Molecular Self-Assemblies: Monte Carlo Predictions for the Structure of the One-Dimensional Translation Aggregate

Jerry Perlstein

Contribution from Copy Products Research and Technology Division, Eastman Kodak Company, Rochester, New York 14650-2021. *Received August 12, 1991*

Abstract: A Monte Carlo method has been developed which quantitatively predicts the structure of one-dimensional translation aggregates assembled from three-dimensional molecules. The molecules are treated as rigid objects interacting with the MM2 forcefield containing only a nonbonded and an electrostatic term. For 13 molecules randomly chosen from the Cambridge Structural Database, one-dimensional translation aggregates within three-dimensional crystal structures are quite predictable when the molecule–molecule energy anisotropy is 1.5/1 or more. These aggregate structures generally lie less than 2 kcal above the predicted global minimum with very little structural variation from either the global minimum or the observed minimum.

Introduction

Molecular self-assemblies are the building blocks for a variety of important aggregate structures in chemistry and physics. They occur naturally in colloidal dispersions, in micelles, as the chlorophyll dimer in photosynthesis, in liquid crystals, in Langmuir–Blodgett films, as layers on silver halide grains for photography, and in organic conductors, photoconductors, superconductors, and nonlinear optical materials. The detailed conformational structures of these assemblies are in many cases poorly understood. In most cases where such information is discernible either from X-ray data or from chemical intuition, the aggregate structure of the assembly is simply taken as a starting point for further analysis of chemical and physical characterization.

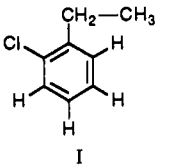
In some cases, however, the mere change of a single chemical bond can have a dramatic effect on structural conformation and consequent physical properties. Control of aggregate conformation has thus been a trial and error process in which the design of specific features has more often then not been thwarted because of a lack of information of the ultimate structure into which the molecules will assemble.

How molecules recognize one another to form assemblies is the subject of this and subsequent papers. We have initiated a program in predicting aggregate structure knowing only the molecular valence geometry of a single molecule. This work is based on the significant efforts of Scaringe[1,2] and Scaringe and Perez[3] (hereafter S&P) to predict the one- and two-dimensional structure of rigid organic molecules in crystalline environments. Their work, which is based on the pioneering efforts of Kitaigorodskii[4] to understand molecular crystal structures from a geometric packing point of view, has come a long way toward placing the theory of organized molecular assemblies on a firm foundation. In this and subsequent papers we will present methods which will allow the complete prediction of aggregate chain structures and layer structures including internal molecular torsion geometries with as few assumptions as possible.

Aggregate Chain Structures

Many types of aggregate structures are known. As a starting point we will consider only those periodic structures which have a single rigid molecule as a building unit excluding for the moment periodic structures built up from multimolecular units such as repeating dimers, trimers, etc. as well as micelles and certain liquid crystalline structures. One-dimensional aggregates made up of three-dimensional molecules are quite common. The methods to be presented here however are not limited to these simple structures but are readily extendible to polymer chain structures and to two-dimensional layer aggregates with flexible molecules or with multiple molecules in the building unit. As suggested by the work of Kitaigorodskii, there is a kind of aufbau principle at work in aggregate structural development in which single molecules come together to form one-dimensional aggregates, the aggregates coupling to form layers and then finally the layers interacting to form full three-dimensional crystals.

S&P have shown that there are four types of one-dimensional aggregate structures which occur most often (92% of the time).[3] These are the translation aggregate, the glide aggregate, the screw aggregate, and the inversion aggregate. We display these four types in Figure 1 using *o*-chloroethylbenzene (I) as a hypothetical example (the aggregate structure for this molecule is not known).



I

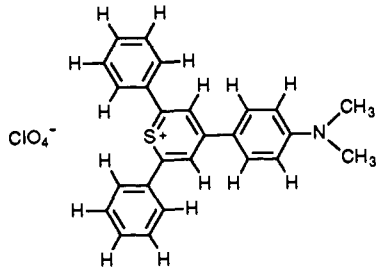
As implied by their names, each aggregate has a specific symmetry property. In the translation aggregate (Figure 1a), all the molecules are spatially identical with the repeat distance equal to the distance between equivalent atoms on neighboring molecules. For the glide aggregate every molecule is the mirror image of the one before it or the one after it so that the repeat distance is the distance between equivalent atoms on every other molecule as indicated in Figure 1b. Similarly, for the screw aggregate, every molecule is rotated 180° about the repeat axis (Figure 1c), whereas for the inversion aggregate, the most complex type, every molecule is related to its neighbor by inverting the coordinates of all atoms through an inversion center (Figure 1d). These aggregates also repeat every other molecule.

To construct an aggregate for a given molecule, pick a symmetry type from the four listed above and construct the aggregate accordingly. Details for the aggregate construction procedures have been worked out and presented by S&P. Not so obvious, however, is the conformation that the molecule should have within the aggregate. For a rigid molecule like *o*-chloroethylbenzene, the translation aggregate for example has about 75 000 possibilities. This number jumps to 55×10^6 for the glide and screw aggregates and to 1.5×10^9 for the inversion aggregate.

