MOLECULAR MODELING

1. Introduction

It can be said that science is the art of building models to explain observations and predict new ones. Chemistry, as the central science, utilizes models in virtually every aspect of the discipline. From the first week of a first chemistry course, students use the scientific method to develop models that explain the behavior of the elements. Anyone who studies or uses chemistry has, in fact, practiced some form of molecular modeling.

A useful method of tracing the origins of molecular modeling is to examine how its offshoot, computer-assisted molecular modeling (CAMM), came to be developed. Molecular modeling today represents a convergence of a number of techniques from different disciplines. Fundamental tools used to accomplish modeling objectives necessarily draw on these. Specific software systems today

assist researchers in the study of molecular systems and provide mechanisms for deriving a rigorous and consistent explanation for the chemical or biological behavior observed or help the researcher to develop a model for predictions.

The literature of molecular modeling continues to expand rapidly. A testimony to this can be seen when examining the "125 Most Cited *JACS* Publications" list on the American Chemical Society's web site. (*JACS* 125th Anniversary: Home Page - http://pubs.acs.org/jacs125th/articles.html.) Five of the top 10 most cited articles involve molecular modeling or computational chemistry. The reasons for this expansion are the following: the tools are becoming easier to use, computers are getting faster, and researchers' awareness of the utility of these tools is growing. Despite this increasing ease of use, however, the development of a good grasp of the origins of the methods and strategies of molecular modeling is worthwhile, as these techniques are more than "black box" computations; and researchers should draw on the many citations from the literature that are included herein to guide them.

2. Historical Perspective

Molecular modeling has evolved as a synthesis of techniques from a number of disciplines: organic chemistry, medicinal chemistry, physical chemistry, chemical physics, computer science, mathematics, and statistics. With the development of quantum mechanics (1,2) in the early 1900s, the laws of physics necessary to relate molecular electronic structure to observable properties were defined. In a confluence of related developments, engineering and the national defense both played roles in the development of computing machinery itself in the United States (3). This evolution had a direct impact on computing in chemistry, as the newly developed devices could be applied to problems in chemistry, permitting solutions to problems previously considered intractable.

Into the late 1940s, Nobel Laureate Robert S. Mulliken, a physical chemist at the University of Chicago, maintained a skeptical view regarding the future of applying the theories of physics to solving practical problems in chemistry (4,5). Subsequently, Mulliken (5) related that "... [it was] only in the '50's that really substantial progress was made ... A major and indeed crucial step beyond the development of formulas for molecular integrals was the programming for large electronic digital computers of otherwise excessively time-consuming numerical computation of these integrals, and of their combination to obtain desired molecular wave functions and related molecular properties."

Whereas many scientists shared Milliken's initial skepticism regarding the practical role of theory in solving problems in chemistry and physics, the work of London (6) on dispersion forces in 1930 and Hückel's π-electron theory in 1931 (7) continued to attract the interest of many, including a young scientist named Frank Westheimer who, drawing on the physics of internal motions as detailed by Pitzer and co-workers (8), first applied the basic concepts of what is now called molecular mechanics to compute the rates of the racemization of *o*-dibromobiphenyls. The 1946 publication (9) of these results would lay the foundation for Westheimer's own systematic conformational analysis studies (10) as well as for many others, eg, Hendrickson's (11) and Allinger and co-workers (12).

These scientists would utilize basic Newtonian mechanics coupled with concepts from spectroscopy (13,14) to develop nonquantum mechanical models of structures, energies, and reactivity.

Researchers in chemistry and chemical physics whose interests were more theoretical were also very active between 1940 and 1965. The Manhattan Project had focused a great deal of public attention on chemistry and physics, garnering as well the energy and interest of both young and accomplished scientists. Coulson's seminal contribution (15) on molecular orbitals, along with studies of the low lying excited states of benzene by Goeppert-Mayer and Sklar (16), laid the groundwork for the subsequent contributions of Dewar (17), Pople (18), and Pariser and Parr (19), whose bodies of work have provided the basis for innumerable citations detailing the application of quantum mechanical methods to organic chemistry since the 1950s, a subject thoroughly treated in several exemplary texts (20–22). Such efforts (17–19) paved the way for Nobel Laureates Woodward and Hoffman (23) to elaborate elegant, orbital-based theories of the relationship between reactivity and molecular electronic structure.

3. Definition of Molecular Modeling and Uses of CAMM

Molecular modeling refers broadly to any study of molecules utilizing physical or theoretical models to explain an observed or predicted behavior. In practice, physical models have expedited the understanding of small molecules, inorganic complexes, proteins, and biopolymers, including such molecules as deoxyribonucleic acid (DNA). The pioneering work of Watson and Crick on DNA would certainly not have been possible without the building of actual physical models of the structure derived from experimental observations of X-ray diffraction patterns. However, physical models have limitations, both in terms of being primarily static representations of dynamic systems, and in being only semiquantitative with respect to scale. Whereas chemists prefer having hands-on experience with actual physical models, their parallel quest to quantify structures and energetics makes it impractical to rely only on such models for all aspects of their work. It is here that computer models, or molecular modeling, enter in.

Molecular modeling may also be defined as the application of computational techniques, grounded in theory, to predict or explain observable biological or physical chemical properties. Whenever molecular modeling is practiced using a computer, the technique then becomes computer-assisted (aided) molecular modeling, or CAMM. This technique is often used synonymously with CAMD, or computer-assisted molecular (materials) design/discovery. CADD refers to computer-assisted drug design–discovery. A computational technique as used herein is a mathematical model derived from principles of chemistry, physics, or statistics, which facilitates molecular modeling. An entire branch of chemistry, ie, computational chemistry, is devoted to developing, benchmarking, and applying computational techniques in order that researchers may be able to better understand and predict properties. Computational chemistry serves as an umbrella under which several disciplines converge to promote the evolution

and integration of better technologies to enhance the understanding of molecules and their reactivities. Some of the properties that may be calculated either exactly or approximately by computational methods include the following:

boiling points
melting points
crystallization energy
heat capacity
heat of formation

heat of fusion
heat of sublimation
heat of vaporization
entropy
molar refractivity
molar volume
partition coefficients

radius of gyration
elasticity
tensile strength
crystal and polymorphic forms
and relative stabilities
free energies of solvation/desolvation
accessible surface area, including
solvent accessible surfaces (SASA)

dipole and inertial moments
quadrupole moments
octupole moments
infrared (ir)-spectra/intensities
nuclear magnetic resonance
(nmr) spectra/chemical shifts
optical rotary dispersion (ord)
Raman spectra
ultraviolet (uv) spectra

ionization potentials
electron affinities
protonation energies and pK_a
ionic strength

conformational energies
Boltzmann distributions

Molecular properties can be classified according to their end-point observables, such as chemical (reactivity, solubility, acid–base), physical (a function of physical state: gas, liquid, solid; thermodynamic), or biological (ligand or enzyme; agonist or antagonist). These properties reflect macroscopic, or bulk, properties, which exist only for the bulk material. Examples are heat of crystallization or microscopic properties, which exist for an ensemble of the molecule. As use of CAMM methods expands to address a broader horizon of applications beyond those in organic, medicinal, and biological chemistry, calculations on metals, semiconductors, and magnetic systems are becoming more common (24).

To gain a proper perspective of the role of computed physical properties, the relationship between estimated and computed properties needs to be understood. Horvath (25) formulated the following definitions of estimating or computing properties.

Interpolating properties. A correlation is found between the desired property and another property or characteristic of related molecules; in this case the desired property may be computed from within the range of application of the correlation, and the interpolated property should be accurately estimated.

Extrapolating properties. In this case, the correlation does not extend to include the molecule of interest, but by extending the correlation, it is possible to estimate the desired property. Since the validity of the correlation in the extrapolated region is unknown, the accuracy of the extrapolated property is difficult to estimate.

Computing properties. In many cases it is possible to compute a property, directly or indirectly, with varying levels of accuracy. Such computed properties can be quite comparable to experimental accuracy and, indeed, may substitute

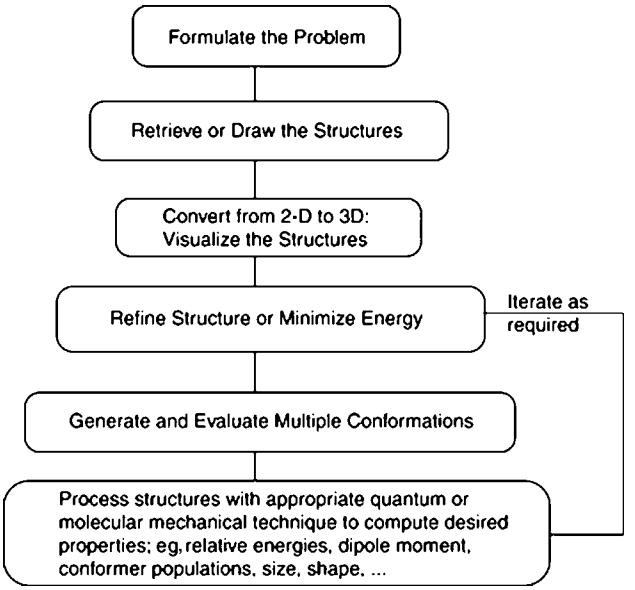

Fig. 1. Flowchart for typical small molecule modeling project.

for the experiment in cases where the experiment would be difficult or impossible to perform.

This last definition should be carefully applied as either an interpolation or an extrapolation, particularly for empirical computational methods based on diverse observations. *It is critical that users of molecular modeling tools understand where it is appropriate to apply a technique and where it is not, and what degree of accuracy can be expected.*

The specific process involved in a given molecular modeling study depends significantly on the nature of the primary objective of the task. However, for many small-molecule and even macromolecular studies, a number of authors have diagrammed the individual steps in the process, and the flowchart in Figure 1 provides a generalized heuristic of their efforts. An initial step that is critical to any CAMM project is the generation or retrieval of the pertinent structures themselves, a subject that is treated in detail herein. Structures may range from simple organic molecules, membrane-bound biopolymers, or monomers of more complex polymers, to full proteins, enzymes, metal surfaces, or zeolites. The modeling process can be influenced by the initial structure and its geometry. Thus, the selection and development of the starting molecular geometry needs to be given particular attention. An important component of the quality control process, as well as in gaining an understanding of the molecules themselves, is the visual examination of structures involved in a modeling study.

3.1. Computer Graphics in Molecular Modeling. The goal of molecular modeling is to define clearly the relationship between chemical constitution, ie, the molecular formula, or a topographic representation thereof, its geometric constitution or three-dimensional (3D) topology (the disposition of its atoms in Cartesian space), and its observed (or predicted) properties. The representation and facile manipulation of 3D arrays of atoms comprise the domain of molecular

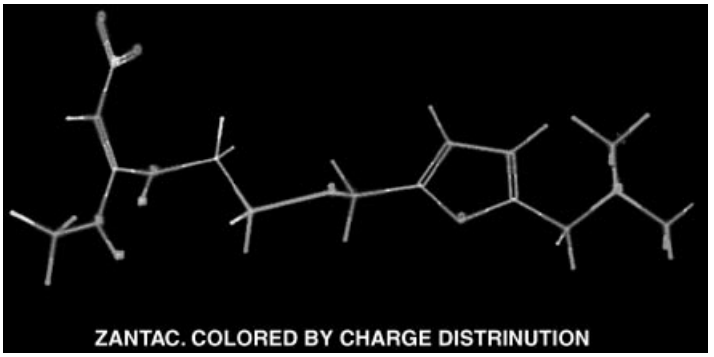

Fig. 2. A stick figure representation of the Zantac™ (ranitidine) molecule, color coded according to atomic charges fit to the electrostatic potential. Red = most negative, Purple/Blue = most positive. From the Molecular Modeling Program, MacroModel V. 3.0 (VAX), W. Clark Still, Columbia University. Photographed from an Evans and Sutherland Color Picture System model PS390. See online version for color.

or, computer graphics. From the time of Levinthal's work in the mid-1960s (26), scientists have endeavored to develop and use graphics software and hardware to expedite molecular modeling studies in both the classroom and the research lab. The growth of graphics tools has paralleled the evolution of computing hardware and, indeed, in some circumstances has even given strong impetus to that evolution (27). An example is the development and use of the Evans and Sutherland computer graphics systems in molecular modeling (27). Computing software and hardware systems have had a profound impact on the ability of modelers to compose a modeling study and address all aspects of the work, ranging from generating two-dimensional (2D) drawings of structures to statistical quality control of computed properties via visualization of multidimensional data.

Illustrating this enhancement in the visualization of structure and properties, Figures 2–5 provide increasingly complex and useful structural

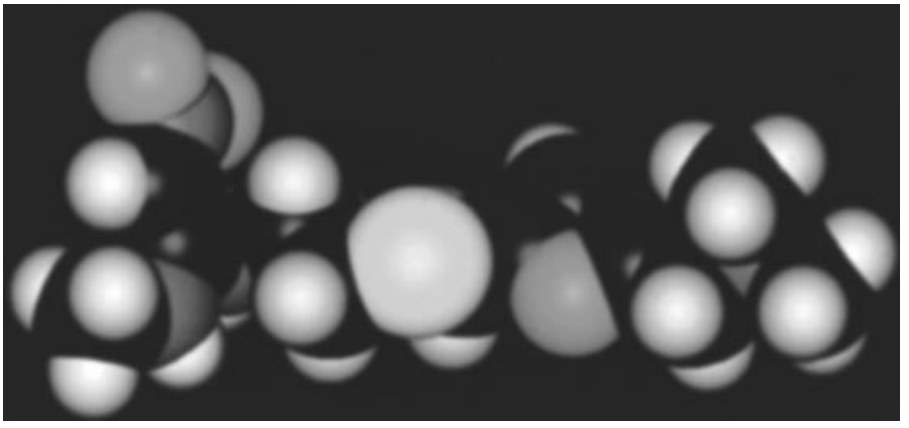

Fig. 3. A CPK rendering of the Zantac™ (ranitidine) molecule. The cream color atoms are hydrogens, black are carbons, yellow is sulfur, blue is nitrogen, white is hydrogen attached to nitrogen, and red atoms are oxygens. Rendered using MacroModel on a E&S PS390. See online version for color.

