Molecular Silverware. I. General Solutions to Excluded Volume Constrained Problems

ARTICLE *in* JOURNAL OF COMPUTATIONAL CHEMISTRY · MARCH 1991

Impact Factor: 3.59 · DOI: 10.1002/jcc.540120214

1 AUTHOR:

Mario Blanco
King Abdullah University of Science and Tech…

64 PUBLICATIONS **1,677** CITATIONS

SEE PROFILE

Electronic Delivery Cover Sheet

Molecular Silverware. I. General Solutions to Excluded Volume Constrained Problems

Mario Blanco

Rohm and Haas Company, 64C, P.O. Box 219, Bristol, Pennsylvania 19007

Received 8 May 1990; accepted 27 July 1990

General mathematical solutions to excluded volume constrained problems in computational chemistry are reported. The solutions have been used to create a new family of molecular modeling algorithms to facilitate the study of molecular interactions in condensed phases. The new algorithms, collectively known as Molecular Silverware, are for the most part interactive and designed for packing, solvating, and sampling molecules embedded in simple or complex topological environments. Multifolded, disconnected, or porous molecular structures are permitted. Molecular Silverware assists the preparation of Monte Carlo and molecular dynamics simulations at a small fraction of the total simulation time. Primary targets for applications include the study of molecular recognition mechanisms and the selective binding of DNA, RNA, peptides, saccharides and other biopolymers in solution as well as the prediction of phase separation behavior and physical properties of non-crystalline condensed phases such as bulk polymers, polymer blends, organic liquids, membranes, micelles, gels, crosslinked networks, glasses, and amorphous heterogeneous catalysts. As a result of this new approach to excluded volume constraints, the computer simulation of noncrystalline condensed phases is no longer hampered by the lack of a general and efficient method for the creation and configurational sampling of small and large molecular assemblies at high densities.

INTRODUCTION

Current computer technology has stimulated academic and industrial research applications of molecular modeling on ever increasing scales. Large size molecules, $N = 10^3$–10^4 atoms, and large numbers of single molecule conformations, 10^5–10^6, can be readily modeled. However, the mechanics and dynamics of noncrystalline molecular assemblies with $M = 10$–10^3 molecules have not been significantly explored in spite of the fact that the calculations are, in some cases, of the same order of magnitude as the large single molecule calculations currently under investigation. Examples include organic liquids, DNA, proteins, oligopeptides and oligosaccharides in solutions, membranes, micelles, gels, polymer blends, cross-linked networks, amorphous glasses, and heterogeneous catalysts among others. It may well be that the modeling of condensed phases presents problems, in addition to size, which are absent in large single molecule computations. One such problem is briefly examined below.

In a molecular computer simulation it is often desirable that

1. The energy distribution among *intramolecular* degrees of freedom is consistent with the average kinetic energy of the system (thermal equilibrium).
2. The interactions involving *intermolecular* degrees of freedom are also in thermal equilibrium.

Much attention has been devoted over the years to the development of robust algorithms to satisfy 1. The modeling of energy distribution among *intramolecular* degrees of freedom is becoming a mature field of research. Metropolis method, Rotational Isomeric State theory, protein folding algorithms are only a few examples of widely studied tools for intramolecular conformational analysis. However, no published general algorithms exist to ensure a priori that *intermolecular* interactions approach thermal equilibrium for the general $M > 1$ case. The lack of automatic and robust methods to prepare noncrystalline molecular assemblies having acceptable *intermolecular* interactions, prior to the actual molecular simulation, has slowed down the progress in computer modeling of condensed phases. It can be said that requirement 2 leads to two of the long standing problems in molecular modeling, namely

1. The automatic creation of noncyrstalline molecular assemblies satisfying excluded volume constraints, and
2. The automatic creation of a priori thermally equilibrated molecular assemblies, i.e., requiring little or no additional equilibration steps prior to the actual computer simulation.

A general mathematical solution to problem 1., the automatic creation of noncrystalline condensed phase molecular assemblies, is reported here for the first time. We present a brief review of past methods

Journal of Computational Chemistry, Vol. 12, No. 2, 237–247 (1991)

CCC 0192-8651/91/020237-11$04.00

which were developed to partly cope with the excluded volume constraint problem. The derivation of the constraint equations as well as their method of solution for the general $M > 1$ case follows. A set of new molecular modeling algorithms is introduced in the following section. We have become accustomed to refer to the collection of such algorithms as *Molecular Silverware* to portray the manner in which one uses them in practical molecular simulations. It is shown that the Lipkowitz–Darden molecular recognition sampling algorithm,[1,2] herein referred to as LD, is a particular case of a more general molecular recognition algorithm. A few applications to current problems in computational chemistry are included. The examples include the molecular modeling of liquid phase systems, water, and organic fluids (CFC's), the sampling of chemical interactions in complex molecular environments (inside a DNA double helix), and the solvation of molecules with arbitrarily complex topologies (Zeolites).

PREVIOUS METHODS

A few methods are in use to cope with the problems brought about by molecular excluded volume constraints. The most common method consists of creating a low density molecular assembly, with large distances in between molecules, to avoid molecular overlaps. Molecular dynamics or Monte Carlo simulations are subsequently used to achieve higher more physically realistic densities. The low density assembly is often created by placing individual molecules at periodic lattice points, allowing for random molecular orientations at each of the lattice sites. A disadvantage of this method is that the achievement of high densities and the destruction of the built-in periodic lattice correlations takes a considerable amount of the total computer simulation time, around 50% in many cases.

An alternative method is to create a high density assembly followed by pruning of molecules with large van der Waals repulsions. This can be done by direct inspection followed by deletion or by spatial displacements of badly overlapped molecules. The method is tedious and time consuming. Consequently, a method has been developed whereby molecular overlaps are resolved by a scaling down of all nonbonded force field parameters.[3] Both, the van der Waals radii and the potential well depth Lennard–Jones parameters are scaled down. The central idea is to resolve the overlaps at a faster rate by temporarily allowing bond crossings to occur during a molecular dynamics run. As the molecular overlap resolution step proceeds, the Lennard–Jones parameters are scaled back to their normal values. This method is effective in resolving van der Waals overlaps but it suffers from a lack of control over the final conformations assumed by individual molecules when resolving their overlaps. Such disadvantage is a great loss considering the variety of widely accepted conformational methods, some mentioned above, for the analysis and preparation of isolated molecular conformations. It is highly desirable to leave the molecular conformations generated by these methods intact during the assembly steps. Any change in the conformational statistics of the assembly should be the result of the specific molecular interactions present in the condensed phase and not the result of artifacts introduced during the molecular assembly steps.

More recently Lipkowitz et al.[1,2] developed an algorithm, herein referred to as LD, for molecular recognition studies for the design of chiral stationary phases in enantiomeric separations. We show that the LD algorithm is part of a more general solution of the $M = 2$ molecular overlap problem. An advantage of the general solution, as well as the LD algorithm, is that during the molecular assembly steps the internal conformations of each molecule, which may have been carefully generated using one of the methods mentioned above, are left undistorted. Each molecule is treated as a rigid body, only translations and rotations are allowed during the assembly steps. The purpose of the next section is to show how a general solution to the excluded volume constraint problem can be found. In addition it will be shown that the new general solution makes possible the formulation of a whole family of molecular modeling algorithms which is physically adapted to the high densities typical of condensed phase systems.

METHOD OF EXCLUDED VOLUME CONSTRAINTS

We first focus attention on the $M = 2$ case. For the purpose of resolving molecular overlaps each molecule is modeled as a rigid body, i.e., as a collection of hard spheres. The size of each sphere is set by the van der Waals radius of the corresponding atom. Of course, covalently bound atoms necessarily have overlapping van der Waals spheres. We are only concerned with the resolution of van der Waals overlaps between nonbonded atoms.

Two rigid bodies require six degrees of freedom to have their relative orientations and positions in free space completely specified. We choose the first three degrees of freedom to be the Euler angles, $\Omega = (\alpha, \beta, \gamma)$, which define the orientation of molecule 2 in a coordinate frame affixed to molecule 1. The vector connecting the centers of mass of both molecules, in polar coordinates (r, θ, ϕ), completes the list of six, initially independent, degrees of freedom. One of these, r, becomes a function of the other five when the excluded volume constraint equations

are enforced. We seek the form and the solution of these equations.

