ELSEVIER

Pharmacology & Therapeutics 85 (2000) 113–121

Pharmacology & Therapeutics

Associate editor: D. Shugar

# Molecular modelling and drug design

Edgar F. Meyer*, Stanley M. Swanson, Jocylin A. Williams

*Biographics Laboratory, Department of Biochemistry & Biophysics, Texas A&M University, College Station, TX 77843-2128, USA*

## Abstract

Drug design is a creative act of the same magnitude as composing, sculpting, or writing. The results can touch the lives of millions, but the creator is rarely one scientist and the rewards are distributed differently in the arts than in the sciences. The mechanisms of creativity are the same, i.e., incremental (plodding from darkness to dawn) or sudden (the "Eureka" effect) realization, but both are poorly understood. Creativity remains a human characteristic, but it is directly related to the tools available, especially computer software and hardware. While modelling software continues to mature, very little new has evolved in terms of hardware. Here, we discuss the history of molecular modelling and describe two novel modelling tools, a haptic device and a program, SCULPT, to generate solid molecular models at atomic resolution. 

*Keywords:* Computer graphics; Molecular modelling; Haptic device; Structure-based drug design; Molecular sculptures

*Abbreviations:* ASCII, American Standard Commission for Information Interchange; CNC, computer numerically controlled; HIV, human immunodeficiency virus; MMP, matrix metalloproteinase; PDB, Protein Data Bank.

**Contents**



## 1. Introduction

There is beauty in the fusion of structure and function. As a creative enterprise, drug design is a synthesis of scientific knowledge, experience, intuition, and aesthetics. However, unlike the arts, this beauty has limited distribution; the general public is severely underinformed about the creative process whereby molecules are designed and created. Indeed, like artists, scientists are hard-pressed to enunciate their intuitive insights. Three crucial components have converged to accentuate structure-based drug design as a product of the end of this century. Because of expanded *databases* (especially from crystallography and NMR), computer graphics *displays*, and linkage by facile *web tools* (Meyer & Funkhouser, 1998), these resources put precise molecular structures before our eyes and at our fingertips. The structural databases, which have been growing exponentially for three decades, now offer a number of interesting targets for structure-based drug design. While graphics hardware and prototypical software for drug design have been available since the 1960s (Meyer, 1980), only during the past decade have they matured to the extent that a syn-

* Corresponding author. Tel.: 979-845-1744; Fax: 979-845-1744 or 409-845-9274

*E-mail address*: E-Meyer@tamu.edu (E. Meyer)

0163-7258/00/$ – see front matter 
PII: S0163-7258(99)00069-8

thetic chemist can now master molecular modelling without having to become a computer scientist; and not very many computer scientists have become synthetic chemists.

Desktop modelling tools are now generally affordable and widely available, but they only partially engage human perceptions (spatial vision:graphics; touch:mouse). Underutilized perceptions (touch, hearing) are valuable, but neglected resources. Virtual reality, while tantalizing, has yet to make an impact on the field. For comparison, an organist utilizes 12 control channels (i.e., degrees of freedom: keys and pedals), whereas the molecular modeller uses only 2 (mouse) or 8 (control dials) input functions. Fundamental limitations also include gaps in databases (knowledge) and deficiencies in parameterization of atomic properties (information). The wisdom to use knowledge and information to ameliorate the suffering of mankind still eludes curing most diseases that arise from molecular (mal-) functions. The material aspects of the health sciences rest firmly on the foundations of structural biochemistry, which is beginning to surmount the challenge. The structure-function relationship is a fundamental component of structure-based drug design, epitomized by Emil Fischer's "lock and key" simile (Fischer, 1894; Meyer, 1995) to describe molecular interactions.

## 2. Modelling challenges

The current explosion in gene databases and "structural genomics" eventually will provide essential sequence information, but the lack of knowledge of the laws of protein folding generally prohibits the inference of structure from sequence. Targeted studies, which seek to understand the basic chemistry and physiology of a disease, are more promising. For example, human immunodeficiency virus (HIV) proteinase inhibitor development (Appelt, 1993; Kaldor et al., 1997) surely must be one of the most spectacular successes in the brief history of structure-based drug design. A novel approach scans a pathological vector using the tools of molecular biology; of the many relevant proteins produced, a few can be isolated, crystallized, and structurally elucidated. For example, *Pyrobaculum aerophilum* is reported to be a cofactor for the Rev and Rex transactivator proteins of HIV-1 and T-cell leukemia virus I. It has been studied (Peat et al., 1998) at 1.75 Å resolution and identified as a target for chemical interdiction. The structures of normal and pathological molecules can be compared and compounds designed to inhibit pathogenic enzymes or receptors selectively. *Defining the target is Challenge #1.*

So with the structure of even one target protein, and the knowledge of function of its receptor or active site, it is now possible to use computer tools to build and dock a ligand or inhibitor ("new leads") prior to investing time and resources for synthesis and testing. Conversely, large-scale screening may detect "new leads" that then must be modelled in order to explore subsequent synthetic analogs. In either case, molecular modelling is essential for understanding and exploring the structure-function relationship. Attractive and repulsive forces can be summed and the fit quantified. Ideally, one seeks a correlated listing of experimental and computational values to give assurance that novel compounds can be evaluated before being synthesized. However, there still are exceptions and unexpected surprises (Meyer et al., 1995) that must temper the enthusiasm of the molecular modeller.

