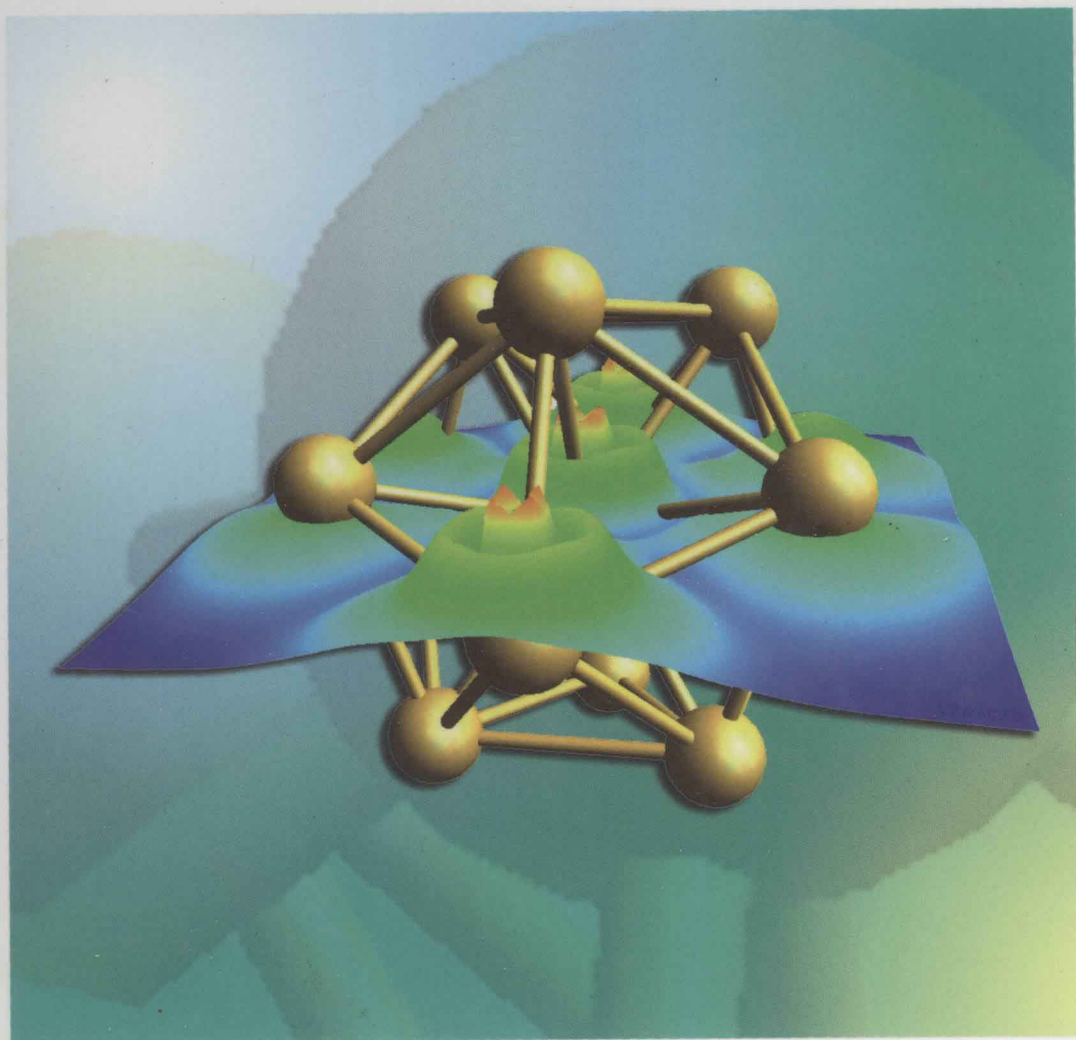


Specialist Periodical Reports

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

© The Royal Society of Chemistry 2006

All rights reserved

Apart from any fair dealing for the purpose of research or private study for non-commercial purposes, or criticism or review as permitted under the terms of the UK Copyright, Designs and Patents Act, 1988 and the Copyright and Related Rights Regulations 2003, this publication may not be reproduced, stored or transmitted, in any form or by any means, without the prior permission in writing of The Royal Society of Chemistry, or in the case of reprographic reproduction only in accordance with the terms of the licences issued by the Copyright Licensing Agency in the UK, or in accordance with the terms of the licences issued by the appropriate Reproduction Rights Organization outside the UK. Enquiries concerning reproduction outside the terms stated here should be sent to The Royal Society of Chemistry at the address printed on this page.