The global coordinate reference frame is depicted in Figure 1. The frame is affixed to the center of mass and oriented along the principal moments of inertia axes of molecule 1. The atoms are number $i = 1, 2, \ldots, N_1$. Atoms in molecule 2 are numbered $j = N_1 + 1, N_1 + 2, \ldots, N_1 + N_2$. In the final self-avoiding assembly the vector location of all atoms in molecule 2 will be given by two rigid body operations, a rotation $R_\Omega(\alpha, \beta, \gamma)$ and a translation $r\,\mathbf{n}$, applied to the original atomic position vectors $\mathbf{r}_j^o$

$$\mathbf{r}_j = R_\Omega(\alpha, \beta, \gamma)\mathbf{r}_j^o + r\,\mathbf{n}; \quad j = N_1 + 1, N_1 + 2, \ldots, N_1 + N_2 \quad (1)$$

$\mathbf{n}$ is a unit vector in the polar direction (θ, ϕ),

$$\mathbf{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \quad (2)$$

$R_\Omega(\alpha, \beta, \gamma)$ is a 3×3 direction cosine matrix[4]

$$R_\Omega(\alpha, \beta, \gamma) = \begin{bmatrix} \cos\gamma\cos\beta\cos\alpha - \sin\gamma\sin\alpha & \cos\gamma\cos\beta\sin\alpha + \sin\gamma\cos\alpha & -\cos\gamma\sin\beta \\ -\sin\gamma\cos\beta\cos\alpha - \cos\gamma\sin\alpha & -\sin\gamma\cos\beta\sin\alpha - \cos\gamma\cos\alpha & \sin\gamma\sin\beta \\ \sin\beta\cos\alpha & \sin\beta\sin\alpha & \cos\beta \end{bmatrix}$$

The original orientation of molecule 2, $\Omega_o = (\alpha = 0, \beta = 0, \gamma = 0)$, is taken to be the orientation which aligns its principal axes with the principal axes of molecule 1. Before the rigid body operations are applied, the centers of mass of both molecules coincide at the origin of the global coordinate reference frame. We now discuss the excluded volume constraint equations and their solutions for several cases in increasing order of complexity.

$M = 2, N_1 > 1, N_2 = 1$

The excluded volume constraint equations can be written in the following general form

$$\Phi_{ij}(r, \theta, \phi) \geq 0 \quad i = 1, 2, \ldots N_1; \quad j = N_1 + 1 \quad (4)$$

where

$$\Phi_{ij}(r, \theta, \phi) = r^2 + 2(\mathbf{r}_{ij}^{\pm} \cdot \mathbf{n})\, r + |\mathbf{r}_{ij}|^2 - (s_i + s_j)^2 \quad (5)$$

and $\mathbf{r}_{ij}$ is the vector difference (refer to Figure 2).

$$\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i \quad (6)$$

s_i and s_j are the van der Waals radii of the corresponding atoms. Because in this first case "molecule" 2 is monoatomic the Euler angles α, β, and γ are all set to zero and the rotation matrix becomes a 3×3 unit matrix.

Figure 2 depicts how the excluded volume constraint expression (5) was obtained. An equation for the magnitude r of the translation vector along the $\mathbf{n}$ line of sight was derived from the expression for the magnitude of the vector addition

$$|\mathbf{r}_{ij} + r\,\mathbf{n}|^2 = (s_i + s_j)^2 \quad (7)$$

Expansion of the left hand side leads to expression (5). The solution and a physical interpretation of expression (5) follows.

The displacement needed to place atom j in contact with atom i along a line of sight given by the direction $\mathbf{n}$ is obtained by setting Φ_{ij} to the minimum value allowed by the constraint eq. (4), i.e., $\Phi_{ij}(r, \theta, \phi) = 0$. Solving for r one gets

$$r_{ij}^{\pm} = -(\mathbf{r}_{ij} \cdot \mathbf{n}) \pm [(\mathbf{r}_{ij} \cdot \mathbf{n})^2 - |\mathbf{r}_{ij}|^2 + (s_i + s_j)^2]^{1/2} \quad (8)$$

r_{ij}^- places atom j on the near side of atom i while r_{ij}^+ places it on the far side. Negative values of r_{ij}^- or r_{ij}^+ are interpreted as positive displacements along the $-\mathbf{n}$ direction. Notice that two complex roots result if

$$(\mathbf{r}_{ij} \cdot \mathbf{n})^2 - |\mathbf{r}_{ij}|^2 + (s_i + s_j)^2 < 0 \quad (9)$$

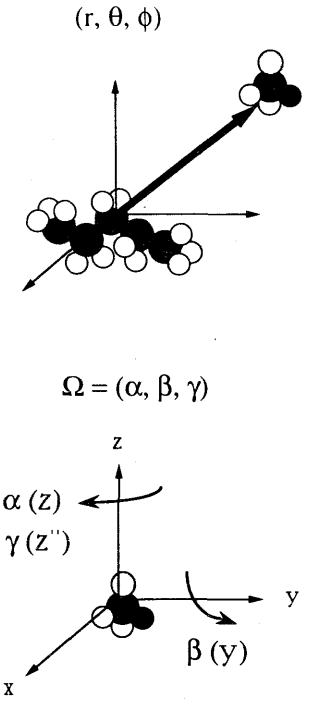
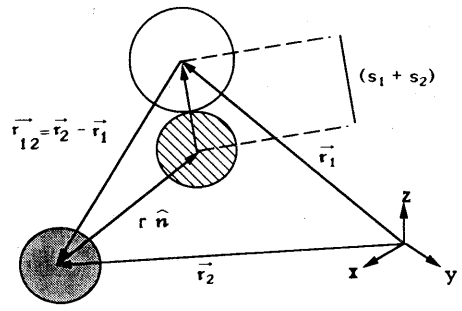


Figure 1. The global coordinates used to solve the excluded volume constraint problem. Molecule 1 is located at the center of the global frame. The unit frame axis coincide with the principal axis of inertia of molecule 1. The position of the center of mass of molecule 2 is given by the polar coordinates (r, θ, ϕ) and its orientation relative to molecule 1 by the three Euler angles given by $\Omega = (\alpha, \beta, \gamma)$. The rotations are carried out in the order $R_{z''}(\gamma)\, R_{y'}(\beta)\, R_z(\alpha)$ from right to left. $\mathbf{y}'$ and $\mathbf{z}''$ are the intermediate and final y and z axis orientations respectively.



$$|\vec{r}_{12} + r\,\hat{n}|^2 = (s_1 + s_2)^2$$

$$r^2 + 2(\vec{r}_{12}\cdot\hat{n})\,r + |\vec{r}_{12}|^2 - (s_1 + s_2)^2 = 0$$

$$r_\pm = -(\vec{r}_{12}\cdot\hat{n}) \pm \sqrt{(\vec{r}_{12}\cdot\hat{n})^2 - |\vec{r}_{12}|^2 + (s_1 + s_2)^2}$$

Figure 2. This figure depicts the vectors employed in casting the excluded volume constraint problem. The cross-hatched atom indicates the final resting position of atom 2, here shown in gray, when atom 1 is located in the path of atom 2 along the **n** line of sight. See text for more details.

Complex roots indicate that no van der Waals overlap exists along the line of sight **n** and thus no real displacement can put atom i and j in contact with each other. Figure 3 depicts all six possible root combinations in terms of the initial positions of atoms i and j and the **n** line of sight.

A more useful interpretation of eq. (4) follows. The real roots of $\Phi_{ij}(r, \theta, \phi)$ define a set of open segments on the real axis

$$\lambda_{qj} = (r_{qj}^-, r_{qj}^+) \qquad q = 1, 2, \ldots, p \qquad (10)$$

where the index q runs only over the p real roots of $\Phi_{ij}(r, \theta, \phi)$. If atom j gets translated through a vector ξ **n**, with ξ a number contained within any λ_{qj} segment, ($r_{qj}^- < \xi < r_{qj}^+$ for some q), it violates constraint eq. (4). The union of all λ_{qj} open segments constitutes the complete set of "forbidden" displacements. Conversely, the *complement* of the union defines all points along the line of sight **n** where atom j can be translated and be free from van der Waals overlaps with molecule 1. Therefore, the solution to the excluded volume constraint eq. (4) is the set

$$\Lambda_j^c(\mathbf{n}) = \{\lambda_{1j} \cup \lambda_{2j} \cup \ldots \cup \lambda_{pj}\}^c \qquad (11)$$

Figure 4 illustrates the effect of the union operation on the possible placement locations for atom j when all van der Waals overlaps are resolved. Note that because the λ_{qj}'s are open segments, their ends, the real roots of $\Phi_{ij}(r, \theta, \phi)$, may be part of the set Λ_j^c. However, the union operation eliminates most of the roots from the solution set. It is relatively simple to write computer code to determine which roots survive the union operation. Two roots which are never eliminated are

$$r_j^- = \mathrm{Min}_q\{r_{qj}^-\} \qquad (12a)$$

$$r_j^+ = \mathrm{Max}_q\{r_{qj}^+\} \qquad q = 1, 2, \ldots, p \qquad (12b)$$

These two immortal roots are always elements of the solution set Λ_j^c because there are no segments to the right and to the left of the right-most and left-most segments respectively. All displacement values in the ranges

$$-\infty \le r \le r_j^-; \quad r_j^+ \le r \le \infty \qquad (12c)$$

are also elements of the solution set Λ_j^c. Incidentally, if one leaves all values greater than r_j^+ and less than r_j^- out of the solution set Λ_j^c one obtains an operational definition of the "probe" accessible surface. If the van der Waals radius s_j is set equal to zero one gets an operational definition of the multiple valued van der Waals surface of molecule 1 in polar coordinates.

