TOPICS IN
STEREOCHEMISTRY

EDITORS

ERNEST L. ELIEL
Professor of Chemistry
University of North Carolina
Chapel Hill, North Carolina

SAMUEL H. WILEN
Professor of Chemistry
City College, City University of New York
New York, New York

VOLUME 20

AN INTERSCIENCE® PUBLICATION
John Wiley & Sons, Inc.
NEW YORK / CHICHESTER / BRISBANE / TORONTO / SINGAPORE
ADVISORY BOARD

STEPHEN J. ANGYAL, University of New South Wales, Sydney, Australia
ALAN R. BATTERSBY, Cambridge University, Cambridge, England
GIANCARLO BERTI, University of Pisa, Pisa, Italy
F. ALBERT COTTON, Texas A & M University, College Station, Texas
JOHANNES DALE, University of Oslo, Oslo, Norway
DAVID A. EVANS, Harvard University, Cambridge, Massachusetts
MEIR LAHAV, The Weizmann Institute of Science, Rehovoth, Israel
JEAN-MARIE LEHN, Collège de France, Paris, France
MARIAN MIKOŁAJCZIK, Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Lodz, Poland
KURT MISLOW, Princeton University, Princeton, New Jersey
MICHI ł ORI ŌKI, Okayama University of Science, Okayama, Japan
VLADIMIR PRELOG, Eidgenössische Technische Hochschule, Zurich, Switzerland
GÜN THER SNATZKE, Ruhruniversität Bochum, Federal Republic of Germany
JOHN B. STOTHERS, University of Western Ontario, London, Ontario, Canada
HANS WYNBERG, University of Groningen, Groningen, The Netherlands
NIKOLAI S. ZEFIROV, Moscow State University, Moscow, U.S.S.R.
INTRODUCTION TO THE SERIES

It is patently impossible for any individual to read enough of the journal literature so as to be aware of all significant developments that may impinge on his or her work, particularly in an area such as stereochemistry, which knows no topical boundaries. Stereochemical investigations may have relevance to an understanding of a wide range of phenomena and findings irrespective of their provenance. Because stereochemistry is important in many areas of chemistry, comprehensive reviews of high quality play a special role in educating and alerting the chemical community to new stereochemical developments.

The above considerations were reason enough for initiating a series such as this. In addition to updating information found in such standard monographs as *Stereochemistry of Carbon Compounds* (Eliehl, McGraw-Hill, 1962) and *Conformational Analysis* (Eliehl, Allinger, Angyal, and Morrison, Inter-science, 1965; reprinted by American Chemical Society, 1981) as well as others published more recently, the series is intended also to deal in greater detail with some of the topics summarized in such texts. It is for this reason that we have selected the title *Topics in Stereochemistry* for this series.

The series is intended for the advanced student, the teacher, and the active researcher. A background of the basic knowledge in the field of stereochemistry is assumed. Each chapter is written by an expert in the field and, hopefully, covers its subject in depth. We have tried to choose topics of fundamental importance aimed primarily at an audience of inorganic and organic chemists. Yet, many of these topics are concerned with basic principles of physical chemistry and some deal with stereochemical aspects of biochemistry as well.

It is our intention to produce future volumes at intervals of one to two years. The editors will welcome suggestions as to suitable topics.

We are fortunate in having been able to secure the help of an international board of editorial advisors who have been of great assistance by suggesting topics and authors for several chapters and by helping us avoid, in so far as possible, duplication of topics appearing in other, related monograph series. We are grateful to the editorial advisors for this assistance, but the editors and authors alone must assume the responsibility for any shortcomings of *Topics in Stereochemistry*.

E. L. Eliehl
S. H. Wilen
PREFACE

The first of the four chapters in this volume of *Topics in Stereochemistry* by William C. Ripka and Jeffrey M. Blaney, deals with applications of computer graphics and molecular modeling. This is an extraordinarily active subject whose growth has been so rapid that, as yet, there seems to be a paucity of textbooks and review articles. The topic, which at its most fundamental level probes the three-dimensional interaction between molecules and between groups and atoms within molecules, is reviewed by two outstanding practitioners of the on-screen manipulation of molecular models. This technique, and the attendant calculations required to insure that the specific conformations examined are of low energy, are increasingly applied to the analysis of biochemical phenomena at the molecular level and to the design and synthesis of new medicinal agents. The authors introduce us to the specialized language that characterizes the field and provide us with a unique overview of the major software packages and of the several modeling techniques presently in use as applied to specific examples.

The second chapter, by David A. Oare and Clayton H. Heathcock, deals with the stereochemistry of uncatalyzed Michael reactions of enamines and of Lewis acid catalyzed reactions of enol ethers with $\alpha,\beta$-unsaturated carbonyl compounds. It is effectively a continuation of their definitive review of base-promoted Michael addition reaction stereochemistry that appeared in the preceding volume of the series.

In the third chapter, Nikolai S. Zefirov and Vladimir A. Palyulin have summarized the conformational behavior of bicyclo[3.3.1]nonanes and their hetero analogs. This review reflects a thoroughly modern viewpoint in which calculations, x-ray crystallographic results and spectroscopic data all are brought to bear on a polycyclic molecular framework that is able to support several relatively stable conformations including some in which boats figure prominently.