Based on Fischer's "lock and key" simile, the mechanical view of molecular interactions can be easily understood and applied to biomolecules. However, (1) even "rigid" molecules have local flexibility (as, indeed, do locks) and (2) fluxional water molecules are usually a structural appendage of both the "lock" and the "key," which means the in vivo structure may differ significantly from that on the display screen. Therefore, modeling software needs to have an option to simulate the presence of pervasive water molecules *(Challenge #2: dynamic hydration).*

Molecular mechanics calculations can only seek the local energy minimum, but are unable to climb the pass into the next (and lower?) energy valley. Molecular dynamics simulations are a powerful tool for inclusion of the fluxional nature of biomolecules and in optimal circumstances, can explore the energetic landscape in search of the energy minimum. Atomic parameters (simplified quantum mechanical values) are approximate and based on a generic, classical atom, whereas these parameters will change in a fluxional structure, so quantum molecular dynamics is needed. This field has yet to mature, and necessary computational resources greatly exceed today's supercomputers, not to mention the desktop computer. Again, how does one treat water rigorously (dielectric constant, ionization state, fluxional H-bonding; bulk vs. microscopic quantities)? *Challenge #3 is a rigorous computational simulation of a biochemical reaction in a form accessible to the synthetic chemist*, as discussed by Professor Ursula Roethlisberger (ETH Zentrum, Zürich, Switzerland) at this symposium, as well as the work (unpublished) of Professor Bogdan Lesyng and colleagues at the ICC, University of Warsaw.

## 3. Gaps in databases

Molecular structure-based drug design is an art and a science. While graphics hardware has matured and software continues to mature, the exponential growth of databases still is insufficient to provide a rigorous basis for many drug design efforts. These gaps have several origins: (1) research yet to be done; (2) research performed, but unpublished or otherwise inaccessible; and (3) missing or incomplete links between databases. While the first gap has been defined as Challenge #1, dramatic changes in storage media can make archival material physically present, but computationally inaccessible (e.g., who can still read archival 800, 1600, or 6250 bpi magnetic tapes?).

The extensive databases of industrial firms may never become available to the academic scientist, even though projects may be completed and patents granted. The third

and especially relevant source of "lost data" is the hundreds, approaching thousands,[1] of incomplete structures that contain a wealth of information on intermolecular interactions, information vital to the design and modelling of novel complexes. Some of these structures may not be fully refined, but each macromolecular complex contains abundant information about atom-to-atom interactions that collectively amount to an enormous database of ligand-protein interactions that begs for attention.

Shelved information is lost information in the scientific world. As an analogy, envision colorful medieval manuscripts in a monastic library, isolated by a sea of security. One can extrapolate the widening cap between industrial (publicly inaccessible) and public (open) databases into the next century; the widening chasm poses a major challenge to academic scholarship and ultimately to the industries that will employ later generations of students.

A need exists for a database archiving dynamic simulations of macromolecules. Even though early protein structures were poorly refined because refinement programs did not exist, and later structures were refined only with available tools, which may not meet today's standards, these structures are valuable historically, as well as being available for refinement with modern tools. Likewise, although dynamic simulations depend upon algorithms and parameters, which are still being improved, the results of such calculations should be archived, together with the parameters used to generate them. In short, who knows to what tools scholars of the future will have access? We owe it to future generations to archive our results.

The physiologist studies the functional properties of an organism. The biochemist studies the feedback mechanism of metabolic signaling and control. The crystallographer studies the structure-function relationship and the crucial role of individual chemical groups. Other disciplines assemble their respective databases. Some of these databases are relatively subjective, others are highly precise. Some databases are critically reviewed and checked for veracity, while others are collections of measurements or observations from uncontrolled experiments. While breadth of knowledge and inspired intuition can link disparate data, we stand at the threshold of having integrated data storage and retrieval capabilities. Relational database methodology could link physical, chemical, structural, metabolic, physiological, pharmacological, genetic, kinetic, synthetic, and other databases. A few are linked now. The question is when and how well will they be integrated. For example, an electronic version of this review would have active, clickable nouns, as in footnotes 3 and 4. The reader could click on a reference and call up the original paper and related papers, or click on a structural diagram and view the structure in three-dimensions, or click on a compound name and cross-correlate pharmacokinetics with organism, or biological age, or metabolic defect. New vistas await learning, thinking, and the intuition of the experimental scientist, as well as the philosopher.

Therefore, in this incomplete, imperfect world, scientific progress can be made and compounds successfully designed, only by using available resources and by designing experiments to close information gaps. However, the creative process may also have an artistic side, which further stimulates the inventive mind to make intuitive leaps across database gaps to link the known with the unknown. Such a leap generates a hypothesis, which calls for an experiment to confirm its validity.