$M = 2, N_1 > 1, N_2 > 1$

The Euler angles are required in this case. There are $N_1 \times N_2$ constraint equations similar to (4)

$$\Phi_{ij}(\alpha, \beta, \gamma, r, \theta, \phi) = r^2 + 2(\mathbf{r}_{ij}^\Omega \cdot \mathbf{n})r + |\mathbf{r}_{ij}^\Omega|^2 - (s_i + s_j)^2 \qquad (13)$$

where

$$\mathbf{r}_{ij}^\Omega = R_\Omega(\alpha, \beta, \gamma)\mathbf{r}_j^o - \mathbf{r}_i^o \quad i = 1, 2, \ldots, N_1; \quad j = N_1 + 1, N_1 + 2, \ldots, N_1 + N_2 \qquad (14)$$

all other quantities have their previous meanings. The general solution to the excluded volume constraint problem is

$$\Lambda^c(\Omega, \mathbf{n}) = \{\Lambda_1 \cup \Lambda_2 \cup \ldots \cup \Lambda_{N2}\}^c \qquad (15)$$

where

$$\Lambda_j = \{\lambda_{ij} \cup \lambda_{2j} \cup \ldots \cup \lambda_{N1j}\} \qquad (16)$$

and

$$\lambda_{ij} = (r_{ij}^{\Omega-}, r_{ij}^{\Omega+}) \qquad (17)$$

For simplicity of notation all roots of eq. (13) were made to appear in the definition of solution (15) but it must be kept in mind that only the real roots are to be included. Just as before we have

$$r_{ij}^{\Omega\pm} = -(\mathbf{r}_{ij}^\Omega \cdot \mathbf{n}) \pm [(\mathbf{r}_{ij}^\Omega \cdot \mathbf{n})^2 - |\mathbf{r}_{ij}^\Omega|^2 + (s_i + s_j)^2]^{1/2} \qquad (18)$$

Expression (15) makes no assumptions regarding molecular shape. All valid excluded volume constraint solutions can be found in the Λ^c set. All physically allowed self-avoiding assemblies can be created regardless of the complexity in the topologies of molecules 1 and 2 by changing the values of

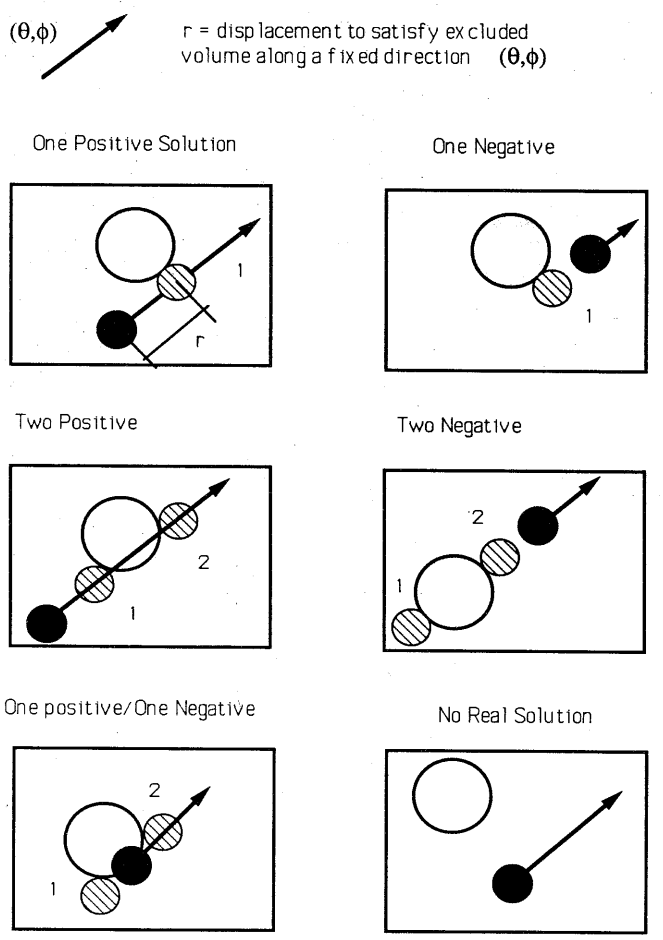


Figure 3. The six possible root combinations for the excluded volume constraint between two atoms are shown. Depending upon the relative position in space of the atoms the solutions are: (1) a single positive displacement, (2) a single negative displacement, (3) two positive, (4) two negative, (5) one positive and one negative when the atoms overlap in the undisplaced coordinates, (6) no real overlap. Note that a quadratic equation is consistent with all these six solutions.

(α, β, γ) and (θ, ϕ) and solving for the Λ^c set following eqs. (16)–(18).

The operational simplicity of solution (15), its generality with respect to molecular topology, and the ability to select appropriate paths through (α, β, γ) and (θ, ϕ) space are important elements for the search and development of more general molecular modeling algorithms appropriate to condensed phase simulations. From such a molecular modeling point of view an important element of the solution set is the value r^c defined such that

$$|r^c| = \text{Min } |\Lambda^c| \tag{19}$$

The vertical bars indicate absolute values. r^c is the shortest possible displacement required by molecule 2 to resolve all atomic overlaps with molecule 1.

$M > 2$

The excluded volume constraint problem can be stated in the most general possible way as follows. Given M molecules with fixed orientations and center of mass directions

$$\Omega_m = (\alpha_m, \beta_m, \gamma_m) \tag{20a}$$

and

$$\mathbf{n}_m = (\sin\theta_m \cos\phi_m, \sin\theta_m \sin\phi_m, \cos\theta_m) \tag{20b}$$

respectively, find the complete set of all molecular displacements (r_m's) which satisfy the excluded volume constraint equations:

$$|R_m^{\Omega}\mathbf{r}_j^o + r_m\mathbf{n}_m - R_{m'}^{\Omega}\mathbf{r}_i^o - r_{m'}\mathbf{n}_{m'}|^2 \geq (s_i + s_j)^2$$
$$\forall\, i < j = 1, 2, \ldots, \sum_m N_m; \quad m > m' = 1, 2, \ldots, M \tag{20c}$$

where r_i^o and r_j^o are the original position vectors for atom i and atom j respectively, m' and m identify the molecules carrying atoms i and j respectively,

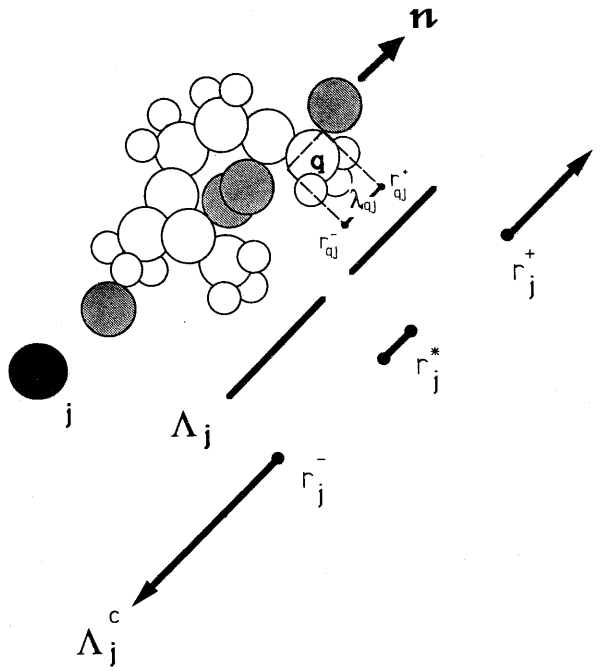


Figure 4. Suitable nonoverlapping locations for atom j along the line of sight n are indicated by the gray colored atoms. The union of the λ_{qj} segments gives the set of forbidden displacements while the complement of the union gives the solution to the excluded volume constraint equation Λ^c_j.

and N_m ($m = 1, 2, \ldots, M$) gives the number of atoms in molecule m.

The elements of the solution set Λ^c are M-tuples of the form $(r_1, r_2, \ldots, r_M)$. There is an infinite number of elements in Λ^c. Most of them belong inside semi-infinite segments of the form (12c). From a molecular modeling point of view, however, the important solutions satisfy eq. (19) also.

Equation (20) defines a nonholonomic constraint problem. There are no general ways of attacking nonholonomic constraint problems. According to Goldstein[5] "the most vicious cases of nonholonomic constraints must be tackled individually." A practical solution to this problem is obtained by extending the method employed in the $M = 2$ case, i.e., the components of the M-tuples are found one at a time.

We begin by placing one of the M molecules at the origin of the global reference frame ($r_1 = 0$). We then select a second molecule and apply the method of the previous section to find the solution set for the two molecule system. In accordance with eq. (19) we set r_2 equal to min $|\Lambda^c|$. Notice that we could have selected the identity of the first and second molecules in M and $M - 1$ different ways respectively. In general there are $M - m + 1$ ways of choosing the mth molecule when the assembly already contains $m - 1$ molecules. Consequently there is a possible total of $M!$ different molecular assemblies. Each of these assemblies is located fully in 3-D space by giving (1) the original coordinates of the atoms in each molecule, $\mathbf{r}^o_i$, (2) the orientations and displacement directions, Ω_m and $\mathbf{n}_m$ respectively, and (3) the M-tuple $(r_1, r_2, \ldots, r_M)$ of molecular centers of mass displacements.