The final chapter in this volume deals with the chemistry of strained (bent and nonplanar) alkenes. Wolfgang Luef and Reinhart Keese have surveyed the recent literature with respect to syntheses and properties. This chapter also reflects the modern tendency to calculate the energies of interesting molecules and to use the calculated values to rationalize properties and to guide syntheses.
With the appearance of this volume we are pleased to welcome two new members to our Editorial Advisory Board: Marian M. Mikolajczyk (Polish Academy of Sciences) and Nikolai S. Zefirov (Moscow State University and Soviet Academy of Science). We also welcome Meir Lahav (Weizmann Institute of Science), who joined the editorial board with the appearance of Volume 19. We hope that these colleagues will help us keep in touch with stereochemical developments in Poland, in the Soviet Union, and in Israel, respectively. We also wish to acknowledge with thanks the advice received over the past decade from Professor Jan Michalski who is relinquishing his position on the Board.

Finally, with the appearance of Volume 20, the *Topics in Stereochemistry* series that began in 1967 marks a minor milestone. In recognition of this milestone, we are pleased to include a cumulative author index. We hope that this index will be useful to readers seeking to locate reviews by the name of the authors who are often leading researchers in the area which they have reviewed in this series.

Ernest L. Eliel
Samuel H. Wilen

*Chapel Hill, North Carolina*
*New York, New York*
*October 1990*
CONTENTS

COMPUTER GRAPHICS AND MOLECULAR MODELING IN THE ANALYSIS OF SYNTHETIC TARGETS 1
by William C. Ripka and Jeffrey M. Blaney
E. I. duPont de Nemours & Co., Inc.
Medical Products Department
Wilmington, Delaware

ACYCLIC STEREOCONTROL IN MICHAEL ADDITION REACTIONS OF ENAMINES AND ENOL ETHERS 87
by David A. Oare and Clayton H. Heathcock
Department of Chemistry
University of California
Berkeley, California

CONFORMATIONAL ANALYSIS OF BICYCLO[3.3.1]NONANES AND THEIR HETERO ANALOGS 171
by Nikolai S. Zefirov and Vladimir A. Palyulin
Department of Chemistry
Moscow State University
Moscow, U.S.S.R.

STRAINED OLEFINS: STRUCTURE AND REACTIVITY OF NONPLANAR CARBON–CARBON DOUBLE BONDS 231
by Wolfgang Luef and Reinhart Keese
Institute of Organic Chemistry
University of Berne
Berne, Switzerland

SUBJECT INDEX 319

CUMULATIVE AUTHOR INDEX, VOLUMES 1–20 335

CUMULATIVE TITLE INDEX, VOLUMES 1–20 339
TOPICS IN
STEREOCHEMISTRY

VOLUME 20
Computer Graphics and Molecular Modeling in the Analysis of Synthetic Targets

WILLIAM C. RIPKA* AND JEFFREY M. BLANEYb

E. I. du Pont de Nemours & Co., Inc., Medical Products Department, Wilmington, Delaware

I. Introduction

II. Computer Graphics Software and Hardware
   A. Small Molecule Construction and Modeling
   B. Macromolecular Construction and Modeling
   C. Molecular Surfaces

III. X-Ray Crystallographic Receptor Structure Determination

IV. Protein Model Building by Homology

V. Docking Small Molecules with Macromolecules

VI. Energy Calculations
   A. Molecular Mechanics
   B. Molecular Dynamics
   C. Free-Energy Perturbation Methods

VII. Distance Geometry as a Modeling Tool
   A. Methodology
   B. Generation of Conformations
   C. Energy Embedding
   D. Generation of Constrained Conformations
   E. Ensemble Distance Geometry

VIII. Binding Forces Critical for Synthetic Design of Ligands

IX. The Synthetic Design Process
   A. Optimal Atom Locations
   B. Geometric Fits of Proposed Ligands
   C. Design of DNA-Binding Drugs
   D. Design of Compounds Against Viruses
   E. Hemoglobin
   F. Dihydrofolate Reductase
   G. Phospholipase A2

*Current address: Corvas, Inc., La Jolla, CA
bCurrent address: Protos, Emeryville, CA

I. INTRODUCTION

X-ray crystallography and molecular modeling provide a detailed view of ligand–receptor interactions and have made possible a new, rational approach where molecules can be designed based on their fit to the three-dimensional structure of a receptor site. Initial research into this approach began in the late 1970s, with widespread application beginning in the early 1980s. The general methodology has been the subject of several recent reviews (1–13).

We survey computer-assisted molecular modeling with a discussion of the selection of receptor targets, the design of small molecule ligands to fit the selected target, computational methods for model building, docking, and energy calculations, and currently available software and hardware. We focus on the several features necessary to meet the demanding requirements of small molecule construction, the first step in most modeling problems dealing with synthetic design, and the docking and fitting of these structures to macro-molecular targets.

Consideration must be given to the target receptor at the outset of a molecular modeling study. Many protein and nucleic acid X-ray structures are available (14), and three-dimensional structures of small to medium sized proteins (< 100 residues) in solution can now be determined by NMR (15–17). These structures can be used directly or can serve as the basis for “homology model building.” The amino acid sequence of the target protein can be used along with one or more X-ray structures of similar (i.e., of the same family) proteins to construct a three-dimensional model of the target protein. Although attempts have been made to predict protein structures in a completely de novo fashion using rule-based approaches (18), they have not been accurate enough to use for designing small molecule ligands.

