计处理数据,按专业要求进行分类判别的方法,适于处理多因素的综合影响。例如,根据人的毛发、血、尿中微量元素质量分数诊断疾病;根据油田水的化学成分探测油矿;根据物性数据设计新的功能材料等。

4. 数据库

数据库是一种综合服务性的软件工程。这里所谓的数据是广义的。在化学数据库中,数

据常数、谱图、文摘、操作规程、应用程序等都是“数据”。数据库能存贮大量信息,并可根据不同需要进行检索。研究者为了查明有关领域的国际现状,并在此水平上进一步提高,通常要耗费大量劳动去查阅文献,常常要求涉及某几个关键词的文献,或某人在某年间的文献等。建成了化学文献库,在使用时可以任意指定领域、要求,能在 1~2 h 内拿到全部打印资料,完成常人半年查阅文献的工作量。

5. 专家系统

专家系统是数据库与人工智能结合的产物,它把“知识规则”作为程序,让机器模拟专家的分析、推理过程,达到用机器代替或部分代替专家的目的。具体例子有:

① 酸碱平衡专家系统,内容包括知识库和检索系统,提出问题时,机器自动查出数据,找到程序,进行计算、绘图、选择判断等处理,并用专业的语言回答问题,例如,任意溶液(包括任意种组分的混合溶液)的 pH 计算,任意溶液用酸、碱进行滴定时操作规程的设计等。

② 定性分析专家系统,用帕斯卡语言编写了阳离子硫化氢系统和阴离子消去法系统,学生拿到未知试样,不用学习和查阅这种古老系统,只须按照机器提示的手续进行操作,所得现象再输入机器,如此逐步处理,就会得出“试样是什么化合物”的结论。

专家系统可以移植,利用一个专家系统的框架,改变其数据库、知识库内容,就可形成另一专业的专家系统。

0.3 计算化学的普及

由于个人计算机上的处理器(Pentium, Pentium Pro)以及外围设备(如高速内存及硬盘)的大幅进步,个人计算机的运算速度已经直逼一些传统的工作站;再加上个人计算机系统无需负担传统多人多任务系统中复杂的作业,使得个人计算机逐渐成为从事量子化学计算的一种经济而有效的工具。早期为个人计算机操作系统所发展的计算化学软件非常有限,因为数十年来大部分科学应用软件都是在 Unix 操作系统下开发出来的,而当时在个人计算机上的 Unix 操作系统都非常昂贵且不易安装与使用。与此同时,逐渐成熟的 Linux 操作系统开始被广泛地使用在个人计算机之上。Linux 是由芬兰人 Torvald Linus 开发出来的在个人计算机上执行的 Unix 操作系统。在经过数年全世界无数人的协助发展下,在 20 世纪 90 年代末期 Linux 已成为一个功能齐全的网络多人多任务操作系统。Linux 最吸引人的地方在其稳定性和不需要昂贵的计算机配备,更重要的是它基本上是免费的。比如说,大家对目前普遍使用的操作系统当机的情况并不陌生,但 Linux 工作站通常可以正常运作几个月而不出现任何问题;此外,一般个人计算机上使用的操作系统、网络、网页、邮件、档案、X-Window 等服务器软件动辄需要上万元,而这些在 Linux 操作系统中几乎都可免费取得。因此,搭配 Linux 操作系统的个人计算机现在已成为量子化学计算的一种很好的选择。当然,近几年来许多可在 Windows

操作系统下运作的计算化学软件也陆续出现,功能也在不断地增强,使得个人计算机成为当今计算化学领域中非常重要的工具。

计算化学普及的另外一个原因是图形接口的发展与使用。传统上计算工作的输入与输出都是以文字方式来表示,输入不但耗时易错,而且许多计算结果的解读也非常不易。近年来图形接口的使用大大地简化了这些过程,使得稍具计算化学知识的人都能够轻易地设计复杂的理论计算,并且能够以简单、直接的视觉效果来分析计算所得的结果。