The solution method consists of the following steps

1. Choose a labeling scheme for the M molecules, $m = 1, 2, \ldots, M$.
2. Define the following operational indexes:

$$N_1^{(m)} = \sum_{k=1}^{m} N_k$$

$$N_2^{(m)} = N_{m+1} \tag{21}$$

3. Set the global coordinate frame to coincide with the center of mass of molecule 1. In effect we are setting $r_1 = 0$. Set $m = 1$.
4. Calculate the value r_{m+1} according to the solution method employed in the $M = 2$ case, i.e., $|r_{m+1}| = \text{Min } |\Lambda^c|$. Equations (14) and (18) now read

$$\mathbf{r}^{\Omega}_{ij} = R_{\Omega m+1}\mathbf{r}^o_j - \mathbf{r}_i \tag{22}$$

$$r^{\Omega\pm}_{ij} = -(\mathbf{r}^{\Omega}_{ij} \cdot \mathbf{n}_{m+1}) \pm [(\mathbf{r}^{\Omega}_{ij} \cdot \mathbf{n}_{m+1})^2 - |\mathbf{r}^{\Omega}_{ij}|^2 + (s_i + s_j)^2]^{1/2} \tag{23}$$

with

$$i = 1, 2, \ldots, N_1^{(m)}; \quad j = N_1^{(m)} + 1, N_1^{(m)} + 2, \ldots, N_1^{(m)} + N_2^{(m)}$$

5. Reset the coordinates of the $m + 1$ molecule according to

$$\mathbf{r}_j = (R_{\Omega m+1}\mathbf{r}^o_j + r_{m+1}\mathbf{n}_{m+1}) \tag{24}$$

   Set m to $m + 1$.
6. Repeat steps 4 and 5 until $m = M - 1$.

One could add another step, a loop around steps 1–6, by requiring a different choice of molecular labeling order. However, empirically we have found that this is not always required. For a set of M chemically and conformationally similar molecules the order in which the assembly operation is performed has little or no effect. In those cases the calculated bulk properties like densities are not very sensitive to the choice of labeling order. On the other hand, for dissimilar molecules different labeling schemes might be required depending on the molecular modeling problem being studied. For example, if one is interested in studying the effects of solvation on the conformational properties of a biologically active molecule one would choose the latter as the first molecule. The ordering of the $M - 1$ identical solvent molecules is not very important according to our former empirical observation. On the other hand, one might be interested in calculating the density of an amorphous polymer by creating assemblies of M

polymer molecules each with a distinct conformation. The Boltzman averaged density can be calculated from the free energies and the densities of each of the assemblies included in such a study. In this case it is important to prepare as many of the $M!$ assemblies as practically possible to sample the true free energy distribution function. It is useful to refer to each of the possible $M!$ assemblies as individual members of a canonical ensemble from which the computation of statistically averaged physical properties is possible.

MOLECULAR SILVERWARE APPLICATIONS

Three molecular modeling algorithms have been developed using as their foundation the solution to the excluded volume constraint problem. We have become accustomed to refer to such algorithms collectively as *Molecular Silverware* to portray the manner in which one employs them in practical molecular simulations, i.e., as desktop serving utensils for molecular manipulations. How is Molecular Silverware used in combination with molecular dynamics, mechanics, or Monte Carlo simulations? We currently use them in tandem but we consider this a question open for research. An introductory overview of their applications is presented below. A separate more detailed description of each algorithm, in particular practical paths through the Euler and polar orientation spaces will be published in the near future.

MS-P: Molecular Packing

In crystalline materials the geometric regularity of the molecular lattice assists the building and packing of self-avoiding molecular assemblies. High densities are easily obtained prior to any molecular dynamics or Monte Carlo simulation. On the other hand noncrystalline condensed phases are perhaps the most difficult systems to model at high densities. The lack of symmetry in a condensed amorphous phase makes the preparation, the running, and the analysis of computer simulations a difficult task. Thus, it is not too surprising that, while most commercially available polymeric materials are amorphous, research on the use of molecular modeling to predict polymer physical properties has been conducted on crystalline polymers almost exclusively.

The general solution, $M > 2$, to the excluded volume constraint equations introduced in the previous section simultaneously defines a molecular packing algorithm which we term MS-P. MS-P should, at the least, ease the preparation steps required prior to any Monte Carlo or molecular dynamics run of condensed amorphous phases. As a bonus, a significant reduction in running times is certain since the computer time required to equilibrate a traditionally prepared molecular assembly, often around 50% of the total simulation time, can now be substantially saved. Using the results of the previous section one easily prepares molecular assemblies with material densities quite close to experimental values in only a small fraction of the time such simulations take to run.

Figure 5 (see color Plate I.) shows two MS-P tightly packed clusters containing 1,200 water molecules. The molecular orientation vectors $\Omega_m = (\alpha_m, \beta_m, \gamma_m)$ and $\mathbf{n}_m$ were chosen randomly and the van der Waals radii for oxygen and hydrogen atoms are fixed at their usual values, 1.5, and 1.1 Å respectively. After the assembly was created the bulk density was calculated by dividing the total mass in the assembly by its volume. The calculated water liquid density radial profile (Fig. 6) converges to the value 1.005 g/cc, in good agreement with the experimental density at room temperature. The density profile shows a maximum around 2.5 Å. The maximum represents the location of the pronounced change in density when the first van der Waals water coordination shell is encountered. The surprising thing about the radial density profile for water is that although no energy minimization calculations were carried out it still bares a striking resemblance to the pair distribution functions obtained for simple liquids using the full Lennard–Jones potential.[11] When the water drop assembly is minimized, using the MM2 force field for example, the largest deviation in the density profile across the drop is always less than 5%.

It has long been believed that, to a first approximation, the properties of nonionic liquids can be predicted solely on the basis of geometric packing considerations. Supporting experimental evidence

---

Figure 5(a). An example of the use of Molecular Silverware packing algorithm MS-P. 1,200 water molecules were tightly packed into a single droplet. Values of van der Waals radii for hydrogen and oxygen were set to 1.1 and 1.5 Å respectively. **(b)** A cross section and close up view of the water droplet depicted in (a) proves that the molecules satisfy excluded volume constraints. No van der Waals overlaps exist throughout the entire droplet.

Figure 8. An example of an assembly of an amorphous polymer created with Molecular Silverware. The assembly was built using one hundred oligomers, 1000 Mw each, of syndiotactic PolyMethyl Methacrylate.

Figure 10. MS-S was used to solvate a DNA sequence containing close to a dozen base pairs. A cross section of the solvated DNA is shown. For clarity of depiction only one solvation shell was employed. Note that water molecules penetrate major and minor grooves achieving a very close wetting of the DNA double helix surface. The explicit solvation of oligonucleotides, oligopeptides, oligosaccharides and other biopolymers containing a few dozen residues takes only a few seconds of computer time with the MS-S algorithm.

(See color Plate I.)

Plate I.

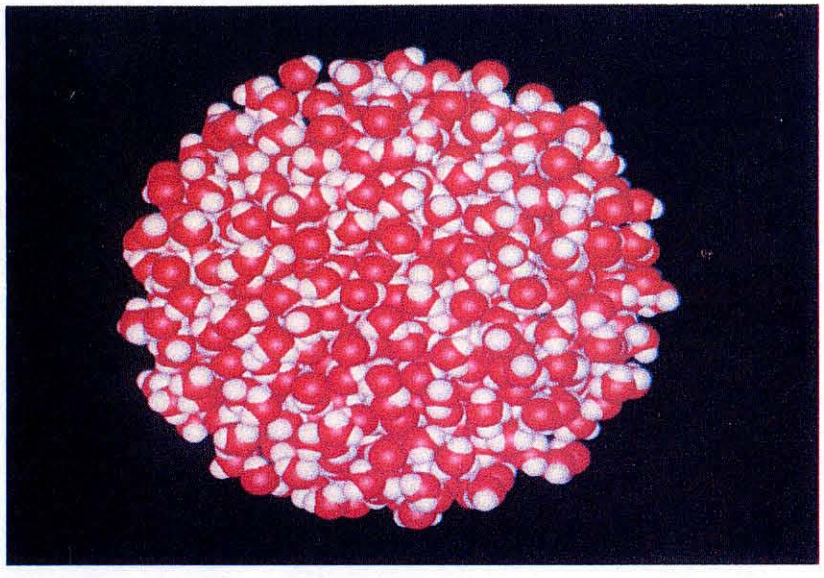


Figure 5(a).

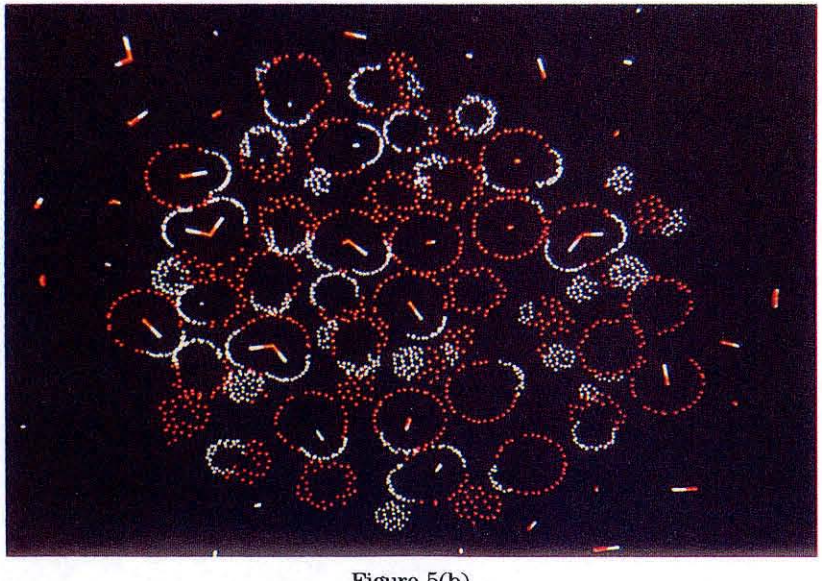


Figure 5(b).