Design of potential ligands can begin after a model for the target receptor becomes available. The cases discussed in this chapter deal primarily with enzymes. Enzymes usually have well-defined “active sites” or pockets, which provide the best opportunities for synthetic design of ligands, and the
effectiveness of the designed ligand can be measured accurately by inhibition of enzymatic activity.

Conceptually, the problem of molecular design seems simple: a ligand must be designed that has a complementary surface to the receptor binding site and positions functional groups so that hydrogen bonding and other electrostatic interactions can occur between the ligand and the walls of the active site. To do this effectively, it is necessary to have some understanding of the binding forces used in macromolecule–ligand interactions. The structures of several protein–ligand and DNA–ligand complexes have been solved by X-ray crystallography and detailed studies of these systems suggest the relative importance of specific intermolecular interactions. When the three-dimensional structural information is coupled with site-specific mutagenesis of the macromolecule (leading to slightly modified structures) and kinetic or binding studies, quantitative estimates can be made of the importance of specific interactions. The design process tries to take advantage of such potentially available interactions and to provide complementary functionality in the designed ligand.

There are currently several commercial molecular modeling software packages available with varying capabilities. All of them will handle simple manipulations of structures. Features beyond these quickly become important in any sophisticated modeling study and the ease with which the software handles them is important. In particular, while the current trend toward a menu-based system satisfies the needs of the beginning and “occasional use” modeler, it can become restrictive in the hands of more experienced users. On the other hand, a powerful command language based on the ability of the program to interpret sensible English language syntax can be extraordinarily powerful, although beginners may find it difficult to use. A significant advantage of a good command language is that commands can be combined and associated in extremely versatile ways to carry out operations that were not anticipated by the software developers and therefore would not be included in menus. In an environment that must meet the needs of both inexperienced and experienced users, software that has both menus and command language must be considered. Unfortunately, much of the recent modeling software has concentrated on reinventing the wheel, ignoring previous developments and experience, and does not always provide important functionality with a good user interface. We hope that in the future new software will incorporate significant advances and that a better educated modeling community will insist on it. State-of-the-art molecular modeling systems provide extensive computational and graphics facilities for analyzing known structures and interactions, but no currently available system is capable of designing molecules by itself, so it is clear that a well-designed system must focus on maximizing the strengths of the key design component—the user.
Currently, there are few systematic approaches to the ligand design problem. Kuntz et al. (19) search for small molecules that provide a complementary steric fit to a receptor site by using a collection of spheres to define the three-dimensional structure of the site. Goodford and co-workers (20, 21) generate probe maps to find favorable placements or "hot spots" of specific functionality (e.g., hydroxyl, amine, carbonyl) in the binding site. The most frequently employed strategy uses interactive, real-time (when you turn a knob or adjust another device, you get immediate, continuous response on the display) modeling in which molecules are designed and constructed based on visual inspection of the target site (with color-coded molecular surface and perhaps probe map displays) using the intuition, experience, and creativity of the chemist-modeler. This latter approach has the advantage that the target will probably be designed with the ease of synthesis kept firmly in mind. Although computer-assisted synthetic analysis programs (e.g., LHASA (22)) can interactively guide experienced synthetic chemists and even suggest new routes, no self-contained algorithms exist that can assess the ease of synthesis of a designed structure; this places severe restrictions on any automated approach to molecular design. Currently, the most effective method to design small molecule ligands for a known binding site appears to be combining the geometric and probe algorithms with interactive modeling to suggest synthetic targets.

Once candidate structures have been designed they must be docked into the binding site. Such docking is useful to refine the initial fit and to look for alternate modes of binding. This can be done by either "off-line" processing or "on the fly" with specialized hardware and software. Another important role of docking is to use it as an initial step to screen a database of known small molecule structures to locate complete or partial structures that fit the site and, in turn, use these structures as the basis for design (Section IX.B). Molecular mechanics (Section VI) can then be used to clean up any close contacts and to estimate the conformational energy of the bound conformation of each ligand. The conformational energies of these bound ligands can then be used to prioritize them, such that the lower energy forms would be considered to be more probable than the higher ones. Current approaches for conformational searching of flexible molecules are described in a recent review article (23). Molecular dynamics and free-energy perturbation methods (Section VI) can be used to impart more flexibility to the fit and, in special cases, to estimate binding energies. Molecular mechanics calculations comparing the conformational energies of free ligand and protein with the bound protein-ligand complex are unlikely to give reasonable estimates of the binding energies except when very close analogs are being compared (24-26).

To date there are extremely few examples of de novo design of ligands based on a known binding site. This suggests the difficulty of the process. However,
several examples are available where modeling has been used to improve the binding of previously known compounds by selective structural changes. Modeling has also been extremely useful in the construction and modification of enzyme substrate analogs. Several case studies, which may give some appreciation for the problems and successes that may be encountered, are discussed later.

Finally, we address some of the newer techniques of constructing three-dimensional pharmacophore maps. When target receptor structures are not available, which is usually the case, the size and shape of the binding site are completely unknown. De novo design of a ligand is impossible without structural details of the binding site. In these cases we must rely on the natural substrate's structure, if known, or the serendipitous discovery of structures that show binding to the target receptor. Judicious synthesis of selected, preferably conformationally restricted, analogs may then provide the basis for constructing a three-dimensional model of the receptor or active site. This model can then be used to improve the binding of the known ligands and, hopefully, design novel ones.