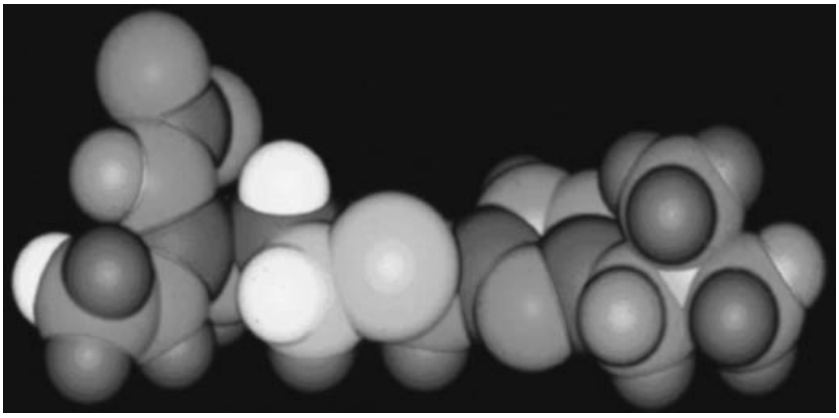

Fig. 4. A CPK rendering of the Zantac™ (ranitidine) molecule, using the same color scheme as in Figure 2. The color ranges from most negative (red) to most positive (purple/blue) following the ROY(W)GBV protocol. White atoms are considered electrically neutral (charges from −0.04e to +0.04e). See online version for color.

representations. Figure 2, a stick drawing of the drug Zantac™ (ranitidine), when viewed in color (not shown here), conveys both 2D structure data and color-coded electronic charge data. Figure 3 shows the CPK (Corey–Pauling–Kolton) shaded solid surface for ranitidine. In Figure 4, atomic color-coded electronic charge data (color not shown here) have been mapped onto the CPK surfaces. In Figure 5,

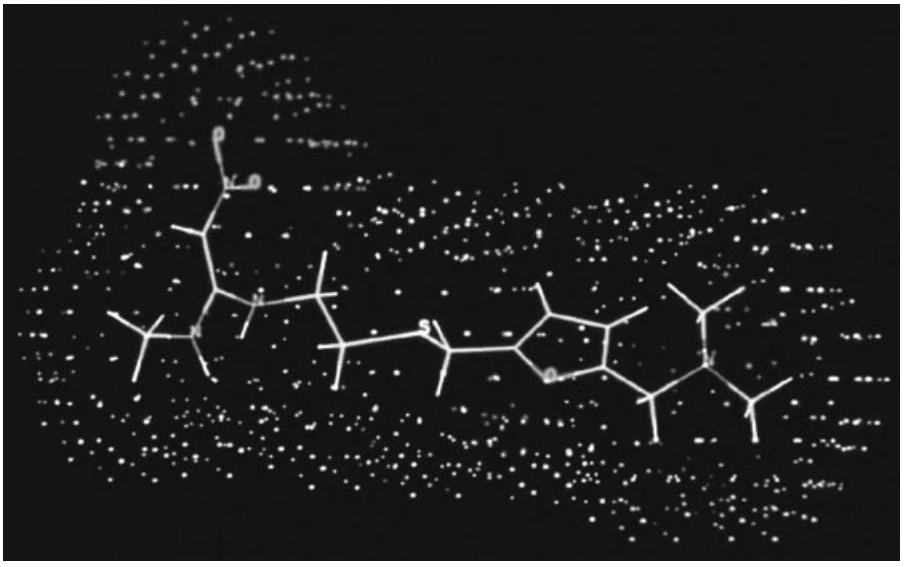

Fig. 5. A atom-type color coded stick figure of Zantac™ (ranitidine) within a four-layer concentric Connolly solvent-accessible dot surface with each dot color coded according to the value of the electrostatic potential as evaluated by the Gaussian80(UCSF) program at the 3-21G* basis set level. Dot color coding as in Figure 4. See online version for color.

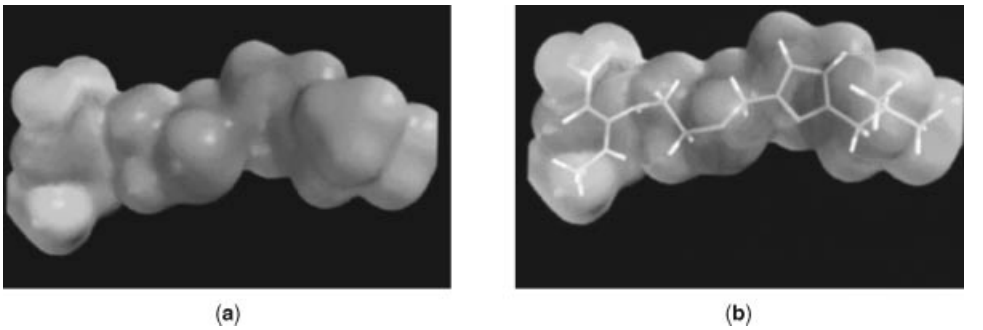

Fig. 6. (**a**). The Zantac^{TM} (ranitidine) molecule rendered at constant value density surface, and color coded according to the electrostatic potential, as displayed using the SPARTAN program (Macintosh version 1.0.1), at the AM1 level. Colors again range from (−) red to (+) blue displaying the charge density. (**b**). Here the same representation as in Figure 6**a** is rendered on a transparent surface, permitting a view of the atom-type color coded tube rendering shown within, again from SPARTAN. See online version for color.

four layers of the Connolly solvent-accessible dot surface, when color-coded (color not shown here), correspond to the energies of the electrostatic potential. In this figure, the highest charge density, when viewed in color, would be indicated by red dots in areas where there is the strongest attraction to an H^+ atom brought to that point. Conversely, at the points that would be shown in purple, the repulsive interaction would be the greatest, signifying regions of maximal positive charge. Figure 6 shows the electrostatic potential of ranitidine mapped, respectively, onto a solid and translucent constant-density surface (0.001e). Figure 6 is actually five-dimensional (5D) when viewed in color (x, y, and z dimensions, augmented by the color and gradients of color not shown here). The added impact of color ordinarily shown in such images considerably augments what the reader is able to view in black-and-white figures. An excellent overview of computer graphics by Hubbard (28) provides details regarding the mechanics of hardware and software systems integral to exemplary methods of computer graphics. Sources of papers on methods and applications in computer graphics and molecular modeling can be found in the *Journal of Molecular Graphics and Modeling* (Elsevier) and the *Journal of Chemical Information and Computer Science* (American Chemical Society).

One facet of chemistry that has benefited greatly from the use of computer graphics to enhance its own development is X-ray crystallography; the use of computers in X-ray structure refinement has also been described (29). Reviews (30) have appeared that cover the development and range of computing methods as adjuncts to crystallographic studies. More recently, details of how computer graphics can provide visual introductions to basic diffraction concepts, including the reciprocal lattice, the Ewald sphere construction, lattices, space group determination, and many other topics have been published (31). One of the most memorable advances in graphics in modeling was showcased on the cover of *Science* magazine in 1981. Researchers (32) in pharmaceutical and physical chemistry as well as computer and information science

at the University of California at San Francisco's Laboratory for Computer Graphics worked together to produce a highly symmetrical graphical representation of the B-DNA molecule viewed along its helix axis. The highly symmetric view produces the effect of a stained glass window. Although this effect was impressive in both a scientific and an artistic sense, it was the work of Connolly (33) representing the solvent accessible molecular surface as a series of dots following Richards (34) who pioneered the surface accessibility concept in 1971, which provided the real advance in molecular graphics. Subsequently, a number of computed properties such as the molecular electrostatic potential (Fig. 5) and the relative hydrophobicity have been mapped onto the Connolly surface, which has become an indispensable part of molecular modeling in the drug industry.

3.2. Computational Methods for Molecular Modeling. *Classical and Quantum Mechanics.* At the beginning of the twentieth century, a revolution was brewing in the world of physics. For hundreds of years, the Newtonian laws of mechanics had satisfactorily provided explanations and supported experimental observations in the physical sciences. However, the experimentalists of the nineteenth century had begun delving into the world of matter at an atomic level. This led to unsatisfactory explanations of the observed patterns of behavior of electricity, light, and matter, and it was these inconsistencies that led Bohr, Compton, deBroglie, Einstein, Planck, and Schrödinger to seek a new order, another level of theory, ie, quantum theory.

Basically, Newtonian mechanics worked well for problems involving terrestrial and even celestial bodies, providing rational and quantifiable relationships between mass, velocity, acceleration, and force. However, in the realm of optics and electricity, numerous observations seemed to defy Newtonian laws. Phenomena such as diffraction and interference could only be explained if light had both particle and wave properties. Indeed, particles such as electrons and X-rays appeared to have both discrete energy states and momentum, properties similar to those of light. None of the classical, or Newtonian, laws could account for such behavior, and such inadequacies led scientists to search for new concepts in the consideration of the nature of reality.

In 1903, when Plank suggested that the energy emitted from heated bodies, ie, black-body radiation, was not composed of waves, but rather discrete particles or quanta, a long-standing physical anomaly was resolved. Similarly, Planck's theory was applied to the photoelectric effect and was subsequently used by Bohr (35) to develop models of atomic structure. By the time Pauling and Wilson published their treatise on quantum mechanics in 1935 (36), the foundations for a workable quantum theory, explaining black-body radiation, electron distributions around nuclei and in chemical bonds, and the wave-particle duality of photons and electrons, had been detailed by a new generation of physicists (37). The extraordinary progress in the theory of matter made during the first three decades of the twentieth century lead Dirac, one of the pioneers of quantum theory (38), to state, "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known" (39).

At the heart of the revolution in quantum theory is Schrödinger's equation, which, in one dimension, for one electron not interacting with its surroundings,

may be written

$$d^2\Psi/dx^2 + 8\pi^2 m/h^2(E - V)\Psi = 0 \quad (1)$$

in which E is the total energy, a constant, and V is the potential energy, which most often is a function of x. The wave function, Ψ provides the solution of this equation and is at the heart of quantum mechanics and its applications to problems in chemistry. After expanding the wave equation to three dimensions, and replacing the second derivatives in equation 1 with the Laplacian operator, V^2, equation 1 can be rearranged to give

$$H = T + V \quad (2)$$

wherein the energy terms have been grouped on the right, and the Hamiltonian operator, H, is used in classical mechanics to provide a function of momenta and coordinates. For a single electron, the Schrödinger equation is

$$H\Psi = E\Psi \quad (3)$$

in which the H is an operator, the constant E is called an eigenvalue, and the function Ψ is called the eigenfunction. Certain constraints must be met if Ψ is to be physically meaningful, ie, that it be continuous and single-valued over the region of interest. The probability of finding the particle in all space must be unity, ie,

$$\int |\Psi|^2 dx\, dy\, dz = 1 \quad (4)$$

The quantity $|\Psi|^2$ represents the probability of finding the electron in the described region. It is also interpreted statistically, in the context of the Heisenberg Uncertainty Principle, as an expectation value. Streitwieser eloquently summarized the importance of such concepts in chemical computations (40): "For organic chemists the importance of quantum mechanics lies not at all in the exact calculations from first principles (*ab initio* calculations), but rather in providing heuristic concepts and insights in establishing qualitative and quantitative semiempirical correlations of experimental data and, especially, in facilitating the application of what has long been the organic chemist's most important tool: reasoning by analogy."

Two important points are made in this statement: first, less important than exact quantification is the development of a heuristic model of chemical behavior; and second, semiempirical correlations are the goal of computations involving applied quantum mechanics. At the time Streitwieser wrote, in 1960–1961, full-scale calculations on systems large enough to be of interest to organic chemists were certainly beyond then-current limits for most levels of theory and the corresponding computer programs. Computing machinery was still quite primitive by today's standards ($\sim 10^8$ slower than a good workstation at the time of this writing, ie, mid-2004), and a full elaboration of semiempirical quantum mechanical methods was in progress (20). Of particular note is the fact that

Streitwieser here suggests that organic chemists are less interested in exact ab initio calculation of properties than in developing qualitative concepts of structure–reactivity relationships. As discussed herein, ab initio quantum mechanical calculations now have become the preeminent choice in the chemist's computational toolkit for molecular modeling of small to medium-sized systems (<200 atoms).

Clearly, most chemistry involves interactions of multielectron systems with each other to yield desired properties for either known chemicals or new chemical entities (NCEs, in pharmaceutical language). At the atomic level, chemical reactions involve structures that have multiple electrons interacting, perhaps even on multiple atomic centers such that the potential energy of these systems is dependent strongly on the relative positions of the electrons with respect to each other as well as with respect to the nuclei. The Born–Oppenheimer approximation is generally applied; this approximation states that the motion of the nuclei and the electrons is independent, and that electron motion is the dominant component of interatomic interactions. Hartree (41) and Fock (42) developed a formalism for reducing the multielectron Schrödinger equation to a sum of single-electron equations, which could be solved to yield what is called a self-consistent field (SCF) approximate method. As an iterative method wherein electrons are distributed into shells based on the Aufbau principle, the SCF method made it possible to develop wave functions that can be approximated by analytical solutions to provide representations of atomic and molecular orbitals that can be readily visualized. Such visualization facilitates the interpretation of chemical reactivity and chemical reactions.

Numerous methods arose utilizing Hartree–Fock SCF techniques, ranging from the simplest, or Hückel π-electron techniques, to the most complete first principles, or ab initio methods. What distinguishes these methods is, in practical terms, which electrons and orbitals are included in the calculations, along with the degree to which the elements of the Fock matrix (representing the operator H) are evaluated explicitly, approximated, or neglected completely. These matrix elements, when they are evaluated in the context of the linear combination of atomic orbitals (LCAO) method, ultimately provide a quantitation of the presence and magnitude of the influence of neighboring electrons on each other (this is the essence of the characteristics or properties that distinguish substances from each other). For systems in which all σ and π electrons are considered, so long as the researcher neglected a selected set of neighboring atom electron–electron interactions LCAO provided the basis for most so-called semiempirical quantum mechanical methods in use in the 1960s.

The Hückel molecular orbital theory (HMO) and its subsequent elaboration, extended HMO theory (EHT), methods provide the simplest quantum mechanical description of π-electron systems. The development and applications of HMO methods were reviewed by Streitwieser in his text (20). Whereas the HMO method neglects σ electrons, it did utilize LCAO for the molecular orbitals. The method worked quite well, particularly for planar conjugated systems and even for certain nonplanar molecules. However, the technique can be best credited as providing the basis for subsequent, more elaborate methods. In Murrell and Harget (22) the development and applications of the Pariser–Pople–Parr (PPP) (19) method is given in a very readable account. It was a

particularly useful technique in that adiabatic ionization potentials as well as singlet–triplet separation energies could be computed accurately. Additionally, the PPP method was used subsequently to compute the energetics of the π-electron component of structures (43,44), in conjunction with an adaptation of the Westheimer method to compute gas-phase conformational energies for planar and nonplanar π-electron systems.

The other significant step in the development of molecular orbital theories involved all-valence-electron methods, wherein the concept of zero differential overlap (ZDO) of two-center integrals involved in the wave functions are refined. In these techniques, which introduced drastically reduced numbers of integrals requiring evaluation, investigators found they could incrementally move toward a more complete set of SCF equations without their being computationally intractable. In 1965, the first group of a series of important papers detailing the complete neglect of differential overlap (CNDO) and neglect of diatomic differential overlap (NDDO) methods were published (45). Full computational details of these methods are also available (21,22). The CNDO and NDDO techniques enabled computation of a broad spectrum of geometric features, such as bond lengths, angles, and related properties, including dipole moments, which could be predicted for singly bonded systems for the first time. The methods continue to be used today (∼2004), although the relatively poor accuracy of the CNDO technique in determining structural and charge distribution properties has led to its being used principally in spectroscopy, ie, the ZINDO program (46).

Through the 1970s and 1980s investigators continued to refine the ZDO technique, especially in increasing the accuracy and range of computed quantities, and in taking advantage of the enhanced computing facilities that were becoming available. Advances made included development of a series of programs utilizing the intermediate neglect of differential overlap (INDO) technique (47). Despite the passing of M.J.S. Dewar, the "Dewar School" continues today to produce semiempirical techniques to permit the computation of properties of organic molecules, organometallics, semiconductors, peptides, and proteins (48,49). Excellent expositions on semiempirical methods derived from NDDO are available (49,50), including a "how to" text (51). However, as Dewar has stated,"MO (Molecular Orbital) theory is not a description of reality. It is only the embodiment of another molecular model, the MO model" (52).