## 4. Haptic interaction

The first 100 plus years of molecular modelling utilized solid, hand-held mechanical models. Computer graphics makes nanoscale (Å) structural models visible to and adjustable by humans. Graphical models can be superimposed to reveal structural differences—the virtues (and limitations) of molecular graphics are well-known. However, graphical depictions commonly in use today frequently discard essential atomic details for an "artistic" representation of secondary/tertiary structures in terms of colorful arrows and swirls. The fruitful efforts required to obtain a precise 2 Å resolution structure are disregarded when the displayed image has an effective resolution of 6 Å.

It is well known that while molecular structures are three-dimensional, the display screen is two-dimensional, so some tricks are necessary for three-dimensional visualizations. Sadly, the mouse is only a two-dimensional device, a poor choice for three-dimensional input. The more input/output available to our senses, the more meaningful the modelling process. We, therefore, have begun to develop a program that makes use of the tactile PHANToM (Fig. 1), a haptic device that provides six-dimensional input (three orthogonal translations and rotations), but most remarkably, also gives tactile, three-dimensional translational feedback. Phase 1 of our research is concerned with implementing the PHANToM as a graphical three-dimensional cursor with force-field feedback, where "real-time" refers to the human clock (milliseconds) rather than the atomic clock (femtoseconds).

Phase 2, which is currently receiving our attention, involves synchronous ("real-time") calculations of molecular interaction forces. The objective is to permit the operator to feel (as well as see) the net attractive or repulsive forces involved in ligand docking. In order to maintain a realistic, functional cycle time, simplifications must be made to include only relevant binding-site atoms.

An obvious application of this method is to let the user sample a number of "what if" binding situations quantitatively, visually, and tangibly. The operator can temporarily give zero weight to forces and move a model through or above energy barriers that are prohibited by molecular mechanics calculations. Promising conformations can be cap-

[1] An estimated 350 structures of complexes of the HIV proteinase (70 structures determined at Agouron alone) are represented by the 36 entries currently in the PDB and 184 in Wlodawer's database. Will the others evaporate?

Fig. 1. Dr. Stanley Swanson juxtaposed with the PHANToM manipulator. Small positional sensors and servomotors are barely visible. This photograph makes it possible to visualize the translational latitude (ca. 20 cm) available for interactive input/output.

tured for off-line evaluation using programs like CHARMM, GROMOS, AMBER, etc. This method thus will permit the creation and evaluation of a large number of trial conformations, more closely approaching actual molecular interactions, but guided by the knowledge and intuition of the chemist, winnowing through a myriad of possibilities to explore likely conformations and configurations.

Phase 3 will involve the optimal inclusion of pervasive water molecules (Meyer, 1992) to aid (or compete with) binding interactions. (Think of a water molecule acting like a grain of sand or as glue interfering with or enhancing binding.)

Phase 4 will involve simulation of a receptor in a reaction intermediate state, and following the suggestion Pauling (1946), the modelling process could involve the ligand-assuming geometric values corresponding to a reaction intermediate (most likely *not* at an energy minimum).

## 5. Tactile models

### *5.1. Permanent versus disposable models*

Molecular models are essential for both teaching and research. In the first years of protein crystallography, protein structures were carefully constructed of scaled (2 cm = 1 Å), brass Watson-Kendrew (Kendrew et al., 1960) components. The investigator enjoyed a tactile, as well as a spatial, perception of these complex structures as they emerged. While these models were difficult to construct, maintain, and export, their initial role in protein crystallography was crucial because they were directly linked to meticulously hand drawn electron density maps, which were the experimental basis for their veracity. Beginning in the 1970s, brass models were replaced by molecular graphics, as facile hardware and software gradually matured and became more widely available (Collins et al., 1975; Morimoto & Meyer, 1976).

Today, interactive molecular modelling enjoys a prominent place in the arsenal of the chemist and biologist because of a number of key components: facile software driving graphics hardware, a rapidly growing Protein Data Bank (PDB[2]; Bernstein et al., 1977a, 1977b, 1978; Meyer, 1997), and responsive networking linkages. Yet, rarely does the architectural grandeur of a macromolecule receive broader distribution as three-dimensional representations beyond pictures of the graphics display. Workstation images are discarded at the rate of 25–30/sec or faster (the interlaced refresh rate of graphics hardware). To restore balance and recover the historical roots of molecular modelling, this project has begun to create scaled, solid molecular models of the macromolecular surface and specifically, the scaled models of ligand-binding sites of selected proteins, as recognized by various substrates and inhibitors. Admittedly, models are only representations of reality, and each model has inherent strengths and limitations. Only rarely can a model convey dynamic characteristics (e.g., a movie); most mechanical models are rigid (e.g., snapshots) that still convey precise spatial relationships based on solid experimental data.