In principle, one could construct all possible aggregates for a given symmetry type, compute the energy of the molecular interactions using a suitable force field, and pick out the lowest energy one. Indeed S&P did this type of conformational analysis for the glide aggregate of the thiapyrylium salt (II) and discovered that for the repeat distance of 11.01 Å the global minimum aggregate for this repeat agreed quite well with the experimental aggregate derived from X-ray data. It should not be expected that the global minimum will always be the structure closest to

(1) Scaringe, R. P. Proceedings—45th Annual Meeting, Electron Microscopy Society of America; Baily, G. W., Ed.; San Francisco Press: San Francisco, 1987; p 472.
(2) Scaringe, R. P. In *Electron Crystallography of Organic Molecules*; Fryer, J. R., Dorset, D. L., Eds.; Kluwer: Dordrecht, 1990; p 85.
(3) Scaringe, R. P.; Perez, S. *J. Phys. Chem.* **1987**, *91*, 2394.
(4) Kitaigorodskii, *Organic Chemical Crystallography*; Consultants Bureau: New York, 1961. For a review see also: Timoseeva, T. V.; Chernikova, N. Yu.; Zorki, P. M. *Russ. Chem. Rev. (Transl.)* **1980**, *49*, 509.

0002-7863/92/1514-1955$03.00/0



II

the observed, and in fact we find that in most cases it is not, although the differences between the global minimum and the best structure are quite small.

The number of possible aggregate conformations is already quite large for the inversion aggregate so that conformational analysis of this symmetry type is problematic at best and becomes impractical if one adds internal bending angles and torsion angles to the analysis. Given S&P's results, however, it seemed reasonable that if a faster search for the global minimum and/or nearby local minima could be found, many more degrees of freedom could be added to the aggregate problem without substantial increase in computation time. Monte Carlo simulation offers this possibility, in that it allows a search for local minima without having to examine all high energy conformations most of which are irrelevant anyway. We will present here the details of the method we use as it applies to the simplest of all aggregate structures, the translation aggregate. Details for the other aggregate types and more complex layer structures with the inclusion of internal molecular degrees of freedom will be presented in future papers.

Construction of the Translation Aggregate

We use a variation of the technique of S&P to construct a translation aggregate of rigid molecules as follows (see Figure 2):

(a) Place the center of mass of the molecule at the origin of an orthogonal coordinate system oriented in some arbitrary way with respect to the unit vectors **i**, **j**, **k** of the coordinate axis x, y, z (Figure 2a). Add some dummy atoms at the origin and at the ends of the unit vectors. These atoms should be allowed to rotate with the molecule but in all other respects contribute nothing to the energy of the system. The dummy atoms define a second coordinate system with unit vectors **i'**, **j'**, **k'** which moves with the molecule (Figure 2b). The direction cosines **i'·k**, **j'·k**, **k'·k** along with the repeat distance will then define the final position state of the aggregate with respect to the original starting conformation.

(b) Rotate the molecule plus its dummy atoms by arbitrary angles θ_x and θ_y about the x and y axis (Figure 2b).

(c) Make a copy of the molecule superimposed on the original.

(d) Translate the copy an arbitrary distance, t, along the z axis. This is the aggregate repeat distance (Figure 2c).

(e) Make additional copies of the original at $-t$, $-2t$, and $+2t$ (Figure 2d).

This completes the construction of a single five-molecule translation aggregate. There are three parameters in this construction procedure. If the range of θ_x is 360° in increments of 5° then that for θ_y need only be 180° in order to cover all of the angle space (for the translation aggregate, rotation about z does not produce a new aggregate conformation, but merely rotates the whole aggregate about z). If we take the range of t to be 7–13 Å in increments of 0.2 Å, then the total number of possible conformations is $72 \times 36 \times 30 = 77\,760$. In a conformational analysis all of these conformations would have to be constructed before the global minimum could be located. The Monte Carlo method finds the global minimum without having to search the entire phase space.

The above construction procedure is taken from S&P and differs from it in that we treat the repeat distance t as a random variable whereas S&P use a construction procedure to compute t by moving the copy along the z axis until the molecules just touch. We found it advantageous in the Monte Carlo simulation to allow the molecules to approach minima from as many pathways as possible and so we have included t as a variable.

The Monte Carlo Method

Monte Carlo methods are random sampling techniques which have been used extensively to determine thermodynamic parameters of complex statistical mechanical systems for which analytical expressions are difficult or impossible to solve.[5] Our use of it here is to reduce the computation time necessary to locate the global energy minimum and to find some nearby local minima by effectively reducing the number of states that have to be examined in the phase space of the aggregate problem. We define the global energy minimum as the lowest energy state found by the Monte Carlo procedure. Within this context, the Monte Carlo algorithm for doing this is very basic and proceeds as follows:

(1) Construct an initial aggregate conformation as described above and compute its energy E_0 (see the next section for the energy computation).

(2) Generate random values for the two angles θ_x and θ_y between $-\theta_{max}$ and $+\theta_{max}$ and one translation, t, between z_{min} and z_{max} using a uniform random number generator. Here θ_{max}, z_{min}, and z_{max} are constraints set by the user (see below).

(3) Construct a new translation aggregate conformation starting from the initial conformation using these values for θ_x, θ_y, and t and compute the energy of this new aggregate E_1.

(4) If $E_1 < E_0$ then save the new structure as the starting point and go back to step 2.

(5) If $E_1 > E_0$ then do importance sampling using the Metropolis algorithm as follows: generate a random number R between 0 and 1 and compare it to the Boltzmann factor $W = \exp[-(E_1 - E_0)/kT]$ where k = the Boltzmann constant and T is the temperature—(a) if $W > R$ than save the new conformation as a starting point and go back to step 2; (b) if $W < R$ than save the old conformation and go back to step 2.

(6) Continue the above process until there is no further change in the energy.

