Molecular Interaction Fields

Applications in Drug Discovery and ADME Prediction

Edited by
Gabriele Cruciani
Molecular Interaction Fields

Edited by

G. Cruciani
### Related Titles

**Methods and Principles in Medicinal Chemistry**

Edited by R. Mannold, H. Kubinyi, G. Folkers
Editorial Board
H.-D. Hölte, H. Timmermann, J. Vacca, H. van de Waterbeemd, T. Wieland

### Previous Volumes of this Series:

<table>
<thead>
<tr>
<th>Title</th>
<th>Volume</th>
<th>ISBN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BioNMR in Drug Research</strong></td>
<td>Vol. 16</td>
<td>3-527-30465-7</td>
</tr>
<tr>
<td><strong>Chemogenomics in Drug Discovery</strong></td>
<td>Vol. 22</td>
<td>3-527-30987-X</td>
</tr>
<tr>
<td><strong>Quantum Medicinal Chemistry</strong></td>
<td>Vol. 17</td>
<td>3-527-30456-8</td>
</tr>
<tr>
<td><strong>Drug Bioavailability</strong></td>
<td>Vol. 18</td>
<td>3-527-30438-X</td>
</tr>
<tr>
<td><strong>G Protein-Coupled Receptors as Drug Discovery</strong></td>
<td>Vol. 24</td>
<td>3-527-30819-9</td>
</tr>
<tr>
<td><strong>Protein Crystallography in Drug Discovery</strong></td>
<td>Vol. 20</td>
<td>3-527-30678-1</td>
</tr>
<tr>
<td><strong>Microwaves in Organic and Medicinal Chemistry</strong></td>
<td>Vol. 25</td>
<td>3-527-31210-2</td>
</tr>
</tbody>
</table>
Molecular Interaction Fields

Applications in Drug Discovery and ADME Prediction

Edited by
Gabriele Cruciani
A Personal Foreword

I was a young organic chemist when I met Prof. Sergio Clementi, but from the very first moment I understood that his guide to the world of Chemometrics would have been a brainwashing for me. And indeed it was so. Sergio was a splendid teacher and what I know in the field of QSAR and Chemometrics is totally due to him.

From Sergio I learned the correct ways to produce mathematical models, and the tricks to interpret them. However, it was immediately clear that Chemometrics (and cheminformatics as well) can do very little when the numerical descriptors are poor or not related to the phenomena under study.

Few years later I had another brainwashing when I met prof. Peter Goodford at the European Symposium on QSAR in Sorrento (Italy). I was fascinated by his presentation and science, and I decided to learn more about. I spent in Oxford one year and Peter was a second scientific father for me. It was fantastic to complete my background working side by side with a scientist who did so much in the field of Structure-Based Drug Design. All I know on force-fields and numerical descriptions of complex phenomena such as (macro)molecular interactions is due to him. From that moment I never stopped to use his software GRID.

When ADME-attritions rate was large and in silico ADME procedures were still unknown, I went to Lausanne to learn pharmacokinetics working with Prof. Bernard Testa. Again another important man in my scientific life. Bernard pushed me deeply in the field of pharmacokinetics, and I was surprised to see how well Peter Goodford’s GRID was working in such a different field.

My scientific career was guided and complemented by these scientists, and the reasons why my interests are so sparse depend on their enthusiasm and imprinting. However, one thing I have always used in all the problems I have encountered, or in all the procedure I developed. I have always used Molecular Interaction Fields to describe the structures of chemical and biological systems. After so may years of work, I’m still fascinated by the amount of information they contain. One never finishes to find new ways to extract information from them. Moreover, combining MIF with chemometric tools, is a powerful approach to all the fields of computer assisted drug development.

This led to the production of different algorithms, all reported in this volume and all based on Peter’s GRID force field. It is noteworthy that GRID-MIF are cur-
rently applied in structure based, pharmacodynamic and pharmacokinetic fields, as well as in metabolism. Another proof, (although not necessary) of the versatility, flexibility and correct bio-parameterisation of Peter’s GRID force field.

This volume reports the MIF theory, and several applications of MIFs in different arena of the drug discovery process. MIFs are decoding the common language of the (macro)molecules, the molecular interaction potential. Using MIF is simple, interpreting them straightforward.

It was a privilege to work on this volume with such a distinguished group of contributors, and I’m sure that this volume will open a window on the fascinating world of Molecular Interaction Fields.

Finally, I want to acknowledge my coworkers at Perugia University, and Prof. Raimund Mannhold and Prof. Hugo Kubinyi for their help, contribution and encouragement to produce this book.