The ultimate goal of quantum mechanical calculations as applied in molecular modeling is the a priori computation of properties of molecules with the highest possible accuracy (rivaling experiment), but utilizing the fewest approximations in the description of the wavefunction. Ab initio, or from first principles, calculations represent the current state of the art in this domain. They are also referred to as nonempirical calculations, although this name is somewhat misleading. Ab initio calculations utilize experimental data on atomic systems to facilitate the adjustment of parameters such as the exponents of the Gaussian functions used to describe orbitals within the formalism. Additionally, these Gaussians are a function of the electron–nucleus distance squared, (r^2), and represent a significant approximation to the Slater type or simple exponential functions of (r), used to describe electron distributions in elemental quantum mechanics. While Gaussian functions do a creditable job at reproducing experimental properties, they were introduced for pragmatic reasons, namely, to

simplify the computation of multicenter integrals. Excellent expositions detailing ab initio molecular orbital calculations, including the very important topic of basis sets, are available (51–55).

The performance of ab initio techniques distinguishes them significantly from their predecessors, semiempirical methods. Their consistent reproduction of data from structural, thermodynamic, and reaction sources to a range falling within the error limits of the experimental values provides scientists with an important tool with which to address various modeling problems. Whereas the absolute value of the relative performance of ab initio techniques varies for each structural or energetic feature examined (and the particular level of theory utilized) it is not unreasonable to suggest that if the quantity can be computed by both semiempirical and ab initio methods, the ab initio value will be closer to experiment or an ideal value than any other method. To get a fuller appreciation of the comparison of these two methods, one need only follow the numerous publications of the Pople group (55) detailing the performance of ab initio methods in comparison to these published by the Dewar (47) group, which continued the refinement of semiempirical techniques. A mathematically rigorous overview of the utility of ab initio calculations is available (56), and many elementary questions about the role of ab initio calculations in molecular modeling studies are addressed in a particularly readable account (57) on the basis of which several important points about both semiempirical and ab initio calculations are worth summarizing herein.

Ab Initio Theory: Caveats And Performance.

- Except for approximations made previously, no others are made.
- All electrons are treated.
- All electron integrals are computed exactly, but with no guarantee of accuracy of the prediction or agreement with experiment.
- Basis sets consist of a finite number of Gaussian functions, introducing an inherent limitation on accuracy.
- Computing time can be proportional to the fourth power of the number (n) of basis functions (at the Hartree–Fock level).
- Whereas correlation energies can be included, in practice it is even more time consuming to use them in the determination of molecular geometries (ie, n^6), and determining the correct basis set to use can be difficult.
- Geometric properties are quite sensitive to the basis set chosen, including the presence or absence of polarization functions (additional s- and p-type functions on H and d-type on heavy atoms).
- Geometric and energetic properties have been shown to be sensitive to the starting geometries, and to the algorithm used for geometry optimization.

The following conclusions apply to organic molecules of ~25 heavy atoms (~60 atoms total), assuming use of the basis set 3-21G*:

- Geometric properties can be reproduced to within 0.15 ± 0.15 nm (0.015 ± 0.015 Å) for bond lengths, $1–2^\circ$ for bond angles, and to $\pm 5^\circ$ for dihedral angles.

- Ionizations energies can be computed to $\sim \pm 0.2$ eV; rotational barriers to $\sim$0.5 kcal/mol; dipole moments to $\sim \pm 0.5$ D; barriers to inversion to $\sim \pm 2.5$ kcal/mol; ir frequencies can be computed with about a 15% error (usually too high); and protonation energies are accurate to $\sim$1 pK unit.
- Hydrogen bond geometries may be reproduced or predicted fairly well with reasonable, but sometimes underestimated, heavy atom–heavy atom distances; radial dependence of the hydrogen bond may be in error.

Semiempirical MO Theory: Caveats and Performance. The same basic theoretical assumptions are made as in ab initio theory.

- Only valence electrons are considered, and the influence of core shell electrons are accommodated by a nuclear screening factor.
- The total number of integrals computed depends greatly on the level of complexity of the method: time cost savings of two orders of magnitude can be realized over ab initio theory (n^2 vs. n^4).
- Some of the elements of the Fock matrix are set to constants that have been empirically adjusted to reproduce certain properties.
- Slater-type orbitals (STOs) are used to represent electron density around an atom.
- Whereas correlation energies can be included, in practice it is more time consuming to use them in the determination of molecular geometries (ie, n^3), and determining the correct approximate method to use can be very difficult.
- Geometric properties are quite sensitive to the method chosen, including the presence or absence of polarization functions (additional *s*- and *p*-type functions on H and *d*-type on heavy atoms).
- Geometric and energetic properties are also sensitive to the starting geometries, and to the technique/algorithm used for geometry optimization (eg, internal vs. Cartesian coordinates).

The following conclusions apply to organic molecules of $\sim$25 heavy atoms ($\sim$60 atoms total), assuming use of the MNDO or AM1 Hamiltonian:

- Geometric properties can be reproduced to within 0.3 ± 0.2 nm (0.03 ± 0.02 Å) for bond lengths, 2–4° for bond angles, and to $\pm 10°$ for dihedral angles.
- Ionization energies can be computed to $\sim \pm 0.3$ eV.
- Rotational barriers to $\sim$1–1.5 kcal/mol.
- Dipole moments to $\sim \pm 0.3$ D.
- Barriers to inversion to $\sim \pm 5$ kcal/mol.
- Protonation energies are accurate to $\sim$1–2 pK units.
- Hydrogen-bond geometries may not be reproduced or predicted well at all, with anomalous heavy atom–heavy atom distances, and no radial dependence of the hydrogen bond. Check that the method being used has been recalibrated for this use if needed.

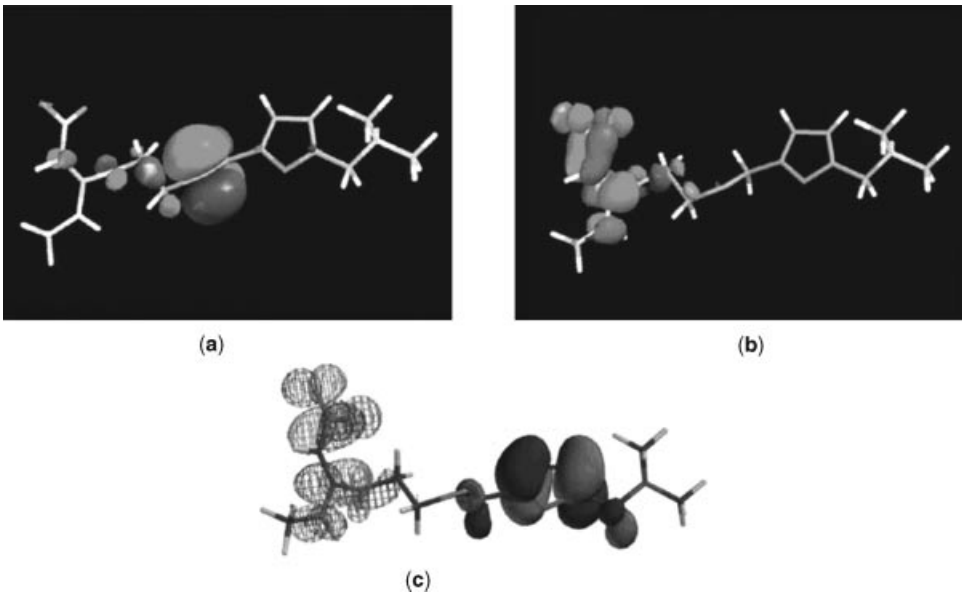

Fig. 7. (**a**). The Zantac™ (ranitidine) molecule showing the Highest Occupied Molecular Orbital or HOMO, as computed using the AM1 method in SPARTAN. Note the concentration of the HOMO on the sulfur atom and adjacent parts of the linker and the furan ring. (**b**). The Lowest Unoccupied Molecular Orbital, or LUMO, for Zantac™ (ranitidine). Note the localization of the LUMO to the nitroguanidine fragment of the molecule. (**c**). In contrast to the previous slide where the AM1 technique does a very good job representing the ESP, here, the 6-31G* HOMO (solid) appears to be significantly different relative to the AM1, indicating the sensitivity of this property to the method used. The 6-31G* LUMO is also shown (mesh). See online version for color.

Thermodynamic properties such as heats of reaction and formation can be computed more reliably by ab initio theory than by semiempirical MO methods (55). However, the literature of the method appropriate to the study should be carefully checked before a technique is selected. Finally, the role of computer graphics in evaluating quantum mechanical properties should not be overlooked. As seen in Figures 2–6, significant information can be conveyed with stick models or various surfaces with charge properties mapped onto them. Additionally, information about orbital energy levels, such as the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), which are indicators of reactivity in electrophilic and nucleophilic reactions, can be plotted readily. Figure 7**a**–**c** and Figure 8**a**–**c** show representations of the HOMO and LUMO, respectively, for the antiulcer drug Zantac™ and the natural product anticancer agent, Camptothecin. Note the sensitivity of the calculated electronic properties to the level of theory used (AM1 vs. STO-3G vs. 6-31G*). This dependence presents a practical dilemma to the practitioner, and care and consistency must be exercised when utilizing such data in developing structure–activity relationships. Further basis set accuracy details are summarized in Ref. (53) (Chapter 10), along with valuable references (p. 90).

Molecular Mechanics and Molecular Dynamics. Background. In the realm of quantum mechanics, researchers deal explicitly with electrons, with

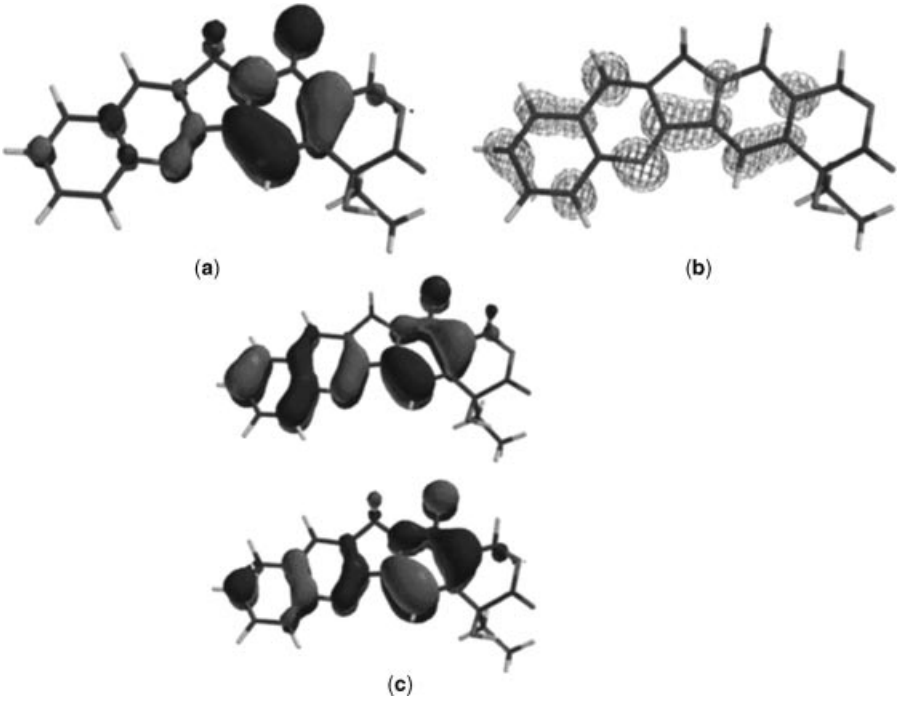

Fig. 8. (**a**). The HOMO for the anticancer agent, Camptothecin, from the STO-3G optimized geometry. (**b**). The LUMO for the molecule Camptothecin, from the STO-3G optimized geometry. (**c**). For comparison, the HOMO for Camptothecin are given for calculations at the 3-21G*//3-21G* and 6-31G*//3-21G* levels, respectively. See online version for color.

their interactions, and their attraction to nuclei, albeit in varying degrees depending on the rigor of the method chosen to solve the particular problem involved. These techniques were somewhat limited in their application prior to the age of large-scale computers and super workstations. In 1946, Westheimer (9) could not readily simulate processes like the conformational interconversion of substituted biphenyls by either quantum or classical mechanics. However, an explicit solution to classical mechanical equations was possible by hand. Quantum mechanics dominated the world of computational chemistry and molecular modeling as it existed in 1950 until ~1975. Quietly, in the mid-1960s researchers in physical, organic, and physical organic chemistry began making incremental advances in computational tools, expanding on Westheimer's technique to study larger, and perhaps, more interesting molecular systems, including proteins. Computers made possible the development of Westheimer's method, also called molecular mechanics, or empirical force field methods and, ultimately, the application of these methods to a full spectrum of studies in structure and energetics.

Molecular Mechanics. Molecular mechanics (MM), or empirical force-field methods (EFF), are so called because they are a model based on equations from Newtonian mechanics. This model initially assumed that atoms are "hard" spheres attached by networks of springs, with discrete force constants. In the past few decades, the more realistic assumption is used that they are deformable.

The force constants in the equations are adjusted empirically to reproduce experimental observations. The net result is a model that relates the "mechanical" forces within a structure to its properties. Force fields are made up of sets of equations each of which represents an element of the decomposition of the total energy of a system (not a quantum mechanical energy, but a classical mechanical one). The sum of the components is called the force field energy, or steric energy, which also routinely includes the electrostatic energy components. Typically, the steric energy is expressed as

$$T_{\text{Total}} = E_{\text{steric}} + E_{\text{electrostatic}} = E_{\text{bonds}} + E_{angles} + E_{vdW} + E_{\text{torsion}} + E_{\text{charge/dipole}}$$

where vdW is van der waals The overall form of each of these equations is fairly simple, ie, energy = a constant times the square of displacement. In most cases the focus is on differences in energy, because these are the quantities that help discriminate reactivity among similar structures. The computational requirement for molecular mechanics calculations grows as n^2, where n is the number of atoms, not the number of electrons or basis functions. Immediately, it can be seen that these calculations will be much faster than an equivalent quantum mechanical study. The size of the systems that can be studied can also substantially eclipse those studied by quantum mechanics.

In a force field calculation, a molecule in three dimensions is constructed using either Cartesian coordinates x, y, and z, or via an internal coordinate matrix consisting of bond distances, bond angles, and dihedral angles to specify the atoms' unique positions (a Z matrix). Then, the initial structure is evaluated to determine the extent to which each degree of freedom (bonds, angles, etc) deviates from the ideal (the zero-energy value) for the particular element and its hybridization. An energy minimization process follows wherein the energy associated with the distortions from ideal is minimized as the individual atomic positions or degrees of freedom are adjusted. Iteratively, this converges on a "minimum energy" or an "optimized" structure. This structure represents the best attempt of the minimization algorithm to render the smallest deviations in position of each of the atoms such that either the derivatives of the change in energy associated with the deviations are the smallest, or, they satisfy either energetic convergence or coordinate change criteria from iteration to iteration. Note that this process is analogous to the geometry optimization process within a quantum mechanical program, except that there the objective is to converge on a stationary point on the potential energy surface that corresponds to a structure which yields the smallest energy derivatives and lowest total energy from solution of the SCF equations. Most simple molecular mechanics force fields include terms (Fig. 8) for

bond stretching	$E_1 = k_1(1_0 - 1)$
bond angle distortion (bending)	$E\theta = k\theta\,(\theta_0 - \theta)$
dihedral angles (torsion)	$E_\tau = V_n/2\,(1 \pm \cos nw)$
van der Waals nonbonded interactions (1 . . . 4 interactions or greater)	$E_{\text{vdW}} = A/r_{kl}^{12} - C/r_{kl}^{6}$
Coulombic interactions (where q_{kl} can be dipoles or charges):	$E_{\text{Coul}} = q_k\, q_1/\epsilon\, r_{kl}$

In these equations, k_l, k_θ, V_n, C, and A represent the empirically adjusted constants associated with changes in bond lengths, angles, dihedrals, and nonbonded interactions. These terms plus 1 ••• 3 nonbonded interactions, and cross-terms such as stretch–bend are considered in the most complex computational models as well as in experimental spectroscopic force fields. Practically all of the early molecular mechanics force fields utilized force constants directly from vibrational spectroscopic studies (see background in Refs. 58 and 59). That is, for a particular interaction included in the force field, the force constant applied to an interaction was one that had been experimentally determined. Although this method can be utilized, it is very difficult to develop a generalized force field for a broad spectrum of molecules, because not all experimental force fields are derived to the same level of accuracy, nor are they consistent, ie, having all force constants derived concurrently. For this reason, Allinger (59) suggested "we must not look at a force field calculation and ask "what interactions are really occurring in the molecule"; rather the question must be "what interactions are really occurring in our model of the molecule." Of course, the hope is that the answers to the last question will, in fact, converge upon the answers to the first question as force fields improve."

