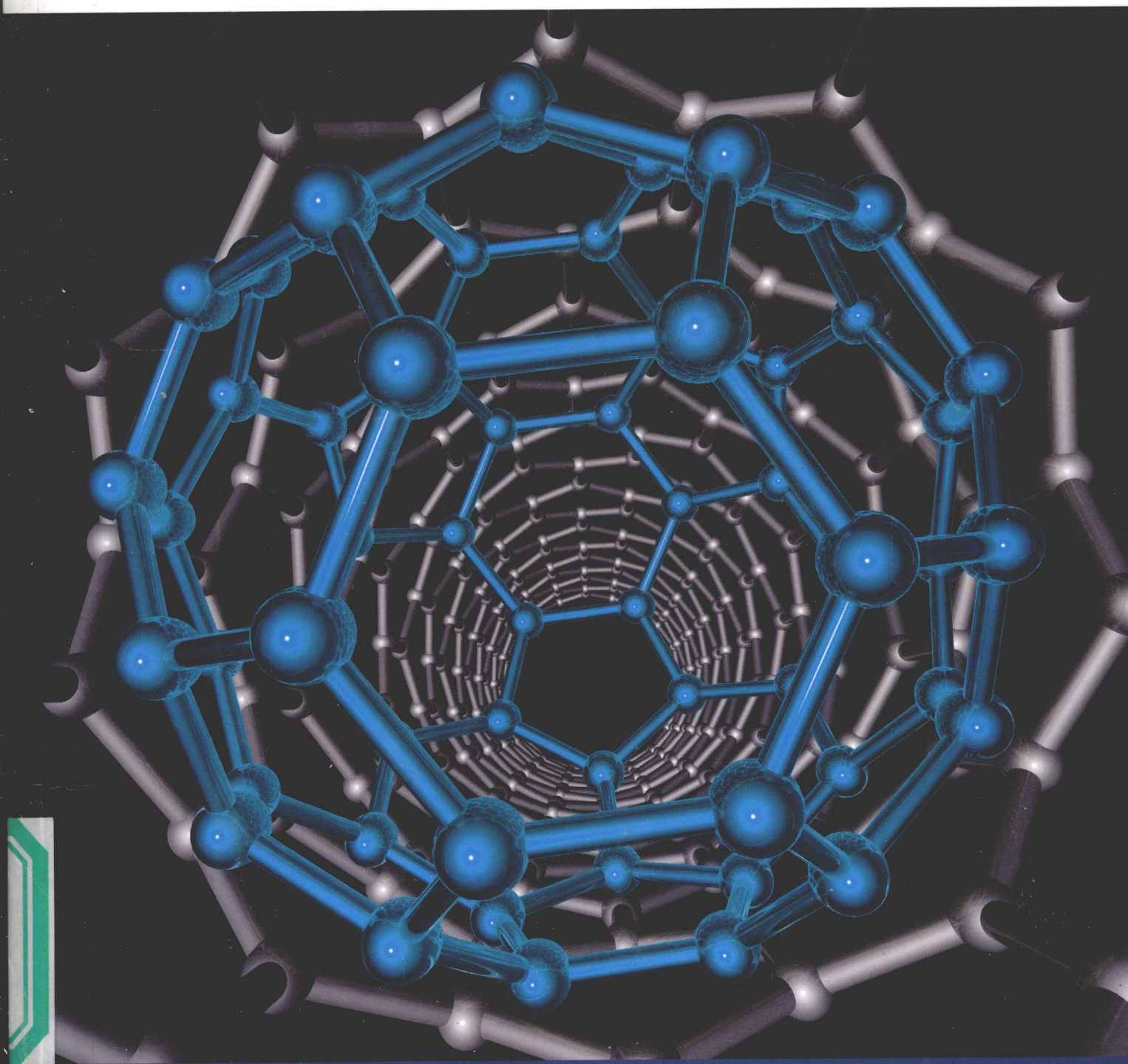
Molecular Modelling

Computational Chemistry Demystified

Peter Bladon, John Gorton and Robert B Hammond



RSC Publishing

Molecular Modelling
Computational Chemistry Demystified

Peter Bladon
Interprobe Chemical Services, Lenzie, Kirkintilloch, Glasgow, UK

John E. Gorton
Gorton Systems, Glasgow, UK

Robert B. Hammond
Institute of Particle Science and Engineering, The University of Leeds, UK

RSC Publishing

ISBN: 978-1-84973-352-6

A catalogue record for this book is available from the British Library

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road,
Cambridge CB4 0WF, UK

Registered Charity Number 207890

For further information see our web site at www.rsc.org

Molecular Modelling
Computational Chemistry Demystified

Preface

'Molecular Modelling - Computational Chemistry Demystified'

While there are many books that deal with *Molecular Modelling* or *Computational Chemistry*, few of them meet the needs of someone who is just starting out, and wishes to set up their own system. Problems arise right at the beginning with the bewildering array of acronyms to do with both chemistry and computing.