II. COMPUTER GRAPHICS SOFTWARE AND HARDWARE

A general and effective molecular modeling system requires capabilities for constructing and manipulating both small molecules and macromolecules and should incorporate features to study their interactions. The ability to model both types of molecule in the same system is essential. Several of the systems currently available were originally designed for handling the regular, repeating polymeric structure of proteins and nucleic acids and deal rather poorly with the more arbitrary structures found in small organic molecules. Other systems, however, were initially designed for modeling small molecules and do not handle macromolecular structures well. Few systems come close to combining the best of macromolecular and small molecule modeling and provide the essential ability to interactively design and build potential ligands directly into a macromolecular binding site. We review the requirements for these two kinds of modeling approaches and suggest benchmarks to evaluate modeling software.

A. Small Molecule Construction and Modeling

For small molecules the system should allow one to construct the molecule and generate a reasonable three-dimensional conformation quickly. The best currently available approach is CONCORD (27), which rapidly (15–30 s) generates a low-energy conformation for most classes of organic compounds from a simple alphanumeric SMILES code (28), a powerful, easily learned
language for encoding chemical structures. Other approaches include AIMB, an artificial intelligence method that rapidly assembles small molecules using rules and fragments from a three-dimensional structure database (29, 30), and those that start from a simple two-dimensional sketch followed by distance geometry (31, 32) or molecular mechanics. The molecular mechanics-based approach usually requires that great care be taken when drawing the initial two-dimensional sketch and often gets stuck in unreasonable conformations. To circumvent this problem, recent systems refine the structure with molecular dynamics (Section VI), which, although time consuming, usually escapes local minima to converge on energetically reasonable conformers.

A good approach for constructing small molecules targeted at a specific site is to design and build the developing ligand piece by piece in the binding site by combining preformed three-dimensional fragments from a library. The library may contain several hundred different ring systems, chains, and functional groups, which should be selected conveniently from within the modeling system. Small molecules can be built very rapidly in this way, and the resulting structures are usually accurate enough for initial fitting or “docking” into the site model. The Cambridge X-ray Database (33) is a particularly useful source of three-dimensional fragments from which to construct small molecules and has the advantage that bond lengths, bond angles, and torsion angles are experimentally determined and represent a local energy minimum. This is particularly useful for flexible rings in which it is difficult to avoid local minima by molecular mechanics energy minimizations. Once the small molecule is completely constructed, it is usually refined with molecular mechanics and/or dynamics (Section VI).

All the above features should be tightly coupled to the graphics display, which should permit one to easily select which parts of the structure are to be acted on by a given command and to see the results in real time.

B. Macromolecular Construction and Modeling

The complexity and size of macromolecules require sophisticated graphics software and hardware to provide real-time, interactive response along with selective display and manipulation (34). A modeling system should be capable of simultaneously handling 20 or more molecules, each with several thousand atoms and thousands of molecular surface points in depth-cued (foreground objects are brighter than background objects) color, with perspective, clipping (cross-sectional display), and stereoscopic display. Each molecule must be individually adjustable in three dimensions, while simultaneously monitoring intermolecular and intramolecular distances and adjusting multiple contiguous or noncontiguous torsion angles—all in real time.

Dials, joysticks, and/or a mouse are usually used to translate and rotate
molecules and rotate bonds. A new control device, "Spaceball" (35), provides one-hand control of all six degrees of freedom (three rotational and three translational) and is a significant improvement over the earlier interactive devices. Most new systems (36) have very fast processors that do complete "bump-checking" (checking for interatomic contacts closer than van der Waals distances), molecular mechanics energy calculations, and even molecular dynamics calculations in real time. These features provide excellent feedback during interactive modeling. Selective control of which molecules or portions of molecules (e.g., an enzyme active site) are displayed, which distances and torsions are "turned on" and can be manipulated at any given time, and so on, should be easily accessible, preferably by a powerful, easily learned command language. Atoms or groups of atoms should be easily selected by interactive "picking" (selection of specific atoms on the screen by positioning the graphics cursor over them) of atoms and bonds with a mouse or stylus. It should be possible to combine interactive picking of molecules, residues, or atoms with the command language to provide full control over specifying complex combinations of molecules, residues, bonds, surfaces, labels, and so on.

Full interactive control over the position (by translation and rotation along the X, Y, and Z coordinate axes) and conformation (by adjustment of torsion angles) of both the macromolecule and the ligand(s) must be independently and simultaneously available. Convenient facilities for adjustment of torsion angles are essential, since optimization of torsion angles is often the most time-consuming aspect of interactive modeling. Several current modeling systems are limited to defining only multiple contiguous torsion angles (i.e., defined by consecutive atoms in a backbone or side chain) and otherwise can only have one torsion angle active at a time. This is a serious limitation and makes complex modeling very slow and tedious, since one usually wishes to adjust several torsion angles in different structures (or parts of the same structure) simultaneously. There should be simultaneous control of as many torsion angles as possible (e.g., 6–24), where the torsions may involve several residues or even several molecules. Choosing bond rotations should be allowed in both a forward and backward direction along a chain and these should be permitted simultaneously. The system should be capable of handling several molecules simultaneously with independent adjustment of rotations, translations, and torsions of each one, allowing the comparison of different ligands in the binding site or of different fits of the same ligand. Molecular surface displays should associate a set of dots with each atom, so that the dots move together with the atom as the molecule is moved or bonds are rotated. Solvent-accessible molecular surface calculations (37) may require long computational times for macromolecules (minutes to hours) and usually must be precalculated for use in a later interactive modeling session. The system should
support using such a precalculated surface in the current modeling session even if the molecule associated with the surface has been translated and rotated from its original position used in the surface calculation. The system should also include the option to rapidly compute van der Waals surfaces (38, 39), which are useful to generate surfaces around small molecules and around small portions of a macromolecule and display these in combination with the solvent-accessible molecular surface.