Step 1 is the starting point for the simulation. Any starting point will do, even one that is very far away from any minima. We use a five molecule aggregate with the molecules spaced 15 Å apart as the starting configuration for all Monte Carlo simulations. The energy computation is described in the next section.

Step 2 generates the random rotations and translation for the construction of a new translation conformation. The maximum size of these variables is determined by $\pm\theta_{max}$ and z_{min}, z_{max} values. Typically, $\theta_{max} = 10°$ and $z_{max} - z_{min} = 6$ Å.

Step 3 does the actual construction of the new translation aggregate from the old translation aggregate using the construction technique described in the last section. We found the search to be more efficient if either the angles or the translation is kept constant during each Monte Carlo cycle. In this step therefore a fourth random number is generated to decide which variable to change.

Steps 4 and 5 are at the heart of the method. In step 4 the system heads downhill in energy toward the local minimum, but sometimes it is allowed to head uphill by the importance sampling technique in step 5. Thus the system tends toward a local minimum but is also able to move out of the local well to a new well by the importance sampling criteria. The rate at which it does this is governed by the temperature T. The choice of temperature thus becomes critical for the method to work. In practice if T is too low it could take a very long time for the system to seek out the lowest energy state because of the low probability of going over barriers and into a new local well. On the other hand if T is too high, the system will jump around a lot from well to well but will never have enough time to seek out the local minima before jumping into a new well. The local minima are then never found.

In order to get around this problem, several methods of cooling the sample have been developed by others[6-8] which satisfy the

(5) Binder, K. *Topics in Current Physics*; Springer-Verlag: New York, 1984; Vol. 46, p 1. For the original work on the Monte Carlo technique see: Metropolis, A. W.; Rosenbluth, M. N.; Teller, A. N.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

(6) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.

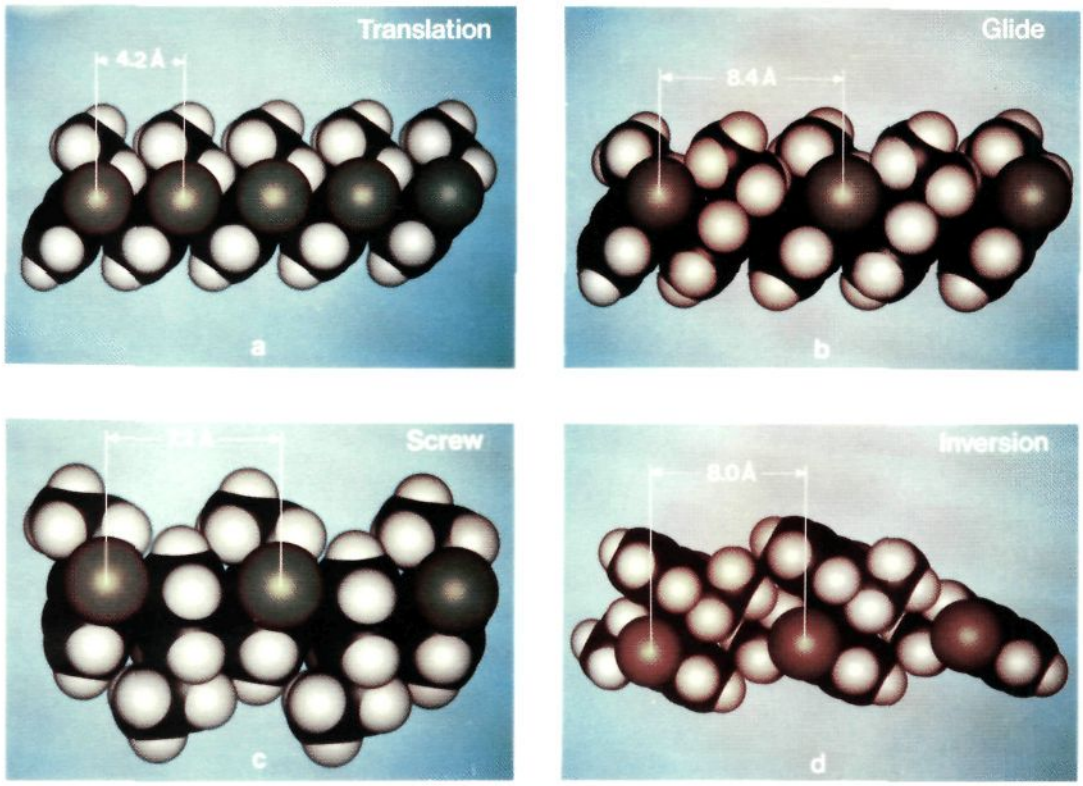


Figure 1. Demonstration of the four types of 1-d aggregates for *o*-chloroethylbenzene with repeat distance, t, lying horizontally: (a) translation type, t = 4.2 Å; (b) glide type, t = 8.4 Å, with the mirror in the plane of the paper; (c) screw type, t = 7.2 Å; and (d) inversion type, t = 8.0 Å. For the glide, screw, and inversion types the molecular mass centers generally do not lie on the repeat axis, but are offset from it. In addition for the inversion type, the molecular mass centers need not be equally spaced along the repeat axis at $t/2$. These are additional variables not present in the translation aggregate.

statistical thermodynamic requirements of reversibility. In principle these methods should find the thermodynamic energy of the system more quickly than with use of the Metropolis algorithm alone. In practice they also take considerable time to implement. Since we are not interested in thermodynamic averages here, we are free to use whatever method finds the global minimum as quickly as possible without concern for system reversibility. We have developed the following heuristic cooling schedule which meets the requirements of finding local wells and examining them in detail at reasonable sampling speed.