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 $n = 0, 2, 4, \dots$. 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'

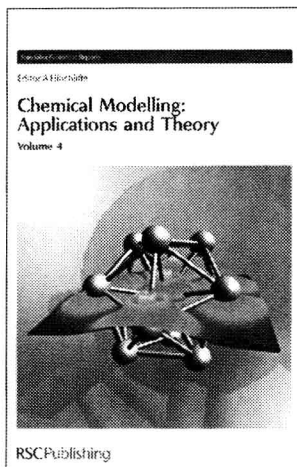
as indeed I am. My colleague J Jerry Spivey is Editor for the Catalysis SPR; he took me at my word and as a result it is a pleasure to welcome our first contribution from David S Sholl on Heterogeneous Catalysis.

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' $W_{Au_{12}}$ reproduced by permission of Pekka Pyykkö, Chemistry Department, University of Helsinki, Finland.

Computer-Aided Drug Design 2003–2005	1
<i>By Bernard Coupez, Henrik Möbitz and Richard A. Lewis</i>	
1 Introduction	1
2 ADME/Tox and Druggability	1
2.1 Druggability and Bioavailability	1
2.2 Metabolism, Inhibitors and Substrates	2
2.3 Toxicity	4
3 Docking and Scoring	4
3.1 Ligand Database Preparation	4
3.2 Target Preparation	5
3.3 Water Molecules	6
3.4 Comparison of Docking Methods	6
3.5 Scoring	7
3.6 New Methods	8
3.7 Application of Virtual Screening	9
4 <i>De Novo</i> , Inverse QSAR and Automated Iterative Design	10

5	3D-QSAR	11
6	Pharmacophores	11
7	Library Design	12
8	Cheminformatics and Data Mining	13
8.1	Scaffold Hopping	13
8.2	Descriptors and Atom Typing	14
8.3	Tools	15
9	Structure-Based Drug Design	15
9.1	Analysis of Active Sites and Target Tracability	15
9.2	Kinase Modelling	16
9.3	GPCR Modelling	16
10	Conclusions	18
	References	18

Modelling Biological Systems	23
-------------------------------------	----

By Adrian J. Mulholland

1	Introduction	23
2	Empirical Forcefields for Biomolecular Simulation: Molecular Mechanics (MM) Methods	24
3	Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods	29
3.1	Interactions between the QM and MM Regions	31
3.2	Basic Theory of QM/MM Methods	34
3.3	Treatment of Long-Range Electrostatic Interactions in QM/MM Simulations	35
3.4	QM/MM Partitioning Methods and Schemes	37
4	Some Comments on Experimental Approaches to the Determination of Biomolecular Structure	41
5	Computational Enzymology	43
5.1	Goals in Modelling Enzyme Reactions	43
5.2	Methods for Modelling Enzyme-Catalysed Reaction Mechanisms	45
5.3	Quantum Chemical Approaches to Modelling Enzyme Reactions: Cluster (or Supramolecule) Approaches, and Linear-Scaling QM Methods	45
5.4	Empirical Valence Bond Methods	47
5.5	Examples of Recent Modelling Studies of Enzymic Reactions	48
6	<i>Ab initio</i> (Car-Parrinello) Molecular Dynamics Simulations	59

7 Conclusions	60
Acknowledgements	60
References	61

Polarizabilities, Hyperpolarizabilities and Analogous Magnetic Properties	69
<i>By David Pugh</i>	

1 Introduction	69
2 Electric Field Related Effects	70
2.1 Atoms	70
2.2 Diatomic Molecules: Non-Relativistic	73
2.3 Diatomic Molecules: Relativistic	73
2.4 Atom-Atom Interactions	74
2.5 Inert Gas Compounds	74
2.6 Water	76
2.7 Small Polyatomic Molecules	87
2.8 Medium Sized Organic Molecules	88
2.9 Organo-Metallic Complexes	93
2.10 Open Shells and Ionic Structures	93
2.11 Clusters, Intermolecular and Solvent Effects, Fullerenes, Nanotubes	95
2.12 One and Two Photon Absorption, Luminescence <i>etc.</i>	95
2.13 Theoretical Developments	95
2.14 Oligomers and Polymers	96
2.15 Molecules in Crystals	96
3 Magnetic Effects	97
3.1 Inert Gases, Atoms, Diatomics	97
3.2 Molecular Magnetisabilities, Nuclear Shielding and Aromaticity, Gauge Invariance	98
References	99