### *5.2. Solid, hand-held models*

More than by seeing, we learn by doing. Today, molecular models are mostly visual (graphical) rather than tactile,

[2]<http://www.rcsb.org/pdb/>.

yet tactile stimuli provide immediate modalities for learning and research. This project provides access to hand-held models for teaching, docking, and manipulation. Physical (hand-held) atomic models of small molecules (carbohydrates, nucleic acids, oligopeptides, lipids, and drug molecules) are commercially available (ca. $0.10–5.00/atom),[3] but a convenient source of physical macromolecular models is lacking. Macromolecular structures containing thousands of atoms can be visualized by color computer graphics and facile web tools (e.g., program RASMOL or CHYME), easily accessible from the PDB. Yet, these images, while possibly three-dimensional on the screen, become two-dimensional when printed or simply disappear at the end of a modelling session. Color raster graphics was first introduced 32 years ago (Meyer, 1968, 1997), and it is the basic technology of today's graphics workstation, desktop, and laptop computers. Now, 32 years later, molecular modeling takes another step towards the development of precise, scaled models that can be handled and docked.

A newly developed program creates scaled molecular models of selected macromolecular domains, such as active sites, receptor sites, pores, or contact regions.

For manipulation and docking exercises, a scale (e.g., 1 cm = 1 Å = 0.1 nm) is first chosen to correspond to a commercially available atomic model set. A new dimension presents itself for teaching and research: docking a ligand with even 5 binary variables (i.e., 32 degrees of freedom) becomes a challenging, three-dimensional puzzle. For example, the simple dipeptide SerGlu, with 9 torsional degrees of freedom and 2 chiral centers ($2^{11}$ = 2048 possibilities) becomes a nontrivial challenge, especially as the complexity increases exponentially. This project, therefore, uses novel software to drive existing technology (computer graphics, computer numerically controlled (CNC) milling machines) and thus, uses computational methods to address a fundamental challenge in chemistry and biology: the manifestation of the "lock and key" simile of Emil Fischer. Historically, although Kekule had visualized benzene, the first documented molecular model was presented to the public in a Friday Evening Discourse at London's Royal Institution on April 7, 1865 (Hofmann, 1865).[4]

### 5.3. Computer numerically controlled milling machine: SCULPT

CNC milling machines produce the building blocks of modern technology and commerce, from gears and tools to massive automobile and airplane components (Fitzpatrick, 1996). Like the simple, "move-draw" commands of the *x–y* plotter, the instruction set of CNC machines is relatively limited, both in terms of dimensions (most typically in two-dimensional) and in terms of functions (linear, circular, spiral machining). CNC machines provide the desirable characteristics of scalability, reproducibility, and precision. Comparable CNC milling machines are found in the workshops of physics or chemistry departments of major universities or research laboratories. Writing a computer program to create models of macromolecular binding domains is based on complex assumptions, both in the choice of algorithms, but especially for optimization of model, materials, and machining methods. The optimal solution varies greatly with the choice of materials (wax, plastic, paper, wood, stone, metals), some of which are well suited as molds for mass-production materials (plastic, rubber).

### 5.4. Method

Program SCULPT (currently 3000 lines of FORTRAN code and comment statements) accepts the output of graphics display programs (atomic coordinates) and outputs an American Standard Commission for Information Interchange (ASCII) file with CNC instructions to create scaled models of molecular surfaces or domains based on standard atomic and bonding parameters. This discussion will be restricted to the experimentally determined Cartesian coordinates of spherical atoms, as found in the PDB. First, the region of interest (e.g., active site) is displayed graphically, oriented, and scaled interactively. An orientation matrix and origin are output to program SCULPT as an ASCII file, along with the list of atoms to be included in the model. Additional input parameters include the scale that relates the input coordinates to the output model, the upper limit being dictated by the size of the "part," which is related to the upper displacement limits of the milling machine. Lower limits are dictated by the size, precision, and strength of cutting tools.[5]

For example, the active site region of an enzyme might be modeled as a block 30 × 20 × 10 cm at a scale of 1 cm/Å in 12–18 mm (1/2–3/4 inch) sections (the approximate size and shape of a shoebox). In **Phase 1**, the CNC milling machine first removes the empty, void material in the binding site in a systematic, linear (step-wise) fashion. An input parameter allows the definition of a buffer region around atomic surfaces, permitting the employment of a larger cutting tool. An end-mill cutting tool looks much like a drill bit, but has a thicker shank. Typical tool bits have flat or rounded (ball) tips, with a typical step size of 0.6 times the tool diameter, thus allowing an increased translational speed without sacrificing precision. Spindle (e.g., drilling) speed and part-positioning speed are dependent upon the choice of material (e.g., 2000 rpm and 100 cm/min for wax; 10,000 rpm and 60 cm/min for hardwoods such as oak).

[3] HGS models <www.maruzen.co.jp/home-eng/HGS.html>, Dreiding models <http://www.brinkmann.com/ster_stermods.html>, CPK models (Harvard Apparatus <http://www.harvardapparatus.com/PDFFiles/s98s54.pdf>), Darling models <http://www.webcom.com/darling/modrep.html>, as found, for example, in the Aldrich catalogue.

[4] <http://ursula.chem.yale.edu/~chem125/>.
<http://ACS.TAMU.EDU/~efm6889/sculpt/sculpt1.htm>.

[5] This method would permit the design of unique nano-scale structures, such as molecular sieves, pipes, or structure-specific cavities. It also could be used to make scaled, solid models of biological receptor cavities defined by diffraction methods, electron microscopy, or atomic force microscopy.