Detailed accounts of representative force fields are also available (59,60).

Force Fields, Molecular Dynamics, and Vibrational Spectroscopy. The details of the relationship between molecular mechanics force fields and spectroscopic vibrational force fields has been discussed (59). Other fundamental papers on molecular mechanics are available as well (58,61–65). The link between molecular mechanics and molecular dynamics comes about through the force field itself. In molecular mechanics, the main interest here is in computing the energy of molecules in the gas phase, at room temperature, in a single, discrete configuration and conformation; time is *not* a variable in the equations. The goal is to know the structure of the molecule, bond lengths, angles, etc; conformational energy relative to some reference structure (eg, *gauche* vs. *anti*); the heat of formation; and, perhaps, approximate vibrational frequencies. From molecular dynamics, the objectives are properties that represent a time-averaged ensemble of states, including, eg, conformationally excited states, rotational states, interconversion rates, or inversion barriers for amines or amides. From this ensemble of states, characterizing the existence of the excited states and their contribution to the total energy of the system (from a Boltzmann distribution) is next. From an understanding of the vibrational spectroscopic roots (66) of molecular mechanical force fields used in dynamics simulations it can easily be appreciated that molecular dynamics may be seen as an extension bridge between theory and experiment, linking "static" molecular mechanical representations of properties with "dynamical" experimental properties.

The accuracy of molecular mechanics and that of molecular dynamics simulations share an inexorable dependence on the rigor of the force field used to elaborate the properties of interest. This aspect of molecular modeling can easily fill a volume by itself. The topic of force field development, or force field parameterization, although primarily a mathematical fitting process, represents a rigorous and highly subjective aspect of the discipline (67). Perspectives behind this high degree of rigor have been summarized (59,68). Briefly put, the different schools of

thought regarding the development of force fields arose principally from the initial objectives of the developers.

For example, in the late 1960s through the 1970s, the Allinger school targeted the computation of the structure and energetics of small organic and medicinal compounds (67,69,70). These efforts involved an incremental development of the force field, building up from hydrocarbons and adding new functional groups after certain performance criteria were met, eg, reproduction of experimental structures, conformational energies, rotational barriers, and heats of formation. Unlike the consistent force field approach of Lifson and co-workers (58,61–63,66), the early Allinger force fields treated a dozen or more functional groups simultaneously, and were not derived by an analytical least-squares fit to all the data (ie, not a purely mathematical treatment to find a least-squares solution, without manual parametric adjustments) (60). However, because the focus of Lifson was the analysis and prediction of the properties of hydrocarbons or peptides, it was not surprising that a consistent force field was possible. The number of variables to be optimized concurrently to permit calculation of all the structure elements, conformational energies, and vibrational spectra concurrently was, and still is, a massive quantity. However, the calculation for a limited number of functional groups could be accomplished, albeit slowly. If the goal is to reproduce and predict vibrational spectra, the full second derivative force constant matrix must be computed. Doing so in the early 1970s meant that a researcher was limited severely by the computing resources then available. By omitting or significantly reducing the number of the second derivatives computed, Allinger and co-workers were able to address a much larger number of problems in conformational analysis while producing force fields that were robust enough to be useful many years after their introduction; information regarding the utility of the very widely used and tested MM2 force field is available (71). Lifson argued that we should have a "consistent force field". Early force fields were certainly not "consistent", simply because they were too simple. The most simple force field is harmonic (squared terms only), and use only the diagonal terms of the force constant matrix. A force field that is going to be competitive with experiment will need both anharmonic terms and off-diagonal force constants. While Lifson's effort to fit vibrational frequencies well was useful, others chose to focus on structures and energies. Knowing the vibrational frequencies does not assure that you have a better force field. It only assures that you have a good representation of the shape of the potential surface near the energy minimum. Both the structure and energy depend on where the minimum is located on that potential surface, and in particular, where the various energy minima are located relative to one another, and their relative energies. Considering that there is a weak coupling between the fit to the stationary points on a surface, and the vibrational frequencies, it is more critical, most likely, to fit the stationary points and energies (the author acknowledges one of the reviewers for pointing out the concepts noted in this paragraph).

What is also not surprising is that many of the schools of force field development have converged significantly. Today (~2004), the most extensively developed force fields rely on ab initio quantum mechanical calculations for both structural and energetic data not available from experiment. Examples include the force fields from the Hagler group (68) (CFF9X/CVFF, Accelrys),

the Halgren MMFF from Merck (72), and Allinger's MM4 (73). The precedent for this approach was independently developed by Allinger and Profeta (67,70) and Hagler and Lifson (63). Unlike what was the case in 1970, today (~2004) the volume of experimental data is vastly dwarfed by the requirements placed on force fields for both molecular mechanics and molecular dynamics simulations. Researchers have endeavored to satisfy the requirements of both breadth of application and depth of rigor for each structural class (68,72,73). However, there remain families of force fields which can be classified as special-purpose force fields; ie, the developers chose to reproduce and predict the properties of specific structural classes, such as peptides and proteins (68,74,75), nucleic acids (74), or organometallic compounds (76). The utility of these force fields should not be underestimated, because they are usually developed by researchers who have significant experience in their respective fields and are designed to serve the special or idiosyncratic requirements of a subspecialty of a particular discipline. Because these techniques are specialized in their focus, care should be exercised in using such methods beyond their intended scope. As has been relevantly noted before, "Models are to be used, not believed" (77).

This leads to the questions as to how well a given force field performs and whether a force field is appropriate for the problem at hand. With regard to the former, a number of perspectives have been given (53,54,60,78,79). A series of publications has detailed how to generate highly accurate force fields, including corrections and extensions to an earlier force field (68,79,80). These studies provided a basis for additional force field development and rigorous application of the MM2 method to the prediction and understanding of chemical reactions (81). Quantitative assessments (82) provide the researcher with data sufficient to evaluate whether the methods in question are suitably accurate for the specific needs. Clearly, if only a reasonable 3D structure is needed, to give a spatial perspective, then highly accurate bond lengths and angles are not necessary. However, if a force field is being used to assist in the refinement of an experimental structure by NMR or X-ray methods, then it is desirable to use the force field that provides the most accurate assessment of structure and energies for the class(es) of structure under evaluation. Boyd (83), in several articles in the series he has co-edited, has discussed extensively the types of molecular mechanics programs available and included background information to help guide the novice and expert alike in the choice of the appropriate computational tools to solve individual problems. Applications of force-field techniques to problems in environmental chemistry, materials science, and molecular biology demand that the methods perform substantially beyond those for which reliable experimental data is available. Only since the time computers have made larger scale ab initio quantum mechanical calculations practical for appropriate model systems have these techniques been reliable for such a broad spectrum of important applications.

Molecular Dynamics and Monte Carlo Simulations. At the heart of the method of molecular dynamics is a simulation model consisting of potential energy functions, or force fields. Molecular dynamics calculations represent a deterministic method, ie, one based on the assumption that atoms move according to laws of Newtonian mechanics. Molecular dynamics simulations can be performed for short time periods, eg, 50–100 ps, to examine localized very high frequency motions, such as bond length distortions, or, over much longer periods

of time, eg, 1000 ps–100 ns, in order to derive equilibrium properties. It is worthwhile to summarize what properties researchers can expect to evaluate by performing molecular simulations:

Conformational states and energetics.
Kinetic properties: rates of reaction and interconversion.
Reaction pathways.
Solubilities.
Diffusion rates (including membrane-based processes).
Binding and complexation data.
Folding processes.
Transition temperatures.
Free energies for point mutations.
Free energies of binding.
As an adjunct to X-ray and NMR for structure refinements.

As noted, force fields are a set of equations relating the total energy of the system to its individual interaction components.

Popular force fields for simulations on organic and biomolecules include the following:

After selection of a force field simulation program which is appropriate to a given problem, the general procedure is as follows:

Program	Principal author(s)
AMBER	Kollman/Case
CHARMM/CHARMm	Karplus/Brooks/Accelrys
CFF9X	Hagler/Accelrys
GROMOS	van Gunsteren
MM2/MM3/MM4	Allinger
BOSS (Monte Carlo)	Jorgensen/Tirado-Rives

1. Initial structure equilibration, wherein bad or close contacts are relieved; this may be done with constraints on bonds, eg, to simplify the process (the premise of the SHAKE technique).
2. Structure refinement (locating energy minima) by energy minimization: this step may take a few hundred picoseconds of simulation time if the original configuration is far from a minimum. Note that the researcher may converge to a local minimum which is significantly higher in energy than the global minimum.
3. Techniques used to find global and local energy minima include: sequential simplex, steepest descents, conjugate gradient and variants (BFGS), and the Newton and modified Newton methods (Newton–Raphson).
4. Set up and solve Newton's equation, equation 5, for each atom in the system:

$$F = m_i a_i(t) = m_1 \partial^2 r_1(t)/\partial t^2 \tag{5}$$

where F = force on atom i at time t a = acceleration of atom i at time t, and r = position of atom i at time t

5. Evaluate the force F as the negative gradient of the potential energy function:

$$F_{\mathrm{i}} = -\partial\nu(r_1,\ r_2, \cdots,\ rN)/\partial r_i \tag{6}$$

6. Compute normal modes. These represent primarily harmonic motions internal to the molecule. There are $3N$-6 displacement eigenvectors, where N is the number of degrees of freedom of the system. The associated eigenvalues are the frequencies.

A variety of techniques have been detailed for handling Newton's equations of motion, equation 5 (84–88). Integration techniques yield atomic positional velocities that are used graphically to display the internal motions and paths of motion computed during the simulation runs. Understanding the behavior of a molecular system as a function of time is a critical element in any simulation, and velocities can be collected into a series of snapshots called a trajectory (atomic coordinate sets). From trajectories, researchers can determine the level of cooperation of motions in the folding or conformational processes that a polymer or biopolymer chain might undergo, eg, the interconversion of forms of DNA prior to and during complexation with a protein.

Although such simulations represent a full-scale challenge to the simulation technique owing to the wide variety of atom types and associated parameters, a different, but no less rigorous type, of challenge exists in the simulations of simple polymers of polyethylene. For example, in studies on polyethylene chains of size C_{10} to C_{100} (89), dynamics were used to elucidate the cooperation of motions at neighboring rotational sites, with a finding that for very short time periods the torsional motions of the chains are effectively Brownian in behavior, as had been found earlier. It should be noted that the operational methodology for applying molecular dynamics does vary from application to application, but it is probably reasonable to assume that simulations involving most polymeric systems other than peptides can be addressed similarly, with appropriate modifications for atomic species and periodicity (90). Even peptide systems are addressable by similar methods, despite the breadth of the literature on it that is available suggesting otherwise. The proliferation of dynamics programs for handling peptides and proteins may be the result of differing perspectives on the method of sampling of excited states, energy refinement algorithms, implementation of constraints, or restraints during the simulation, or force field. The situation is similar to understanding that a car is a vehicle for transporting people and cargo from point A to point B. Whereas the relative comfort and efficiency of the process may vary with the specific vehicle manufacturer, the fact that cars transport people and cargo is common to all manufacturers. All of the subsequent levels of structure and function in peptides and proteins depend intrinsically on the primary structure of the system.

Unlike simulations on homopolymers or rigid systems with regular periodic sets of atoms at essentially fixed distances, as in the case of zeolites, the complexity of the simulation problem increases as a function of the variety of the types

of structural motifs (helices, sheets, turns, barrels, coils) present, because each motif can behave both as individual atomic entities and as a single collective entity. Numerous investigators have detailed approaches to dealing with the analysis of these systems (68,86–88,91,92). Dynamical characterization of biopolymers and related systems remains one of the most challenging and stimulating aspects of molecular modeling.

In the context of molecular dynamics (MD), some additional pointers may help give a fuller perspective on the MD simulation operational process:

1. Initializing the initial kinetic energy and temperature of the system: It is necessary to start the motion at some level, eg, assume a Boltzmann (random) distribution of atomic velocities, at 300 K.
2. Time steps for the simulation need to be short enough to capture high frequency motions such as bond stretching, eg, 10^{-15} s.
3. Simulations need a substantial number of timesteps to sample configurational states such that desired properties are represented well enough to either confirm the experiment or establish valid prediction. Initial equilibrations take 50–100 ps. For systems of >100 atoms, run for 2000–10,000 ps if computing resources permit.
4. Langevin Dynamics: A technique to reduce the total number of equations of motion that are solved. Utilize the coupled heat bath, wherein the method models the solvent effect by incorporating a friction constant into the overall expression for the force.
5. The SHAKE method for bond constraints reduces the number of degrees of freedom during the initial stages of simulations; it is good for minimizing solvent bath overhead.
6. Simulated annealing is a technique that heats up the system and slowly cools it, perhaps to multiple different minima (conformational searching).
7. Use nonbonded (NB) truncation methods to reduce size of NB pairlist; it is a dominating term in the calculation! It is important to remember the pairlist increases as N^2, consider truncation of NBs at 100–120 nm (10–12 Å), and to experiment with electrostatic cutoffs independently of van der Waals.
8. Update the NB list every 25–50 time steps.
9. Utilize periodic boundary conditions, which permit reduction of the number of nonbonded interactions at greater distances by involving only the "nearest-neighbor" atoms from copies of the system that are in different but adjacent cells.
10. Use a solvent or water bath to capture the influence of solvent on the solute. Use the fewest shells of water–solvent possible, but no fewer than two, if compute resources are limited. Use a formal "box" of water, if possible, to reduce the influence of edge effects.

Monte Carlo (MC) techniques for molecular simulations have a long and rich history, and have been used to a great extent in studying the chemical physics of polymers. The majority of molecular modeling studies today do not involve the

use of MC methods; however, the sampling capability provided by MC methods has gained some popularity among computational chemists as a result of various studies (93–95). Relevant concepts of MC are summarized herein.

