

•教学研究与管理•

doi: 10.3866/PKU.DXHX201807028

www.dxhx.pku.edu.cn

大学本科化学信息学课程的设计与优化

翟红林*, 张晓昀, 姚小军, 李书艳, 贾学庆, 曹晶晶

兰州大学化学化工学院, 兰州 730000

摘要: 在分析了化学信息学课程基本内容与特点的基础之上, 结合近年来我院相关课程的教学改革与实践, 提出了大学化学信息学教学的两个层次; 重点探讨了本科生化学信息学课程体系的设计与优化, 为推进化学信息学教学改革抛砖引玉。

关键词: 化学信息学; 课程设计与优化; 教学改革与实践

中图分类号: G64; O6

Design and Optimization of the Undergraduate Cheminformatics Course

ZHAI Honglin*, ZHANG Xiaoyun, YAO Xiaojun, LI Shuyan, JIA Xueqing, CAO Jingjing

College of Chemistry & Chemical Engineering, Lanzhou University, Lanzhou 730000, P. R. China.

Abstract: In this paper, the basic implications and characteristics of the cheminformatics course are first analyzed. Combined with the teaching reform and practice of related courses in our college, we propose that the teaching of the cheminformatics course should be divided into two levels. The design and optimization of the cheminformatics course for undergraduates are discussed emphatically, which is only expected to stimulate the further work of the teaching reform of cheminformatics course.

Key Words: Cheminformatics; Design and optimization of course; Teaching reform and practice

自从化学学科出现以来, 信息的传承及对其深入的分析与挖掘在化学学科的发展中起着越来越重要的作用, 已经成为化学学科的一个重要组成部分。随着现代科学研究的不断深入、学科的多重交叉及计算机与网络等相关技术的迅猛发展, 特别是在信息时代走向大数据时代的今天, 通过对于海量数据的深入分析以发掘隐藏在其中的内在本质与规律成为包括化学学科在内的诸多学科中新兴的研究热点。化学信息学正是在这种大环境中应运而生、并快速发展起来的。2018年5月在杭州召开的中国化学会第31届学术年会上指出, 实验、理论与计算是化学的三大支柱; 随着科学与技术发展的不断深入及大数据时代的到来, 进一步表明了信息技术在化学及相关领域中应用的能力已经成为现代化学专业人才应必备的基本素质之一。

化学信息学课程的教学改革一直是一个热点, 有不少与化学信息学教学相关的教材、论文业已出版, 很多学校也将化学信息学课程列为本科生的必修课程之一, 教学安排多为36–54学时。然而, 对于化学信息学不同的理解形成了各有侧重、深浅不一的教学体系。如有的课程教学仅限于化学文献的检索; 有的则讲授理论计算化学; 有的侧重于各种统计分析技术与方法; 有的甚至还包括化合

收稿: 2018-07-12; 录用: 2018-08-21; 网络发表: 2018-09-21

*通讯作者, Email: zhahl@lzu.edu.cn

基金资助: 兰州大学教学研究项目(2017015; 2018003)

物英文命名等。这些本科生化学信息学课程的教学体系难以反映化学信息学的基本内容与特征, 教学内容或因偏颇而形成误导、或因超出本科生的接受能力而成为鸡肋, 既不利于引导学生掌握基本的信息学等技术与方法开展化学相关的科学研究, 更无助于化学信息学的健康发展。

早在 2002 年, 徐筱杰^[1]撰文提出了化学信息学的基本涵义及教学建议; 随后, 邵学广等^[2]对化学信息学及其课程建设进行了探讨, 提出了较为完整的课程体系。田正芳等^[3]针对地方高校化学信息学课程从信息资源、计算机软件、计算机能力、学生就业多元化等 4 个方面进行了思考。应该说, 这些探索与实践推动了化学信息学教学的不断深化。但值得注意的是, 现在的很多大学生的基本信息能力较之以往已经大幅提升; 同时, 随着科技发展, 化学信息学教学的要求也日益提高。

化学信息学具有多重学科交叉的特点, 其研究内容十分丰富、涉及面极为宽广。如何在有限的教学课时内使学生既能掌握化学信息处理的基本技术与方法从而更加有效地开展学习与研究, 又可以开拓学科视角、完善现代科技所需的知识结构以适应学科发展、能力提高的基本需求, 成为新形势下教学研究迫在眉睫的重要议题之一, 其关键就在于能否正确理解化学信息学、合理构建大学生化学信息学课程教学体系。