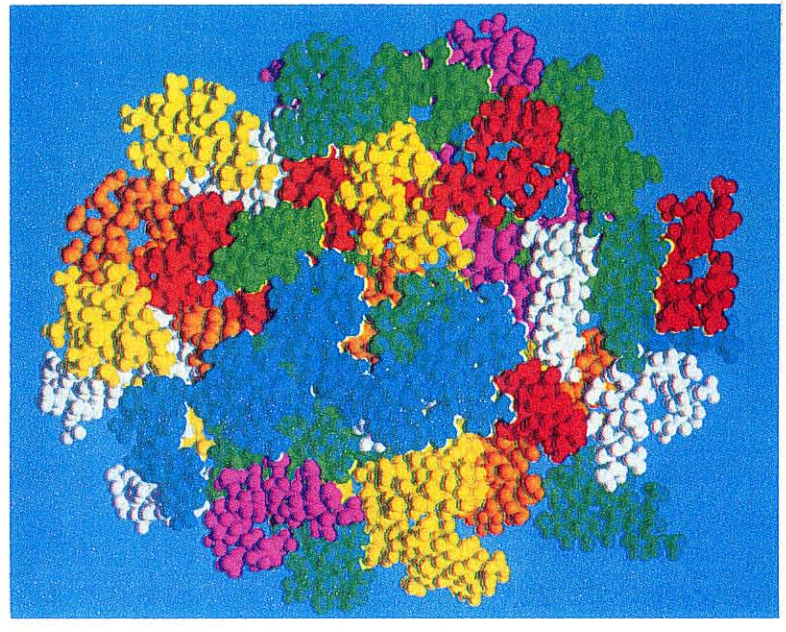


Figure 8.

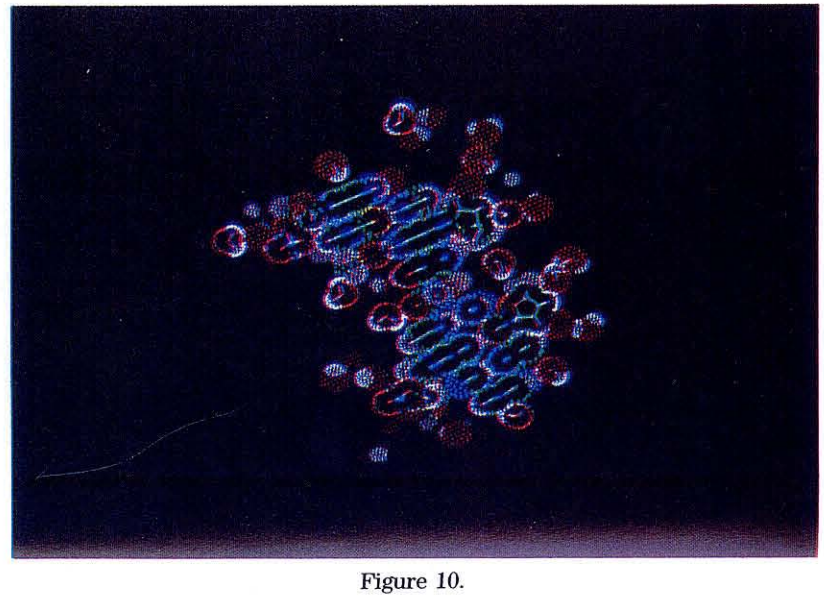


Figure 10.

Plate I.

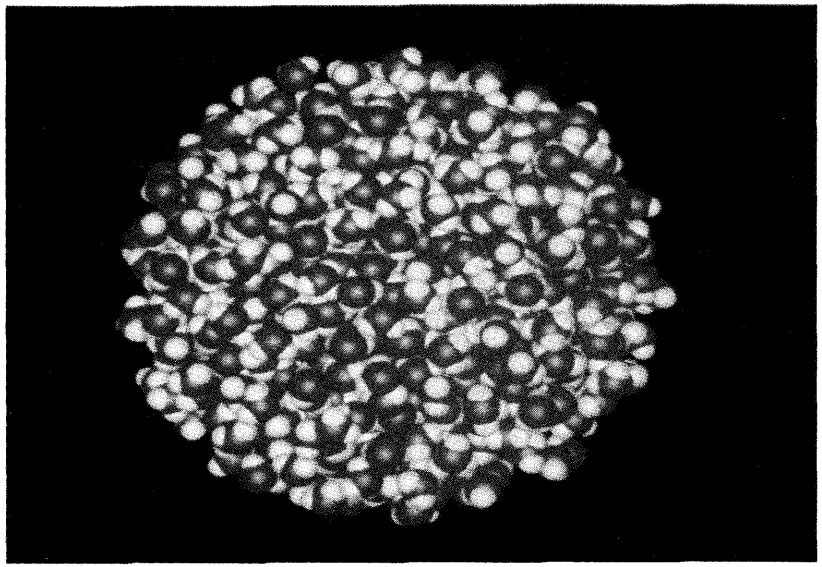


Figure 5(a).

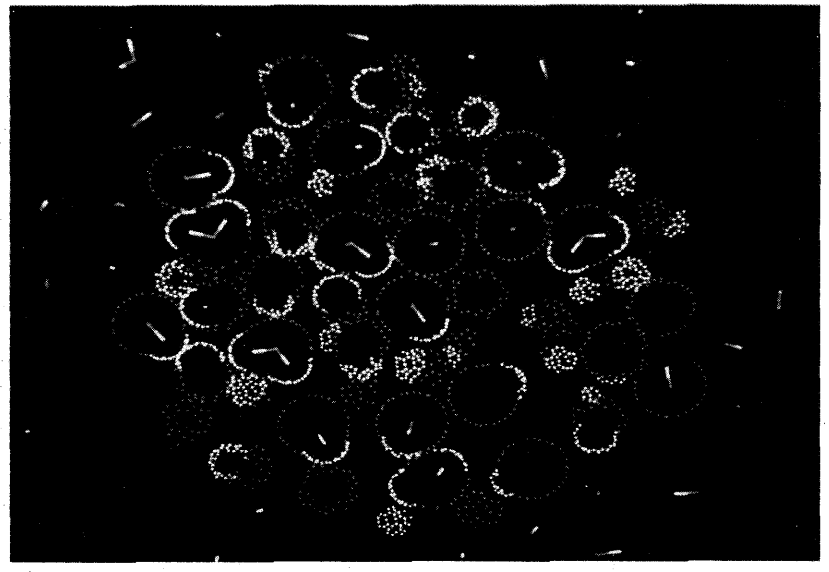


Figure 5(b).

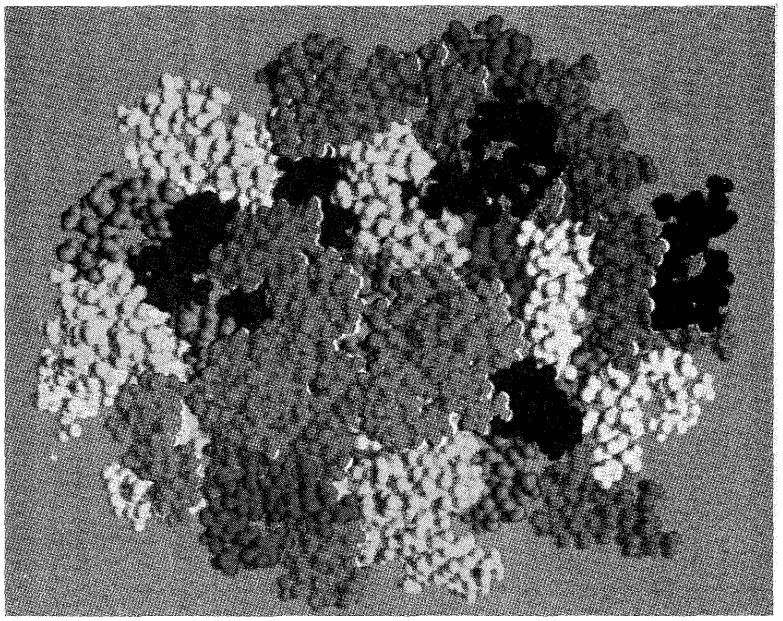


Figure 8.

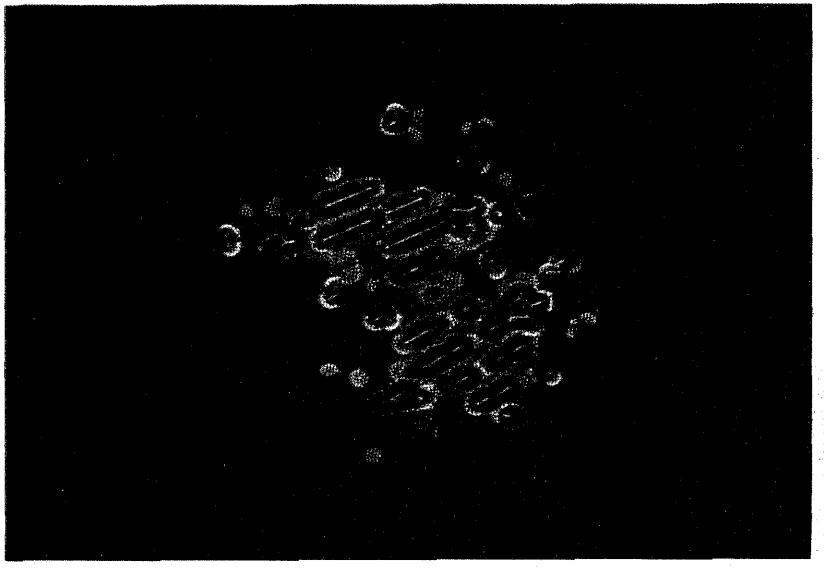


Figure 10.