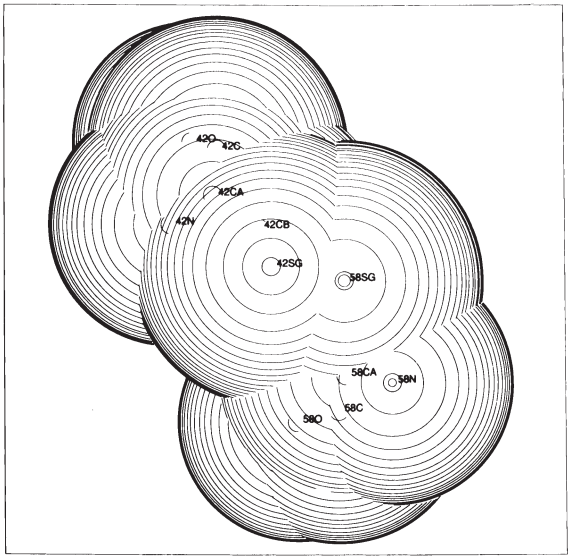


Fig. 2. The path traversed by the cutting tool in preparing a *bas-relief* model of the Cys-Cys disulfide linkage. For clarity, only every fourth step is shown. Atoms are labelled and a small arc or circle at the atomic center indicates the size of the cutting bit (<http://cmdnmr.tamu.edu/meyerlab/mlab/index.html>).

While the CNC instructions are output as an ASCII file by program SCULPT in a few minutes, the time needed by the milling machine to execute the instructions (typically hours) is directly related to the product of translation speed, part dimensions, and incremental step size. For debugging and visualization purposes, and especially for safety, industry-standard software (e.g., EZ-MILL) is used to check each processing step visually. In addition, an option in program SCULPT writes a PostScript (1986) file with an image (Fig. 2) of the paths of the cutting tool as a useful visual check of specific processing steps. Besides being printed, this output, properly scaled, can be used for "Xerox-type" three-dimensional rapid prototyping machines, which produce laminated paper or plastic objects.

The CNC milling machine reads an ASCII file containing a sequence of internationally recognized and accepted commands defining the cutting tool position and velocity (CNC command G01) and rotation (M15, M16) direction to remove the void region by translating, stepwise, the emerging "part" under the cutting tool in millimeter (G71) or inch (G70) units. Contact of the tool with the emerging surface is anticipated by calculating the vector components of positional parameters; dimples and crevices can thus be machined to the effective radius of the cutting tool. The typical result is a smooth or textured surface.

Interior concave regions cannot be carved directly by a two and one-half-dimensional CNC milling machine. In order to carve out concavities, the tiltable carriage characteristics of a full, three-dimensional machine (G72, G73) are required. This feature greatly increases the cost of the CNC machine and is replaced here by the "bottom-up"-"top-down" two-pass machining of sections, which subsequently are assembled as the model grows in the vertical direction. For the standard two and one-half-dimensional CNC machine, program SCULPT, therefore, causes the maximum spherical radius of an atom to be projected downward on lower surfaces for later machining. Optimal positioning of the binding site on the graphics display sometimes can be used to minimize overlapping atoms and the number of buried concavities.

**Phase 2** milling employs the circular (G02, G03) and spherical (G72, G73) interpolation CNC commands to create spherical surfaces with finer definition. Cutting tool diameter and related step size can be reduced for finer definition, but require proportionally increased processing times. When needed, **Phase 3** machining can be used to create color-coded or textured surfaces (e.g., indicating atom type or other locus-dependent characteristics) to convey chemical or functional information. Currently, special wood-compatible color pens are being developed to be used in the place of an end-mill tool bit to color and, therefore, help identify heteroatoms (red = O, blue = N, etc.).

For example, SCULPT can create surface dimples (like a golf ball), ridges, stripes, spirals, etc., which are especially useful for Selective Laser Sintering rapid prototyping. Rapid prototyping (Stereo Lithography, Selective Laser Sintering) processes use tiny resin-coated spheres (nylon, glass, etc.) to construct plastic models with 4 μ precision, with an upper size limit of ca. 12 inches (30 cm) and at a cost of ca. $100/model. These also could be used as molds for mass production of models for teaching. Like the early days of computer graphics, these models are monochromatic, so one of the major challenges is the introduction of texturing to distinguish heteroatoms.

To a first approximation, ligand/inhibitor-binding sites are either exposed or buried. Models of exposed sites are readily created, an example being the extended binding site of the proteolytic enzymes trypsin (Marquart et al., 1983) and thrombin (Bode et al., 1989). There are any number of examples of buried binding sites with pharmacological significance (e.g., aspirin or ibuprofen [Picot et al., 1994], acetylcholinesterase [Harel et al., 1993], the proteosomes [Lowe et al., 1995], ion channels [Wiener et al., 1997], etc.). Models of buried sites can be created by slicing through the binding site to make paired segments (right-left or top-bottom) of the binding cavity.