Monte Carlo methods as applied to chemical problems owe their popularity to the work of Metropolis (96), who first utilized them on very early computers to evaluate properties of simple molecular systems. Typically, a set of configurations in a given thermodynamic ensemble is generated by the random sampling of configuration space. By configurations is meant sets of atomic coordinates corresponding to discrete geometries, including those with substantial distortions in bonding or other periodic structural behaviors. In MC calculations, no time relationship exists between successively calculated configurations. Unlike in a molecular dynamics simulation, there is no path or trajectory which the system follows as interconversions occur and states are sampled. Rather, in MC simulations random steps are taken, necessitating that a very large sampling be done to optimize the sampling of the desired configurations. Unfortunately, with the complexity of the potential energy surface of large polymers or proteins, it is not always possible to be assured that one has sampled sufficiently. This open-endedness may be the primary reason that molecular modelers have not embraced Monte Carlo techniques as freely as they have molecular dynamics (MD) simulations, despite the steep computing requirements of the latter.

For many systems, the ensemble that is used in an MC simulation refers to the canonical ensemble (N, V, T). This ensemble permits a rise and fall in the pressure of the system, P, because the temperature and volume are held constant. Thus, the probability, $\wp$x, that any system of N particles, in a volume V at temperature T is found in a configuration x is proportional to the Boltzmann weighted energy at that state, E_x, and it is given by$_x$

$$\wp_x = \exp(-E_x/kT)/(N!Z)$$

where Z is the configurational partition function

$$Z = \left[(N!)^{-1}\right] \int \exp(-E_x/kT)dx = \exp(-A/kt)$$

and A is the Helmholz free energy. The average value of a property is

$$\langle F \rangle = \int F_x P_x dx$$

where F_x is the property evaluated at configuration x. Another popular ensemble used in protein and DNA MC simulations (93) is the (N, P, T) ensemble.

One method used to enhance the efficiency of sampling is biased sampling (96). An algorithm utilizing biased sampling allows low energy configurations to be sampled more often, and is usually more efficient than random sampling at sampling those configurations that contribute more significantly to the true average of a system. For example, to simulate an MC run (90), an algorithm might involve the following steps: (*1*) calculate the energy of the current configuration, E_1; (*2*) assign a new configuration, with a new energy, E_2; (*3*) calculate the

weight, Ω, of $\exp(-[E_2 - E_1]/kT)$; (*4*) compare the result to a uniform, random value in the range of [0,1]; (*5*) if the weight Ω is greater than the random value, reset E_1 to E_2, and restart the cycle. The Markov chain that results from this process produces a probability of a given configuration that is proportional to the calculated weight.

The list that follows gives an outline of the properties of a Monte Carlo simulation used in the context of molecular modeling studies for sampling either multiple conformations of smaller, flexible structures or multiple local minima of larger macromolecules or polymers (ie, conformational searching):

Probabilistic, as opposed to deterministic molecular dynamics.

Essentially random atomic moves—uses the Metropolis method (96).

Calculate new energy and compare to previous configuration(s).

Keep or discard new configuration.

Develop a statistical ensemble of energetically accessible states.

Usually sample many millions of states; but not always easy to sample space well.

Use Boltzmann or biased-sampling techniques.

Umbrella sampling can give free-energy differences, but not absolute free energies.

Usually done in NPT: isothermal-isobaric ensembles, including a water box.

Useful for efficient sampling of conformational space in systems such as polymers (90) and peptides (91–93).

Relatively recently, another method for conformational sampling/searching, based on genetic algorithms (GA) has been explored by an increasing number of researchers. For useful papers introducing the reader to this technique see Ref. (97).

To conclude, the reader is directed to Refs. (98) and (99), which focus on the critical role of force fields in the area of general property prediction. These studies provide an excellent survey of the performance of force fields for structures, energetics, and thermodynamic properties, such as phase equilibria, for CFF, CHARMm, AMBER*, MM3*, MM3, CVFF, CFF91, and the U(niversal)FF.

Combined Quantum and Molecular Mechanical Simulations. A recently developed technique is one wherein a molecular dynamics simulation includes the treatment of some part of the system with a quantum mechanical technique. This approach, QM/MM, is similar to the coupled quantum and molecular mechanical methods introduced by Warshel and Karplus (45) and at the heart of the MMI, MMP2, and MM3 programs by Allinger (46,58). These latter programs use quantum mechanical methods to treat the π-systems of the structures in question separately from the sigma framework. The results are combined at the end to render a structure that is optimized and energy refined to satisfy both SCF and force-field energy convergence. Newer QM/MM approaches treat the bulk of the molecule via a molecular mechanical force field to give its dynamical behavior, whereas a selected portion of the structure(s) is treated quantum mechanically to yield information about interactions

between the selected segments. This technique is particularly appropriate when applied to systems such as ligands bound to biological receptors or molecular "guests" trapped by "hosts".

In the 1980s, the most comprehensive implementation of this method was the program Gaussian80 (UCSF) developed at the University of California at San Francisco (97–99), which combined the molecular mechanics and dynamics code AMBER (100) and the ab initio quantum mechanics program Gaussian (101). The program permitted researchers to generate structures of interest, then run initial steps of energy minimization to resolve close contacts and significantly distorted bond lengths and angles. The quantum mechanical atoms of the enzyme–substrate complex would be defined next. The dynamics simulation would be run over several hundred picoseconds, with the majority of the atoms dealt with using the MM component of the program; substructures including the quantum mechanical atoms are treated by a full-scale ab initio calculation at the STO-3G level, eg, at selected intervals. The process would be captured in a trajectory file, and the quantum mechanical data would permit researchers to investigate reaction mechanisms, proton shuttling, salt-bridge formation, and other intermolecular associations pertinent to the dynamical processes. The general method has been elaborated by a number of researchers, a published survey (102) provides further details and examples of the application of the method.

Treatment of QM/MM calculations would not be complete without the mention of the use of a newer quantum mechanical method called density functional techniques. Although density functional theory DFT is not new, recent implementations have made it significantly more popular than in the past. In particular, the potential of DFT as the QM component of combined QM/MM efforts has focused more attention on DFT methods; an introduction to DFT and its applications is available (103) which cites the primary literature in DFT, including a classic monograph (104). The comprehensive survey of QM/MM (102) also gives references pertinent to the application of the DFT method to QM/MM studies.

Briefly put, DFT methods depend on the use of a functional of the electron density rather than the wave function functional, the advantage being that the density is an observable entity whereas the wave function is not. Further simplifying the calculations in DFT is the fact that the electron density has three spatial coordinates, regardless of the number of electrons in the chemical system. This makes it possible to readily compute properties directly from the electron density for systems of hundreds of electrons, a feat not easily accomplished routinely within wave functional theory. Just as occurs in the case of ab initio Hartree–Fock theory, DFT utilizes basis sets, and many of the more popular ones, along with the variants of DFT functionals, are available in commercial QM programs such as the Gaussian series (101) or SPARTAN (105).

Structure Generation, QSAR, and CoMFA Modeling Methods. Rule-Based Structure Generation Techniques. Figure 1 summarized a paradigm for small molecule modeling. An important component of that process is the generation of the 3D coordinates of the molecules to be studied. Many modeling program suites have their own 2D drawing-to-3D coordinates conversion routines. However, it was not always so simple or routine to make the 2D-to-3D conversions. During the 1980s, significant progress toward simplifying this aspect of modeling was made, particularly in the area of knowledge-based methods. One

of the most rigorous, and by far the most successful, of the conversion programs is CONCORD (106), produced at the University of Texas at Austin. CONCORD is a rule-based algorithm that evaluates atomic connectivity tables and maps reasonable bond lengths, angles, and dihedrals to the structures based on connectivity information and other geometric features permissible for a given atom and its normal valence states. The tool became the de facto standard for the conversion of databases of hundreds to millions of 2D structure representations into 3D structures in the 1990s. Because the geometric parameters for structures were derived from AM1 (Austin Model 1, a successor to MINDO) semiempirical quantum mechanical optimized geometries for families of model functionalities, the method is quite extensible and reliable, although it does suffer from some of the same deficiencies that are characteristic of AM1 calculations (eg, underestimated dipole moments, incorrect conformations for hydroxyl groups and alkyoxyl groups). A thorough review of methods for generating 3D structures is available (107).

Distance Geometry. Another technique that utilizes both rule-based and computational geometry methods for generating 3D structures is called distance geometry. This technique, which owes its origins to Cayley (108) in 1841, has been popularized as a result of pioneering work (109,110) and by the DGEOM program (111) and applications (112,113) developed by researchers at the University of California at San Francisco. The technique is unique among structure or conformation generation methods in that it applies equally well to both small and large molecular ensembles. The distance geometry algorithm has been interfaced with a graphical display engine (113) and the method used to generate multiple conformations of flexible and complex ring systems to facilitate conformational analysis. The use of distance geometry (DG) in conformational searching has been eloquently summarized (114). There, the idea is to generate multiple, geometrically feasible conformations for energy refinement by MM or QM methods. The mix of random and systematic structure generation permits a fairly thorough search of conformational space. Others have also detailed the use of DG in molecular modeling, particularly in the context of macromolecular structure generation and docking (115), wherein the basics of the mathematical elements of the technique, how distance constraints are developed for determining the distance bounds that determine structural features, the metric matrix technique for developing the geometries, and additional techniques from computational geometry that are integral to DG are all given. Of particular interest are the summaries of the applications of DG to the generation of polymers and biomolecules (alone or in conjunction with NMR or X-ray) data, ligand-receptor docking studies, and pharmacophore modeling. Ghose and Crippen (116) have provided a summary of the spectrum of distance geometry applications and the utility of distance geometry in molecular modeling, a portion of which is quoted in a slightly adapted form in the following list:

1. Generate an approximate 3D structure of a ligand.
2. Generate the low energy conformations.
3. Evaluate the minimum and maximum distances between the ligand atoms in (the structure's) energetically allowed conformations.

4. [Using a computerized search method, develop] a hypothesis [of] the binding mode of the ligand at the receptor site.
5. [Evaluate the alternative ligand] binding modes due to rigid rotations.
6. Classify the site pockets into different types to differentiate [the level of] interaction [with ligands].

In addition, researchers can survey the surface and clefts of a macromolecule for potential receptor sites based on ligand distance ranges; "use the common distance range of the superimposed atoms" (116) to facilitate the development of a "pharmacophore" or critical contact assembly; besides the ligand binding modes due to rigid rotations (116), evaluate those due to conformational flexibility, ie, by using highly constrained ligands and only considering the lowest energy conformations of flexible ones, the scope of this process can be reduced; evaluate the binding data with a training set and develop a regression model to predict the binding of other known ligands; and finally, "if the predictive power is acceptable, use it to predict [or screen sets of] structure" (116) from databases or combinatorial libraries.

Quantitative Structure–Activity Relationships (QSAR). Quantitative Structure–Activity Relationships (QSAR) is the name given to a broad spectrum of modeling methods that attempt to relate the biological activities of molecules to specific structural features, and do so in a quantitative manner (see ENZYME INHIBITORS). The method has been extensively applied. The concepts involved in QSAR studies and a brief overview of the methodology and applications are given here.

Historically, QSAR has been applied primarily to drug molecules; however, more recently, Quantitative Structure–Property (Toxicology) Relationships [QSP (T) R] have been elaborated by a number of researchers (117). Regardless of the context of application, QSAR has its roots in the use of linear free energy relationship (LFER) methods of deriving an equation that correlates the observed effect and the structural features responsible for the activity. The work of Hansch in relating hydrophobicity to biological activity is exemplary as a pioneering application of LFER in QSAR (118). It is interesting to note, however, that many of the early investigations in QSAR involved analysis of the relationship of hydrophobicity to activity, the nature of which relationship is more often parabolic rather than linear (119,120). The QSARs are usually best derived from a series of compounds (typically differing only at one or two substitution points) for which the activities are well determined by a stable biological assay. A QSAR table can be established wherein the columns are assigned to activities (the *y*s), and the metrics or properties (the *x*s), which can be either observable properties such as log *P*, high performance liquid chromatography (HPLC) retention times; nmr chemical shifts; computed values such as shape and size descriptors; dipole moments; atomic charges; or conformational energies. Each row represents an individual compound or conformation. Statistical relationships can then be developed from this table by means of univariate or multivariate techniques such as linear or multiple linear regression (MLR), or partial least squares (PLS). If the statistical significance of the relationship is sufficiently high, then this relationship is robust enough to be used to predict

or assess the activity of untested compounds. Usually the known data is divided into two groups, a training set and a test set to establish the statistical model. As a rule of thumb, there should be between 3 and 10 times the number of observations (rows) as x-variables in order to derive a model that would have predictive power and be able to minimize chance correlations. Under these circumstances, there would be some constraints against pursuing QSARs if only a few observations (molecules for which activities or properties are known) are available. Often this is the case, yet investigators have proceeded with the development of the QSAR. A discussion of the use of partial least squares (PLS) as a potential means of resolution for this difficulty is treated herein in the context of the CoMFA paradigm. It should also be recognized that many more molecular conformations and property descriptors than ever before can now be computed. For example, in the QSAR and Diversity modules of the Cerius2 software from Accelrys, Inc. (121), there are in excess of 200 shape, size, and electrostatics descriptors that can be computed. This abundance is in sharp contrast to the relatively small numbers of descriptors available to early investigators, who worked diligently with classical Hammett σs and log P values to derive LFERs (122–125). A particularly lucid account of the Hansch approach to application of LFERs in QSAR, which is illustrated with numerous examples is given by the Hansch collaborator and specialist in QSAR, T. Fujita (126). An example of the classical linear equation is represented by equation 7:

$$\mathrm{PED}_{50} = 0.606(\pm 0.184)E_5^c + 1.518(\pm 0.265) \tag{7}$$

where $n = 16$, $r = 0.884$, and $s = 0.204$. This equation was derived for a series of fungicide compounds of the type N-substituted aminoacetonitriles, $\mathrm{RNHCH_2CN}$, wherein it was found that the potency ($\mathrm{ED_{50}}$ = effective dose to kill 50%) was dependent on the corrected steric parameter, $E_s{}^c$ (127). This variant of the Taft steric parameter, E_s, emphasizes the effect of branching in addition to steric bulk of the R group. Another example of a simple linear equation has been derived for the enzymatic hydrolysis of esters by the serine hydrolase, trypsin (128). For a series of esters, $\mathrm{X{-}Ph{-}OCOCH_2NHCOPh}$, hydrolysis yields equation 8, where X = 4-$\mathrm{SO_2NH_2}$, 4-$\mathrm{NH_2}$, 4-CN, 4-$\mathrm{NO_2}$, 4-$\mathrm{NHCONH_2}$, 4-$\mathrm{OCH_3}$, H, 4-$\mathrm{CH_3}$, 4-Cl:

$$\log(1/K_m) = 0.71(\pm 0.17)\sigma + 3.31(\pm 0.09) \tag{8}$$

where $n = 10$, $r = 0.961$, and $s = 0.100$. These examples are not atypical of the hundreds that can be found in the literature. More exemplary of the relationships involving the log P are given by equation 9:

$$\log(1/C) = a(\log P) + b(\log P)^2 + c \tag{9}$$

where C is the equipotent concentration or dose, and a, b, and c are the coefficients of the linear, quadratic, and constants terms, respectively. Often, the linear term is zero, and a purely parabolic relationship between activity and log P is observed (126). More complex multivariate linear, bilinear, and parabolic equations can be valid for QSARs. However, it is important to perform a critical evaluation of both the biological data and the statistical paradigms before

publishing such results. The treatment and testing of both the biological data and the QSAR model equations have been detailed (129–132). More recent studies on this subject can be found in the chapter by Pleiss and Unger (133), and in the journal, *Quantitative Structure–Activity Relationship*s (Wiley-VCH).