A student studying chemistry at a well endowed university, and needing to fulfil requirements in computational chemistry, would find the necessary equipment and programs already set up. All (s)he has to do is to follow the prescribed parts of the course and pass the tests. The same student, having graduated, and now proceeding to a higher degree would be able to use the same or similar facilities when needed. But there are those who, for various reasons, choose to work alone, or are obliged to do so. Catering for these is what this book is all about. Molecular modelling can provide an entry into chemistry, that in former years was available by experimentation, now unfortunately, but understandably, frowned on by authority. For someone in this category who is impatient and wants to get started right away, we offer a first chapter that allows just this.

Demystification is one of our goals. But it does not require that Molecular Modelling be trivialized. So we have tried to provide a book that would be valuable to people at many stages of learning about Chemistry. Is this possible? We believe it is, since even the business of writing the book and the software has provided us with insights into aspects of chemistry that were new to us.

"First catch your hare..." This quotation is wrongly attributed to Mrs Beeton in her recipe for jugged hare. In the present context we can paraphrase it "First buy/acquire/beg/borrow... your computer". We show in the second chapter how to use such a computer, that you maybe already have, and that is running a version of the Windows® Operating system, to run Molecular Modelling and Computational Chemistry programs. Not only that - the Interprobe programs (INTERCHEM, PRESTO, *etc.*) are included with this book, and sections are devoted to setting them up and using them!

The rest of the book is devoted to explaining the mystique and jargon that surrounds computational chemistry. There is a chapter devoted to modelling the crystalline state, a feature not catered for elsewhere. Other chapters deal with modelling biopolymers such as proteins that impinge on medicinal chemistry. There is an underlying heuristic aim throughout the book; problems for the reader to solve are provided.

The authors are conscious of the difficulties of explaining in detail how to perform tasks on a personal computer. The expectation of users today is that all the tasks will be started as the results of user - computer interaction. This means that ***every*** request from the computer and ***every*** response from the user shall be specified in detail. Such requirements can result in text that is difficult to read; we have done our best in this regard, and hope that the parts of the book that are not so circumscribed will make easier reading.

We have framed the instructional examples and the questions in the context of organic chemistry of recent decades, with an emphasis on structures. Modelling chemical *reactions* presents significant challenges and, it may be argued, can only be done in terms of starting, intermediate, and end *structures* and the stereochemistry and thermodynamics that relates them. To aid this we provide a chapter dealing with stereochemistry, and conformational analysis.

The book aims to be a practical guide to the use of computers in aiding structural chemistry, and when we have digressed into the chemistry itself, we have done so to explain the relevance of the modelling techniques or computations. The book is neither to be regarded as a textbook on crystallography nor a textbook on medicinal chemistry. However with these two topics, and the topic of protein structures, we hope to provide an introduction to fields of chemical science that are both challenging and worthy of serious study.

The reasons for issuing software with the book

In coupling the publication of this book with an offering of Open-Source software, we are making a bold statement. This comes from a belief that openness is the key to scientific progress. It is possible that a scientist in 2061, referring to one of the many papers concerning the docking of small molecule structures into protein cavities (a subject we touch upon in Chapter 9), would not be able to make any sense of the paper itself, and would have difficulty in reproducing the experiments. This would be in part due the complexity of the science, but a contributing factor would be the fact that the software being used is proprietary and secret. The case for software being open-source now largely rests on contemporary issues (who should own the software?), but we believe that this other aspect of future traceability is a more important reason for the algorithms and source code based on them to be completely disclosed. The present situation in computational chemistry is in stark contrast to that existing in conventional experimental work, where results obtained today can be (and regularly are) compared with results obtained in the last two centuries, if necessary by repeating the original experiments.