Cooling Schedule

The cooling schedule for the translation aggregate is as follows:

(1) Start the Monte Carlo simulation at T_0 = 4000 K using large values for θ_x and θ_y.

(2) After 20 cycles at T_0, cool the system to $T = 0.9T_0$. This becomes the new T_0.

(3) Continue the Monte Carlo simulation for another 20 cycles but with small values of θ_x and θ_y.

(4) Go back to step 2 until $T \leq 300$ K.

(5) When $T \leq 300$ K save the resulting structure and go back to step 1.

Step 1 allows the system to make a large jump to a new local well. In this step θ_{max} is set to a large value of 45°. The temperature has to be very high for this large jump in order for importance sampling to have a high probability of accepting the jump as a new state. Steps 2 through 4 examine this well in detail using importance sampling to find the local minimum. The number of cycles at each temperature is arbitrary, but typically we do at least 10 cycles for the angle variables and 10 cycles for the translation variable. We then check the "time" derivative of E for the last 10 cycles to make sure the system is heading downhill before changing the temperature. The size of the angle variables at this stage is kept small, usually less than 10°. This gives the simulation enough time to examine the local well in some detail at each T.

In step 5 the simulation is continued at room temperature as long as the energy continues to head downhill. The lowest energy conformation as a function of repeat distance is then saved. In practice it is possible that the lowest energy conformation would be found well above room temperature, so as a matter of course, we examine the low-energy conformations after each temperature cycle and save the lowest energy as a function of repeat distance before continuing the simulation at the next lower T.

Energy Computation

The binding energy of the aggregate is the energy necessary to separate all the molecules to infinity. The energy necessary to remove one molecule from the center of the aggregate is just twice this.[9] We thus only need to pick a test molecule in the aggregate and compute the energy of its interaction with all the others. S&P suggest that only the nearest neighbors need be included as this accounts for 85% of the total energy, the remaining 15% will not effect the structure. We find this to be usually true except in cases where the ionic contribution to the energy is sufficiently large that next nearest neighbors need to be included. As a matter of practice we thus include nearest and next nearest neighbors in all aggregate computations (hence a 5 molecule aggregate).

For rigid molecules the molecular interaction energy consists of two terms, (a) a nonbonding van der Waals contribution E^{nb} and (b) an electrostatic term E^{el}. We have specifically excluded a hydrogen bonding term at this stage of the problem.

A. Nonbonding Term. We have tried a number of nonbonding terms and have settled on using that formulated by Allinger[10] for the MM2 force field as it appears to give the most consistent

(7) Szu, H.; Hartley, R. *Phys. Lett. A* **1987**, *122*, 157.
(8) Matsuba, I. *Phys. Rev. A* **1989**, *39*, 2635.
(9) Bushing, W. R. *J. Phys. Chem. Solids* **1978**, *39*, 691.
(10) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.