Perugia, June 2005

Gabriele Cruciani
Contents

A Personal Foreword  V

Preface  XIV

List of Contributors  XVI

I Introduction  1

1 The Basic Principles of GRID  3
Peter Goodford
1.1 Introduction  3
1.2 Philosophy and Objectives  3
1.3 Priorities  4
1.4 The GRID Method  5
1.4.1 GRID Probes Are Anisometric  6
1.4.2 The Target “Responds” to the Probe  8
1.4.3 The Target is Immersed in Water  10
1.5 The GRID Force Field  10
1.5.1 The Lennard-Jones Term  11
1.5.2 The Electrostatic Term  11
1.5.3 The Hydrogen Bond Term  12
1.5.4 The Other Terms  12
1.6 Nomenclature  14
1.6.1 “ATOM” Records  14
1.6.2 “HETATM” Records  15
1.7 Calibrating the GRID Force Field  16
1.7.1 Checking the Calibration  17
1.7.2 Checking Datafile GRUB  17
1.8 The Output from GRID  18
1.8.1 GRID Maps from Macromolecules  19
1.8.2 GRID Maps from a Small Molecule  24
1.9 Conclusions  25
2 Calculation and Application of Molecular Interaction Fields 27
Rebecca C. Wade

2.1 Introduction 27
2.2 Calculation of MIFs 27
2.2.1 The Target 27
2.2.2 The Probe 28
2.2.3 The Interaction Function 29
2.2.3.1 Van der Waals Interactions 29
2.2.3.2 Electrostatic Interactions 30
2.2.3.3 Hydrogen Bonds 31
2.2.3.4 Entropy 32

2.3 Selected Applications of MIFs 33
2.3.1 Mapping a Ligand Binding Site in a Protein 33
2.3.2 Deriving 3D-QSARs 34
2.3.3 Similarity Analysis of a Set of Related Molecules 36
2.4 Concluding Remarks and Outlook 38

II Pharmacodynamics 43

3 Protein Selectivity Studies Using GRID-MIFs 45
Thomas Fox

3.1 Introduction 45
3.2 GRID Calculations and Chemometric Analysis 46
3.2.1 Source and Selection of Target Structures 46
3.2.2 Selection and Superimposition of Binding Sites 47
3.2.3 Calculation of the Molecular Interaction Field 47
3.2.4 Matrix Generation and Pretreatments 50
3.2.4.1 Region Cut-outs 51
3.2.5 GRID/PCA 51
3.2.5.1 Score Plots 52
3.2.5.2 Two-Dimensional Loading Plots 53
3.2.5.3 Loading Contour Maps 54
3.2.5.4 Problems of GRID/PCA 54
3.2.6 GRID/CPCA 55
3.2.6.1 Block Unscaled Weights 56
3.2.6.2 CPCA 58
3.2.6.3 Identification of Important Variable Blocks for Selectivity 59
3.2.6.4 Contour Plots 59

3.3 Applications 60
3.3.1 DNA Minor Groove Binding – Compare AAA and GGG Double Helix 60
3.3.2 Dihydrofolate Reductase 61
3.3.3 Cyclooxygenase 61
3.3.4 Penicillin Acylase 62
6 Alignment-independent Descriptors from Molecular Interaction Fields 117
Manuel Pastor

6.1 Introduction 117
6.1.1 The Need for MIF-derived Alignment-independent Descriptors 117
6.1.2 GRIND Applications 119
6.2 GRIND 120
6.2.1 The Basic Idea 120
6.2.1.1 Computation of MIF 121
6.2.1.2 Extraction of Highly Relevant Regions 122
6.2.1.3 MACC2 Encoding 124
6.2.2 The Analysis of GRIND Variables 128
6.3 How to Interpret a GRIND-based 3D QSAR Model 130
6.3.1 Overview 130
6.3.2 Interpreting Correlograms 131
6.3.3 Interpreting Single Variables 133
6.3.4 GRIND-based 3D QSAR Models are not Pharmacophores 134
6.4 GRIND Limitations and Problems 135
6.4.1 GRIND and the Ligand Conformations 135
6.4.2 The Ambiguities 137
6.4.3 Chirality 139
6.5 Recent and Future Developments 139
6.5.1 Latest Developments 139
6.5.1.1 Shape Description 139
6.5.1.2 Anchor GRIND 140
6.5.2 The Future 140
6.6 Conclusions 141

7 3D-QSAR Using the GRID/GOLPE Approach 145
Wolfgang Sippl

7.1 Introduction 145
7.2 3D-QSAR Using the GRID/GOLPE Approach 147
7.3 GRID/GOLPE Application Examples 149
7.3.1 Estrogen Receptor Ligands 149
7.3.2 Acetylcholinesterase Inhibitors 158
7.4 Conclusion 165
III Pharmacokinetics 171

8 Use of MIF-based VolSurf Descriptors in Physicochemical and Pharmacokinetic Studies 173
   Raimund Mannhold, Giuliano Berellini, Emanuele Carosati, and Paolo Benedetti

   8.1 ADME Properties and Their Prediction 173
   8.2 VolSurf Descriptors 174
   8.3 Application Examples 179
     8.3.1 Aqueous Solubility 180
     8.3.2 Octanol/Water Partition Coefficients 184
     8.3.3 Volume of Distribution (VD) 190
     8.3.4 Metabolic Stability 192
   8.4 Conclusion 193