The ethical aspects of molecular modelling have been raised before. In a paper entitled *Guidelines for Publications in Molecular Modeling Related to Medicinal Chemistry*,[1] the members of Working Party on Computer Assisted Molecular Modeling, commissioned by IUPAC made several significant recommendations. In relation to proprietary algorithms they suggested that ".... the commercializing companies have the obligation to submit for publication articles giving at least general descriptions of algorithms and databases used, with sample calculation results to aid calibration and evaluation."

The reasons why the Interprobe software came about

The straightforward answer is economic necessity. Twenty five years ago commercially available modelling packages were expensive and of limited scope. Subsequently mergers, takeovers, company failures, and 'rationalisations' have reduced the number of choices for purchasing software without reducing costs. What has resulted from twenty years of development is a set of programs that fills most of the needs of entry- level molecular modelling. We would be foolish to claim that it is the best, but we hope that what is provided may stimulate purchasers/users to

experiment. We would also welcome constructive comments aimed at its improvement.

The software owes its origin to earlier versions that were designed to work on Digital Equipment Vax computers and on SGI Irix systems. These versions were developed in the Department of Pure and Applied Chemistry at The University of Strathclyde where one of the authors was employed. The version of the software that is supplied with this book, and that works on systems running Microsoft Windows operating systems, derives from these programs. It is a reworking of these earlier systems and has been developed independently and without any outside financial aid.

We understand that in some circumstances complete openness of software has disadvantages; the chief being that 'modified' versions become available! In this connection, the Interprobe software is 'protected' by the fact that the compilers used are not open software; if you wish to modify the code for and distribute it, this is possible only if you acquire the Salford software currently licensed by the firm Silverfrost. This firm does provide, at no cost, a demonstration version of the compilers *etc.* that does allow you to make modifications to our programs for your own use, *but not redistribute them*. Furthermore, this enabling software only allows the production of versions for Microsoft Windows operating systems. We do not hide the fact that we would have preferred, in an ideal world, to have a system working under Linux. Maybe that will come.

Chemistry needs computing needs computers need chemistry!

That chemistry depends on computation and therefore on computers is not much in doubt. But let it not be forgotten that the reverse is also true. Two facts stand out:

(1) All modern computers depend on devices that require highly purified elemental silicon. The key to this is the chemistry of silanes, an area of chemistry opened up by the work of Alfred Stock in the first third of the 20th Century, and at the time considered of only academic interest.

(2) Any laptop computer that you own will rely on a *liquid crystal* screen for a display, and most desktop machines will also have screens of this type. In 1924 Daniel Vorländer wrote on page 89 of his monograph on liquid crystals:[2]

"Man hat mir wohl die Frage gestellt, ob sich die kristallin-flüssigen Substanzen technisch verwerten lassen? Ich sehe keine Möglichkeit dazu"

["I have indeed asked myself whether liquid crystal substances are capable of technical application? I see no possibility of that."]

He expressed his honest opinion, but how wrong he was!

These are just two examples, but what better testimony and reasons could we want, for the continued support of blue-sky research?

[1] P. Grund, D. C. Berry, J. M. Blaney, and N. C. Cohen, *J. Med. Chem.*, 1988. **31**, 2230.

[2] D. Vorländer, *Chemische Kristallographie der Flüssigkeiten.*, Akademische Verlagsgesellschaft, Leipzig , 1924.

Acknowledgements

J.G. acknowledges the help from many students and colleagues who have given great help and simplified his convoluted explanations.

P.B. acknowledges the help of many former colleagues at the University of Strathclyde; in particular, Dr. Mark Dufton for stimulating interest in protein chemistry, Dr. Robin Breckenridge for collaboration in the early days of Interprobe and for subsequently keeping him on track on drug design politics and economics, Dr. David Pugh for putting him right when dealing with crystal structures, and Professor Douglas McGregor for guidance on the humane treatment of computers. The molecular mechanics that is used in the Interprobe software is based on code that was generously supplied by Dr. Armin Widmer of Novartis, Basel and to whom a great debt is due.

R.B.H. acknowledges the help and support of colleagues at the University of Leeds in particular, Dr. Christopher Hammond for permission to reproduce a Table from his book. He also wishes to thank his parents Dorothy and Bryan for their steadfast support at all times.

We are also grateful to the Royal Society for permission to quote from Dirac's paper that is referred to in Chapter 1.