Additional useful features include the ability to enter new molecules into an ongoing modeling session at any time and to “save” individual molecules at any time. Since the conformation of the macromolecule is usually not changed during the initial modeling, it should be necessary to store its updated coordinates along with each saved “docked” ligand. Saving each macromolecule–ligand complex eventually results in confusion due to the accumulation of multiple copies of the same macromolecule coordinate set saved in different orientations relative to the screen. It is much more convenient to store each “docked” ligand conformation in a fixed orientation relative to the initial macromolecule coordinates (some systems provide an automatic facility to do this), so that only one copy of the macromolecule needs to be saved. Finally, a facility to associate arbitrary three-dimensional graphical objects with individual molecules is very useful; such objects might be electrostatic potential maps, molecular orbital plots, and electron density maps. The ability to display a molecular dynamics simulation by animation (rapidly switching from one saved coordinate set to the next) is essential; dynamics simulations produce an enormous amount of data that are difficult to interpret without a graphics display.

For peptides and nucleic acids, the system should provide rapid generation of a model from sequence data in any of the commonly observed conformations (e.g., α-helix, β-sheet, β-turn, B-DNA, Z-DNA). For peptides, it should be possible to make insertions or deletions in the sequence easily and to mutate side chains for homology model-building applications, where the sequence of the unknown structure is mapped onto the three-dimensional structure of a sequentially homologous protein whose structure has previously been determined by X-ray crystallography.

Raster graphics (used in conventional television) is now the dominant technology in interactive molecular modeling. Raster graphics technology has advanced rapidly during the last decade to the point where its price/performance is competitive with the best calligraphic (vector) systems, as demonstrated by the latest high performance workstations. In fact, only one vector display system is still commercially available, the Evans and Sutherland PS390. Vector and dot images (on raster displays) still provide the best approach for interactive molecular modeling due to their ability to provide full transparency and clipping while displaying a complex, color-coded molecular
surface and bonds in real time. These features are essential for studying interactions deep inside a macromolecular binding site (34). Stereoscopic viewing, where the left and right eye views are alternately displayed and viewed either through a mechanical or liquid crystal shutter synchronized to the display, provides a very convincing three-dimensional illusion and is extremely helpful for modeling complex interactions (40). The best currently available stereo display system places a liquid crystal polarizing screen over the graphics scope, allowing the user(s) to wear circularly polarized plastic glasses (40). A variation of this device uses battery-powered liquid crystal glasses that communicate with the monitor via an infrared sensor.

C. Molecular Surfaces

The simultaneous development of real-time interactive color graphics (34) and Connolly's molecular surface program (37) in 1980 revolutionized macromolecular computer graphics modeling. Connolly's original program implemented Richards' definition (41) of the molecular surface by rolling a probe sphere (usually a radius of 1.4 Å, the effective radius of a water molecule) over the surface of the macromolecule (Figure 1), resulting in a smooth surface of dots which represents the surface accessible to a water molecule, including
internal cavities (see color insert, Figure 2a). Bash et al. (38) and Pearl and Honegger (39) independently developed very fast van der Waals surface programs that are several orders of magnitude faster than Connolly's original program. However, they are not as effective at eliminating buried surface (the surface area on each of two atoms in close proximity to each other that is inaccessible to a probe sphere) and produce a more complicated surface display (see color insert, Figure 2b). A combination of molecular and van der Waals surface calculations provides a good compromise. Thus, it is usually more advantageous to calculate the more computationally demanding molecular surface for the macromolecule before the modeling session and to quickly calculate the van der Waals surface for the ligand and any side chains which may be adjusted in the protein during the modeling session.

When surfaces are generated around both a receptor site and a ligand to be docked, it is often difficult to visually determine how well these surfaces match. Barry (42) introduced the very useful concept of "extra radius" surface, which is calculated one van der Waals radius beyond the normal surface, collapsing the surface of the binding site onto the stick model of the ligand and eliminating the need for displaying the ligand's surface. With the receptor site surface at two van der Waals radii away from the site atoms, it is only necessary to fit the stick structure of the ligand onto this surface to obtain a good fit such that the atoms of the ligand and the site are at or beyond the sum of their van der Waals radii. This simple graphics trick makes it much easier to visualize the "docking" of a ligand into a binding site. For example, the specificity of chymotrypsin for aromatic amino acid side chains is not immediately apparent from a conventional molecular surface of its active site, while the "extra radius" surface reveals an almost perfectly planar pocket (see color insert, Figure 2c) which is obviously complementary to an aromatic ring. The "extra radius" surface can also be color-coded according to hydrophobicity or electrostatic potential.