9 Molecular Interaction Fields in ADME and Safety 197
   Giovanni Cianchetta, Yi Li, Robert Singleton, Meng Zhang, Marianne Wildgoose, David Rampe, Jiesheng Kang, and Roy J. Vaz

   9.1 Introduction 197
   9.2 GRID and MIFs 198
   9.3 Role of Pgp Efflux in the Absorption 199
     9.3.1 Materials and Methods 199
     9.3.1.1 Dataset 199
     9.3.1.2 Computational Methods 199
     9.3.2 Results 200
     9.3.3 Pharmacophoric Model Interpretation 202
     9.3.4 Discussion 203
   9.4 HERG Inhibition 204
     9.4.1 Materials and Methods 204
     9.4.1.1 Dataset 204
     9.4.1.2 Computational Methods 204
     9.4.2 Results 205
     9.4.2.1 Nonbasic Nitrogen Subset 205
     9.4.2.2 Ionizable Nitrogen Subset 206
     9.4.2.3 Interpretation of Pharmacophoric Models 208
   9.5 CYP 3A4 Inhibition 209
     9.5.1 Materials and Methods 209
     9.5.1.1 Dataset 209
     9.5.1.2 Computational Methods 210
     9.5.1.3 Ligand GRIND Descriptors 210
     9.5.1.4 Protein GRIND Descriptors 210
     9.5.1.5 Overlap of Structures 211
     9.5.2 Results 211
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.5.2.1 Distances in the Protein Pocket</td>
<td>214</td>
</tr>
<tr>
<td>9.5.3 Discussion</td>
<td>215</td>
</tr>
<tr>
<td>9.6 Conclusions</td>
<td>216</td>
</tr>
<tr>
<td>10 Progress in ADME Prediction Using GRID-Molecular Interaction Fields</td>
<td>219</td>
</tr>
<tr>
<td>Ismael Zamora, Marianne Ridderström, Anna-Lena Ungell, Tommy Andersson, and Lovisa Afzelius</td>
<td></td>
</tr>
<tr>
<td>10.1 Introduction: ADME Field in the Drug Discovery Process</td>
<td>219</td>
</tr>
<tr>
<td>10.2 Absorption</td>
<td>223</td>
</tr>
<tr>
<td>10.2.1 Passive Transport, Trans-cellular Pathway</td>
<td>225</td>
</tr>
<tr>
<td>10.2.2 Active Transport</td>
<td>227</td>
</tr>
<tr>
<td>10.3 Distribution</td>
<td>228</td>
</tr>
<tr>
<td>10.3.1 Solubility</td>
<td>228</td>
</tr>
<tr>
<td>10.3.2 Unspecific Protein Binding</td>
<td>230</td>
</tr>
<tr>
<td>10.3.3 Volume of Distribution</td>
<td>230</td>
</tr>
<tr>
<td>10.4 Metabolism</td>
<td>232</td>
</tr>
<tr>
<td>10.4.1 Cytochrome P450 Inhibition</td>
<td>233</td>
</tr>
<tr>
<td>10.4.2 Site of Metabolism Prediction</td>
<td>233</td>
</tr>
<tr>
<td>10.4.3 Metabolic Stability</td>
<td>234</td>
</tr>
<tr>
<td>10.4.4 Selectivity Analysis</td>
<td>235</td>
</tr>
<tr>
<td>10.5 Conclusions</td>
<td>242</td>
</tr>
<tr>
<td>11 Rapid ADME Filters for Lead Discovery</td>
<td>249</td>
</tr>
<tr>
<td>Tudor I. Oprea, Paolo Benedetti, Giuliano Berellini, Marius Olah, Kim Feigin, and Scott Boyer</td>
<td></td>
</tr>
<tr>
<td>11.1 Introduction</td>
<td>249</td>
</tr>
<tr>
<td>11.2 The Rule of Five (Ro5) as ADME Filter</td>
<td>250</td>
</tr>
<tr>
<td>11.3 Molecular Interaction Fields (MIFs): VolSurf</td>
<td>251</td>
</tr>
<tr>
<td>11.4 MIF-based ADME Models</td>
<td>253</td>
</tr>
<tr>
<td>11.5 Clinical Pharmacokinetics (PK) and Toxicological (Tox) Datasets</td>
<td>254</td>
</tr>
<tr>
<td>11.6 VolSurf in Clinical PK Data Modeling</td>
<td>256</td>
</tr>
<tr>
<td>11.7 ChemGPS-VolSurf (GPSVS) in Clinical PK Property Modeling</td>
<td>257</td>
</tr>
<tr>
<td>11.8 ADME Filters: GPSVS vs. Ro5</td>
<td>261</td>
</tr>
<tr>
<td>11.9 PENGUINS: Ultrafast ADME Filter</td>
<td>264</td>
</tr>
<tr>
<td>11.10 Integrated ADME and Binding Affinity Predictions</td>
<td>267</td>
</tr>
<tr>
<td>11.11 Conclusions</td>
<td>268</td>
</tr>
</tbody>
</table>
12 GRID-Derived Molecular Interaction Fields for Predicting the Site of Metabolism in Human Cytochromes 273
   Gabriele Cruciani, Yasmin Aristei, Riccardo Vianello, and Massimo Baroni