1 化学信息学课程的基本内容与特点

化学信息的运用伴随着化学及相关学科的发展而展开。1987 年度诺贝尔化学奖获得者莱恩 (Jean-Marie Lehn) 教授在其获奖报告中首次提出了“化学信息学”概念。从那时起, 化学信息学的内涵便不断地演化与更新, 从分子、原子的各种结构表示到化学文献的检索与管理, 从化学现象的表观描述到内在化学规律的深入挖掘, 化学信息学的研究内容不断地扩展, 可应用的技术与方法也越来越丰富。

目前, 化学信息学的基本定义已形成共识, 即将信息学方法、计算机及网络技术等应用于化学及相关领域的研究之中, 主要涉及化学信息的获取与表达、管理与传播, 并在此基础上进行化学信息的分析与挖掘, 进而实现知识的创新^[2,4,5]。因此, 化学信息学研究的基本内容可概括为:

- (1) 化学信息的表达(编码、可视化、虚拟化等);
- (2) 化学信息的获取(实验设计与测定、理论计算、信息检索等);
- (3) 化学信息的收集、存储、管理与提供(计算机、网络与数据库技术等);
- (4) 化学信息的解析、性质预测(数据处理与分析、图谱模拟与解析等);
- (5) 分子设计、合成路线设计、化学反应(作用)及化工过程等的模拟;
- (6) 化学信息分析中的新技术与方法研究等。

从化学信息学的基本内容与发展不难看出, 化学信息学涉及了化学、数学、信息学、计算机科学等多重学科的交叉。从原理到技术、从算法到应用, 化学信息学呈现出内容广博、多层次、多梯度、多视角的特点, 其本质就是利用信息学相关技术与方法解决化学相关问题。

2 化学信息学课程教学体系的定位与构建

尽管化学信息学在现代科技人才培养中的重要作用和地位毋庸置疑, 然而作为新兴交叉学科的化学信息学所具有的自身特点对授课教师本身就具有很大的挑战性。另一方面, 大多数化学专业学生的数理基础等相对薄弱, 加之教学课时所限, 教学内容的取舍难免呈现诸多差异。

通过不断的教学改革与实践, 我院已经形成了以大一的“大学信息技术基础”课程为入门、以大二的“化学信息学”必修课程为核心、以大三的“化学化工应用软件”等及大四的“生物信息学入门”等多门选修课程为扩展的化学专业大学生信息技术与应用三阶段教学体系, 并根据学生已有信息技术应用能力的实际及一流化学学科人才培养目标的需求不断优化、更新大学信息技术教学体系及相关课程的教学内容。由于科技的进步, 知识是讲不完、学不完的, “授之以渔”是教学的根本所在; 基础是相对的, 并非一成不变。因此, 教学内容设计的关键就在于要针对学生已有的知识

结构和认知水平,以具体知识点为支撑、技术与方法为突破口、实际应用为指引,进而达到有效拓展学生思维与应用能力的培养目标。

我们认为,化学信息学的教学必须基于学生的实际接受能力而梯度扩展;简化繁琐的数理推导、突出方法的思路及其在相关软件中的实现;内容要精选、避免大而全。因此,在教学实践中,我们将大学化学信息学教学内容分解为以下两个层次。

(1) 本科生阶段,重点在于化学信息学的基础及初步应用:

化学信息学的定义、发展及其主要内容;

化学信息的表达、文献检索及文献管理软件的使用;

常用的数据统计与分析、试验设计方法;

分子力学、分子动力学及分子模拟简介;

化学相关软件的初步使用。

(2) 研究生阶段,重点在于化学信息学研究的深入与发展:

化学信息表达新方法及其应用研究;

建模分析等数据挖掘新方法及其应用研究;

图谱模拟与解析、化学模式识别、计算机辅助合成路线设计等;

理论化学计算、分子模拟及其应用的深入研究。

大学化学信息学教学体系的界定有利于厘清本科生与研究生阶段化学信息学教学各自的目标和重点,即在化学本科生层次了解并能够应用化学信息学的基本方法与技术,而在研究生层次(特别是化学信息学相关专业)则更加深入地开展相关理论与算法的学习和研究。这样的设计不但充分考虑到学生不同阶段的知识结构与认知能力,也缓解了教学内容与教学时数的矛盾,有利于提高教学的实际效果。在夯实基础、激发兴趣的同时,给学生留下了自由扩展的线索和空间。