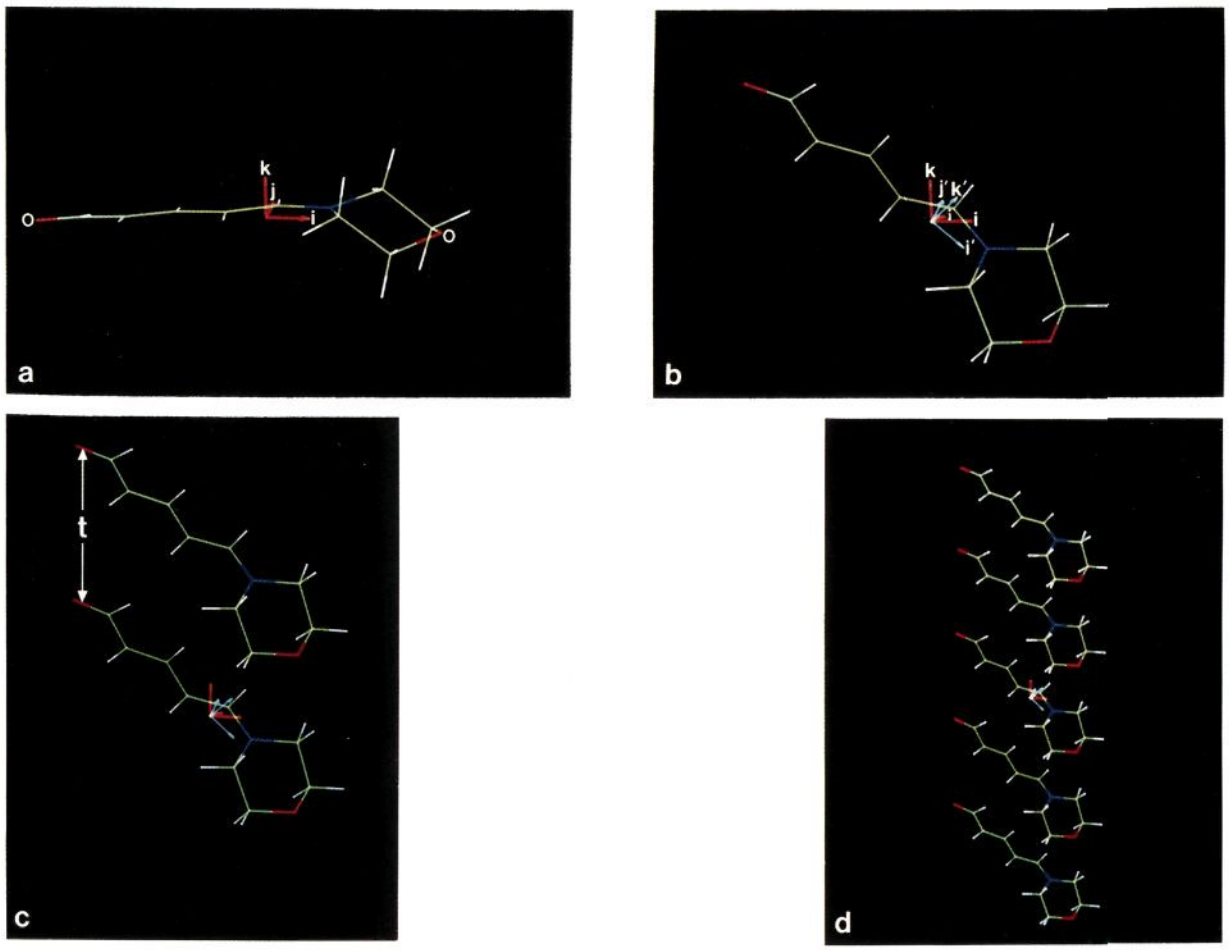


Figure 2. Procedure for constructing a translation aggregate of MORPCY (see Figure 3): (a) molecular mass center placed in an arbitrary orientation at the origin of an orthogonal coordinate system with unit vectors, **i**, **j**, **k**; (b) molecule plus dummy atoms defining unit vectors **i'**, **j'**, **k'** is rotated by a random amount about the x and y axis (a duplicate copy is then made (not shown) in the new orientation); (c) the copy is translated along z a random distance, t; and (d) additional copies are translated along z a distance $-t$, $-2t$, and $+2t$, completing a 5 molecule aggregate.

results for a variety of structural types. There are some problems with it and these are being addressed in MM3.[11] The nonbonded energy between any two atoms i and j separated a distance r_{ij} has the form

$$E_{ij}^{nb} = A\left[(2.90 \times 10^5)\exp\left(\frac{-12.50B}{r_{ij}}\right) - 2.25\left(\frac{B}{r_{ij}}\right)^6\right] \quad (1)$$

where the parameters A and B are given by

$$A = (A_iA_j)^{1/2} \quad (2)$$

$$B = B_i + B_j \quad (3)$$

The A and B parameters for the atoms used in this study were taken from the molecular modeling program MACROMODEL[12] which uses the MM2 force field and are listed in the Appendix. For the case of a bond to hydrogen, MM2 first shrinks the bond to hydrogen by 8.5% before computing r_{ij}. The nonbonded contribution to the total energy is then the sum of eq 1 over all atom–atom interactions between one molecule and its neighbors.

B. Electrostatic Term. The electrostatic interaction between molecules is difficult to compute accurately since the atomic charges due to bond polarization are not well-known nor well-understood. Any computation will at best be crude. The correct method would be to do an Ewald lattice summation, but this is a lengthy computation and is not justified considering approximations in the atomic charge densities that have to be made anyway. We simplify the procedure by using empirical charges of Gasteiger[13] at each atomic site and then compute the atom–atom electrostatic potential with a linear distant dependent dielectric constant. The energy has the form

$$E_{ij}^{el} = \frac{q_iq_j}{\epsilon r_{ij}} \quad (4)$$

where q_i and q_j are the partial Gasteiger charges on atomic sites i and j and separated a distance r and

$$\epsilon = \epsilon_0 r_{ij} \quad (5)$$

with dielectric constant $\epsilon_0 = 1.0$.

For neutral molecules, this contribution is usually quite small compared to the nonbonded term. For ionic systems, however, it can grow to be the major contribution to the energy although it is not necessarily structure determining even when large.[14] The above heuristic form for the electrostatic term seems to be satisfactory for the present problem.

Experimental Section

A. Initial Translation Aggregate Construction. The coordinates for the single rigid molecule were obtained from the Cambridge Structural

(11) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(12) Still, W. C.; Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Lipton, W.; Liskamp, R.; Chang, G.; Hendrickson, T.; DeGunst, F.; Hasel, W.; Macromodel V2.5, Department of Chemistry, Columbia University, New York, NY 10027.

(13) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.

(14) For a contrary point of view concerning the importance of the electrostatic term see: Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525.