Applications of Density Functional Theory to Heterogeneous Catalysis	108
<i>By David S. Scholl</i>	

1 Introduction	108
2 Success Stories	111
2.1 Success Story Number One: CO Oxidation over RuO ₂ (110)	111

2.2	Success Story Number Two: Ammonia Synthesis on Ru Catalysts	114
2.3	Success Story Number Three: Ethylene Epoxidation	122
3	Areas of Recent Activity	129
3.1	<i>Ab initio</i> Thermodynamics	130
3.2	Catalytic Activity of Supported Gold Nanoclusters	134
3.3	Bimetallic Catalysts	142
4	Areas Poised for Future Progress	146
4.1	Catalysis In Reversible Hydrogen Storage	146
4.2	Electrocatalysis	147
4.3	Zeolite Catalysis	148
5	Conclusion and Outlook	152
	Acknowledgements	152
	References	153

Numerical Methods in Chemistry 161

By *T.E. Simos*

1	Introduction	161
2	Partitioned Trigonometrically-Fitted Multistep Methods	163
2.1	First Method of the Partitioned Multistep Method	163
2.2	Second Method of the Partitioned Multistep Method	167
2.3	Numerical Results	172
3	Dispersion and Dissipation Properties for Explicit Runge-Kutta Methods	176
3.1	Basic Theory	176
3.2	Construction of Runge-Kutta Methods which is Based on Dispersion and Dissipation Properties	177
3.3	Numerical Results	181
4	Four-Step P-Stable Methods with Minimal Phase-Lag	185
4.1	Phase-Lag Analysis of General Symmetric $2k$ – Step, $k \in N$ Methods	185
4.2	Development of the New Method	186
4.3	Numerical Results	189
5	Trigonometrically Fitted Fifth-Order Runge-Kutta Methods for the Numerical Solution of the Schrödinger Equation	190
5.1	Explicit Runge-Kutta Methods for the Schrödinger Equation	190
5.2	Exponentially Fitted Runge-Kutta Methods	191

5.3 Construction of Trigonometrically-Fitted Runge-Kutta Methods	191
6 Four-Step P-Stable Trigonometrically-Fitted Methods	194
6.1 Development of the New Method	194
6.2 Numerical Results	198
7 Comments on the Recent Bibliography	200
References	209
Appendix A Partitioned Multistep Methods – Maple Program of Construction of the Methods	211
Appendix B Maple Program for the development of Dispersive-fitted and dissipative-fitted explicit Runge-Kutta method	216
Appendix C Maple Program for the development of explicit Runge-Kutta method with minimal Dispersion	223
Appendix D Maple Program for the development of explicit Runge-Kutta method with minimal Dissipation	230
Appendix E Maple Program for the development of the New Four-Step P-stable method with minimal Phase-Lag	237
Appendix F Maple Program for the development of the Trigonometrically Fitted Fifth-Order Runge-Kutta Methods	238
Appendix G Maple Program for the development of the New Four-Step P-stable Trigonometrically-Fitted method	244

Determination of Structure in Electronic Structure Calculations	249
<i>By Michael Springborg</i>	
1 Introduction	249
2 Determining the Global Total-Energy Minima for Clusters	256
2.1 Random vs. Selected Structures	256
2.2 Molecular-Dynamics and Monte Carlo Simulations	258
2.3 The Car-Parrinello Method	260