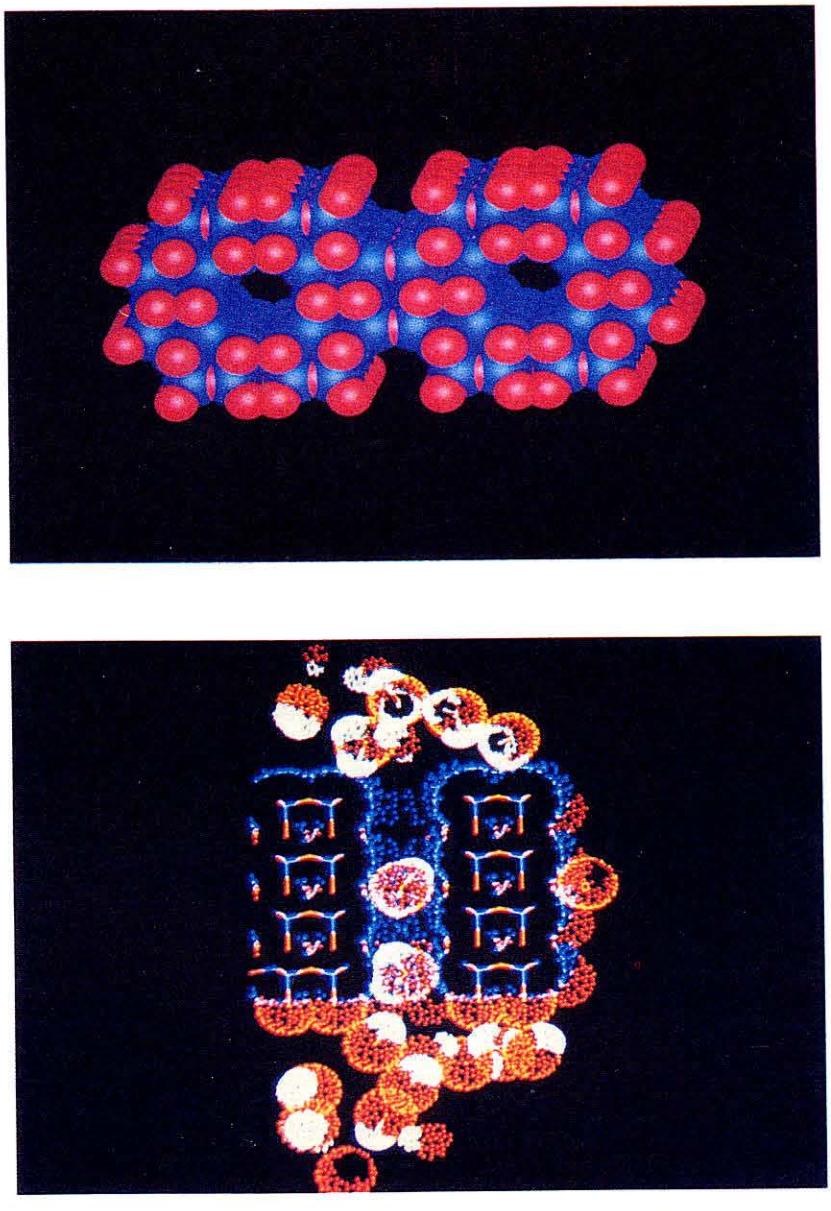


Figure 11. Water solvation of a Zeolite catalyst.[9] **(a)** ZSM5 Zeolite is a molecular sieve containing open channels. Two such channels are depicted before solvation. **(b)** A cross section top view of the Zeolite structure shows that the Molecular Silverware solvation algorithm, MS-S, is able to find and solvate the open channels with no additional information but the molecular structure of ZSM5. Two water molecules can be seen residing completely inside the molecular sieve channel.

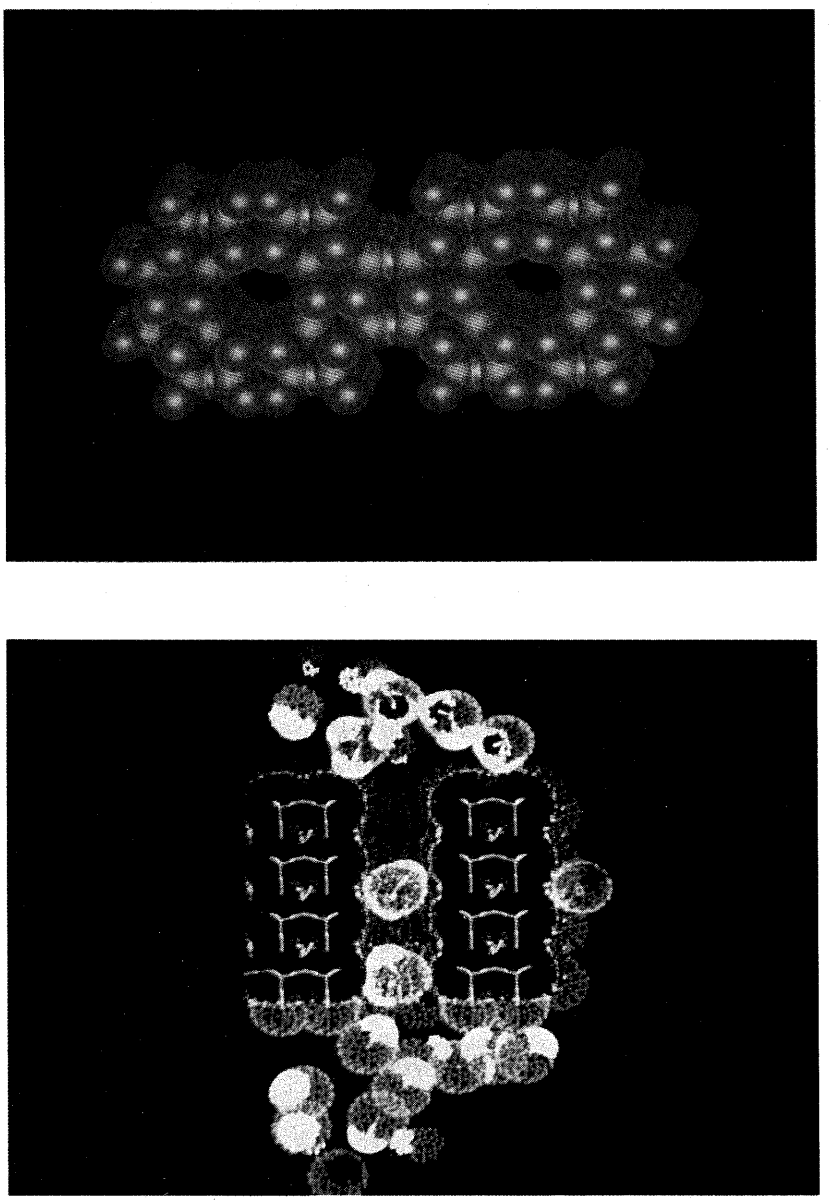


Figure 11. Water solvation of a Zeolite catalyst.[9] **(a)** ZSM5 Zeolite is a molecular sieve containing open channels. Two such channels are depicted before solvation. **(b)** A cross section top view of the Zeolite structure shows that the Molecular Silverware solvation algorithm, MS-S, is able to find and solvate the open channels with no additional information but the molecular structure of ZSM5. Two water molecules can be seen residing completely inside the molecular sieve channel.

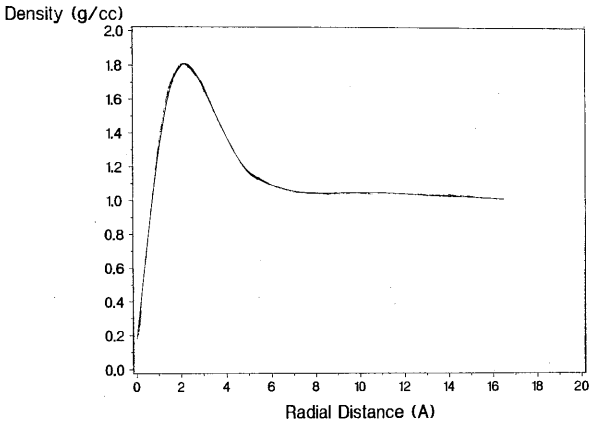


Figure 6. The calculated density of a packed water droplet depicted in Figure 5 as a function of distance from the center of the droplet. The asymptotic, large droplet, density value is 1.005 g/cc.

has recently appeared in the literature.[8] It is not surprising therefore that for quite some time Percus-Yevick-like equations of state give a good description of the local structure of molecular liquids represented by fused hard spheres. The calculated versus experimental 25°C liquid densities for a series of chlorofluorocarbon molecules with general formula $C_1H_nX_mY_{4-(n+m)}$ are shown in Figure 7. Notice that to a good approximation experimental densities of nonionic organic liquids can be linearly correlated with values calculated solely on the basis of molecular packing considerations. A notable exception is CHF_3 whose experimental density is extraordinarily much lower than what one would expect solely on the basis of molecular packing effects. We consider the peculiar low experimental density of liquid CHF_3 worthy of a more in-depth study. We can only speculate on two possible explanations for the discrepancies between experimental and geometric packing densities, (1) the actual van der Waals atomic radii in CHF_3 are much larger than the standard values used in the packing calculations (2) more likely, the close proximity of the critical point in the liquid-vapor phase diagram of CHF_3 (critical temperature T_c = 25.9°C) introduces large local density fluctuations which are unaccounted for in the current packing algorithm. Density fluctuations could be incorporated by packing additional free volume with the correct size distribution function. One possibility is to pack "holes" of different sizes together with the CHF_3 molecules. This approach has not been attempted here.