Formal permission has not been obtained for our quotation (in the Preface) from Vorländer's monograph; investigation has failed to reveal who is the current owner of the copyright.

We are grateful to the Cambridge Crystallographic Data Centre for permission to access, reproduce, and comment on selected files from that repository.

Typographical conventions used in this book

(1) The fount that we use **generally** in the text is

Times New Roman 12 point

(2) The fount that we use in the **endnote**s is

Times New Roman 11 point

(3) When we wish to indicate that text is printed by a computer we show it in

```
Courier New 12 point
```

This fount is also used when we wish to indicate that you should type something in response to a computer.

(4) When we want to show a complete line of text produced by a computer, without wrapping, we will use a fount as small as `Courier New 8 point.`

(5) We use the convention of enclosing the **type** of response needed by the computer by enclosing the name describing the entity in chevron brackets thus:

```
<file name>
```

In this case your response should be an actual file name (**without** the brackets!).

(6) In displaying SMILES strings in the text, we would have preferred to have used an equal spaced fount like `Courier New.` Unfortunately this had the effect of seeming to introduce spaces before and after brackets thus:

```
c2cc(-c1ccccc1)ccc2CCC
123456789012345678901 2
```

So we settled for Arial

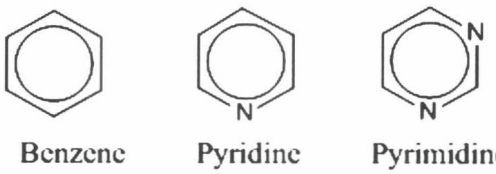
c2cc(-c1ccccc1)ccc2CCC
123456789012345678901 2

This gets rid of the spaces but characters do not have equal widths.

(7) In the instructions for using the Interprobe programs INTERCHEM and PRESTO that have graphical user interfaces (GUIs), we have adopted a convention of referring to the Menu Buttons by their Row and Column coordinates, usually following the wording on the button. The Menu Name may also be included, thus:

Dual Mode Menu, Invert Structure (R1/C6)

(8) In displaying line drawings of 6-membered aromatic rings the presence of the six pi-electrons is shown by a circle inscribed within the hexagon. We prefer this notation to that showing alternating double and single bonds (the Kekulé notation). We do not extend this notation to the 5-membered rings of heterocyclic compounds such as pyrrole, furan and thiophen, preferring in these cases to show the double bonds.



Benzene Pyridine Pyrimidine

BUT

Pyrrole Thiophene Indole Cyclopentadienyl

Contents

Chapter 1
Introduction

1.1 The beginnings – Some history

In 1929, in a paper entitled "Quantum Mechanics of Many-Electron Systems", the physicist P. A. M. Dirac made the following provocative statement:[1]

> "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too difficult to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation."

Presumably, the physicists who read this paper were mollified to some extent because some of their number would still be employed, but it seems that the paper was (fortunately) ignored by chemists, who carried on as usual. Indeed, in the eighty two years that have followed the appearance of Dirac's paper, it is likely that more new chemical compounds have been recorded than in all the previous years. What also happened was the "approximate practical methods", envisaged by Dirac, appeared, because of something that he only hinted at: "computation". Dirac shared the Nobel Prize for Physics in 1933 with Erwin Schrödinger.

While the majority of chemists were doing their own things ("Stamp Collecting" according to Ernest Rutherford – another physicist, who ironically got a Nobel Prize for Chemistry!) some physical chemists and chemical physicists did take Dirac's words to heart, and so 'Theoretical Chemistry' and then (when computers were invented) 'Computational Chemistry' came about.

1.2 About the book

This book is about "Molecular Modelling". We have given it a subtitle "Computational Chemistry Demystified" because we believe that *modelling* provides an easy entry point into computational chemistry, a branch of chemistry that has hitherto been misunderstood and undervalued, and thence into the rest of chemistry. While chemists have used many sorts of models for a long time, we concentrate our attention on *computer based modelling* because now this is the method of choice, and the necessary equipment is, if not already available, easily affordable.

We follow the present introductory chapter with one that deals with the choice of computer equipment. We do this because we believe that success in modelling and computation can be critically dependent on making the right choice. However, if you have already purchased your computer for other purposes, do not despair, what we have written may still help you in improving its performance, and to adapt it for molecular modelling.