12.1 Introduction 273
11.2 The Human Cytochromes P450 274
12.3 CYPs Characterization using GRID Molecular Interaction Fields 275
12.4 Description of the Method 279
12.4.1 P450 Molecular Interaction Fields Transformation 280
12.4.2 3D Structure of Substrates and Fingerprint Generation 281
12.4.3 Substrate–CYP Enzyme Comparison: the Recognition Component 282
12.4.4 The Reactivity Component 283
12.4.5 Computation of the Probability of a Site being the Metabolic Site 284
12.5 An Overview of the Most Significant Results 285
12.5.1 Importing Different P450 Cytochromes 287
12.6 Conclusions 289
12.7 Software Package 289

Index 291

CD-ROM Information 305
Volume 27 of our series “Methods and Principles in Medicinal Chemistry” is dedicated to “Molecular Interaction Fields” and their impact on current drug research.

In the early 1980s Peter Goodford developed the GRID force field for determining energetically favorable binding sites on molecules of known structure. The GRID force field has always been calibrated as far as possible by studying experimental measurements, and the calibration is then checked by studying how well GRID predicts observed crystal structures. Crystal packing is determined by free energy considerations rather than by enthalpy alone. The force field includes entropic terms; GRID can detect the hydrophobic binding regions which are so important when high-affinity ligands are being designed, and it can also detect sites for the polar groups which determine ligand selectivity. GRID may be used to study individual molecules such as drugs, molecular arrays such as membranes or crystals, and macromolecules such as proteins, nucleic acids, glycoproteins or polysaccharides.

Moreover GRID can be used to understand the structural differences related to enzyme selectivity, a fundamental field in the rational design of drugs. GRID maps can also be used as descriptor input in statistical procedures like CoMFA, GOLPE or SIMCA for QSAR or 3D-QSAR analyses.

The GRID force field represents the basis for several software packages specifically developed for application to pharmacodynamic aspects of drug research, including the programs ALMOND, Pathfinder, and FLAP or, in the ADME field, the programs VolSurf and MetaSite.

Correspondingly, the present volume is quite logically divided into three sections. An introductory section contains two chapters dealing with the theoretical background. The chapter of Peter Goodford, who originally developed the GRID software, focuses in detail on the basic principles of GRID, whereas the chapter by Rebecca Wade is dedicated to “Calculation and Application of Molecular Interaction Fields”.

The second section refers to pharmacodynamic aspects and contains chapters on “Protein selectivity studies using GRID-MIF” by Thomas Fox, “The Complexity of Molecular Interaction: Molecular Shape Fingerprint by PathFinder Approach” by McLay, Hann, Carosati, Cruciani, and Baroni, “Alignment-Independent Descriptors from Molecular Interaction Fields” by Manuel Pastor, “FLAP: 4-point pharma-
cophore fingerprints from GRID” by Perruccio, Mason, Sciabola, and Baroni as well as a chapter on “3D QSAR using the GRID/GOLPE approach” by Wolfgang Sippl.

The third and last section is dedicated to pharmacokinetics including chapters on “Molecular Interaction Fields in ADME and Safety” by Cianchetta, Li, Singleton, Zhang, Wildgoose, Rampe, Kang, and Vaz, “MIF-based VolSurf descriptors in Physicochemical and Pharmacokinetic studies” by Mannhold, Berellini, Carosati, and Benedetti, “Progress in ADME prediction using GRID-Molecular Interaction Fields” by Zamora, Ridderström, Ungell, Andersson, and Afzelius, “Rapid ADME filters for Lead Discovery” by Oprea, Benedetti, Berellini, Olah, Fejgin, and Boyer and finally a chapter on “GRID-Derived Molecular Interaction Fields for Predicting the Site of Metabolism in Human Cytochromes” by Cruciani, Aristei, Vianello, and Baroni.

A remarkable peculiarity of this volume is the inclusion of a CD-ROM containing some software packages used in the three sections of the book.

The series editors believe that this book is unique in its topic and presentation and adds a fascinating facet to the series. We are indebted to all authors for their well-elaborated contributions and we would like to thank Gabriele Cruciani for his enthusiasm in organizing this volume. We also want to express our gratitude to Renate Doetzer and Frank Weinreich from Wiley-VCH for their valuable contributions to this project.