3 大学本科化学信息学课程设计与优化

我院本科生化学信息学课程教学安排为讲授 36 学时、上机实习 18 学时(2 学分),各章学时分配见表 1。

表 1 化学信息学课程学时分配

教学内容章目	讲授学时数	上机实习学时数
第一章 化学信息学绪论	2	
第二章 化学结构的计算机表示	4	4
第三章 化学文献检索与管理	8	8
第四章 常用数据处理方法	10	4
第五章 实验设计与优化	6	
第六章 分子力学与分子模拟初步	6	2
合 计	36	18

课堂讲授内容主要包括以下六个部分:

(1) 化学信息学绪论。

化学信息学的基本概念、研究内容与方法、发展及应用。

以化学信息学产生的历史必然、发展及取得的成就为线索,全面准确地介绍化学信息学的概念、研究内容、研究方法及应用;引导学生建立从数据获得信息加工再到知识创新的思维方式;阐明化学信息学与实验科学的关系,培养学生自觉应用化学信息学的意识。

教学中注意穿插如基于实验观测发现开普勒三大行星定律、常用现代药物设计成功案例及 2013

年诺贝尔化学奖第二次授予计算化学科学家等具有代表性的事例，引导学生思考当今的化学学科已从传统的实验科学迈向了实验、理论与计算并举的新时代，而化学信息学则成为化学研究中重要的组成部分。

(2) 化学结构的计算机表示。

化学结构的计算机表示：特点、线性表示(以 SMILES 编码为例)、连接表示(以 MOL 格式为例)、输入与可视化方法等，并简要介绍 ChemOffice、HyperChem 软件的基本功能与应用。

从已经学过的各种化学结构信息表示方法出发，总结其具有的特点与不足；引入化学结构计算机表示的特点与方法。以 ChemOffice 及 HyperChem 软件的基本功能与应用为实现手段，掌握化学结构信息的计算机输入与可视化的方法，了解常见分子结构文件格式及其相互转换，并为后续的化学文献检索、分子模拟等教学提供便利。

(3) 文献检索与学术期刊。

文献基础知识：文献的类型、特点及其对资料搜寻的指导，专利基础知识等；

重要文献数据库检索(SciFinder、Web of Science、CNKI 等)：文献数据库检索系统的特点(如有无结构与反应的检索等)、检索的两大策略(关键字检索与引文检索)、一般检索方法及对于检索结果的分析与应用等；

搜索引擎(谷歌学术、百度学术)：工作原理、特点及其应用；

期刊影响因子与分区：布尔德福文献定律，正确认识影响因子与分区对期刊及研究工作的评价作用；

重要化学学术期刊：国内外重要的综合类及各专业学术期刊简介；

文献管理软件：重点介绍 EndNote 的特点与使用等；

论文 DOI 号、Open Access 出版模式、学术规范等基础知识。

随着计算机与网络技术的广泛应用，在这一部分教学内容设计中，我们舍弃了传统纸质文献检索方法的讲授，介绍基于网络的文献数据库系统等的检索与应用，强调引文索引在科学研究中的重要作用。教学的关键定位于将学生原有的资料查找水平提升到专业文献检索的高度，形成学生在学习及今后学术研究中基本的信息获取技能，增强学生的科研信息素养等。

(4) 常用的数据统计与分析。

数据基本分析：数据基本描述(均数、中位数、四分位数、频数与分布、方差、精确性与准确性等)及相关性分析等；

常用数据预处理方法与意义：异常值剔除、插值、规格化、中心化、标准化、除噪与降维等；

线性回归建模与评价分析：以一元线性回归、多元线性回归、逐步回归、主成分分析(PCA)为主，并对因子分析(FA)与平行因子分析(PARAFAC)、偏最小二乘(PLS)、交替最小二乘-多元曲线分辨(MCR-ALS)等常见建模统计与分析方法及其应用加以简介。

