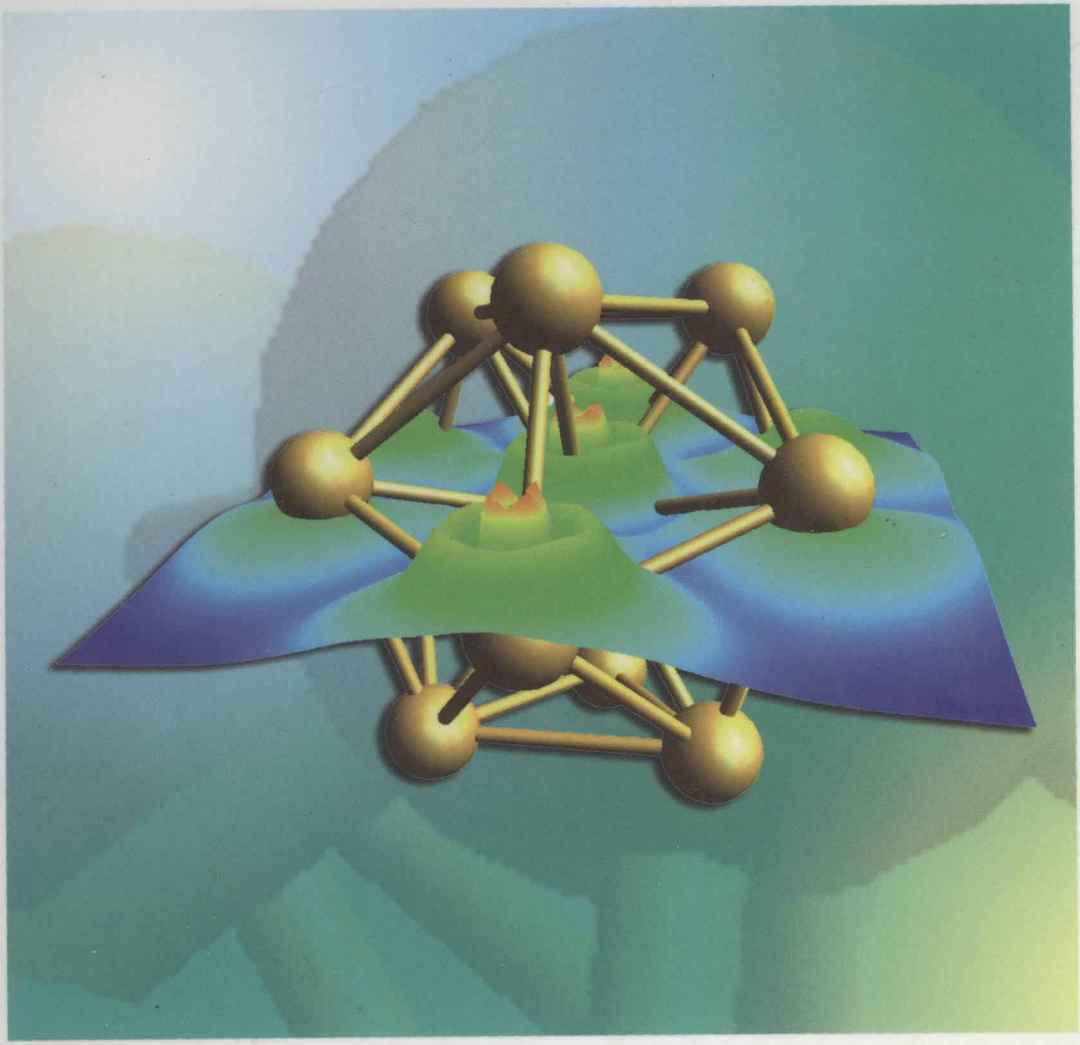
Specialist Periodical Reports

Editor A Hinchliffe

Chemical Modelling: Applications and Theory

Volume 4



RSC Publishing

A Specialist Periodical Report

Chemical Modelling
Applications and Theory

Volume 4

A Review of Recent Literature Published between June 2003 and May 2005

Editor
A. Hinchliffe, *School of Chemistry, The University of Manchester, Manchester, UK*