Connolly (43, 44) and Richmond (45) also developed analytical methods for calculating molecular surface area and volume, which provide nearly exact values for the surface area and enclosed volume. Richmond's method provides analytical derivatives for surface area with respect to the cartesian coordinates of the atoms, which may be useful for docking (Section V). Connolly's algorithm also produces spectacular shaded raster graphics images (46), which give a very different "feel" for a macromolecular surface than conventional space-filling displays.

Color-coded molecular surfaces can provide qualitative or quantitative displays of hydrophobic and hydrophilic regions, neutral and charged amino acid side chains, electrostatic potential, and conformational mobility of side chains (based on the temperature factors from X-ray crystallographic refinement or molecular dynamics simulation). Color-coding by hydro-
phobicity and by electrostatic potential is particularly useful in drug design applications, where the goal is to design a molecule that is complementary in shape, hydrophobicity, and charge to a binding site. Hydrophobic color-coding originally colored all surface points associated with carbon "hydrophobic" (e.g., red) and all nitrogen and oxygen surface points "hydrophilic" (e.g., blue); a more detailed approach (47) included "neutral" or "semi hydrophilic" surface (e.g., yellow) for sulfur, \(\alpha\)-carbon atoms of amino acids, the carbon between the imidazole nitrogens in histidine, and carbonyl carbon atoms. A recent approach is based on color-coding by "hydrophobic potential" (48), calculated using partial atomic hydrophobicities (49) (analogous to partial atomic charge), and a function similar to the classical coulombic electrostatic interaction. While this approach is not based on a physically meaningful calculation, it appears to provide a qualitatively useful display of relative hydrophobicity and hydrophilicity. Electrostatic potential molecular surfaces (50) are calculated using quantum mechanically derived partial atomic charges for each atom (51, 52). The potential is typically calculated one probe sphere radius above the molecular surface, which should provide a reasonable estimate of what an incoming ligand "sees" as it approaches the macromolecule. The molecular surface is then color-coded according to the value of the electrostatic potential at each point. The electrostatic potential surface for superoxide dismutase (53) is shown in Figure 3 (see color insert). The electrostatic potential gradient can also be displayed graphically, where the gradient at each point on a grid above the molecular surface is displayed as a short vector. This method was used to locate the probable trajectory for superoxide anion as it approaches superoxide dismutase (53). More accurate estimates of electrostatic potential are available in recent methods that directly solve the Poisson–Boltzmann equation (54).

III. X-RAY CRYSTALLOGRAPHIC RECEPTOR STRUCTURE DETERMINATION

The X-ray crystallographic structure of the specific macromolecular receptor is the best starting point for designing a ligand for it. Over 300 X-ray crystal structures of proteins and nucleic acids have now been solved, including several ligand–macromolecule complexes (55); most of these are available in the Brookhaven Protein Data Bank (14). NMR is also now providing the equivalent of medium (~3 Å) resolution structures for proteins up to about 100 residues (15–17, 56).

The rate-limiting step in crystallography is still the complicated art of macromolecular purification and crystallization, which may take years of effort to find conditions that produce crystals that diffract X-rays well.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Source</th>
<th>Disease</th>
<th>Target</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemoglobin</td>
<td>Human</td>
<td>Sickle cell anemia</td>
<td>Hemoglobin S</td>
<td>180, 275–278</td>
</tr>
<tr>
<td>Phospholipase A₂</td>
<td>Pancreas, venom</td>
<td>Inflammation</td>
<td>Intracellular phospholipase A₂</td>
<td>210–212, 279–282</td>
</tr>
<tr>
<td>Dihydrofolate reductase</td>
<td>Bacteria, chicken</td>
<td>Cancer, bacterial infection</td>
<td>Dihydrofolate reductase</td>
<td>283</td>
</tr>
<tr>
<td>Carboxypeptidase A</td>
<td>Bovine, bacteria</td>
<td>Hypertension</td>
<td>Angiotensin converting enzyme</td>
<td>284–287</td>
</tr>
<tr>
<td>Thermolysin</td>
<td>Bacteria</td>
<td>Hypertension</td>
<td>Angiotensin converting enzyme</td>
<td>220, 288–291</td>
</tr>
<tr>
<td>Aspartyl proteases</td>
<td>Bacteria, human</td>
<td>Hypertension</td>
<td>Renin</td>
<td>83, 84, 292</td>
</tr>
<tr>
<td>Elastase</td>
<td>Porcine</td>
<td>Emphysema</td>
<td>Leukocyte elastase</td>
<td>261, 293, 294</td>
</tr>
<tr>
<td>Subtilisin</td>
<td>Bacteria</td>
<td>Emphysema</td>
<td>Leukocyte elastase</td>
<td></td>
</tr>
<tr>
<td>α 1-Proteinase inhibitor</td>
<td>Human</td>
<td>Emphysema, thrombosis</td>
<td>α-1-Proteinase anti-thrombin</td>
<td>295</td>
</tr>
<tr>
<td>Calmodulin</td>
<td>Rat</td>
<td>Psychosis</td>
<td>Calmodulin</td>
<td>296, 297</td>
</tr>
<tr>
<td>Interleukin-1</td>
<td>Human</td>
<td>Immune system</td>
<td>Interleukin-1</td>
<td>298</td>
</tr>
<tr>
<td>Interleukin-2</td>
<td>Human</td>
<td>Cancer</td>
<td>Interleukin-2</td>
<td>299</td>
</tr>
<tr>
<td>c-H-ras Oncogene protein</td>
<td>Human</td>
<td>Cancer</td>
<td>ras-Oncogene protein</td>
<td>197</td>
</tr>
<tr>
<td>Major histocompatibility complex</td>
<td>Human</td>
<td>Immune system</td>
<td>MHC protein</td>
<td>300</td>
</tr>
<tr>
<td>Rhinovirus HRV14</td>
<td>Human</td>
<td>Cold, hepatitis</td>
<td>HRV14</td>
<td>102</td>
</tr>
<tr>
<td>Thymidylate synthetase</td>
<td>Bacteria</td>
<td>Cancer</td>
<td>Human thymidylate synthetase</td>
<td></td>
</tr>
<tr>
<td>Hemagglutin</td>
<td>Human</td>
<td>Influenza</td>
<td>Hemagglutin</td>
<td>302</td>
</tr>
<tr>
<td>Neuraminidase</td>
<td>Human</td>
<td>Influenza</td>
<td>Neuraminidase</td>
<td>230</td>
</tr>
<tr>
<td>RSV protease</td>
<td>Human</td>
<td>AIDS</td>
<td>HIV protease</td>
<td>303</td>
</tr>
<tr>
<td>HIV protease</td>
<td>Human</td>
<td>AIDS</td>
<td>HIV protease</td>
<td>304, 305</td>
</tr>
</tbody>
</table>
Fortunately, once a parent macromolecular structure has been solved, new
structures of the macromolecule complexed with different ligands can often be
solved very quickly (within a few days in some cases (57)). These new structures
are determined by cocrystallization of the ligand–macromolecule complex or
by soaking protein crystals in a solution of the ligand and allowing the ligand
to diffuse into the binding site.