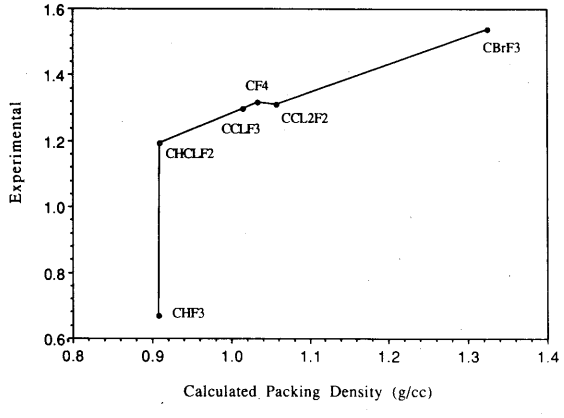


Figure 7. The experimental (25°C) liquid density versus calculated packing densities for a series of chlorofluorocarbons. Notice that to a good approximation densities of nonionic organic liquids are linearly correlated with the densities calculated solely on the basis of molecular packing considerations. One notable exception is CHF_3 whose experimental density is extraordinarily low when compared to that calculated on the basis of molecular packing considerations. See main text for details.

Figure 8 (see color Plate I.) shows an assembly of an amorphous polymer. The packed density of 0.85 g/cc is somewhat lower than the experimental value, ca. 1.18 g/cc for high molecular weight (>30,000) pMMA. The assembly was built using one hundred oligomers, 1000 Mw each, of syndiotactic pMMA. To obtain a higher packing density a larger number of molecules, on the order of 1,000, with higher molecular weight and Boltzman weighted conformations are required. Such calculations require much longer runs and are preferably performed in batch mode. A more complete study will be presented separately.

During the assembly steps the molecules are treated as rigid bodies. Consequently, the MS-P molecular packing algorithm is conformation preserving, a property which should not be overlooked. In reaching for a higher level of understanding, e.g., the conformational statistics of molecular assemblies, one should not discard the significant gains in understanding the conformational statistics of isolated molecules which have been made over the last two decades.[7] In conclusion, the new molecular packing algorithm described here should greatly facilitate the study of noncrystalline condensed phases, further the understanding of the changes in molecular conformational statistics in going from isolated to bulk systems, probe the reason for the past success of approximate equations of state for liquids, and contribute to elaborate new more exact expressions.

MS-R: Molecular Recognition

To a great extent it can be said that computational chemistry is a sampling problem. Several techniques, classified as molecular recognition algorithms, are available to investigate specific chemical interactions between two or more molecules. Molecular recognition algorithms sample the space of possible orientations of two molecules, with varying degrees of human intervention, in search for such special interactions. Lipkowitz, Darden et al.[1,2] developed an algorithm for molecular recognition studies and the design of chiral stationary phases for enantiomeric

separations. An important part of the LD algorithm consists of expressions needed to compute free energies. These free energy expressions rigorously follow from the definitions of statistical mechanical averages in a canonical ensemble. An equally important element in LD is the automatic sampling of bimolecular interactions. The sampling algorithm admits one generalization, the extension to molecules having non-convex surfaces, e.g., having one or more foldings.

The LD algorithm employs as the solution to the excluded volume constraint equations the displacement given by the largest positive root of the quadratic constraint equations. This root is easily identified as the positive immortal root in the $M = 2$ excluded volume constraint case (see eq. (18) above). By analogy with the $N_2 = 1$ case depicted in Figure 4 such root can be labeled r^+. One must notice that three other roots, r_j^- and the two labeled r_j^*, can also give rise to valid displacements where the two molecules just touch without overlapping. In the general case there are two such displacements for every segment λ_{ij} which survives the union operation in the definition of the general solution $\Lambda^c(\Omega,\hat{n})$ given by eq. (15).

We propose eq. (15) as the basis for a more general molecular recognition algorithm applicable to molecules of arbitrary topology. We term this algorithm MS-R. MS-R, and LD give equivalent results when the two molecules in question are convex with respect to each other. We understand a molecule to be convex with respect to one another if the excluded volume constraint set $\Lambda(\Omega,\hat{n})$ contains one and only one continuous segment on the real axis for all possible values of Ω and $\mathbf{n}$. For example, the molecule depicted in Figure 4 is not convex with respect to atom j because Λ_j contains two continuous segments. Consequently there are positions inside the "convex hull" of molecule 1 where molecule 2 can be located free of van der Waals overlaps.

Because a good number of known biopolymers, e.g., enzymes, DNA, RNA, etc ..., having varying degrees of folding it is hoped that the new algorithm will find significant uses in the field of molecular recognition in biologically active molecules. A systematic grid search of all relative molecular orientations, the $(\Omega,\mathbf{n})$ space, and a semiautomatic "binding motif" search[6] have both been implemented to create MS-R, a very general molecular recognition algorithm. A complete description of MS-R will be published separately.

We have chosen indole-3-carbinol, I3C, and a very short DNA four paired base sequence, ACGT, to illustrate the use of the Molecular Silverware recognition algorithm. I3C has been widely investigated as a possible inhibitor of hepatocarcinogenesis in trout and other species induced by aflatoxin B_1 (AFB_1) and polycyclic aromatic hydrocarbons. The most recent studies[10] indicate that I3C reaction products produced under acidic conditions in vitro act as inhibitors in the metabolic activation of AFB_1 preventing its binding to DNA. An alternative, but perhaps less likely inhibition route might be the direct but reversible blocking of potential binding sites for AFB_1 in DNA by I3C or its reaction products. A complete investigation of this question would require the identification of potential I3C binding sites along DNA segments of comparable molecular size, or about four base pairs. To illustrate this point several hundred orientations of I3C-ACGT complex were automatically created using Molecular Silverware recognition algorithm MS-R. Each van der Waals complex was subsequently minimized using the AMBER force field. The most stable I3C-ACGT complex found is shown in Figure 9(a). A histogram of all the binding energies is shown in Figure 9(b). Such histograms represent the density of states, $g(E)$, of the appropriate canonical interaction ensemble partition function and are quite useful for calculating ensemble averaged properties such as binding energies, dipoles, and averaged ligand distances. The arrow points at the binding energy of the most stable van der Waals complex. Of course, a complete computational study of the possible binding properties of I3C must include all unique four-base sequences of DNA (256) as well as binding interactions with AFB_1 and its metabolic activators. A similar but more complete study of quinolones, Norfloxacin and Ofloxacin, binding to DNA will be reported shortly.

MS-S: Molecular Solvation

It has been a standard practice to use a spherical probe to identify the solvent accessible surface of a molecule. In place of a spherical probe one could actually use the atomic coordinates and van der Waals radii of the solvent molecule. MS-S implements this explicit solvation approach. The available solvation sites can be simultaneously identified and filled with solvent molecules. The resulting solvated assembly can be used in subsequent Monte Carlo or molecular dynamics simulations.

MS-S is general enough to allow any desired solvent to interact with solutes of any kind including porous media, zeolite-like structures, and cross-linked networks. Explicit solvation should be recommended in those cases where special interactions between solvent and solute molecules are not taken into account in the more efficient, but also more approximate, implicit solvation methods. Most likely explicit solvation methods will become more widely used as the ease with which explicitly solvated molecular structures of arbitrary complexity can be prepared becomes more widely known.