Authors
B. Coupez, *Novartis Institutes for Biomedical Research, Basel, Switzerland*
R.A. Lewis, *Novartis Institutes for Biomedical Research, Basel, Switzerland*
H. Möbitz, *Novartis Institutes for Biomedical Research, Basel, Switzerland*
A.J. Mulholland, *University of Bristol, Bristol, UK*
A. Miličević, *The Institute of Medical Research and Occupational Health, Zagreb, Croatia*
D. Pugh, *University of Strathclyde, Glasgow*
D.J. Searles, *Griffith University, Brisbane, Australia*
D.S. Sholl, *Carnegie Mellon University, Pittsburgh, PA, USA*
T.E. Simos, *University of Peloponnese, Athens, Greece*
M. Springborg, *University of Saarland, Saarbrücken, Germany*
B.D. Todd, *Swinburne University of Technology, Victoria, Australia*
N. Trinajstić, *Rudjer Bošković Institute, Zagreb, Croatia*
S. Wilson, *Rutherford Appleton Laboratory, Chilton, Oxfordshire*

RSC Publishing

If you buy this title on standing order, you will be given FREE access to the chapters online. Please contact sales@rsc.org with proof of purchase to arrange access to be set up.

Thank you.

ISBN-10: 085404-243-1
ISBN-13: 978-0-85404-243-2
ISSN 0584-8555

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

Typeset by Macmillan India Ltd, Bangalore, India
Printed and bound by Henry Ling Ltd, Dorchester, Dorset, UK

Chemical Modelling
Applications and Theory

Volume 4

Preface

Welcome to Volume 4 of the 'Chemical Modelling' SPR. Naturally, I want to start by thanking my team of authors for the hard work they have put into making this the best and most comprehensive volume so far.

It seems a long time since I wrote the following in my Preface to Volume 1 (1999) . . .

'*Starting a new SPR is never easy, and there was the problem of where the contributors should start their accounts; since time began? five years ago? An SPR should be the first port of call for an up-to-the-minute account of trends in a specialist subject rather than a dull collection of references. My solution was to ask contributors to include enough historical perspective to bring a non-specialist up to speed, but to include all pertinent references through May 1999. Volume 2 will cover the literature from June 1999 to May 2001 and so on. In subsequent Volumes, I shall ask those Contributors dealing with the topics from Volume 1 to start from there. New topics will be given the same generous historical perspective opportunity as Volume 1 but will have to cover the literature to 2001 + n where n = 0, 2, 4, ... This process will continue until equilibrium is reached.*'

I think we have now reached equilibrium; some topics have reached maturity and so don't need cover every Volume, whilst a casual monthly glance at the content pages of JACS, JCP, JPC, CPL, THEOCHEM, Faraday Transactions (to name my favorites, not given in order of merit) reveals growth areas.

As an example of a 'mature' topic, consider Density Functional Theory (DFT). DFT is far from new and can be traced back to the work of John Slater and other solid state physicists in the 1950's, but it was ignored by chemists despite the famous papers by Hohenberg/ Kohn (1964) and Kohn/ Sham (KS) (1965). The HF-LCAO model dominated molecular structure theory from the 1960's until the early 1990s and I guess the turning point was the release of the rather primitive KS-LCAO version of GAUSSIAN. DFT never looked back after that point, and it quickly became the standard for molecular structure calculations. So this Volume of the SPR doesn't have a self contained Chapter on DFT because the field is mature.

As an example of a 'perennial' topic, consider the theory of liquids. Almost every undergraduate physical chemistry text tells us that gases

and solids are easy to understand because in the first case we have random motion, whilst in the second rigid structures. The gist of this argument is that liquids are really tricky, as indeed they are. The first computer simulation of a liquid was carried out in 1953 at the Los Alamos National Laboratories. The MANIAC mainframe was much less powerful than the PC I am using to write this Preface but the early work by Metropolis et. al. laid the foundations for modern liquid modeling. David Heyes (Volume 2) and Karl Travis (Volume 3) told you how things were in a few years ago, and the story is continued by Billy Todd and Debra Bernhardt in Volume 4.

