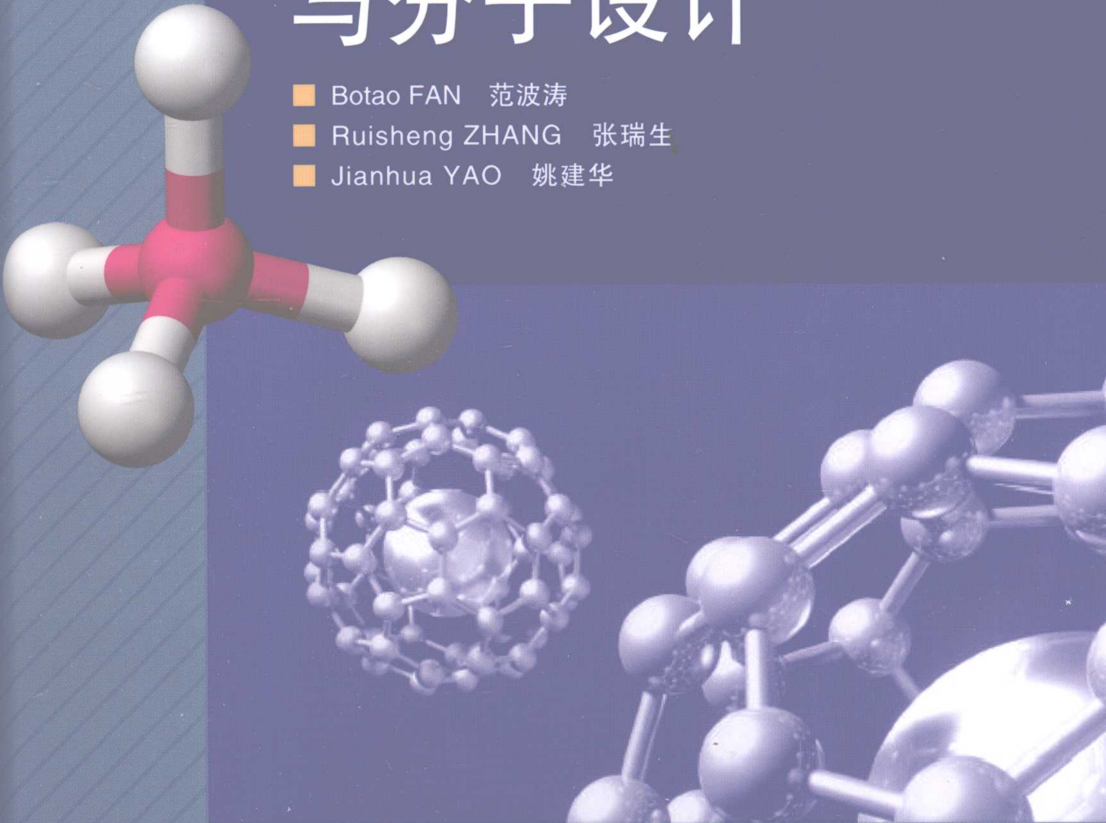
Computer Chemistry and Molecular Design

计算机化学与分子设计

■ Botao FAN　范波涛

■ Ruisheng ZHANG　张瑞生

■ Jianhua YAO　姚建华



高等教育出版社
Higher Education Press

Computer Chemistry and Molecular Design

Botao FAN, Ruisheng ZHANG, Jianhua YAO

Higher Education Press

图书在版编目（CIP）数据

计算机化学与分子设计=Computer Chemistry and Molecular Design: 英文/范波涛，张瑞生，姚建华著．—北京：高等教育出版社，2009.3
ISBN 978-7-04-018296-5

Ⅰ．计…　Ⅱ．①范…②张…③姚…　Ⅲ．①计算机应用-化学-英文 ②分子结构-设计-英文　Ⅳ．O6-39　O561.1

中国版本图书馆 CIP 数据核字（2008）第 179705 号

策划编辑	李冰祥	**责任编辑**	柳丽丽	**封面设计**	王凌波	**责任绘图**	尹　莉
版式设计	马敬茹	**责任校对**	胡晓琪	**责任印制**	陈伟光		

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	100120	**网　　址**	http://www.hep.edu.cn
总　　机	010-58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经　　销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印　　刷	涿州市星河印刷有限公司	**畅想教育**	http://www.widedu.com