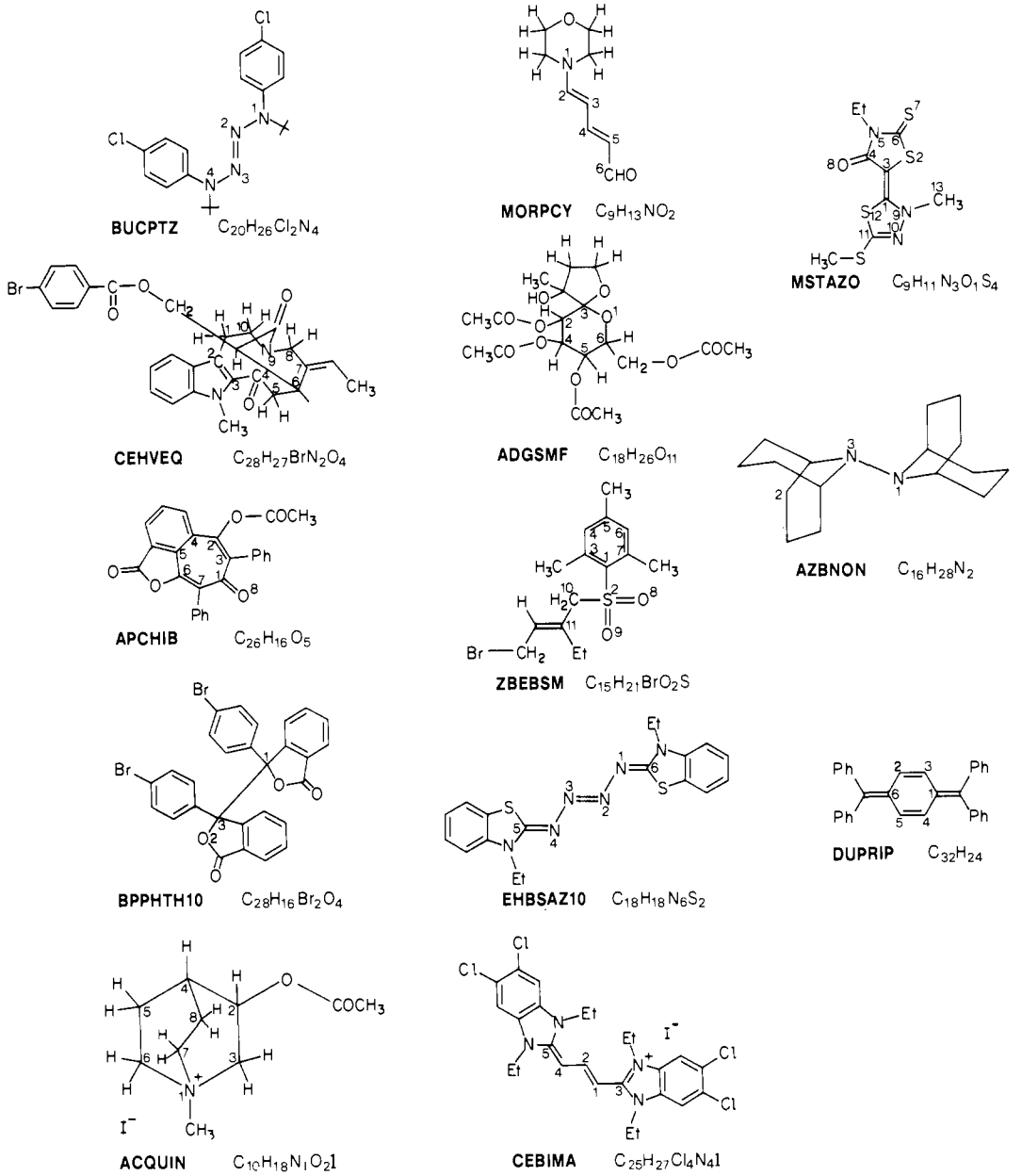


Figure 3. Molecular structures from the Cambridge Crystallographic Database identified by reference code and chemical formula. Numbered atoms were used to compute the centroid of the orthogonal coordinate system. For BPPHTH10 and AZBNON the centroid was determined by atom numbers 1 and 3 only.

Database[15] and transferred to CHEM-X,[16] a molecular modeling package. Where necessary, hydrogen atoms were added at a distance of 1.08 Å to complete the valence structure. The molecule was then centered at the origin of an orthogonal coordinate system with unit vectors **i**, **j**, **k** along the x, y, and z axis. Dummy atoms were added at the molecular centroid and at the end of the unit vectors **i**, **j**, **k** to complete a molecular coordinate system with unit vectors **i'**, **j'**, and **k'** which rotates with the molecule. An initial conformation consisting of five molecules was then constructed. The entire construction procedure was automated in a FORTRAN subroutine which could be linked directly to the CHEM-X package using CHEM-X's CHEMLIB interface. This initial construction was done on a VAX8600. After computation of the empirical Gasteiger charges, the final coordinates were then transferred to a Stardent STELLAR GS1000 workstation to do the Monte Carlo simulation.

B. Uniform Random Number Generator. Uniform random numbers r between 0 and 1 were generated using the IMSL[17] routine GGUBS, and these were mapped into the range $\pm\theta_{max}$ and z_{min}, z_{max} to give random angles or distances (RAN) using the relation

$$\mathrm{RAN} = br + a(1 - r) \tag{6}$$

where $a = -\theta_{max}$, $b = +\theta_{max}$ for the angles or $a = z_{min}$, $b = z_{max}$ for the translation.

C. Rotation Variables. Positive rotation of a point about an arbitrary vector is carried out by using the generalized rotation matrix **R**[18]

$$\mathbf{R} = \begin{bmatrix} \cos\alpha+n_1^2(1-\cos\alpha) & n_1n_2(1-\cos\alpha)-n_3\sin\alpha & n_1n_3(1-\cos\alpha)+n_2\sin\alpha \\ n_1n_2(1-\cos\alpha)+n_3\sin\alpha & \cos\alpha+n_2^2(1-\cos\alpha) & n_2n_3(1-\cos\alpha)-n_1\sin\alpha \\ n_3n_1(1-\cos\alpha)-n_2\sin\alpha & n_3n_2(1-\cos\alpha)+n_1\sin\alpha & \cos\alpha+n_3^2(1-\cos\alpha) \end{bmatrix} \tag{7}$$

(15) Allen, F. H.; Kennard, O.; Taylor, R. *Acc. Chem. Res.* **1983**, *16*, 146. The Cambridge Crystallographic Database contains coordinates derived from X-ray and neutron diffraction studies on organic and inorganic molecules and is available from the Cambridge Crystallographic Data Centre, Lensfield Road, Cambridge CB2 1EW, U.K.