September 2005

Raimund Mannhold, Düsseldorf
Hugo Kubinyi, Weisenheim am Sand
Gerd Folkers, Zürich
List of Contributors

Lovisa Afzelius
AstraZeneca R&D Mölndal
DMPK and Bioanalytical Chemistry
431 Mölndal
Sweden

Tommy Andersson
AstraZeneca R&D Mölndal
DMPK and Bioanalytical Chemistry
431 Mölndal
Sweden

Yasmin Aristei
University of Perugia
Chemistry Department
Via Elce di Sotto 8
06123 Perugia
Italy

Massimo Baroni
Molecular Discovery
215 Marsh Road
HA5 5NE Middlesex – London
UK

Paolo Benedetti
University of Perugia
Chemistry Department
Via Elce di Sotto 8
06123 Perugia
Italy

Giuliano Berellini
University of Perugia
Chemistry Department
Via Elce di Sotto 8
06123 Perugia
Italy

Scott Boyer
AstraZeneca R&D Mölndal
Global Safety Assessment
43183 Mölndal
Sweden

Emanuele Carosati
University of Perugia
Chemistry Department
Via Elce di Sotto 10
06123 Perugia
Italy

Giovanni Cianchetta
Università di Perugia
Dipartimento di Chimica e Tecnologia del Farmaco
Via del Liceo 1
06123 Perugia
Italy
Gabriele Cruciani
University of Perugia
Chemistry Department
Via Elce di Sotto 8
06123 Perugia
Italy

Kim Fejgin
Göteborg University
Department of Pharmacology
Institute of Physiology and Pharmacology
Box 431
40530 Göteborg
Sweden

Thomas Fox
Boehringer Ingelheim Pharma KG
Department Chemical Research / Structural Research
J51-00-01
88397 Biberach
Germany

Peter Goodford
Oxford University
Laboratory of Molecular Biophysics
South Parks Road
Oxford OX1 3Q
UK

Mike Hann
GlaxoSmithKline Medicines Research Centre
Structural and Biophysical Sciences
Gunnels Wood Road
Stevenage
Hertfordshire SG1 2NY
UK

Jiesheng Kang
Sanofi-Aventis Pharmaceuticals
1041 Route 202/206 N
Bridgewater, NJ 08807
USA

Yi Li
Sanofi-Aventis Pharmaceuticals
1041 Route 202/206 N
Bridgewater, NJ 08807
USA

Raimund Mannhold
Heinrich-Heine-Universität
Department of Laser Medicine
Molecular Drug Research Group
Universitätsstr. 1
40225 Düsseldorf
Germany

Jonathan S. Mason
Pfizer Global R&D
Molecular Informatics, Structure & Design
Sandwich, Kent CT13 9NJ
UK

Iain McLay
GlaxoSmithKline Research & Development
Gunnels Wood Road
Stevenage, SG1 2NY
UK

Marius Olah
University of New Mexico
School of Medicine
Division of Biocomputing
MSC08 4560
Albuquerque, NM 87131-0001
USA

Tudor I. Oprea
Sunset Molecular Discovery LLC
1704 B Llano St., S-te 140
Santa Fe, NM 87505
USA
Manuel Pastor  
IMIM/UPF  
Computer-Assisted Drug Design Laboratory  
Passeig Marítim de la Barceloneta, 37-49  
08003 Barcelona  
Spain  

Francesca Perruccio  
Pfizer Global Research & Development  
Ramsgate Road  
Sandwich, Kent CT13 9NJ  
UK  

David Rampe  
Sanofi-Aventis Pharmaceuticals  
1041 Route 202/206 N.  
Bridgewater, NJ 08807  
USA  

Marianne Ridderström  
AstraZeneca R&D Mölndal  
DMPK and Bioanalytical Chemistry  
431 Mölndal  
Sweden  

Simone Sciabola  
University of Perugia  
Chemistry Department  
Laboratory of Chemometrics  
Via Elce di Sotto 10  
06123 Perugia  
Italy  

Robert Singleton  
Sanofi-Aventis Pharmaceuticals  
1041 Route 202/206 N.  
Bridgewater, NJ 08807  
USA  

Wolfgang Sippl  
University of Düsseldorf  
Universitätsstr. 1  
40225 Düsseldorf  
Germany  

Anna-Lena Ungell  
AstraZeneca R&D Mölndal  
DMPK and Bioanalytical Chemistry  
431 Mölndal  
Sweden  

Roy K. Vaz  
Sanofi-Aventis Pharmaceuticals  
1041 Route 202/206 N.  
Bridgewater, NJ 08807  
USA  

Riccardo Vianello  
Molecular Discovery Ltd.  
215 Marsh Road  
Pinner, Middlesex  
HA5 5NE London  
UK  

Rebecca Wade  
EML Research gGmbH  
Molecular and Cellular Modeling Group  
Schloss-Wolfsbrunnenweg 33  
69118 Heidelberg  
Germany  

Marianne Wildgoose  
Sanofi-Aventis Pharmaceuticals  
1041 Route 202/206 N.  
Bridgewater, NJ 08807  
USA  