开　　本	787 × 1092　1/16	**版　　次**	2009 年 3 月第 1 版
印　　张	32	**印　　次**	2009 年 3 月第 1 次印刷
字　　数	540 000	**定　　价**	68.00 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。

物料号　18296-00

To my father and my mother
To my wife and my children
Botao Fan

Preface 1

The exponential development of computer technology since the late 1960s, not only provided chemists with extended computational capabilities but also gave rise to a new field: Chemoinformatics. Chemoinformatics which gathers all computerised treatments of chemical systems, at the confluence of several disciplines, knew how to take advantage of advances in such varied areas as: Computer Graphics, Theoretical Chemistry, Information Technologies, and rapidly became in the 1970s a well identified and worldwide recognized discipline with its own specific features.

High level quantum chemical calculations on the electronic structure of molecules are now possible in reasonable time. Time dependent evolution of complex systems (including solvent) can be characterized by Molecular Dynamics on a time scale allowing to grasp conformational changes, at least for medium-size molecules. Advances in Computer Graphics give varied, highly aesthetic, displays of molecular systems with interactive capabilities making the analyses more easily understandable and more user-friend. Faced to the huge flow of data stored in databases or generated by combinatorial chemistry, Information Technologies provide efficient tools for exchanging, extracting and managing information.

To investigate chemical (or biological) properties or phenomena, numerous turnkey programs for molecular modeling and computational chemistry are now easily available, thanks particularly to the Internet. But to develop reliable and creative applications, it is essential that chemists have a good understanding of the models, their capabilities and physicochemical meaning, they must also be aware of their limitations and the approximation they are based on.

The authors, Prof. Fan B.T., Zhang R.S. and Yao J.H. joining their expertise in Physical Chemistry, Mathematics and Computer Science, present an excellent synthesis of complementary facets of molecular modeling, from theoretical background to applications.

The first two chapters of the book concern some relevant mathematical tools: fundamentals on linear algebra and differential equations and on another hand, minimization methods that is a key step in situations as diverse as searching for the preferred geometry of a molecular system or adjusting the parameterization of a model.

Chapters 3 to 5 introduce the basis of the quantum chemistry: *Ab Initio* Methods (including post-SCF treatments) are first presented. Density Functional Theory, and Semi-Empirical Approaches, at different levels of sophistication (from CNDO to PM3), are then developed and compared to the more traditional *Ab Initio* Methods.

Empirical Force Fields and Molecular Mechanics are then presented. Such methods are very rapid and now highly reliable and of great interest in conformational analysis of macromolecules (polypeptides, fragments of proteins, DNA$\cdots$), which is a basic step in many applications.

The following chapters examine two powerful methods for the simulation of molecular systems: Molecular Dynamics and Monte Carlo Methods. They are widely used for the calculation of a lot of thermodynamic properties, such as free energies changes for varied processes (binding of a drug to a receptor, chemical reactions $\cdots$), or to take into account the influence of solvation.

The second part of the book is dedicated to applications of graph theory in Chemistry. Indeed Chemistry largely relies on the concept of structural formulae, that are not simple drawings but implicitly convey important structural information. Perception of these formulae as molecular graphs is fundamental for computer treatments and is at the very hearth of a number of approaches.

Extraction of substructures in a molecule, searching for a substructure common to several molecules in a set, recognition of ring systems are basic steps for efficient exploitation of databases. Detection of symmetry and equivalent sites in a structural formula is also a valuable tool which is the subject of chapters 10 to 13. This leads to a presentation of modern management systems for Chemical Information, a critical need on account of the huge flow of data. Formerly, particularly in drug design, the notion of similarity was essential: if two molecules are similar, it may be expected that they share the same properties. Now, as much as chemists more frequently work with large populations of molecules, rather than on limited collections, the notion of diversity takes increased interest. How comparable are two sets of structures? Does one of them more widely cover the structural space?

The third part of the book is dedicated to recent methods, relying on artificial intelligence that afford extended resources in many areas of chemical modeling: Artificial Neural Networks, Support Vector Machine and Genetic Algorithms. For example, in drug design, Neural Networks (NN) are able to extract from raw data (possibly with noise) the underlying (and sometimes complex) relationship between structures and an associated property, without need to define in advance a type of model. The various types of artificial neural networks are presented in details. Hopfield networks have been recently used for structural recognition. For classification tasks, Kohonen self organizing maps give a visual display making easier the perception of the results.