Although relatively few structures of pharmaceutically important enzymes
or receptors have been determined (Table I), the rate of solving these
structures has increased steadily during the last few years and will continue to
increase due to improvements in crystallographic and NMR methods and the
availability of new proteins through recombinant DNA approaches. Unfortu-
nately, the rate of release of the three-dimensional coordinates of newly solved,
bio logically important macromolecular structures to the Brookhaven Protein
Data Bank is decreasing, so much of the potential benefit of X-ray
crystallography for drug design is unrealized. This counterproductive situ-
ation has recently been reviewed by Richards (58).

Although X-ray crystallography represents a static, time-averaged model of
a dynamic structure, crystal structures are often good starting models for the
biologically active solution conformations. One reason for this is that the
crystals usually have very high solvent content (30–78%) and therefore mimic
the solution state quite well (59). It has, in fact, been found experimentally that
many enzymes even retain catalytic activity in the crystalline state. If one
suspects that a target protein has regions of high mobility, some information
about the flexibility of the macromolecule is provided by the crystallographic
temperature factors (B-values). Surface side chains are frequently very mobile,
as indicated by high-temperature factors, so the X-ray positions for these
atoms represent a time average. Molecular dynamics simulations (Section VI)
can estimate these intramolecular motions, although the observed extent of
these motions is limited by the time period over which the dynamics
calculations can currently be extended (up to a few hundred picoseconds). For
the above reasons and because, in practice, it is extremely difficult to hit a
moving target, virtually all design efforts begin with the static X-ray model,
although it should be kept in mind that limited conformational changes in the
protein are possible.

Accuracy of the molecular models derived from X-ray crystallography
depends on both the level of resolution and refinement (60). Refined structures
with resolutions at 2.5 Å or higher will typically have uncertainties in atomic
coordinates of up to 0.5 Å, although the average uncertainty is only about
0.25 Å. Resolution of 3 Å or poorer will usually be sufficient to trace the path of
the peptide backbone but will reveal few details about the side chains and may
contain errors. Protein structures solved by NMR currently appear to be
comparable to approximately 3 Å resolution X-ray structures.
IV. PROTEIN MODEL BUILDING BY HOMOLOGY

Although the number of protein structures defined at atomic resolution has increased in the last several years from application of improved X-ray crystallographic analysis to large proteins and of two-dimensional NMR techniques to small ones, they represent only a small fraction of the total number of proteins that have been isolated and sequenced. For rational synthetic design of ligands to be successful, information about the target macromolecules is crucial. Recent advances in computer graphics, computational techniques, and database technology have allowed approximate models to be constructed based on analogies between the protein to be modeled and other proteins of known three-dimensional structure.

Protein amino acid sequences are available from the Protein Identification Resource Data Bank at the National Biomedical Research Foundation (61). Any of the several DNA sequence data banks can also be accessed and the gene sequences converted to protein amino acid sequences (62, 63). Sequence homology between a target protein and a structurally related one can be determined using sequence alignment algorithms (64-66). Although it is beyond the scope of this chapter to describe in detail the methodology in comparative model building, the general approach is outlined here. First, correct alignment of the sequences of the structurally unknown target and the structurally known protein is essential for success and even minor errors can have serious consequences (67). Once the basic alignment is accomplished, insertions, deletions, and replacement corrections must be made in the known three-dimensional structure to transform it into the target protein. This is a critical step since insertions and deletions often occur in loop regions and these, in turn, are frequently at or near the active sites of interest and thus constitute the focal point of ligand design. The problems of modeling loop regions have been summarized by Blundell et al. (68). One particularly attractive approach to this problem searches a database of well-resolved protein structures to find all possible loops of the correct length using a method based on the distance between the two "end" x carbons of the loop (69). An alternate and effective method in the absence of satisfactory fits from the database search is to use distance geometry (Section VII) coupled with molecular mechanics and/or molecular dynamics (70). Once a satisfactory backbone structure has been obtained, appropriate side chain replacements are constructed using interactive computer graphics, with care being taken to avoid close contacts of adjacent groups. Finding acceptable rotamers of side chains can be aided by the library of rotamers compiled by Ponder and Richards (71). The entire structure is then energy minimized with molecular mechanics programs such as AMBER (72) or CHARMM (73, 74).