Figures 10 and 11 (see color Plates I. and II.) show two cases where explicit solvation has been used. Figure 10 shows a cross section of a DNA double

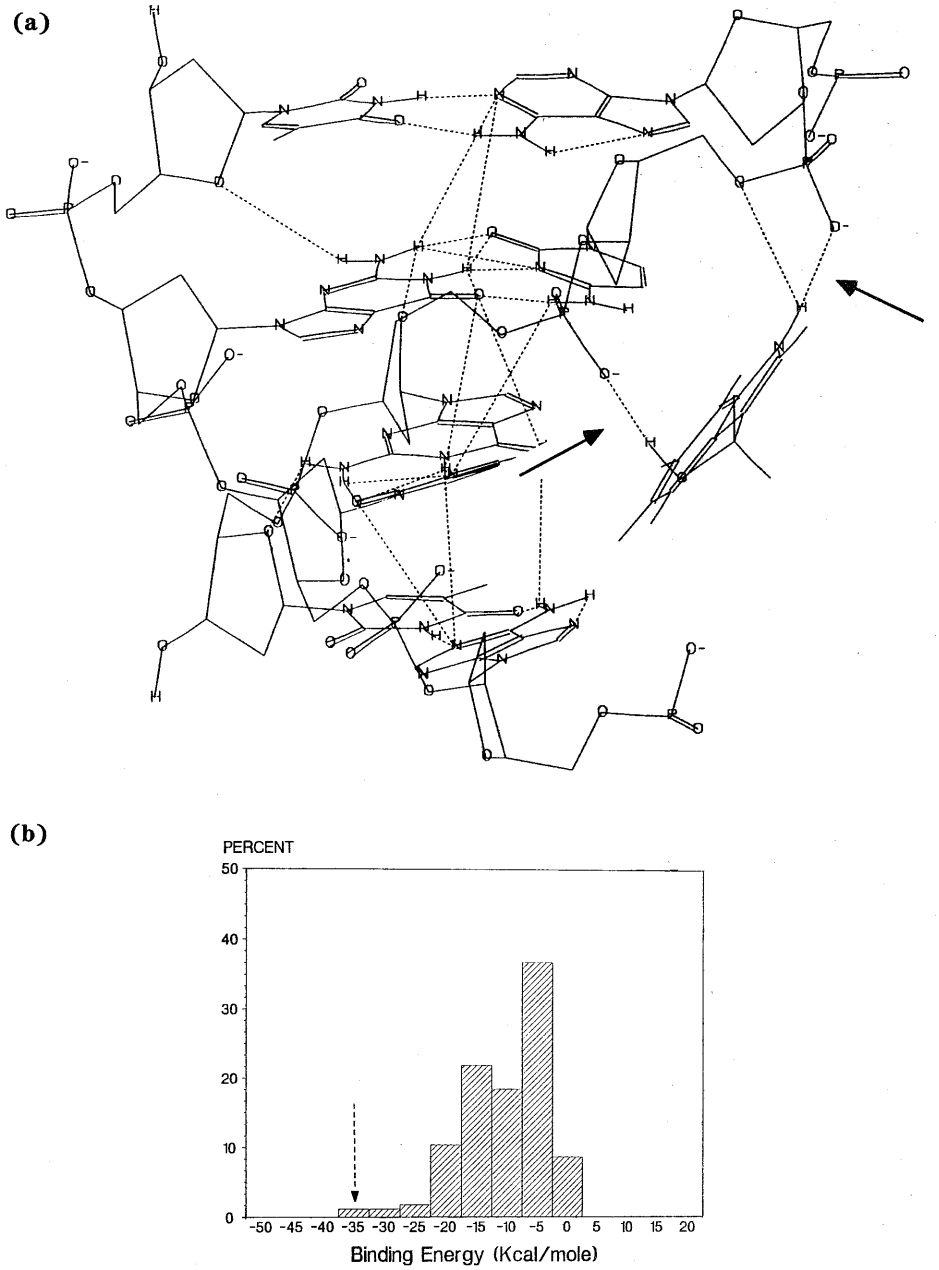


Figure 9(a). The preferred interaction geometry of Indole-3-carbinol (I3C) with ACGT-DNA sequence was identified with the help of the Molecular Silverware Recognition algorithm MS-R. The small arrows indicate the presence of hydrogen bonding between I3C and the ACGT's sugar phosphate flanking segments inside the major groove of DNA. **(b)** The binding energies for all I3C-ACGT complexes are displayed in histogram form. The dotted arrow points at the most exothermic binding energy found by MS-R. The corresponding complex is the one depicted in (a).

helix solvated with one van der Waals layer of water. Any number of layers are allowed in the present algorithm but for simplicity of depiction only one solvent layer was added. The cross section of the solvated DNA molecule shows that water molecules closely wet the surface of DNA, even inserting themselves among the major and minor grooves of the double helix. The explicity solvated structures can be used as input to programs designed to compute free energies of solvation from Monte Carlo or Molecular dynamics simulations. Conversely, a large number of such solvated structures, differing from each other in the positions of the solvent molecules, followed by single point energy calculations can be used to calculate a canonical ensemble averaged free energy of solvation.

Figure 11 shows the solvated channels in a Zeolite catalyst.[9] Such channels are automatically found by MS-S whether they have openings to the external van der Waals surface of the solute molecule or not.

Proteins with internal "pockets" could be equally solvated or if preferred a logical switch in MS-S can prevent such internal moieties from being solvated. The ability to handle topologically complex molecules, including multifolded, disconnected, or porous structures is a property of all Molecular Silverware algorithms and a direct consequence of enforcing excluded volume constraints in the most general possible manner.

DISCUSSION

A few words about computer timings. MS-R, molecular recognition, and MS-S, molecular solvation are essentially interactive algorithms. Only for large systems (>1000 atoms), batch execution is preferable. The solvation of the Zeolite example, approximately 500 atoms, took less than one minute of execution time, on a micro VAX class machine, per solvation shell added. A typical molecular recognition problem, which creates several hundred molecular configurations, takes on the order of one to three minutes for molecules containing up to 50 atoms. Close to 90% of the time is spent in writing out to disk the required molecular geometry deck files.

MS-P, molecular packing, can produce results interactively if lower than tight packing is allowed. For densities approaching normal values at room temperature, however, batch execution is preferable because several dozen molecular orientations are tested before each molecule gets added to the assembly. Even then, the calculations take only a fraction of the total time required by the subsequent molecular dynamics or Monte Carlo simulation. The tight packing of 100 chlorofluorocarbon molecules ($C_1X_nY_{4-n}$), 1500 atoms, takes on the order of 10 minutes on a VAX 8500 class machine. The tight packing of 1000 chlorofluorocarbon molecules, 15,000 atoms, takes on the order of five hours. Topologically complex molecules add only 2% overhead to the molecular packing times. Because independent molecular orientations can be evaluated concurrently the algorithm was coded to run close to 100% in parallel on a CRAY YMP machine. Execution speeds in excess of 1.5 Gigaflops were achieved.

CONCLUSIONS

A general solution to the long standing problem of enforcing excluded volume constraints in molecular modeling has been found. The solution has allowed the formulation of molecular modeling algorithms which are better adapted to the high physical densities typical of condensed phase systems. For example, large molecular assemblies with densities close to experimental and devoid of any van der Waals overlaps are easily put together using such algorithms. As a result more time can now be spent collecting data during a Monte Carlo or molecular dynamics simulation instead of achieving high densities or resolving molecular overlaps. Individual molecules in an assembly can have arbitrarily complex topologies. The native conformation of each molecule is left intact during the assembly steps. The algorithms also allow the solvation of the true solvent accessible surface of multifolded and porous molecular structures, and the automatic sampling of chemical interactions required for free energy calculations in molecular recognition problems.

The new solution and algorithms should stimulate and facilitate the study of condensed phase systems. Immediate areas of application include research on the binding selectivity of biopolymers in solution, phase separation behavior and physical properties of polymers and polymer blends, organic liquids, membranes, micelles, gels, lyotropic liquid crystals, cross-linked networks, amorphous glasses, and heterogeneous catalysts among others.*

My thanks to John P. Flory and Kris Tufto of Cray Research Inc., Pittsburgh office, for their valuable help in optimizing and porting Molecular Silverware to the CRAY YMP.

References

1. K.B. Lipkowitz, D.A. Demeter, R. Zegarra, R. Larter, and T. Darden, *J. Am. Chem. Soc.*, **110,** 3436 (1988).
2. K.B. Lipkowitz and R. Zegarra, *J. Comp. Chem.*, **10,** 595 (1989).
3. D.N. Theodorou and U.W. Suter, *Macromolecules*, **18,** 1206, (1985).
4. G. Arfken, *Mathematical Methods for Physicists*, Academic Press, New York, chap. 4, p. 178, 1970.
5. H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, MA, 1959.
6. R. Däppen, H.R. Karfunkel, and F.J.J. Leusen, *J. Comp. Chem.*, **11,** 181 (1990).
7. P.J. Flory, *Statistical Mechanics of Chain Molecules*, Interscience, New York, 1969.
8. H. Nomura and M. Ohba, *J. Phys. Chem.*, **93,** 8101 (1989).
9. L. Leherte, J.-M. Andre, D.P. Vercauteren, and E.G. Derouane, *J. Mol. Catal.*, **54,** 426 (1989). R. Vetrivel, C.R.A. Catlow, and E.A. Colbourn, *J. Phys. Chem.*, **93,** 4594 (1989).
10. A.T. Fong, H.I. Swanson, R.H. Dashwood, D.E. Williams, J.D. Hendricks, and G.S. Bailey, *Biochemical Pharmacology*, **39,** 19 (1990).
11. E. Clementi, Ed., *Modern Techniques in Computational Chemistry: MOTECC-89*, ch. 8, p 415, ESCOM Science Pub. B.V., Leiden, The Netherlands, 1989. See also M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids.* Clarendon Press, Oxford, ch 2, p. 55, 1989.

*Software inquiries are welcome by the author. Suggestions for collaborative work on potential application of Molecular Silverware can be directed to the author's address, or electronically to address rs0xmb@rohvm1.BITNET or at INTERNET address rs0xmb@sc.msc.edu