Genetic Algorithms mimicking the biological evolution process are essential for various applications. In the area of Quantitative Structure Activity Relationships (QSAR), the selection of relevant elements among the huge number of structural descriptors now available is a difficult and time-consuming problem, with often subjective decisions, whereas GAs offer a rational approach. Support vector machine (recently introduced) is designed for robust classifications or correlations with high predictive ability.

No doubt this book will provide undergraduate and graduate students with a sound training in molecular modeling, allowing them to master the actual state of the art and to easily adapt to the development of this rapidly evolving discipline.

Jean Pierre DOUCET
Professor. ITODYS. University Paris 7- Denis Diderot
September 2008, Paris

Preface 2

In 1997, the Chinese scientists working in France organized an activity to support the western development of China. We arrived at Lanzhou University. I had the opportunity to meet Professors HU Zhide, JIA Zhongjian, ZHENG Rongliang, LIU Mancang, and other teaching staffs of the Chemistry Department. The scientific exchange between us reached to a common point in research and scientific domains. Based on this same point of view, we signed a series of cooperation agreements, which include their requirement to open a summer class in Computer Chemistry for graduated students. This proposal obtained the support of the Ministry of Education of China. Professors DOUCET J.P. and PANAYE A. of University Paris 7-Denis Diderot expressed also their concerns and supports. Next year, 1998, the summer class was opened, this course of speciality was officially inserted in the list of teaching plan.

Encouraged by this successful beginning, Professor HU Zhide and Professor LIU Mancang suggested me to publish my course in order to solve the problem that the students lack teaching materials. After the discussion with Professor ZHANG Ruisheng, by considering the fact that the time is too short, we thought that we can not publish a book covering a large scope of this domain. Therefore we can only publish a book with limited contents.

Thanks to Lanzhou University, this teaching material was published in 1998. But because of the short time, it lacks a global plan. Moreover, the contents covered by this book are also limited, so we are not satisfactory for this version.

I'm now Professor of University Paris 7, and lead a research group, the laboratory of “Molecular Simulation and Molecular Information” , in ITODYS institute. At the same time, I'm also the Director of Institute of Chemoinformatics of Lanzhou University. I have almost 20 years of experience in research and teaching of computer chemistry and published more than 120 papers and reviews in this area. My teaching experience covers almost all aspects of this area. Therefore, I hope to write a book which gathers the contents of our research and teaching work, in order to provide a reference to all researchers and graduated students worked in this domain.

Great thanks to High Education Press of China. Their heartily invitation gave me a good opportunity to realize my hopes. After receiving the invitation, I invited two colleagues who work long time in Chemoinformatics, to work together for writing this book. One is Professor ZHANG Ruisheng of Lanzhou University, another is Associate Professor YAO Jianhua of Shanghai Institute of Organic Chemistry. The modern computer chemistry covers a large scope of contents. It includes the traditional computational chemistry, molecular graph theory and applications, chemoinformatics, molecular modeling, molecular design (including drug design), molecular simulation, QSAR/QSPR, and so on.

In this book, we cite and refer a lot of literatures reported by specialists. For example, in molecular modeling, personally I consider that the best book is "Molecular Modeling: Principles and Applications" , written by A. Leach. In some related chapters of our book, we cited some contents of this reference book. About the mathematical principle of artificial neural network, we referred the book "Des réseaux de neurones", EYROLLES, 1990, written by E. Davalo and P. Naim. Massart *et al.* developed RBFNN (Massart is a great specialist in this area). The important statements about the theory and applications were reported in their published papers. These papers are the basis of RBFNN. In our book, some of their remarkable works have been cited and referred.

We cited also the works of our laboratory. These results are all the research works of my colleagues and PhD students, including Professor Doucet, Professor Panaye, Professor HU Zhide, Professor LIU Mancang, Dr. Barbault, Dr. Petitjean, Dr. Maldonado, Dr. CHEN Haifeng, Dr. YAO Xiaojun, Dr. XIA Hairong, Dr. LIU Huangxiang, Professor GAO Kun, and so on. I would like to mention specially Professor Doucet and Dr. Maldonado, because a very important part of their works has been cited in this book. I express here my sincere thanks to these colleagues and students. I thank also my family, my wife and children. Without their supports, encouragement and concerns, without a good environment of writing, I can not certainly finish the writing of this book.

This work was supported in part by National Natural Science Foundation of China through Grants 20572120.

I thank again the invitation and supports from Higher Education Press.

Botao FAN
August 2006, Paris

Contents