Ismael Zamora  
Lead Molecular Design  
Francesc Cabanes i Alibau, 1-3 2-1  
08190 Sant Cugat del Valle, Barcelona  
Spain  

Meng Zhang  
Sanofi-Aventis Pharmaceuticals  
1041 Route 202/206 N.  
Bridgewater, NJ 08807  
USA
The Basic Principles of GRID

Peter Goodford

1.1 Introduction

One cannot go out and buy a computer program in the confident expectation that it will do its job exactly as expected. Of course there are some things, like a lawn mower, where a relatively quick and easy test can be made to discover if it is good enough. Can it cut long grass? Cut wet grass? Does it pick up all the clippings? Will it leave beautiful light and dark stripes on the lawn? However, a molecular interaction field (MIF) is a good deal more complicated than a lawn mower, and it is not at all easy to establish which MIF programs work in a satisfactory way. Each program must be assessed very carefully before deciding what software should be used for any particular task, and many different factors must be taken into account. Some are obvious, like the available computer hardware; its speed; the size of its memory; and the amount of disk space on the user’s system. Some are less apparent, such as the objectives, priorities and overall philosophy of the people who wrote the software, and the way in which they devised and calibrated their MIF. The most important factor is to be certain in one’s own mind about the precise jobs which one wants the program to do.

1.2 Philosophy and Objectives

Even the most superficial study of molecular interaction fields shows that each MIF has its own particular characteristics. This field may put great emphasis on the accurate computation of the individual atomic charges. A different MIF may give more attention to the way in which those charges are distributed between an atom and its bonds, and a third may place some of each atom’s charge onto its lone pair electrons. Another MIF may attempt to make accurate predictions of the $pK_a$ of every polar atom, in order to be certain that each one is appropriately protonated before the MIF computations begin. Some fields may require the system under investigation to have zero overall charge. Other fields will happily do com-
putations on a couple of phosphate ions, for example, with none of the oxygens protonated and no counter cations so that the two anionic phosphates move remorselessly apart until they vanish at the edge of the universe! One field may always compute the local pH, and another may need pH information as part of the input data. This field may give detailed attention to estimating the local dielectric environment and how it changes from place to place, while that one may assume an arbitrary overall dielectric constant.

It is not only the electrostatic treatment which is different in each MIF, but also other molecular characteristics. Many fields require all the hydrogen coordinates to be defined, but some only need the location of hydrogen-bonding hydrogens, and others take no specific account of any hydrogen positions. Some fields use simple harmonic motion to describe bond vibrations, but others attempt to consider deviations from harmonicity. Some have dedicated computations which deal with hydrogen bonds, and others pay no particular attention to hydrogen bond geometry. Most fields do not allow for tautomeric changes, but some can take tautomerism into account and a few can cope with alterations in the hybridisation of an atom. Some deal exclusively with enthalpy, but others can take account of entropy which is a major component of the hydrophobic effect.

Whenever another research group begins to study MIFs, they introduce a new perspective and a new set of ideas, so the extension and improvement of force fields has been a matter of continuously improving approximations. There will never be an absolutely correct MIF, but even the very earliest work was surprisingly valuable. Huggins and Pauling [1] introduced their atomic radii seventy years ago, but they immediately extended the understanding of crystal packing and of many other properties. However no force field is perfect, and one can only hope that the approximations will continue to improve in the years ahead.

### 1.3 Priorities

The priorities of the people who create any MIF are a concrete manifestation of their scientific philosophy and overall objectives, and seven requirements seemed particularly important when the GRID force field [2] was being designed:

1. This force field was explicitly intended for use with the GRID method.
2. The overall objective was to predict where ligands bind to biological macromolecules, and so gain a better understanding of the factors involved in binding. However that improved understanding should also help in the design of improved ligands.
3. The GRID force field could have been calibrated either by using theoretical calculations, or by studying experimental observations, and after much discussion the experimental approach was adopted whenever possible.
4. The input data must always be thoroughly checked before every computation, and an associated program called GRIN was written to do this job.
5. The equations used for the computation must be reasonably straightforward, so that anybody working in the drug-discovery field (biologist, pharmacologist, medicinal chemist, crystallographer, clinician, statistician, patent expert, administrator etc.) could easily discover exactly how program GRID had calculated any particular result.

6. It must be relatively easy for anybody working in the field to interpret the output from GRID.

7. An annual reappraisal policy was established so that the worst features in the current version of GRID would always be identified, and could be dealt with appropriately in each successive year.

It is conventional to write impersonally about scientific research, but subjective decisions are made when one decides which features are worst, or which objectives are most important for a program. The personal pronoun “we” will therefore be used in this article, whenever it is appropriate to draw attention to subjectivity. “We” are still discussing priorities for the forthcoming release of GRID, but before describing the GRID force field in detail we must first describe how the GRID method actually works.

1.4 The GRID Method

There are many programs which can be used compute the electrostatic potential around a molecule. A computer model is first prepared from the x, y, z coordinates of the atoms, and this model is then surrounded by an imaginary orthogonal grid.