0.4 计算化学未来的发展

在硬件上,由于计算机科技进步的速度非常快,因此不易预测长远的计算化学的发展。比如说,1990年前,当 Apple II 带起个人计算机旋风时,所用的处理器仅是频率为 1 MHz 的 8 位 Motorola 6502,内存只有 48~64 KB,资料只能存在录音带中,不到 100 KB 的软盘是一种奢侈品。1995 年前,第一代的 IBM PC XT 是使用 4.77 MHz 的 Intel 8088 处理器,内存“高达”640 KB。当时恐怕没有人能想象如今最新一代的个人计算机可配有超过 1 GHz 的 Pentium III、Pentium 4 或 Athlon 等 32 位处理器,内存可达 1 GB 以上,而超过 30 GB 的高速硬盘也已非常普遍。2000 年,Intel 又推出了 64 位的 Itanium 处理器,而且支持此处理器的 Red Hat Linux 7.1 也已发行,借由更有效率的计算与数据处理能力与对海量存储器与档案系统的支持,这可大幅提升 PC 上的科学运算能力。15 年来个人计算机在功能上有超过 1 000 倍的提升。而未来 15 年的发展又是如何呢?依照专家的估计,由于物理定律的限制,类似现有的处理器架构在功能上大概只有 10 倍的成长空间。短期来看,平行处理的技术可大幅提升运算的效率;长远而言,或许光学计算机甚至于量子计算机将会提供现今无法想象的计算速率。

然而计算化学要有真正突破性的发展,除了硬件的进步外,理论上的研发似乎更为重要。目前对于大分子的计算限于理论的复杂性只能使用分子力学或半经验法;而且就算计算机功能上能有 1 000 倍的提升,距离准确的量子仿真仍有一段距离。目前的理论方法,仅能对小于 10 个原子的系统达到化学误差(1~2 kcal^①/mol)内的准确度;而且这些准确的计算方法的计算量大约是与系统大小的 7 次方成正比。因此,计算机计算功能的提升通常无法将可准确仿真的系统加大多少。一般认为要能以计算化学准确仿真各种生物及材料系统,理论化学家需要研发出计算量仅与系统大小的平方甚至一次方成正比的准确量子化学方法。

0.5 结 语

1998 年的诺贝尔化学奖颁给了对计算化学有卓越贡献的 Walter Kohn 及 John Pople。

① 1 cal=4.185 5 J。

这是对计算化学领域的一大肯定,也显示了计算化学在现今化学研究中的重要性。数十年来,借由计算化学的研究使得我们大幅增加了对许多化学现象的了解。随着计算能力的大幅增加,在将来的化学研究中计算化学将扮演更加重要的角色。比如说以往计算化学所研究的系统主要局限在气态中的中小型分子,但近来国际上一些主要化学期刊中已陆续出现准确仿真大型生化系统的计算化学研究。国内在这方面的起步较晚,虽然研究人才不算少,但受到重视的程度仍有待加强,并且从目前状况来看,一般做计算化学工作的大多集中于理论部分的研究和探讨,却没有自己的实验来补充到这个工作中去,而做实验测试的工作者往往不将自己的精力用于深入的理论分析,甚至认为是一种“数字游戏”,没有多大意义。对于此,我们认为科学没有贵贱之分,更不能说它有用还是没用,任何学科既然它存在就有它的价值,这种价值也许并不是在短时期内所能预见的,但是并不是说不能预见就没有意义。如何将计算化学研究工作中理论和实践统一起来,除了需要消除不同群体的观念差异之外,更需要增进不同领域工作者的交流和协作,并逐渐将二者结合起来,最终成为一个有机的整体。我们相信,目前正处于艰难摸索阶段的计算化学终将有一天能够以适当的姿态站上历史的舞台,发挥它应有的作用,为人类的进步和发展做出巨大贡献。

第 1 章 数理统计基础

1.1 误 差

1.1.1 误差的定义

误差是指测量值与真值之差,误差可表示为测量值偏离真值的程度,也可以反映测量值的不确定度。

1.1.2 误差的来源

误差必然存在于一切科学实验和测量过程中,这一点已被科学工作者所公认。误差的来源比较广泛,总体来说可以分为测量装置误差、环境误差、方法误差、人员误差、数据处理误差等几大类。

1. 测量装置误差

测量装置误差包括标准器误差、仪器误差、附件误差、结构误差、调整误差、变化性误差等六种。

2. 环境误差

许多仪器设备对使用环境都有较为严格的要求,如对温度、湿度、噪声、灰尘、加速度、电磁场、气压、光、震动等。如果使用仪器设备的环境与要求的环境状态不一致,就会引起环境误差。

3. 方法误差

方法误差是由于测量方法不完善,对标准方法进行了某些省略或测量所依据的理论不完善而引起的。

4. 人员误差

测量人员操作过程中,一些主观因素引起的误差。这些主观因素包括测量人员的心理和