(16) Chem-X is a molecular modeling program developed and distributed by Chemical Design Ltd., 7 Westway, Oxford OX2 OJB, U.K.

(17) IMSL Problem Solving Software Systems, 2500 City West Blvd., Houston, Texas 77042-3020, is a collection of mathematical routines available for most mainframe computers and UNIX based workstations.

(18) Jeffreys, H.; Jeffreys, B. S. *Methods of Mathematical Physics*; Cambridge, 1972; p 122.

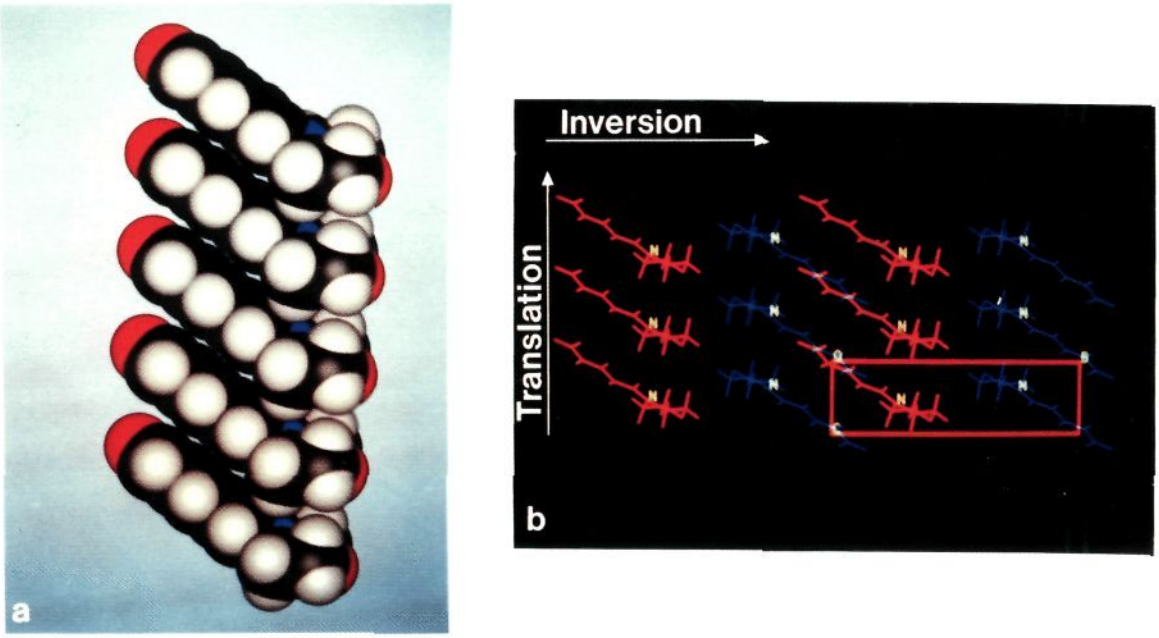


Figure 4. Monte Carlo prediction for the structure of MORPCY: (a) the global minimum prediction (Table I, structure **5**); (b) the X-ray structure projected down the *a* axis. The Monte Carlo prediction is equivalent to the translation aggregate (red) running along the *c* axis. There are also higher energy inversion aggregates (alternating red, blue) along the *b* axis as well as diagonally through the *bc* plane.

where n_1, n_2, n_3 are the direction cosines of the vector with respect to the x, y, and z axis and α is the angle of rotation. The rotation is clockwise as viewed along the vector from the origin. For the translation aggregate with rotation first about x by angle θ_x and then about y by angle θ_y, **R** simplifies to $\mathbf{R}_y\mathbf{R}_x$ where

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \tag{8}$$

$$\mathbf{R}_y = \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \tag{9}$$

$$\mathbf{R}_y\mathbf{R}_x = \begin{bmatrix} \cos\theta_y & \sin\theta_x \sin\theta_y & \cos\theta_x \sin\theta_y \\ 0 & \cos\theta_x & -\sin\theta_x \\ -\sin\theta_y & \sin\theta_x \cos\theta_y & \cos\theta_x \cos\theta_y \end{bmatrix} \tag{10}$$

The values of θ_x and θ_y derived from the uniform random number generator (eq 6) were inserted into (10) and used to compute the rotation of a molecule with initial coordinates (x_i, y_i, z_i) for all atoms i to final coordinates (x'_i, y'_i, z'_i) since

$$\begin{bmatrix} x'_i \\ y'_i \\ z'_i \end{bmatrix} = \mathbf{R}_y\mathbf{R}_x \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \tag{11}$$

D. Monte Carlo Simulation. The code for the simulation was written in FORTRAN 77 on a Stardent STELLAR GS1000 workstation under the UNIX operating system. Specific pieces of the code were vectorized for enhanced speed. These included subroutines for the rotation of the molecule and the computation of the energy which took up approximately 50% of the CPU time. A single Monte Carlo cycle which includes construction of a new conformation and computation of its energy took 0.4 s of CPU time. The entire simulation was usually run from 15 to 24 h, which was well beyond the time needed to find the translation global minimum plus nearby local minima.

E. Structure Visualization. Final coordinates for global minimum and nearby local minima were saved and transferred back to the VAX 8600 for viewing using CHEM-X with and Evans and Sutherland 3-D graphics terminal. Visual comparison could then be made between the various minima and the experimental X-ray structures. The visualization technique was also extremely helpful in the code development for the Monte Carlo simulation. Irregularities in the structure were easily traceable to code errors which could then be corrected and retested.