2.4	Eigenmode Methods	261
2.5	GDIIS	263
2.6	Lattice Growth	264
2.7	Cluster Growth	265
2.8	Aufbau/Abbau Method	265
2.9	The Basin Hopping Method	266
2.10	Genetic Algorithms	267
2.11	Tabu Search	268
2.12	Combining the Methods	270
3	Descriptors for Cluster Properties	271
3.1	Energetics	271
3.2	Shape	272
3.3	Atomic Positions	272
3.4	Structural Similarity	273
3.5	Structural Motifs	274
3.6	Phase Transitions	276
4	Examples for Optimizing the Structures of Clusters	278
4.1	One-Component Lennard-Jones Clusters	278
4.2	Two-Component Lennard-Jones Clusters	282
4.3	Morse Clusters	283
4.4	Sodium Clusters	284
4.5	Other Metal Clusters	288
4.6	Non-Metal Clusters	297
4.7	Metal Clusters with More Types of Atoms	299
4.8	Non-Metal Clusters with More Types of Atoms	304
4.9	Clusters on Surfaces	307
5	Determining Saddle Points and Reaction Paths	308
5.1	Interpolation	309
5.2	Eigenmode Methods	309
5.3	The Intrinsic Reaction Path	310
5.4	Changing the Fitness Function	310
5.5	Chain-of-States Methods	311
5.6	Nudged Elastic-Band Methods	312
5.7	String Methods	312
5.8	Approximating the Total-Energy Surface	314
6	Examples for Saddle-Point and Reaction-Path Calculations	314
7	Conclusions	318
	References	320

Simulation of Liquids	324
<i>By B.D. Todd and D.J. Searles</i>	
1 Introduction	324
2 Classical Simulation Techniques	325
2.1 Statistical Mechanical Ensembles and Equilibrium Techniques	325
2.2 Nonequilibrium MD Simulations and Hybrid Atomistic-Continuum Schemes	328
3 Potential Energy Hypersurfaces for Liquid State Simulations	332
3.1 Quantum Mechanical Interaction Potentials for Weak Interactions	334
3.2 Three-Body Interactions	336
3.3 Potential Energy Functions for Confined Fluids	337
4 Quantum Mechanical Considerations	339
4.1 Born-Oppenheimer, Car-Parrinello and Atom-Centred Density Matrix Propagation Methods	339
4.2 Hybrid Methods	340
4.3 Cluster Calculations	341
4.4 Dynamical Quantum Effects	341
5 Lyapunov Exponents	343
6 Thermodynamic and Transport Properties	344
6.1 Thermodynamic Properties	344
6.2 Free Energies and Entropy Production	347
6.3 Transport Properties	350
7 Phase Diagrams and Phase Transitions	355
7.1 Bulk Fluids	355
7.2 Phase Transitions in Confined Systems	358
8 Complex Fluids	360
8.1 Colloids, Dendrimers, Alkanes, Biomolecular Systems, <i>etc.</i>	361
8.2 Polymers	367
9 Confined Fluids	376
9.1 Nanofluidics, Friction, Stick-Slip Boundary Conditions, Transport and Structure	377
9.2 Confined Complex Fluids	384
9.3 Simple Models	389
10 Water	391
11 Conclusions	392
References	392

Combinatorial Enumeration in Chemistry

405

By A. Miličević and N. Trinajstić

1	Introduction	405
2	Current Results	405
2.1	Isomer Enumeration	405
2.2	Kekulé Structures	421
2.3	Walks	436
2.4	Structural Complexity	442
2.5	Other Enumerations	450
3	Conclusion	457
	Acknowledgment	459
	References	459

Many-Body Perturbation Theory and its Application to the Molecular Structure Problem

470

By S. Wilson

1	Introduction	470
2	Computation and Supercomputation	472
2.1	The Role of Computation	473
2.2	Supercomputational Science	475
2.3	Literate Programming	476
2.4	A Literate Program for Many-Body Perturbation Theory	482
3	Increasingly Complex Molecular Systems	510
3.1	Large Molecular Systems	511
3.2	Relativistic Formulations	511
3.3	Multireference Formalisms	512
3.4	Multicomponent Formulations	514
4	Diagrammatic Many-Body Perturbation Theory of Molecular Electronic Structure: A Review of Applications	514
4.1	Incidence of the String “MP2” in Titles and/or Keywords and/or Abstracts	514
4.2	Comparison with Other Methods	517
4.3	Synopsis of Applications of Second Order Many-Body Perturbation Theory	519
5	Summary and Prospects	523
	References	524

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 Davis² claim that molecular weight, flexibility, the number of O and N atoms and hydrogen-bond acceptors have risen, by up to

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 study³ 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 Å²) 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 score⁵ 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. Wegner⁶ 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 recently⁷). 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 training 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 group¹¹ 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 Å² 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 Beck¹⁵ 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