The range of application of QSAR is extensive. Additional aspects of QSAR that are important and have recently reached the production level of application in industrial QSAR studies (134) include molecular similarity searching, 2D fingerprinting, 3D substructure searching, molecular superpositioning, pharmacophore identification, pharmacophore searching, 3D databases, structural alignments, receptor–ligand binding energetics/modes, 3D QSAR, molecular shape analysis, 1D, 2D, and 3D descriptors, including shape/size, charge, and hydrophobic fields, de novo ligand design, and hypothetical active site lattices.

The entire domain of "new-lead" discovery has expanded considerably. This development has affected what have been traditionally divergent approaches, namely, QSAR and structure-based design, leading them to become integrated so as to provide a more powerful approach (133). Several recently published comprehensive volumes capture the state of the art and can be consulted to determine precedents relevant to any particular study (134–138).

3.3. Comparative Molecular Field Analysis (CoMFA). Another method for molecular modeling is one which was developed in the mid-1980s, but came into frequent and practical use in the 1990s owing to the rapid advances in workstation computing power and in techniques for aligning diverse sets of structures. This technique, developed by Cramer and co-workers (139) and frequently referred to as 3D-QSAR, is called comparative molecular field analysis or CoMFA (see ENZYME INHIBITORS). The CoMFA is a computational technique that attempts to mimic the interaction of a "ligand" with a "receptor" by means of a lattice of points (receptor) within which the molecule of interest (ligand) is placed, and the interactions between the molecule and the grid points are evaluated. The points of the lattice that are inclusive of the volume of the molecule of interest are discarded. The researcher must propose an alignment (superpositioning) of the structures of interest. This can be done by rigid or flexible fitting, in either case an additional QSAR descriptor, the distortion energy of superpositioning, can be added to the classical QSAR regression. All other points are assigned both steric and electrostatic properties, such as Csp^3 carbon van der Waals steric properties, and a +1 charge for electrostatics. The interaction energies for both property types are then evaluated for the ~2000 points that remain in the lattice. These data (the *x*s) are then processed by Wold's technique of partial least-squares (PLS) regression analysis (140) against the activity values (the *y*s) supplied by the researcher. The program produces a model that will both reproduce the training set of data values (bioactivity, or similar property) and have predictive power as well.

The PLS model is cross-validated by successively eliminating observations, rederiving the model, and predicting the eliminated observations. As new, diverse data are added to the training set, the predictive power of the model is enhanced. Additionally, the CoMFA model is visually displayed, indicating regions either where steric bulk is favored or unfavored, as well as where changes in the electrostatic field enhance or diminish activity. The power of CoMFA is made possible by the PLS technique, because without

PLS to reduce the dimensionality of the xs, it would not be possible to derive a believable regression model when presented with a data matrix that is 10–20 rows by 2000 columns wide. Studies on factor analysis and principal components regression (141), on chemometric tools (142), and on the performance of biased regression techniques (143) have helped in understanding the complexity of regression techniques, their pitfalls, and their importance of QSAR. Indeed, a set of recommendations and caveats regarding the use of CoMFA has been published (144), and as with any computational or modeling technique, its capabilities and limitations become better understood as the frequency and breadth of its use grows.

An excellent complement to CoMFA is seen in the RECEPTOR technique that permits the easy visualization of properties such as hydrogen bonding, lipophilicity, and electrostatic gradients. The method "shrink wraps" a solvent accessible surface around one or more active compounds, and allows one to dock candidate structures within the surface, and score the degree of complementarity to the surface. Examples of CoMFA, Figure 9, and RECEPTOR, Figure 10, representations are given below.

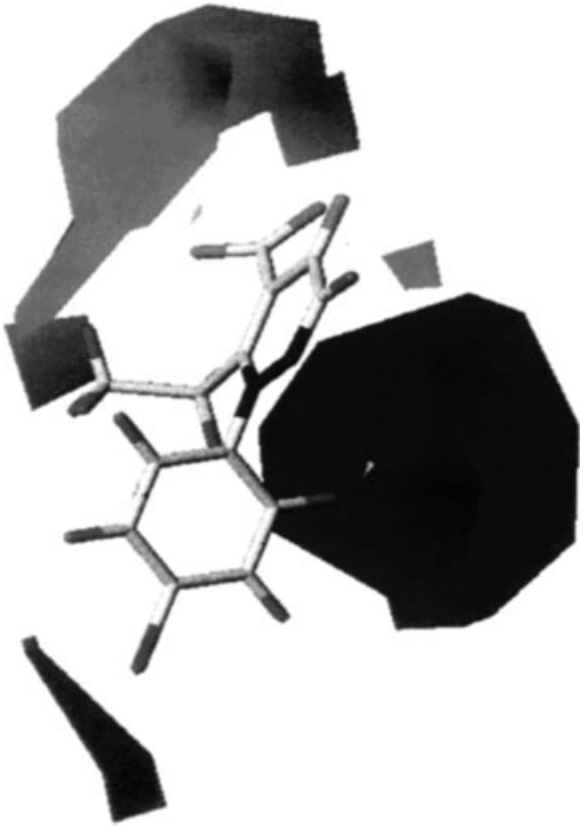

Fig. 9. An example of a CoMFA analysis plot. The red surface indicates areas of reduced tolerance to additional steric bulk, whereas the blue areas indicate additional steric bulk should enhance the activity. See online version for color.

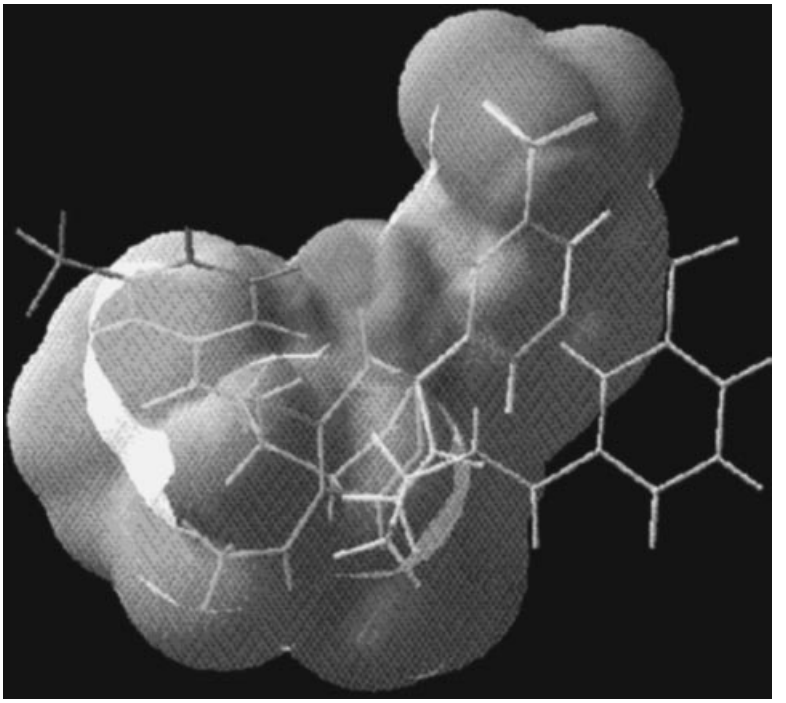

Fig. 10. An example of a RECEPTOR surface, created by three template molecules, and with a good fitting candidate structure docked into it (purple), and two poorer fitting candidates (green and yellow), which protrude through the surface, indicating regions of undesired steric bulk. See online version for color.

In summary, QSAR techniques give researchers a wide range of approaches to the problem of quantitating the relationship between chemical structures, changes in chemical structures, and observed physical and biological properties.

4. Summary

Clearly, whereas molecular modeling as a practice has its roots in the development of quantum theory at the turn of the twentieth century, it has been the exponential growth in computing power between the mid-1970s and the mid-1990s that has catalyzed the development and application of molecular modeling methods during that period. The spectrum of software systems available (145) covers all aspects of modeling. A sampling is given in Table 1. Although it is not possible to duplicate the exhaustive survey of available systems given in Refs. (53,54,146), Table 1 furnishes a starting point from which to search for the right tool to address specific problems.

An understanding of the precedents in both methods development and applications citations in the literature is thus critical to the researcher working in fields that employ molecular modeling as a tool. With it, the varied application

Table 1. **Molecular Modeling Software Systems**

Databases and DB Management Systems
Daylight Chemical Information Systems
Molecular Design Limited, Inc. (MDLI)
CCDC - Cambridge Crystallographic Data Center X-ray Database
PDB - Brookhaven Protein Data Bank
CAST-3D - The Chemical Abstracts Service 3D structure DB
UNITY - Part of the SYBYL suite from Tripos
Desktop Modeling and Data Management
ISIS Draw/SAR/Excel - Molecular Design Ltd., Inc.
ISIS Add-ins - Microsimulations, Inc.
Statistics - SAS/JMP, MINITAB,
ChemDraw/Chem3-D/ChemOffice - CambridgeSoft
Stand Alone Software Programs
Quantum Mechanical - NBO, GAUSSIAN, SPARTAN, MOPAC, PROAIM
PEOE Methods for Charges - Gasteiger, Scheraga, Gombar
Molecular Mechanical - MM3/MM4/MMFF
Molecular Dynamics - AMBER, GROMOS, QUANTA/CHARMm
Monte Carlo Simulations - BOSS
Molecular Modeling Software Suites
Hyperchem
InsightII/Cerius2- Accelrys
MacroModel - Schrödinger
Sybyl - Tripos
MOE - Chemical Computing Group
2-D to 3-D Converters
CONCORD
CORINA
WIZARD
DGEOM
QSAR
ADAPT - Pattern Recognition Toolkit from P. C. Jurs (MDLI)
APEX - Full range of statistical treatments for QSAR (Accelrys)
CHEMEST - Technical Database Services
CLOGP - Biobyte, Inc. (Pomona MedChem Project)
MOLCONN-X - L. Hall (Kier & Hall Indices)
QSAR-PC - Biosoft, Inc.
SigmaStat - Stat Analysis package (Jandel Scientific Software)
TOPKAT - Toxicology prediction tool
TOPMOST - Computes charges, descriptors for QS(APT)R
Docking and Pharmacophore Applications
AutoDOCK - Monte Carlo docking of ligands to receptors - A. Olson at Scripps Institute
CAVEAT - Database generator for new ligands - P. Bartlett at UC Berkeley
CATALYST - A pharmacophore generation program from Accelrys
DISCO - Tool for deriving pharmacophore from active compounds from Tripos
DOCK - Ligand docking/active site probe tool from UCSF
GRIN/GRID - Non-Bonded force probe of active sites - Molecular Discovery, Ltd.
HINT - adjunct to CoMFA, computes hydrophobic fields - EduSoft, Inc.
LEAPFROG - Generates new leads from fragments from Tripos.
LUDI - De Novo ligand design from Accelrys.

and untapped potential of molecular modeling may be used more profitably in individual researchers' specific fields of interest.

4.1. General References. *Web Resources*

MODELING SOFTWARE	URL
Chem software	http://www.chemistry-software.com/
Univ of FL	http://www.che.ufl.edu/www-che/topics/software.html
NetSci	http://www.netsci.org/Resources/Software
NIH software list	http://cmm.info.nih.gov/software.html
MDL software	http://www.mdli.com/
Univ of Potsdam	http://www.chem.uni-potsdam.de/linkcenter/sofz.html
Univ of Utah	http://www.cs.utah.edu/classes/cs5630/vis_link.html
Cerius II	http://www.accelrys.com/doc/
	http://www.icrs.tohoku.ac.jp/molsys/man/msi/quanta98/user_patches.html
DOCK	http://dock.compbio.ucsf.edu/
Chem-X	http://www.chemx.com/
FlexX	http://www.biosolveit.de/FlexX/flexx-intro.htm
GRAMMv1.03	http://reco3.ams.sunysb.edu/gramm/
MMTK	http://starship.python.net/crew/hinsen/MMTK/
Raptor	http://www.bioinformaticssolutions.com/products/raptor.php
MODELLER	http://salilab.org/modeller/modeller.html
NanoCAD	http://willware.net:8080/ncad.html
Spartan	http://www.wavefun.com
STALK	http://www-fp.mcs.anl.gov/ccst/research/reports_pre1998/comp_bio/stalk/docking.html
SYBYL	http://www.tripos.com/sciTech/inSilicoDisc/moleculeModeling/sybase.html
AMBER	http://www.amber.ucsf.edu/amber/amber.html
CHARMM	http://yuri.harvard.edu/
GROMOS	http://www.igc.ethz.ch/gromos/
Mdynamix	http://www.fos.su.se/physical/sasha/md_prog.html
MM2/MM3	http://europa.chem.uga.edu/allinger/mm2mm3.html
VMD	http://www.mathtools.net/Applications/Biotechnology/Molecular_Biology/Software/Molecular_Modelling/
Alignment	http://www.bork.embl-heidelberg.de/
Comp Chem	
MacroDox	http://pirn.chem.tntech.edu/macrodox.html
UHBD (Univ of Houston Brownian Dyn)	http://mccammon.ucsd.edu/uhbd.html
Crystallography	
George Phillips, Jr.	http://phillips-lab.biochem.wisc.edu/tools.html
ALB	http://sdpd.univ-lemans.fr/index.html
CaRlne	http://perso.wanadoo.fr/carine.crystallography/
education & general	
NASA	http://www.paloweb.com/science/physics/
	http://spacelink.nasa.gov/Instructional.Materials/NASA.Educational.Products/.index.html

UK Centre for Materials	http://www.materials.ac.uk/resources/index.asp
IUCr	http://journals.iucr.org/cww-top/crystal.index.html
VASP (Vienna Ab-initio Simulat'n Pckg)	http://cms.mpi.univie.ac.at/vasp/
	http://vaspview.sourceforge.net/
chemistry sofwares	http://hackberry.chem.trinity.edu/Chemistry Software.html
	http://www.chimicasoft.com/html/body_links.html
Docking	
Scripps/Art Olson	http://www.scripps.edu/pub/olson-web/doc/autodock/
The Genetic Algorithms Archive	http://www.aic.nrl.navy.mil/galist/
Molecular modeling	
Paul W. Chunn at UFL	http://www.med.ufl.edu/biochem/pchun/#pro3d
UofManchester Inst of Sci & Tech BmS	http://sjh.bi.umist.ac.uk/
chemistry-software.com	http://www.chemistry-software.com/software_guide/modelling_index.htm
Molecular Surface	
MSMS	http://www.scsb.utmb.edu/cgi-bin/get_a_form.tcl
	http://www.scripps.edu/pub/olson-web/people/sanner/html/msms_home.html
MSV	http://www.biohedron.com/
	http://www.scripps.edu/pub/olson-web/people/sanner/html/msv.html
visualization with Chime and RasMol	http://www.chem.leeds.ac.uk/stuartg/marching/surface.html
Molecular Vol-Calc	
QSAR	ftp://hobbes.gh.wits.ac.za/pub/steric
	http://www.ndsu.nodak.edu/qsar_soc/
	http://www.netsci.org/Science/Compchem/feature19.html
	http://mmlin1.pha.unc.edu/~jin/QSAR/
	http://clogp.pomona.edu/medchem/chem/qsar-db/
	http://www.iainm.demon.co.uk/indexnew.htm
Quantum Chem	
GAMESS-US	www.msg.ameslab.gov/GAMESS/GAMESS.html
GAMESS-UK	http://www.dl.ac.uk/CFS/cfs.html
QCPE	http://qcpe.chem.indiana.edu/
MOLPRO	http://www.molpro.net/
ACES II	http://www.qtp.ufl.edu/Aces2/
ATMOL	http://tc5.chem.uu.nl/ATMOL/
DISCO	http://hcc.keldysh.ru/~fock/codes/Quantum.html
HyperChem	http://www.hyper.com/
ADF, MOLF-DIR, etc	http://theochem.chem.rug.nl/
ADF, (BAND?)	http://www.scm.com/