A number of homology-built protein models have been constructed,
including models of \( \alpha \)-lactalbumin (75–77), relaxins (78, 79), insulin-like growth factors (80), serine proteinases (81), HLA-DR antigens (82), aspartic proteinases, for example, renin (83–86), nicotinic acetylcholine receptor (partial model) (87), immunoglobulins (88, 89), human-liver alcohol dehydrogenase (90), sorbitol dehydrogenase (91), retinol binding protein (69), dimer of sea lamprey hemoglobin (92), and frog lens \( \beta\alpha_1 \) crystallin (93). In a particularly interesting approach to the model-building problem, Jones and Thirup (69) showed that a protein can be built up from a small number of large substructures taken from unrelated proteins. The several possible techniques available for model building and the future of this approach have been summarized in a recent review (68).

While molecular mechanics can be useful to clean up bad interactions in homology-built models, care must be taken not to overinterpret the results of such calculations. Novotny et al. (94) constructed two incorrectly folded proteins and showed that energy minimizations gave potential energy values comparable to the correct structures. The analysis of the incorrectly folded structures showed no bad nonbonded contacts, which suggests that their absence is a necessary but not sufficient condition for correct folding. The modelling must be accompanied by a thorough evaluation of additional factors, such as solvent accessible surface, the fraction of nonpolar side chains exposed to solvent, and other experimentally observed packing characteristics of proteins (95).

Although one would prefer the most exact models possible, even approximate models of receptors can be useful in the design of potential ligands and inhibitors. Problematic fits of proposed inhibitors can often be recognized and eliminated and reasonable candidate structures can be suggested. Because of the approximate nature of a homology-built model, the ligand "fits" will be less precise and less reliable than in those cases where an actual X-ray structure of the enzyme or enzyme–inhibitor complex is available. Rather than immediately attempting to design novel ligands for these crude models, one might take the intermediate step of proposing binding modes of linear peptide substrates or inhibitors that are known to bind to these proteins. These fits may then suggest more rigid cyclic structures which would be entropically favored over the more flexible linear peptide ligands and these, in turn, may be useful in suggesting nonpeptide mimics. Sham et al. (96) used this approach to design inhibitors of renin. They first constructed a model of the target protein based on the amino acid sequence of renin and the known three-dimensional structures of three, structurally related, fungal enzymes and a related mammalian porcine pepsin. The resulting model was used to propose the binding mode of a known linear hexapeptide inhibitor which had been synthesized from the hexapeptide substrate for this enzyme by substituting a reduced amide for the scissile bond at the peptide’s cleavage site. Several
conformationally constrained cyclic peptide inhibitors were designed to fit the active site model based on suggested cyclized versions of the bound linear peptide inhibitor, such that the preferred bound conformation of this linear peptide was not altered. Modeling was also used to explain the lack of potency of a 10-membered ring compound; its lack of activity was traced to a cis-peptide bond that forced the 10-membered ring into a conformation unacceptable for binding.

V. DOCKING SMALL MOLECULES WITH MACROMOLECULES

The "docking problem" is the positioning of a target macromolecule and a ligand so that one is a geometric and electrostatic complement of the other and there is a favorable interaction energy between them. Docking is typically done interactively with molecular surface displays (e.g., "extra radius" surface) and color coding based on hydrophobic or electrostatic potential used to guide the fit. The binding site of the protein is initially treated as being completely rigid, while the conformation of the ligand is adjusted interactively.

Physically impossible models of molecular complexes are easily built with current systems, which allow molecules to collide and pass through each other; the visual cues provided by molecular surface displays are essential for realistic modeling to avoid "close contact" problems. Some current hardware is fast enough to calculate molecular mechanics energies in real time during docking and use this information to provide feedback, thus preventing collisions or high-energy conformations. A method for real-time docking using graphics and high-speed calculations of the interaction energies between a ligand and a receptor site was developed by Pattabiraman et al. (97). A three-dimensional grid enclosing the receptor site is built prior to the docking and the van der Waals and electrostatic energies, in the absence of ligand, are calculated at each grid point. As the ligand is moved within the receptor site, the interaction energy between the ligand atoms and the grid points is calculated and updated in real time. This approach requires a close grid spacing (0.25–0.50 Å), which in turn requires substantial computer memory to store the precalculated energy grid map. The much faster workstations available today can usually perform the docking calculation without resorting to this approximate grid technique.

Evans and Sutherland developed an energy coprocessor board for their PS300 graphics system to provide real-time docking energetics (98). This coprocessor rapidly computes the pairwise steric and electrostatic interaction energies between receptor and ligand on the fly as the ligand is moved. The coprocessor handles approximately 250,000 atom pairs/s and displays the