The next step is to compute the work needed to bring a unit electrostatic charge from infinity to the first point on the grid, and the total work required for this job is a measure of the electrical potential at that particular grid point. The same procedure is then repeated for each of the other grid points, including those which are actually inside the molecule, until the potential has been calculated for every position.

At this stage it would be possible to print out the individual potential values as a table of numbers for detailed study. The findings could then be used as input for further computations, but studying a printed data table would be a rather clumsy way of displaying the results, and a much better method is to create a three-dimensional computer plot showing a contour surface surrounding the molecule. This contour surface defines a single user-selected value of the electrostatic potential, and the final picture usually shows the molecule together with something looking rather like a child’s balloon!

Program GRID works in very much the same way, but the objective is to obtain chemically specific information about the molecule. An electrostatic potential does not normally allow one to differentiate between favorable binding sites for a primary or a secondary or a tertiary amine cation, or tetramethyl ammonium or
pyridinium or a sodium cation, and the GRID method is an attempt to compute analogous potentials which do have some chemical specificity. The generic name “target” is given to the molecule (or group of molecules) being studied by GRID, and the object used to measure the potential at each point is called a “probe”. The individual potential values are called “GRID values” and the final computer plot is called a “GRID map”. Many different probes can be used on the same target one after the other, and each probe represents a specific chemical group so that chemically specific information can be accumulated about the way in which the target might interact favorably with other molecules.

The GRID method differs in three critical ways from traditional programs which just display electrostatic potentials:

1. GRID probes are often anisometric.
2. The target “responds” when the probe is moved around it from place to place.
3. It is assumed that both the target and the probe are immersed in water.

These differences must now be considered in more detail.

1.4.1
GRID Probes Are Anisometric

Most GRID probes are anisometric because each probe represents an atom or a small group of atoms. For example a carbonyl oxygen probe is one oxygen atom with a couple of sp2 lone pairs. It has a size and a polarizability and an electrostatic charge, and each lone pair can accept one hydrogen bond. The center of the oxygen is placed at the first grid point, and a check is then made for unacceptably bad close contacts. If none is found the program then searches for nearby hydrogen-bond donor atoms on the target, and a list of those donors is made and sorted. Target atoms are rejected from the list if their donor hydrogens are pointing the wrong way, and the probe is then rotated (keeping its oxygen fixed at the grid point) so that its lone pairs will be oriented until they make the best possible hydrogen bonds to nearby target atoms. When this has been done the GRID force field is used to compute a GRID value for that particular probe at that particular point, and the whole process is repeated systematically until the potential for carbonyl oxygen is known for every grid point on the map.

An aromatic sp2 hydroxy probe differs in several ways from carbonyl oxygen. The oxygen atom of the hydroxy is placed at the grid point as before, but the probe has a larger polarizability and makes hydrogen bonds of a different strength. It can accept only one hydrogen bond, but the oxygen is bonded to a hydrogen atom which can donate. If both the donor and acceptor hydrogen bonds are made simultaneously they will be mutually constrained towards the sp2 angle of 120°. The bond length from the oxygen to its hydrogen is about 1 Å, and the probe’s donor hydrogen moves round the grid point at this distance when the probe is rotated. Figure 1.1 shows the target with an sp2 hydroxy probe placed at the first point, ready for the computation to begin.
1.4 The GRID Method

The sp2 carboxyl oxygen probe differs from both sp2 carbonyl and sp2 hydroxy, having a much greater polarizability and much greater negative charge than either. The sp3 aliphatic hydroxy probe is distinguished by making its hydrogen bonds at the sp3 angle of 109° instead of 120°, and by accepting at two lone pairs instead of just one. “Multi-atom probes” can also be used, such as aromatic carboxylate which represents the anion of a complete benzoic acid molecule. This multi-atom probe has two sp2 carboxy oxygens both bonded to the carboxy carbon.

Figure 1.1. The set up for GRID. See text Section 1.4.1.

Figure 1.2. The initial orientation of an sp2 hydroxy probe at its GRID point.

The sp2 carboxyl oxygen probe differs from both sp2 carbonyl and sp2 hydroxy, having a much greater polarizability and much greater negative charge than either. The sp3 aliphatic hydroxy probe is distinguished by making its hydrogen bonds at the sp3 angle of 109° instead of 120°, and by accepting at two lone pairs instead of just one. “Multi-atom probes” can also be used, such as aromatic carboxylate which represents the anion of a complete benzoic acid molecule. This multi-atom probe has two sp2 carboxy oxygens both bonded to the carboxy carbon.
which is bonded to the aromatic ring. Each oxygen has a couple of lone pairs, all appropriately oriented and of appropriate strength. Both oxygens are deprotonated, and they both have a substantial negative charge which is partly counterbalanced by a modest positive charge on the carboxy carbon, so the whole probe has an overall charge of $-1$. Its oxygens are both identical, and one of them is fixed as usual at the grid point. The whole multi-atom probe is then rotated to find all the orientations in which it can make good hydrogen bonds to the target, and good electrostatic interactions, while avoiding steric clashes. The chosen oxygen always stays on its grid point, and the GRID potential for that point is computed when the best orientation of the whole multiatom probe has been established.