PSI	Software-Univ of Potsdam (in the list of)
AllChem	http://ws2.theochem.uni-hannover.de/AllChem/
CASTEP	http://www.tcm.phy.cam.ac.uk/castep/
DGauss	http://www.cachesoftware.com/cache/dgauss/features.shtml
DMol	http://www.accelrys.com/cerius2/dmol3.html
WIEN97	http://info.tuwien.ac.at/theochem/wien97/
Argus	http://www.planaria-software.com/
YAeHMOP	http://yaehmop.sourceforge.net/
Simulation annealing	http://www.taygeta.com/annealing/simanneal.html
Visualization & Animation	
BallRoom	http://www.fisica.uniud.it/~ercolessi/ballroom.html
ChemScape, Chime	http://www.mdli.com/products/framework/chemscape/
Molscript	http://www.avatar.se/molscript/
MSV OpenGL	http://www.scripps.edu/pub/olson-web/people/sanner/html/msv.html
PDBtool	http://www.sdsc.edu/CCMS/Packages/PDBtool.html
PovChem	http://www.chemicalgraphics.com/PovChem
RasMol	http://www.umass.edu/microbio/rasmol/
SciAn	??? http://www.ccl.net/chemistry/resources/software/AIX/
Viewmol2.0	??? http://ccl.osc.edu/pub/chemistry/software/SOURCES/C/viewmol/
ViewerPro 5.0	http://www.accelrys.com/dstudio/ds_viewer/
CRYSTALLOGRAPHY	
webmineral	http://webmineral.com/crystall.shtml
COD (Cryst Open Database)	http://www.crystallography.net/
CCDC (Cambridge-CrysDB Centre)	http://www.ccdc.cam.ac.uk/
PDB	http://www.rcsb.org/pdb/
CADPAC	http://www-theor.ch.cam.ac.uk/software/cadpac.html

BIBLIOGRAPHY

"Molecular Modeling" in *ECT* 4th ed., Suppl. Vol., pp. 315–352, by Salvatore Profeta, Jr., Monsanto; "Molecular Modeling" in *ECT* (online), posting date: December 4, 2000, by Salvatore Profeta, Jr., Monsanto.

CITED PUBLICATIONS

1. W. Heisenberg, *Z. Phys.* **33**, 879 (1925).
2. E. Schrödinger, *Ann. Phys.* **79**, 361, 489; **80**, 437; **81**, 109 (1926).
3. J. D. Bolcer and R. B. Hermann, in K. B. Lipkowitz and D. B. Boyd, eds., *The Development of Computational Chemistry in the United States, Reviews in Computational Chemistry*, Vol. 5, VCH Publishers, Inc., New York, 1994, pp. 1–63.

4. R. S. Mulliken, in D. A. Ramsey and J. Hinze, eds., *Selected Papers of Robert Mulliken*, University of Chicago Press, Chicago, 1975, pp. 39–42.
5. R. S. Mulliken, *Life of a Scientist*, Springer-Verlag, Berlin, 1989, pp. 136–161.
6. F. London, *Z. Phys.* **63**, 245 (1930).
7. E. Hückel, *Z. Phys.* **70**, 204 (1931).
8. J. D. Kemp and K. S. Pitzer, *J. Chem. Phys.* **14**, 733 (1936).
9. F. H. Westheimer and J. E. Meyer, *J. Chem. Phys.* **14**, 733 (1946).
10. F. H. Westheimer, in M. S. Newman, ed., *Steric Effects in Organic Chemistry*, John Wiley & Sons, Inc. New York, 1956, p. 523.
11. J. B. Hendrickson, *J. Am. Chem. Soc.* **83**, 5537 (1961).
12. N. L. Allinger, J. Allinger, and M. A. DaRooge, *J. Am. Chem. Soc.* **86**, 4061 (1964).
13. G. Herzberg, *Molecular Spectra and Molecular Structure*. II. *Infrared and Raman Spectra of Polyatomic Molecules*, Van Nostrand Reinhold, New York, 1945. E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations*, McGraw-Hill, New York, 1955.
14. T. L. Hill, *J. Chem. Phys.* **14**, 465 (1946).
15. C. A. Coulson and G. S. Rushbrooke, *Proc. Cambridge Philos. Soc. Math. Phys. Sci.* **36**, 193 (1940).
16. M. Goeppert-Mayer and A. I. Sklar, *J. Chem. Phys.* **6**, 645 (1938).
17. M. J. S. Dewar, *J. Chem. Soc.* 2329 (1950). M. J. S. Dewar, *J. Chem. Soc.* 3532 (1952).
18. J. A. Pople, *Trans. Faraday. Soc.* **49**, 1375 (1953). J. A. Pople, *J. Phys. Chem.* **61**, 6 (1957).
19. R. Pariser and R. Parr, *J. Chem. Phys.* **21**, 466, 767 (1953).
20. A. Streitwieser, Jr., *Molecular Orbital Theory for Organic Chemists*, John Wiley & Sons, Inc., New York, 1961.
21. J. A. Pople and D. L. Beveridge, *Approximate Molecular Orbital Theory*, McGraw-Hill Book Co., New York, 1970.
22. J. N. Murrell and A. J. Harget, *Semi-empirical self-consistent-field molecular orbital theory of molecules*, Wiley-Interscience, London, England, 1972.
23. R. Hoffman and R. B. Woodward, *Acc. Chem. Res.* **1**, 17 (1968). R. B. Woodward and R. Hoffman, *Conservation of Orbital Symmetry*, Verlag Chemie, Weinheim, Germany, 1970.
24. E. Wimmer, *Chem. Design Auto. News* **8**(4), 1 (1993). R. C. Merkle, *Chem. Design Auto. News* **8**(9/10), 1 (1993).
25. A. L. Horvath, *Molecular Design*, Vol. 75, Elsevier, Amsterdam/New York, 1992.
26. C. Levinthal, *Sci. Am.* **214**, 42 (1966).
27. R. Langridge and T. Klein, *Introduction to Computer Graphics and Its Use for Displaying Molecular Structures*, in C. A. Ramsden, ed., *Comprehensive Medicinal Chemistry*, Vol. 4, Quantitative Drug Design, Pergamon Press, New York, 1990, pp. 413–430.
28. R. E. Hubbard , in N. C. Cohen, ed., *Molecular Graphics and Modeling: Tools of the Trade, Guidebook on Molecular Modeling in Drug Design*, Academic Press, San Diego, Calif., 1996, pp. 19–54.
29. T. Jones, *J. Appl. Crystallogr.* **11**, 268 (1978).
30. D. Sayer, ed., *Computational Crystallography*, Oxford University Press, New York, 1982.
31. G. N. Phillips, Jr., *Biophys. J.* **69**, 1281 (1995).
32. R. Langridge, T. E. Ferrin, I. D. Kuntz , and M. L. Connolly *Science* **211**, 661 (1981).
33. M. L. Connolly, *Science* **221**, 709 (1983). M. L. Connolly, *J. Appl. Crystallogr.* **16**, 548 (1983).
34. F. M. Richards, *Annu. Rev. Biophys. Bioeng.* **6**, 151 (1977).

35. N. Bohr, *Philos. Mag.* **26**, 476 (1913).
36. L. Pauling and E. B. Wilson, Jr., *Introduction to Quantum Mechanics*, McGraw-Hill Book Company, New York, 1935.
37. L. deBroglie, *The Revolution in Physics*, Noonday Press, Inc., New York, 1953. H. Eyring, J. Walter , and G. E. Kimball, *Quantum Chemistry*, John Wiley & Sons, Inc., New York, 1944. W. Kauzmann, *Quantum Chemistry*, Academic Press, New York, 1957.
38. P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford University Press, New York, 1958.
39. P. A. M. Dirac, *Proc. Soc. (London)* **A123**, 714 (1929).
40. A. Streitwieser, Jr., *Molecular Orbital Theory for Organic Chemists*, John Wiley & Sons, Inc. New York, 1961, pp. 6–7.
41. D. R. Hartree, *Proc. Cambridge Philos. Soc.* **24**, 89, 111, 426 (1927).
42. V. Fock, *Z. Phys.* **62**, 795 (1930).
43. A. Warshel and M. Karplus, *J. Am. Chem. Soc.* **94**, 5612 (1972).
44. N. L. Allinger and J. T. Sprague, *J. Am. Chem. Soc.* **95**, 3893 (1973).
45. J. A. Pople, D. P. Santry, and G. A. Segal, *J. Chem. Phys.* **43**, S129 (1965). J. A. Pople and G. A. Segal, *J. Chem. Phys.* **43**, S136 (1965).
46. M. Zerner, in K. B. Lipkowitz and D. B. Boyd, eds., *Semiempirical Molecular Orbital Methods, Reviews in Computational Chemistry*, Vol. 2, VCH Publishers, Inc., New York, 1991, pp. 313–356.
47. R. C. Bingham, M. J. S. Dewar, and D. H. Lo, *J. Am. Chem. Soc.* **97**, 1285 (1975).
48. M. J. S. Dewar, E. F. Healy, J. J. P. Stewart, and W. Thiel, in Ref. (46). See also: K. Jug and F. Neumann, *Encycl. Comput. Chem.* **1**, 507 (1998); J. J. P. Stewart, *Encycl. Comput. Chem.* **2**, 1513; **3**, 2000 (1998); A. J. Holder, *Encycl. Comput. Chem.* **1**, 8; **4**, 2542 (1998); J. J. P. Stewart, *J. Comput. Chem.* **10**, 209, 221 (1989); B. H. Bezler, K. M. Merz, Jr., and P. A. Kollman, *J. Comput. Chem.* **11**, 431 (1990).
49. J. J. P. Stewart, *J. Comput.-Aided Mol. Design* **4**, 1 (1990).
50. J. J. P. Stewart, in K. B. Lipkowitz and D. B. Boyd, eds., *Semiempirical Molecular Orbital Methods, Reviews in Computational Chemistry*, Vol. 1, VCH Publishers, Inc., New York, 1990, pp. 45–82.
51. T. Clark, *A Handbook of Computational Chemistry*, John Wiley & Sons, Inc., New York, 1985, Chapters 3 and 4; D. Feller and E. R. Davidson, in K. B. Lipkowitz and D. B. Boyd, eds., Basis Sets for *Ab Initio Molecular Orbital Calculations and Intermolecular Interactions, Reviews in Computational Chemistry*, Vol. 1, VCH Publishers, Inc., New York, 1990, pp. 1–44; N. R. Kestner and J. E. Combariza, in D. B. Boyd and K. B. Lipkowitz, eds., *Reviews in Computational Chemistry*, Vol. 13, VCH, New York, 1999, pp. 99–132.
52. M. J. S. Dewar, *J. Am. Chem. Soc.* **106**, 669 (1984).
53. D. Young, *Computational Chemistry, A Practical Guide for Applying Techniques to Real World Problems*, Wiley-Interscience, 2001.
54. A. Leach, *Molecular Modeling, Principles and Applications*, Longman, Essex, 1996.
55. W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, Wiley-Insterscience, New York, 1986, and references therein. W. J. Hehre, *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, Wavefunction, Irvine, Calf., 2003, pp. 21–55.
56. C. E. Dykstra, J. D. Augspurger, B. Kirtman, and D. J. Malik, in K. B. Lipkowitz and D. B. Boyd, eds., *Properties of Molecules by Direct Calculation, Reviews in Computational Chemistry*, Vol. 1, VCH Publishers, Inc., New York, 1990, pp. 83–118.
57. D. B. Boyd, in K. B. Lipkowitz and D. B. Boyd, eds., *Aspects of Molecular Modeling, Reviews in Computational Chemistry*, Vol. 1, VCH Publishers, Inc., New York, 1990, pp. 321–354.

58. L. S. Bartell, *J. Am. Chem. Soc.* **99**, 1977, 3279 (1977), and references therein.
59. U. Burkert and N. L. Allinger, *Molecular Mechanics*, ACS Monograph Series, Vol. 177, American Chemical Society, Washington, D.C., 1982. pp. 38, 17–58, 157.
60. N. L. Allinger, *Adv. Phys. Org. Chem.* **13**, 1 (1976); N. L. Allinger, *Encycl. Comput. Chem.* **2**, 1013 (1998); T. A. Halgren, *Encycl. Comput. Chem.* **2**, 1033 (1998); J. R. Maple, *Encycl. Comput. Chem.* **2**, 1015, 1025 (1998); A.K. Rappé and C.J. Casewit, *Molecular Mechanics across Chemistry*, University Science Books, Sausalito, Calif., 1997; see also Refs. 53, 54.
61. O. Ermer and S. Lifson, *J. Am. Chem. Soc.* **95**, 4121 (1973). O. Ermer, *Structure Bonding* **27**, 161. (1976).
62. S. Lifson and A. Warshel, *J. Chem. Phys.* **49**, 5116 (1968); A. Warshel and S. Lifson, *Chem. Phys. Lett.* **4**, 255 (1969); A. Warshel and S. Lifson, *J. Chem. Phys.* **53**, 582 (1970); A. T. Hagler and S. Lifson, *Acta Crystallogr. Sect. B* **30**, 619 (1974); A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.* **96**, 5319 (1974); A. Warshel and M. Levitt, *Nature (London)* **253**, 694 (1975); M. Levitt and A. Warshel, *J. Mol. Biol.* **106**, 421 (1976); M. Levitt and A. Warshel, *J. Am. Chem. Soc.* **100**, 2607 (1978).
63. A. T. Hagler, L. Leiserowitz, and M. Tuval, *J. Am. Chem. Soc.* **98**, 4600 (1976). S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.* **101**, 5111 (1979).
64. J. E. Williams, P. J. Stang, and P. v. R. Schleyer, *Annu. Rev. Phys. Chem.* **19**, 531 (1968).
65. K. B. Wiberg and R. H. Boyd, *J. Am. Chem. Soc.* **94**, 8426 (1972). K. B. Wiberg, *Computers Chem.* **1**, 221 (1977).
66. D. H. Andrews, *Phys. Rev.* **36**, 544 (1930).
67. S. Profeta, Jr., Conformational Analysis of Aliphatic Amines by Quantum and Molecular Mechanical Methods, Ph.D. Dissertation, University of Georgia, 1978. Diss. Abstr. 79–14051 (1979).
68. U. Dinur and A. T. Hagler, in K. B. Lipkowitz and D. B. Boyd, eds., *New Approaches to Empirical Force Fields, Reviews in Computational Chemistry*, Vol. 2, VCH Publishers, Inc., New York, 1991, pp. 99–164. J. R. Maple, U. Dinur and A. T. Hagler, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5350 (1988).
69. Ref. (59), pp. 253–283.
70. N. L. Allinger and S. Profeta, Jr., *J. Comput. Chem.* **1**, 181 (1980); S. Profeta, Jr. and N. L. Allinger, *J. Am. Chem. Soc.* **107**, 1907 (1985).
71. D. B. Boyd, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 6, VCH Publishers, Inc., New York, 1995, pp. 317–354.
72. T. Halgren, *J. Comp. Chem.* **17**, 490, 520, 553, 587 (1996).
73. N. L. Allinger and co-workers, *J. Comp. Chem.* **17**, 642, 669, 695, 730, 747 (1996) and subsequent papers in this series.
74. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. K. Weiner, *J. Am. Chem. Soc.* **106**, 765 (1984). S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comp. Chem.* **10**, 982 (1989).
75. J. A. McCammon and M. Karplus, *Annu. Rev. Phys. Chem.* **311**, 29 (1980). M. Karplus and J. A. McCammon, *Annu. Rev. Biochem.* **53**, 263 (1983). W. F. van Gunsteren and H. J. C. Berendsen, *Angew. Chem. Int. Ed. Engl.* **29**, 992 (1990).
76. C. R. Landis, D. M. Root and T. Cleveland, in Ref. (71), pp. 73–148.
77. F. M. Menger and co-workers, *J. Am. Chem. Soc.* **100**, 4676 (1978).
78. E. K. Davies and N. W. Murrell, *Computers Chem.* **13**, 149 (1989).
79. S. Profeta, Jr., R. J. Unwalla, B. T. Nguyen, and F. K. Cartledge, *J. Comp. Chem.* **7**, 528 (1986); S. Profeta, Jr., R. J. Unwalla, and F. K. Cartledge, *J. Org. Chem.* **51**, 1884 (1986); R. J. Unwalla, S. Profeta, Jr., and F. K. Cartledge, *J. Org. Chem.* **53**, 5658 (1988); S. Profeta, Jr., R. J. Unwalla, and F. K. Cartledge, *J. Comp. Chem.* **10**, 99 (1989).