While the surface of a binding or receptor cavity can be described by the convergence of atomic surfaces, the choice of atomic radii markedly affects the shape of the surface: covalent radii give a "bumpy" surface that is defined by clusters of discrete atoms. The van der Waals radius better describes the impenetrable surface characteristic of tight ligand-binding interactions. The double van der Waals radius surface is still smoother and represents the confluence of atomic surfaces to form the binding site, as well as anticipates the linear interactions of ligand and receptor atoms (e.g., useful with Dreiding models). One, therefore, has considerable latitude in defining compatible surfaces and employing these selected parameters to create solid, tactile models of complex biological systems. A large number of

variables requires optimal decisions for optimal results, but also provides the opportunity to create sculpted models that depict various characteristics of atomic and molecular interactions. While results are obtained in hours rather than fractions of a second (as with interactive graphics), the resulting models have enduring value as a complement to ephemeral computer-generated images. These models are useful for teaching undergraduate biochemistry and can be used equally well for the teaching of high school biology and chemistry. While millions of individuals can benefit from a miracle drug, most are uninformed about the drug's structure. A molecular sculpture should be both aesthetically pleasing ("museum quality"), as well as structurally precise. Such sculptures could help open the eyes of the public to the beauty and intricacies of nature at the molecular level.

### *5.5. Examples for teaching and research*

Our serendipitous study of haemorrhagic rattlesnake collagenase, Ht-d (Zhang et al., 1994), has been directly linked to human tumor necrosis factor-α convertase (Maskos et al., 1998) and other mammalian matrix metalloproteinases (MMPs) in the ADAM family with significant pharmacological implications (Gomis-Rüth et al., 1998). Mammalian, reptilian, and bacterial MMPs share common features, even though their primary sequences differ greatly outside the signature Zn-binding site sequence, HExxHxxGxxH.[6] The primary specificity (S1′) site is exceptionally interesting because it is both much deeper (Botos et al., 1996, see p. 2752) than the length of any amino acid side chain in collagen (indeed, the S1′ site is a tunnel that extends completely through the enzyme structure), but also because it is universally conserved (bacterial, snake venom, and mammalian collagenases).

Program SCULPT has been used to create solid models of the active sites of several enzymes using a CNC milling machine in the teaching laboratory of the Department of Engineering Technology and Industrial Distribution. Program SCULPT produces structurally significant, geometrically precise models of selected regions of macromolecules scaled to the dimensions of commercially available atomic models.

A model of the S1′- and Zn-binding sites of mouse collagenase-3 (Botos et al., 1999) has been constructed with dimensions 17 × 12 × 6 cm, based on the atomic coordinates of our 2.0 Å resolution structure. The model (Fig. 3) was constructed from 4 slabs of oak with dimensions 7.1 × 5.5 × 3/4 inches (18 × 14 × 1.8 cm) to a scale of 1 cm/Å. Depending on void volume and atomic surface area, each slab required ca. 6 hr to machine. The next bottom-up slab was glued to the growing model before being machined top-down. While conventional wood glues set in a few hours, it is convenient to machine 1 slab/day, so, at best, a 3 inch (7.5 cm) thick model requires 24 hr of machining over 4 working days.

In practice, one starts with a slab (e.g., of wood) rather than a block. Program SCULPT, therefore, can make two passes (bottom-up and top-down) of each slab in order to create three-dimensional concavities. The thickness of the slabs is dictated not only by the scale (slab thickness <atomic radius) but is especially limited by the cutting depth of end mill tools. The lateral forces of a milling machine can be so great that only rarely can one use a standard conical-tipped drill bit. Conveniently available end mill (e.g., 6 mm; 1/4 inch) cutting tools have a thick shank (10 mm; 3/8 inch) and a cutting depth