学生之前所接触的简单化学数据处理的相关知识大多是散落在其他不同的课程之中。因此，这部分教学内容的重点在于除对原有的知识进行系统化、规范化外，引入建模分析的思想；要特别注意对原有错误认知的纠正。教学的难点在于学生对以数学方法(模型)表征化学信息(数据)之间的关系并进行挖掘、分析的理解。为此，我们简化数学推导、突出方法的思路及强调在相关软件(如 Excel、Matlab 等)中的实现，并结合学生熟悉的化学问题及热点问题(如全国各地统计数据分析与排名、奥运会成绩分析等)讲解模型与方法的应用。

化学数据的处理与分析是化学信息学的重要组成部分，也是化学专业人才最重要的基本素质之一。教学中，不但要让学生掌握基本的方法与应用，还应结合相关学术成果(论文)向学生展示新方法、新技术在解决化学问题中的应用，以扩展学生的视野。

(5) 试验设计方法。

单因素优化：黄金分割法、Fibonacci 分数法；

多因素优化：单因素轮换法、正交设计、均匀设计。

试验设计属优化问题。各种试验设计方法都有其优势与局限，可成为优化设计方法选择的主要依据。对于多因素试验，因素间的相互作用是导致单因素轮换法难以获得理想结果的根本原因；在正交设计中，应强调自由度及其应用、正交表的选取还与结果分析所采用的方法有关、因素列的安排在考虑交互作用时须根据其交互作用表来设计等；均匀设计无正交性，其试验方案与获得的实验结果常作为建模分析的基础数据(样本)，建模方法也并不限定于线性。教学中，可将均匀设计与之前讲授的建模分析方法相关联。

(6) 分子力学与分子模拟初步。

分子力学的基本原理、势能组成及分子力场；

分子力学应用及分子动力学简介；

分子模拟基本方法及其应用简介等。

分子模拟技术是在分子水平上开展化学反应研究、功能新材料设计、生物大分子结构与功能探索，已经成为化学研究的重要组成部分。然而，因其涉及更多更深的理论化学、生物化学、药物化学等相关知识，在教学中宜采取点面结合、以问题为向导的方法。重点从分子力学的基本假定、基本原理、分子势能的组成、分子力场的组成及力场参数的获得途径讲解分子力场的特点及其适用体系；通过对比分子力学与量子化学的特点与异同，构建分子模拟的理论基础。在分子模拟教学中，可介绍化学结构建模与优化、定量构效关系(QSAR)、分子对接、分子力学、分子动力学等基本方法、相关软件及应用示例。

本科生的化学信息学是基础课程，内容不仅涉及一些公共课、专业课的知识点，还体现出不同于传统实验学科的新方法、新思维。因此，在化学信息学教学中，视点要高、眼界要宽，不仅需要关联相关课程的知识点，还应将化学信息处理技术应用于学生熟悉、关心的问题，以提高学生对化学信息学的理解和应用能力。根据化学信息学教学的基本内容与要求，适时介绍学生易于学习、可操作的相关软件，使学生能够随时将理论学习与实验实践相结合。基于化学信息学的特点，在具体的知识点教学活动中，应注意讲透核心内容利于学生理解、落实操作实现便于学生掌握、给出相关文献资料为感兴趣的学生提供进一步扩展的空间。

4 结语

从信息时代走到了大数据时代的今天，信息技术的飞速发展正在逐渐地改变着人们的生活和思维方式！充分运用信息技术已经成为现代科技人才必备的最重要的基本技能之一。正是顺应时代的进步，化学信息学才得以产生并快速发展起来，不仅成为整个化学学科发展的重要组成部分，而且有力地推动了相关学科群的发展，从而成为充满活力、颇具影响的新兴交叉领域。我们愿借此与各位同仁交流、切磋，共同提高本科生化学信息学课程教学水平，为化学专业一流人才的培养及化学信息学的健康发展添砖加瓦。

参 考 文 献

- [1] 徐筱杰. 大学化学, **2002**, 17(1), 38.
- [2] 邵学广, 蔡文生. 大学化学, **2002**, 17(3), 12.
- [3] 田正芳, 蒋小春, 熊绪杰. 黄冈师范学院学报, **2017**, 37(3), 76.
- [4] 陈泓, 曹庆文, 李梦龙. 化学研究与应用, **2004**, 16(4), 453.
- [5] 邵学广, 蔡文生. 化学信息学. 第3版. 普通高等教育“十二五”规划教材. 北京: 科学出版社, 2013.