My final sentence for Volume 1 was

'*I am always willing to listen to convincing ideas for new topics*'

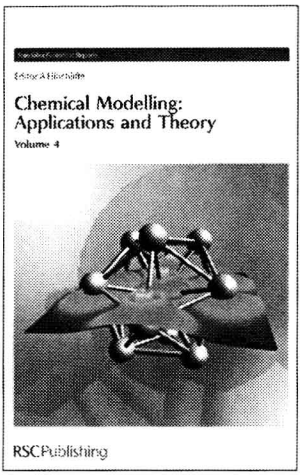
as indeed I am. My colleague J Jerry Spivey is Editor for the Catalysis SPR; he took me at my word and as a result it is a pleasure to welcome our first contribution from David S Sholl on Heterogeneous Catalysis.

I haven't space to give glowing descriptions of the remaining contributions from each colleague. We hope you will derive benefit and perhaps even pleasure from our efforts.

On a rare personal note, I should tell you that UMIST and the Victoria University of Manchester recently decided to merge to become the UK's largest University; I'm still sitting at the same desk in the same office but my employer is now 'The University of Manchester' and my e-mail has changed to alan.hinchliffe@manchester.ac.uk

Alan Hinchliffe
Manchester 2006

Contents



Cover
The icosahedral 'golden fullerene' WAu_{12} reproduced by permission of Pekka Pyykkö, Chemistry Department, University of Helsinki, Finland.

1

Computer-Aided Drug Design 2003–2005

BY BERNARD COUPEZ, HENRIK MÖBITZ AND RICHARD A. LEWIS
Novartis Institutes for Biomedical Research, Basel CH-4002, Switzerland

1 Introduction

The themes for this review again have been driven strongly by the need of the Pharmaceutical industry to make the discovery process quicker and more reliable. Virtual screening in all its forms is at the heart of most research, from bioavailability filters through to rigorous estimations of the free energy of binding. Two areas of relative heat have been docking/scoring, and ADME/Tox. On the other hand, 3D-QSAR and pharmacophores have become quiet. Part of the reason for this may arise from the successes in high-throughput crystallography, delivering more targets and complexes, the relative failure of HTS, and the increase in the amount of high quality data coming from late-phase research/early-phase development concerning the fate of clinical candidates. These trends look set to continue in the future, and the next two years should yield many new breakthroughs.

2 ADME/Tox and Druggability

There has been a fresh impetus to the modelling of ADME, Toxicity and druggability phenomena, partly driven by a desire to understand why such complex phenomena can, apparently, be described so simply, and partly to see if better models can be built, to improve the attrition rate in medicinal chemistry still further.

2.1 Druggability and Bioavailability. – In the continuing debate over what physicochemical properties are required for bioavailability, Vieth *et al.*[1] have surveyed 1729 marketed drugs with respect to their route of administration, h-bonding capability, lipophilicity and flexibility. One conclusion they draw is that these properties have not varied substantially over time, implying that oral bioavailability is independent of target or molecular complexity. Compounds with lower molecular weight, balanced lipophilicity and less flexibility tend to be favoured. Leeson and Davis[2] claim that molecular weight, flexibility, the number of O and N atoms and hydrogen-bond acceptors have risen, by up to

Chemical Modelling: Applications and Theory, Volume 4

29%. This may be partly due to the choice of 1983 as the reference year, or the advent of more complex targets with greater selectivity needs (*e.g.* kinases). In the same vein, a study[3] re-examined the correlation of flexibility and polar surface area (PSA) with bioavailability proposed by Veber *et al.*[4] One conclusion is that there are significant differences in the ways of defining flexibility and PSA, and the correlations depend markedly on the method used (this is not surprising, as neither quantity is precisely definable). A second conclusion was that the limits defined (Number of rotatable bond < 10, PSA < 140 Å^2) excluded a significant number of compounds with acceptable rat bioavailability. In the authors' words, "This observation underscores the potential danger of attempting to generalise a very complicated endpoint and of using that generalisation in a prospective selection application". Despite this, another bioavailability score[5] has been devised, to predict the probability that a compound has $> 10\%$ bioavailability in the rat. Compounds are grouped by ionisation class (anions, cations, neutral). It was found that the standard rule-of-5 does well for cations and neutrals (88% of the compounds predicted to have low bioavailability are observed as such). Anionic compounds were better described by PSA limits. Some simple rules are given to compute the bioavailability score. In Abbott laboratories, this score is now routinely computed for all compounds and is used for hit-list triaging. It will be interesting to see if the results can be repeated on other data sets; the paper has certainly sparked much interest in the modelling community. Wegner[6] provides support for the idea that human intestinal absorption correlates with PSA, by generating a classification model. The justification is that the error in the experimental data is 25%, and 80% of the observations occur in the top and bottom quartiles, that is, the data is more binary than evenly spread. In addition to PSA, other descriptors that reflect the electronic character of atoms and their environment also came to the fore.