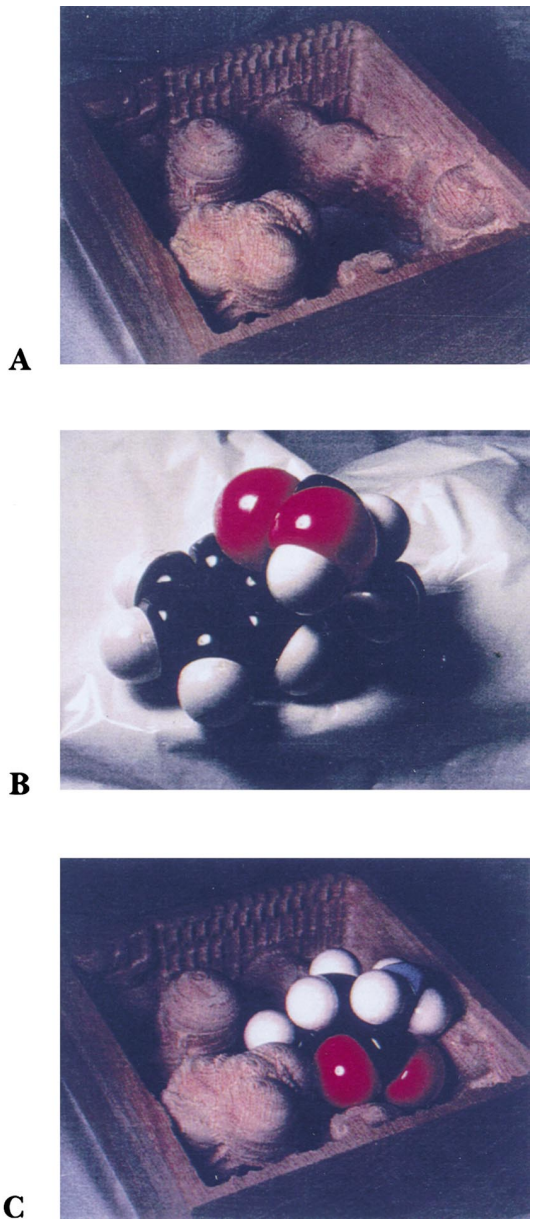


Fig. 3. A scaled model of the active site of the MMPs. (a) The Zn-binding site and empty S1′ primary specificity pocket of mouse collagenase-3 (Botos et al., 1999). (b) An HGS model of phenylalanine (scale of 1 Å = 1 cm). (c) The primary specificity pocket of mouse collagenase-3 easily contains the phenyl side chain of Phe while permitting the carboxylic acid O atoms to chelate with the active site Zn atom. This view illustrates the binding mode of BB-94 (Botos et al., 1996).

[6] <http://cmdnmr.tamu.edu/meyerlab/mlab/index.html>.

of 12–25 mm (1/2–1 inch), the latter being required for machining 18 mm (3/4 inch) slabs of wood. Program SCULPT anticipates the forces on these longer tools by reducing initial attack velocities, but the large variety of conditions involved in machining a complex three-dimensional object preclude a comprehensive solution at this time. This tangle of prose attempts to convey the levels of complexity and necessarily sequential order in the choices and limitations of materials and machining methods, which must be solved on an ad hoc basis through experience and trial and error methods.

## 6. Conclusions

Since the publication of the first three-dimensional structure, myoglobin, 40 years ago and the development of 3 crucial components, the PDB, three-dimensional computer graphics, and networks, a fundamental understanding of molecular structure and function has deepened and enriched our understanding of the life sciences. While many molecular-based diseases exist, only very few (e.g., thrombin inhibitors as anticoagulants, HIV proteinase inhibitors to arrest acquired immunodeficiency syndrome) have yielded thus far to structure-based drug design. The challenge is as great as is the potential to meet the challenge.

Originally, structure analysis required hands-on construction of tactile (Kendrew) models, a tedious, but information-rich, procedure, which was replaced by graphics techniques 25–30 years ago. Subsequently, visual perceptions have dominated structural biology. This review describes two efforts to restore tactile manipulations of complex molecular interactions to enhance both teaching and research in structural biology.

We stand at the threshold of a new century, a new millennium, and a new age of scientific research. Whereas computers made peripheral impact in the past, they are likely to be central to every aspect of research in the future, even—or especially—the creative aspects. The first successes show where gaps and weaknesses persist, to be rectified in the century ahead. Will the intellect become the slave of the computer in the new century? The father of cybernetics gave the answer (Wiener, 1964) in his paraphrase, "render unto man the things which are man's and unto the computer the things which are the computer's."

## Acknowledgments

Professor Nancy Amato has collaborated and made available the PHANToM haptic device in the Department of Computer Science and Mr. Tom Beck has provided helpful advice and access to the CNC milling machine in the Department of Engineering Technology and Industrial Distribution at this University. Professor J. D. Dunitz pointed to Professor Michael McBride's Chemistry 125 web page[7] at Yale University (New Haven, CT, USA) for the reference to the first molecular model. We are especially grateful to the Robert A. Welch Foundation, the Texas Imported Fire Ant Research & Management Plan (Dr. B. M. Drees), the TAMU College of Engineering Summer Research Grant, and the Texas A&M Honors Program for financial support.

[7] <http://ursula.chem.yale.edu/~chem125/>.

## References

Appelt, K. (1993). Crystal structures of HIV-1 protease-inhibitor complexes. *Perspec Drug Discov Des 1*, 23–48.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr. Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977a). The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur J Biochem 80*, 319–324.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr. Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977b). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol 112*, 535–542.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr. Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1978). The Protein Data Bank: a computer-based archival file for macromolecular structures. *Arch Biochem Biophys 185*, 584–591.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R., & Hofsteenge, J. (1989). The refined 1.9 angstroms crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J 8*, 3467–3475.

Botos, I., Scapozza, L., Zhang, D., Liotta, L. A., & Meyer, E. F. (1996). Batimastat, a potent matrix metalloproteinase inhibitor, exhibits an unexpected mode of binding. *Proc Natl Acad Sci USA 93*, 2749–2754.

Botos, I., Meyer, E., Swanson, S. M., Lemîatre, V., Eeckhout, Y., & Meyer, E. F. (1999). The crystal structure of recombinant mouse collagenase-3 (MMP-13). *J Mol Biol 292*, 837–844.

Collins, D. M., Cotton, F. A., Hazen, E. E., Jr., Meyer, E. F., Jr., & Morimoto, C. N. (1975). Protein crystal structures: quicker, cheaper approaches. *Science 190*, 1047–1053.