A multiatom probe usually finds pairs of minima which would correspond in this example to the two oxygens of the carboxylate group. Of course the computation for a multi-atom probe takes somewhat longer than the map for a simpler probe, but the force field was written explicitly for the program and so GRID computations are never particularly time consuming.

1.4.2
The Target “Responds” to the Probe

Figure 1.2 shows in more detail how an aromatic sp2 hydroxy probe might be placed on its grid point at the start of a cycle of computation. In this figure the hydrogen of the probe happens to be pointing by chance towards a nearby serine residue of the target. The orientation of the probe is completely random at this early stage of the job, but with a slight readjustment GRID can make the probe’s hydrogen point directly at the serine’s side chain sp3 hydroxy oxygen. A hydrogen bond could then be formed and that would be quite a good arrangement, but a better one is shown in Fig. 1.3. Program GRID has to search and find the better alternative, and must do three things to make this happen:

1. GRID has to rotate the probe, while keeping its oxygen firmly anchored at the grid point, until its hydrogen is redirected towards the nearby backbone sp2 carbonyl oxygen as shown in Fig. 1.3.
2. The probe then has to spin about an imaginary sp2–sp2 axis (A in Fig. 1.3) which links it to the backbone carbonyl oxygen, until the probe’s own lone pair points as directly as possible towards the sp3 hydroxy oxygen of the serine.
3. The sp3 hydroxy group of the serine must finally spin about bond B which links it to the to the rest of the protein, until its hydrogen points as well as possible towards the probe’s lone pair.

This rotation of the serine oxygen is called the “response of the target to the probe”, and finding the best response is often a much more complicated job than it appears in Fig. 1.3. There are usually many different hydrogen bonding groups on or near the surface of the target, reasonably close to the probe, as shown in Fig. 1.4, and they must all be taken into account.
Methods are provided in GRID so that the user can adjust the size of the “response.” For instance he could prevent the serine hydroxy from rotating on its axis, if he knew that it was already making another strong hydrogen bond which would be broken if the probe interacted as described above. There is also a lysine side chain shown near the top of Figs. 1.2, 1.3 and 1.4, but the NH$_3^+$ group of that lysine cannot reach the probe at its grid point as things are shown in the figures. However resetting one of the directives would allow the lysine side chain to swing down, and perhaps make a useful hydrogen-bond interaction with the probe. The directives are always set by default so that things like this do not happen, and long side chains like the lysine do not normally search around during a regular GRID
1 The Basic Principles of GRID

run, unless the user has made a positive decision to release them. That kind of decision can only be taken after a thorough study of the binding site of the protein. The user must understand some or all of its properties, and this enhancement of the user’s understanding was a major objective when program GRID was being written.

1.4.3 The Target is Immersed in Water

The concept of electrical potentials was developed by physicists in the 19th century, and they quite naturally took a vacuum as their reference state. The dielectric constant of a vacuum is 1.0 by definition, and many of the early experiments on electrostatics were made in air which has a dielectric constant very close to unity. However biological systems are full of water, and biologists must invoke a dielectric constant of up to 80 in order to make traditional electrostatic calculations. It is therefore hardly surprising that MIF computations in biological systems tend to give unstable results, when such a large dielectric correction factor must be used.

The GRID force field was designed on a more appropriate basis for biology. It is assumed a priori that the environment surrounding the target has a bulk dielectric of 80, and that the dielectric diminishes towards 4 in the deep center of a large globular macromolecule. These are the default values which were used in calibrating the MIF, but of course each user can alter them to any reasonable alternative during his own GRID runs. It has been reported [3] that a value between 10 and 20 gives results which agree better with experiments on small molecules.

Some years ago a large oil company wanted to use program GRID for calculations on zeolites. These are minerals, and it was first necessary to calibrate several elements such as silicon which had not previously been used in GRID runs. Preliminary computations were then started, but the results from zeolites were misleading. A bulk dielectric of 80 would clearly be inappropriate in this case, because zeolites are used at approximately 300 °C for oil refining and are therefore completely dry. However it was impossible to find any dielectric values which yielded satisfactory results for zeolites, and this seems to demonstrate that one should not expect a single MIF to work for every system. Each force field should be calibrated for the job in hand, and much more sophisticated methods are needed if one wishes to study all 100 elements in all experimental conditions. GRID and its force field must be restricted to the wet biological environment for which they were calibrated.

1.5 The GRID Force Field

The target is always prepared and checked by an associated program called GRIN which is used before the actual GRID run begins, and perhaps the most important job of GRIN is the amalgamation of every nonpolar hydrogen atom of the