2.2 Metabolism, Inhibitors and Substrates. – The field of cytochrome modelling is becoming more mature as we begin to understand the limitations of the experimental data and the subtleties of the mechanisms (the whole field of cytochrome P450 modelling, including homology, pharmacophore and 3D-QSAR models has been reviewed in detail recently[7]). Empirical models are still preferred, especially for rapid evaluation of large libraries. In one case, use of a jury system improved prediction accuracy to over 90%.[8] Chohan *et al.*[9] have developed 4 models for Cytochrome P450 (Cyp) 1A2 inhibition, and identified the expected descriptors as being important to the QSAR (lipophilicity, aromaticity, HOMO/LUMO energies). Perhaps a more interesting result in this paper was the use of the k index to assess predictive powers of the models using test data.

$$k = \frac{\text{observed agreement-chance agreement}}{\text{total observed-chance agreement}}$$

This index should prove useful for data sets that are diverse and noisy. The validity of QSAR model predictions has also been studied by Guha and Jurs.[10]

The protocol is quite straightforward. The initial QSAR models were built, and the residuals of the compounds in the training set were used to classify the trains set predictions into good and bad. The threshold for the classification is arbitrary. Test compounds were predicted, and the predictions were grouped by substructural similarity to the nearest neighbour in the training set. It was seen that test compounds that had neighbours with low/good residuals were themselves well-predicted, with the reverse being the case for neighbours with high residuals. The success rate for classifying the strength of the prediction was 73% to 94%. The Merck group[11] performed a retrospective study of in-house data sets, and concluded that the distance to the nearest neighbour, and the number of nearest neighbours (local density) were the two most useful measures for predicting prediction quality. They also concluded that distance does not have to be measured in the same descriptor space as was used to build the QSAR model. Topological descriptors combined with a Dice coefficient worked equally well.

A number of groups have been active in the prediction of the most likely sites of metabolism of molecules that are substrates for cytochromes. Singh *et al.*[12] developed a semi-quantitative method based on the energy barrier to the creation of hydrogen radicals as calculated by AM1. Using a set of 50 substrates for Cyp 3A4, they were able to show that only hydrogens with a solvent-accessible surface area over 8 $Å^2$ are susceptible to attack. The expensive quantum mechanic calculations could be approximated by local neighbourhood descriptors which could be well correlated to the energies ($R^2 = 0.98$), offering a fast and practical method for screening large libraries. An extension of this concept is embodied in the MetaSite program,[13] which uses propensity to react, accessibility and GRID molecular interaction fields as descriptors. The methodology is more general, and can be applied to any cytochrome structure: in validation experiments, an accuracy of 80% is claimed. It is also important to be able to predict which compounds will be inhibitors as well as substrates, to avoid drug-drug interactions. A classifier based on a support vector machine (SVM)[14] has been created that correctly predicts compounds into high, medium and low affinity at 70% accuracy, even with simple 2D descriptors. The improved accuracy was obtained through a systematic variation and optimisation of the SVM parameters.

Considering the success of surprisingly simple, semiempirical methods in ADME modelling, it is interesting to see whether more advanced methods could bring further improvements. A recent paper of Beck[15] provides a link to the rich literature of DFT studies of hemes and cytochromes. The author uses Fukui functions to gauge the site of highest nucleophilicity of a number of known drugs. The predictions give mixed results and demonstrate that the implicit assumption of Fukui functions, *i.e.* an isotropic electrophilic attack, is flawed, not to mention that their MO-like shape does not allow a ranking of single atoms. In conclusion, the study suggests that it is more important to have an accurate description of the cytochrome-ligand complex than to invest in a high-level description of the chemical reactivity. De Visser *et al.*[16] have used DFT on 10 C–H barriers with reference to bacterial cytochromes, and claim an