Fischer, E. (1894). Einflus der Configuration auf die Wirkung der Enzyme. *Ber Deutschen Chem Gesellschaft 27*, 2984–2993.

Fitzpatrick, M. (1996). *Computer Numerical Control Simplified.* Westerville: Glencoe/McGraw-Hill.

Gomis-Rüth, F.-X., Meyer, E. F., Kress, L. F., & Politi, V. (1998). Structure of adamalysin II with peptidic inhibitors. Implications for the design of tumor necrosis factor α convertase inhibitors. *Protein Sci 7*, 283–292.

Harel, M., Schalk, I., Ehret-Sabattier, L., Bouet, F., Goeldner, M., Hirth, C., Axelsen, P., Silman, I., & Sussman, J. (1993). Quaternary ligand binding to aromatic residues in the active-site gorge of acetylcholinesterase. *Proc Natl Acad Sci USA 90*, 9031–9035.

Hofmann, A. W. (1865). On the combining power of atoms. *Proc R Inst 4*, 401–430.

Kaldor, S. W., Kalish, V. J., Davies, J. F., Shetty, B. V., Fritz, J. E., Appelt, K., Burgess, J. A., Campanale, K. M., Chirgadze, N. Y., Clawson, D. K., Dressman, B. A., Hatch, S. D., Khalil, D. A., Kosa, M. B., Lubbehusen, P. P., Muesing, M. A., Patick, A. K., Reich, S. H., Su, K. S., & Tatlock, J. H. (1997). Viracept (nelfinavir mesylate, AG1343)—a potent, orally bioavailable inhibitor of HIV-1 protease. *J Med Chem 40*, 3979–3985.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., & Shore, V. C. (1960). Structure of myoglobin. *Nature 185*, 422–427.

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., & Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science 268*, 533–539.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr 39B*, 480–490.

Maskos, K., Fernandez-Catalan, C., Huber, R., Bourenkov, G. P., Bartunik, H., Ellestad, G. A., Reddy, P., Wolfson, M. F., Rauch, C. T., Castner, B. J., Davis, R., Clarke, H. R., Petersen, M., Fitzner, J. N., Cerretti,

D. P., March, C. J., Paxton, R. J., Black, R. A., & Bode, W. (1998). Crystal structure of the catalytic domain of human tumor necrosis factor-alpha-converting enzyme. *Proc Natl Acad Sci USA 95*, 3408–3412.

Meyer, E. F. (1968). Annual Report, Brookhaven National Laboratory: First use of color graphics for interactive molecular modelling, p. 81.

Meyer, E. F. (1980). Interactive graphics in medicinal chemistry. In E. J. Ariens (Ed.), *Drug Design*, Vol. IX (pp. 267–289). New York: Academic Press.

Meyer, E. F. (1992). Internal water molecules and H-bonding in biological macromolecules: a review of structural features with functional implications. *Protein Sci 1*, 1543–1562.

Meyer, E. F. (1995). Emil Fischer: then and now. *Pharm Acta Helv 69*, 177–183.

Meyer, E. F. (1997). The first years of the Protein Data Bank. *Protein Sci 6*, 1591–1597.

Meyer, E. F., & Funkhouser, N. F. (1998). A brief history of networking in the U.S. *J Chem Inf Comput Sci 38*, 951–955.

Meyer, E. F., Botos, I., Scapozza, L., & Zhang, D. (1995). Backwards binding and other structural surprises. *Perspect Drug Discov Des 3*, 168–195.

Morimoto, C. N., & Meyer, E. F. (1976). Information retrieval, computer graphics and remote computing. In F. R. Ahmed (Ed.), *Crystallographic Computing Techniques* (pp. 488–496). Copenhagen: Munksgaard.

Pauling, L. (1946). Modern structural chemistry. *Chem Eng News 24*, 1788.

Peat, T. S., Newman, J., Waldo, G. S., Berendzen, J., & Terwilliger, T. C. (1998). Structure of translation initiation factor 5a from *Pyrobaculum aerophilum* at 1.75 angstrom resolution. *Structure 6*, 1207–1214.

Picot, D., Loll, P. J., & Garavito, R. M. (1994). The X-ray crystal structure of the membrane protein prostaglandin H2 synthase-1. *Nature 367*, 243–249.

PostScript (1986). *PostScript Language Tutorial and Cookbook.* Reading: Addison-Wesley.

Wiener, M., Freymann, D., Ghosh, P., & Stroud, R. M. (1997). Crystal structure of colicin Ia. *Nature 385*, 461–464.

Wiener, N. (1964). *God & Golem, Inc.* Cambridge: The M.I.T. Press.

Zhang, D., Botos, I., Gomis-Rüth, F. X., Doll, R., Blood, C., Njoroge, F. G., Fox, J., Bode, W., & Meyer, E. F. (1994). Structural interaction of natural and synthetic inhibitors with the venom metalloproteinase, atrolysin C (Ht-d). *Proc Natl Acad Sci USA 91*, 8447–8451.