80. M. R. Frierson, M. R. Imam, V. B. Zalkow, and N. L. Allinger, *J. Org. Chem.* **53**, 5248 (1988)
81. S. G. Cho, R. J. Unwalla, F. K. Cartledge, and S. Profeta, Jr., *J. Comp. Chem.* **10**, 832 (1989); S. G. Cho, F. K. Cartledge, R. J. Unwalla, and S. Profeta, Jr., *J. Mol. Struct.* **204**, 79 (1990); S. G. Cho, F. K. Cartledge, R. J. Unwalla, and S. Profeta, Jr., *Tetrahedron* **46**, 8005 (1990).
82. I. Pettersson and T. Liljefors, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 9, VCH Publishers, Inc., New York, 1996, pp. 167–190.
83. D. B. Boyd, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 1–7, VCH Publishers, Inc., New York, 1990-1995.
84. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford Univeristy Press, London, 1987.
85. J. M. Haile, *Molecular Dynamics Simulation: Elementary Methods*, Wiley-Interscience, New York, 1992.
86. C. L. Brooks III, M. Karplus, and B. M. Pettit, *Proteins. A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, 1988.
87. J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, 1987.
88. W. van Gunsteren and P. K. Weiner, eds., *Computer Simulation of Biomolecular Systems. Theoretical and Experimental Applications*, ESCOM, Leiden, The Netherlands, 1989.
89. I. Zuniga, I. Bahar, R. Dodge, and W. L. Mattice, *J. Chem. Phys.* **95**, 5348 (1991).
90. V. Galiatsatos, in K. B. Lipkowitz and D. B. Boyd, eds., *Computational Methods for Modeling Polymers: An Introduction, Reviews in Computational Chemistry*, Vol. 6, VCH Publishers, Inc., New York, 1996, pp. 149–208.
91. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.* **110**, 1657 (1988).
92. A. T. Hagler, in Victor J.Hruby, ed., *Theoretical Simulation of Conformation, Energetics and Dynamics of Peptides, The Peptides. Analysis, Synthesis and Biology: Conformation in Biology and Drug Design*, Vol. 7, 1985, pp. 213–299; H. Meirovitch, in K.B. Lipkowitz and D.B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 12, VCH Pubs, New York 1998, pp. 1–74.
93. W. L. Jorgensen and C. Ravimohan, *J. Chem. Phys.* **83**, 3050 (1985). W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
94. M. Saunders, K. N. Houk, Y.-D. Wu, W. C. Still, M. Lipton, G. Chang, and W. C. Guida, *J. Am. Chem. Soc.* **112**, (1990).
95. J. Skolnick and A. Kolinski, *Annu. Rev. Phys. Chem.* **40**, 207 (1989).
96. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
97. J. Devillers, ed., *Genetic Algorithms in Molecular Modeling*, Academic Press, San Diego, 1996; R. Judson, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 10, VCH Publishers, New York, 1997; D. B. McGarrah and R.S. Judson, *J. Comput. Chem.* **14**, 1385 (1993); R. S. Judson, W. P. Jaeger, A. M. Treasurywala, and M.L. Peterson, *J. Comput. Chem.* **14**, 1407 (1993); R. Judson, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 10, VCH Publishers, New York, 1997, pp. 1–73.
98. S. OskJonsdottir, W. J. Welsh, K. Rasmussen, and R. A. Klein, *New J. Chem.* 153 (1999).
99. I. Pettersson and T. Liljefors, in K. B. Lipkowitz and D.B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 9, VCH Publishers, New York, 1996, pp. 167–189; M. Jalaie and K.B. Lipkowitz, in D.B. Boyd and K.B. Lipkowitz, eds. RCC, Vol. 14,

VCH Publishers, New York, 200, pp. 441–486. (d). U. C. Singh and P. A. Kollman, Gaussian80 (UCSF), *Quantum Chemistry Program Exchange*, Prog. #446; QCPE Bull. 2, 121 (1982). See: U. C. Singh and P. A. Kollman, *J. Comp. Chem.* **7**, 718 (1986).

100. P. A. Kollman and co-workers, AMBER: Assisted Model Building and Energy inement, University of California, San Francisco 1980-present; S. J. Weiner and co-workers, *J. Am. Chem. Soc.* **106**, 765 (1984); S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *J. Comp. Chem.* **10**, 982 (1989); subsequent to the death of Professor Kollman, Professor D. A. Case has assumed responsibility for the program. J. Wang, R.M. Wolf, J. W. Caldwell, P.A. Kollman, and D.A. Case, *J. Comput. Chem.* **25**, 1157 (2004) and references therein.
101. Gaussian98, Rev. A.9, M. J. Frisch and co-workers, Gaussian Inc., Pittsburgh, Pa 1998.
102. J. Gao, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 7, VCH Publishers, Inc., New York, 1996, pp. 119–186.
103. L. Bartolotti and K. Flurchick, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 7, VCH Publishers, Inc., New York, 1996, pp. 187–216.
104. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York, 1989.
105. SPARTAN, A Quantum Mechanical and Molecular Modeling Program Suite, available from Wavefunction, Inc., Irvine, Calif.
106. R. S. Pearlman, R. Balducci, A. Rusinko, J. M. Skell, and K. M. Smith, CONCORD: A Program for Generating Three-Dimensional Coordinates, available from Tripos, Associates, Inc. St. Louis, Mo.
107. R. S. Pearlman, in H. Kubinyi, ed., *3-D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM Science Publishers, Leiden, The Netherlands, 1993, pp. 41–79.
108. A. Cayley, *Cambridge Math. J.* **II**, 267 (1841).
109. G. M. Crippen, *J. Comput. Phys.* **26**, 449 (1978).
110. G. M. Crippen and T. F. Havel, *Acta Crystallogr., Sect. A* **34**, 282 (1978). G. M. Crippen and T. F. Havel, *Distance Geometry and Molecular Conformation*, Chemometrics Research Studies Series 15. John Wiley & Sons, New York, 1988.
111. J. M. Blaney, G. M. Crippen, A. Dearing, and J. S. Dixon, DGEOM, Program # 590 QCPE Bull. 10, (1990). Quantum Chemistry Program Exchange, Indiana University, Bloomington, India.
112. T. F. Havel, I. D. Kuntz, and G. M. Crippen, *Bull. Math. Biol.* **45**, 665 (1983).
113. P. K. Weiner, S. Profeta, Jr., G. Wipff, T. Havel, I. D. Kuntz, R. Langridge, and P. A. Kollman, *Tetrahedron* **39**, 1113 (1983).
114. A. R. Leach, in K. B. Lipkowitz and D. B. Boyd, eds., *A Survey of Methods for Searching the Conformational Space of Small and Medium-Sized Molecules, Reviews in Computational Chemistry*, Vol. 2, VCH Publishers, Inc., New York, 1991, pp. 1–55.
115. J. M. Blaney and J. S. Dixon, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 5, VCH Publishers, Inc., New York, 1994, pp. 299–335.
116. A. K. Ghose and G. M. Crippen, in C. A. Ramsden, ed., *The Distance Geometry Approach to Modeling Receptor Sites, Comprehensive Medicinal Chemistry*, Vol 4, Quantitative Drug Design, Pergamon Press, New York, 1990, pp. 715–734.
117. QSPR: R. S. Pearlman, *Chem. Design Automat. News* **7**(3), 1 (1992); QSTR: V. K. Gombar, K. Enslein, and B. W. Blake, *Chemosphere* **31**, 2499 (1995); D. F. V. Lewis, in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 3, VCH Publishers, Inc., New York, 1992, pp. 173–222.
118. A. J. Leo and C. Hansch, *J. Org. Chem.* **36**, 1539 (1971); C. Hansch, in C. J. Cavallito, eds., *Structure-Activity Relationships*, Vol. 1, Pergamon Press, Oxford,

U.K., 1973, p. 150; C. Hansch and A. J. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, D.C., 1995, Chapts. 4 and 5.
119. J. W. McFarland, *J. Med. Chem.* **13**, 1192 (1970).
120. C. Hansch, A. R. Steward, S. M. Anderson, and D. Bentley, *J. Med. Chem.* **11**, 1 (1968); C. Hansch and J. M. Clayton, *J. Pharm. Sci.* **62**, 1 (1973).
121. Cerius^2 a Software Suite for Molecular Modeling, Accelrys, Inc., San Diego, CA.
122. R. W. Taft, in M. S. Newman, ed., *Steric Effects in Organic Chemistry*, John Wiley & Sons, Inc., New York, 1956, pp. 556–675.
123. P. N. Craig, *J. Med. Chem.* **14**, 680 (1971).
124. C. Hansch, S. H. Unger, and A. B. Forsythe, *J. Med. Chem.* **16**, 1212 (1973).
125. A. Leo, P. Jow, C. Silipi, and C. Hansch, *J. Med. Chem.* **18**, 865 (1975).
126. T. Fujia, in C. A. Ramsden, ed., *The Extrathermodynamic Approach to Drug Design, Comprehensive Medicinal Chemistry*, Vol 4, Quantitative Drug Design, Pergamon Press, New York, 1990, pp. 497–560; C. Hansch C. Hansch and T. Fujita, Eds., *Classical and Three-Dimensional QSAR in Agrochemistry*, ACS Symposium Series No. 606, American Chemical Society, Washington, D. C., 1995.
127. O. Kirino, H. Ohshita, T. Oishi, and T. Kato, *Agric. Biol. Chem.* **44**, 31 (1980).
128. C. D. Selassie, M. Chow, and C. Hansch, *Chem.-Biol. Interact.* **68**, 13 (1988).
129. Y. C. Martin, *Quantitative Drug Design*, Marcel Dekker, New York, 1978.
130. D. Cuthbert and F. S. Wood, *Fitting Equations to Data*, John Wiley & Sons, Inc., New York, 1971.
131. A. Cammarata, R. C. Allen, J. K. Seydel, and E. Wempe, *J. Pharm. Sci.* **59**, 1496 (1970).
132. J. G. Topliss and R. J. Costello, *J. Med. Chem.* **15**, 1066 (1972).
133. M. A. Pleiss and S. H. Unger, *The Design of Test Series and the Significance of QSAR Relationships*, in C. A. Ramsden, ed., *Comprehensive Medicinal Chemistry*, Vol 4, Quantitative Drug Design, Pergamon Press, New York, 1990, pp. 561–587.
134. H. Kubinyi, ed., *3-D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM Science Publishers, Leiden, The Netherlands, 1993;
135. See Ref. (115), from *Reviews in Computational Chemistry*, G. M. Downs and P. Willett, in D.B. Boyd and K.B. Lipkowitz, eds., Vol. 7, VCH Publishers, New York, 1995, pp. 1–66; A.C. Good and J.S. Mason, *Ibid*., pp. 67–117; J. R. Damewood, Jr., in D.B. Boyd and K.B. Lipkowitz, eds., Vol. 9, VCH Publishers, New York, 1996, pp. 1–79; T.I. Oprea and C.L. Waller, in D.B. Boyd and K. B. Lipkowitz, eds., Vol. 11, VCH Publishers, New York, 1997, pp. 127–182; G. Greco, E. Novellino and Y.C. Martin, in D.B. Boyd and K.B. Lipkowitz, eds., Vol. 11, VCH Publishers, New York, 1997, pp. 183–240; R.A. Lewis, S.D. Pickett and D. E. Clark, in D.B. Boyd and K.B. Lipkowitz, eds., Vol. 16, VCH Publishers, New York, 2000, pp. 1–51; M. A. Murcko in D.B. Boyd and K.B. Lipkowitz, eds., Vol. 11, VCH Publishers, New York, 1997, pp. 1–66;
136. A. Itai, Y. Mizutani, Y. Nishibata, and N. Tomioka, in N. C. Cohen, ed., *Computer-Assisted New Lead Design, Guidebook on Molecular Modeling in Drug Design*, Academic Press, San Diego, Calif., 1996, pp. 93–138; P.S. Charifson and I.D. Kuntz, in P.S. Charifson, ed., *Practical Application of Computer-Aided Design*, Dekker, New York, 1997, p. 1.
137. P. S. Magee, D. R. Henry, and J. H. Block, *Probing Bioactive Mechanisms*, ACS Symposium Series No. 413, American Chemical Society, Washington, D.C., 1989.
138. C. H. Reynolds, M. K. Hollaway, and H. K. Cox, *Computer-Aided Molecular Design. Applications in Agrochemicals, Materials and Pharmaceuticals*, ACS Symposium Series No. 589, American Chemical Society, Washington, D.C., 1995.
139. R. D. Cramer, D. E. Patterson, and J. D. Bunce, *J. Am. Chem. Soc.* **110**, 5959 (1988).

140. S. Wold, A. Ruhe, H. Wold, and W. J. Dunn, *SIAM J. Sci. Stat. Comput.* **5**, 735 (1984).
141. S. Wold, *Technometrics* **20**, 397 (1978).
142. M. Baroni, G. Costantino, G. Cruciani, D. Riganelli, R. Valigi, and S. Clementi, *Quant. Struct.-Act. Relat.* **12**, 9 (1993); M. Baroni, D. Bonelli, S. Clementi, G. Cruciani, C. Ebert, and B. Skagerberg, *Quant. Struct.-Act. Relat.* **9**, 101 (1990).
143. K. G. Kowalski, *Chemomet. Intel. Lab Syst.* **9**, 177 (1990).
144. U. Thibaut, G. Folkers, G. Klebe, H. Kubinyi, A. Merz, and D. Rognan, in H. Kubinyi. ed., *Recommendations for COMFA Studies and 3-D QSAR Publications, 3-D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM Science Publishers, Leiden, The Netherlands, 1993, Appendix A and B, pp. 711–728.
145. D. B. Boyd, in Ref. (102), pp. 303–380.
146. M. A. Hahn, *J. Med. Chem.* **38**, 2080; **2091** (1995).

SALVATORE PROFETA JR.
University of South Carolina