F. Comparisons with X-ray Data. 1. Energy Comparisons. The experimental aggregate conformation was extracted from the X-ray crystal structures in the following way. Coordinates for the asymmetric molecular unit were extracted from the Cambridge Crystallographic Database and loaded into CHEMX. The structure was then packed using the space group for the structure. Typically a 3 × 3 × 3 unit cell translation was sufficient to give enough molecules in all directions to compute the crystal energy. After computing the Gasteiger atomic charges, the crystal binding energy was computed using the same MM2 energy function used in the Monte Carlo simulation by selecting a test molecule in the center of the structure and computing the contribution to the total energy of its interaction with every other molecule in the structure. A single aggregate could be extracted by finding the two lowest energy neighbors of the test molecule. The three molecules then form an aggregate which could be expanded to five molecules by including the next nearest neighbors along the translation axis. The aggregate energy was then computed for comparison with the Monte Carlo results.

Table I. Monte Carlo Simulation Results (kcal) for MORPCY[a]

structure	repeat, Å	nonbonded	electrostatic	total
1	3.47	−11.23	−1.92	−13.15
2	3.69	−8.91	1.06	−7.85
3	3.90	−14.73	0.79	−13.94
4	4.10	−15.80	0.57	−15.23
5	4.12	−15.78	0.53	−15.24
6	4.30	−15.02	0.35	−14.68
7	4.50	−13.86	0.19	−13.68
8	4.70	−12.32	−0.74	−13.06
9	4.90	−11.49	−0.78	−12.27

[a] Acceptance ratio 0.524.

Care was taken with ionic lattices to make sure that the anion in the original asymmetric unit was the closet one to the cation. If it was not, the structure was repacked starting with the closest cation–anion pair in the assymetric unit.

2. Geometry Comparisons. For comparison with the X-ray structure it is only necessary to compute the direction cosines of the vectors defined by the dummy atoms at unit vector positions **i**′, **j**′, **k**′ with the z axis unit vector **k**, of the aggregate and compare this with the same direction cosines from the X-ray structure. These along with the repeat distance make up the total quantitative comparison of the Monte Carlo predictions with the observed values. If desired, the angles θ_x and θ_y needed to rotate the starting molecule to the finished orientation can be computed by inserting the three direction cosines into the 3rd row of the rotation matrix $\mathbf{R}_y\mathbf{R}_x$ (see the Appendix). Two sets of angles will result which are related by 180° about the z axis. Since the translation aggregate is invariant with respect to rotation about the z axis, both sets of angles will generate the same aggregate structure.

Table II. Tilt Angles and Deviations Observed for MORPCY

structure	t^a	$\theta_{i'\cdot k}{}^b$	$\theta_{j'\cdot k}$	$\theta_{k'\cdot k}$	deviations $\Delta(t)$	$\Delta(\theta_{i'\cdot k})$	$\Delta(\theta_{j'\cdot k})$	$\Delta(\theta_{k'\cdot k})$	RSS
1	3.47	71.88	89.54	18.14	−1.04	9.23	−4.54	−9.57	14.08
2	3.69	68.13	91.53	21.94	−0.82	5.48	−2.55	−5.77	8.40
3	3.90	67.24	90.04	22.77	−0.61	4.59	−4.04	−4.94	7.88
4	4.10	66.07	92.02	24.03	−0.41	3.42	−2.06	−3.68	5.44
5	4.12	65.52	91.73	24.55	−0.39	2.87	−2.35	−3.16	4.89
6	4.30	62.58	91.08	27.44	−0.21	−0.07	−3.00	−0.27	3.02
7	4.50	59.00	89.07	31.02	−0.01	−3.65	−5.01	3.31	7.03
8	4.70	59.53	65.45	40.97	0.19	−3.12	−28.63	13.26	31.71
9	4.90	57.09	63.88	44.39	0.39	−5.56	−30.20	16.68	34.95
obsd	4.51	62.65	94.08	27.71					

[a] Repeat distance in angstroms. [b] Angles in degrees.

Table III. Tilt Angles with Respect to the Translation Axis, Repeat Distance, and Deviations from the Observed X-ray Aggregate

	Monte Carlo prediction				deviation from observed[c]					
	tilt angles[a]			repeat						
ref code	$\theta_{i'\cdot k}$	$\theta_{j'\cdot k}$	$\theta_{k'\cdot k}$	t^b	$\Delta(\theta_{i'\cdot k})$	$\Delta(\theta_{j'\cdot k})$	$\Delta(\theta_{k'\cdot k})$	$\Delta(t)$	RSS[d]	$\Delta(E)$[e]
BUCPTZ	80.54	59.67	147.93	6.71	−0.17	−0.17	−0.22	−0.06	0.33	0.80
CEHVEQ	124.39	35.02	83.96	7.51	0.26	0.14	0.52	0.22	0.63	1.87
APCHIB	57.28	108.26	38.68	5.00	0.31	0.72	0.15	−0.06	0.80	0.12
BPPHTH10	161.13	98.51	73.29	7.85	−1.60	0.66	1.43	−0.23	2.26	1.93
ACQUIN	99.57	26.58	65.45	6.65	−1.80	−1.86	1.14	−0.96	2.99	3.16
MORPCY	62.58	91.08	27.44	4.30	−0.07	−3.00	−0.27	−0.21	3.02	0.56
ADGSMF	62.12	113.43	37.88	5.80	−1.04	1.82	2.24	0.23	3.07	0.83
ZBEBSM	61.89	105.42	147.23	4.60	−2.26	0.86	−2.55	−0.21	3.52	1.80
EHBSAZ10	25.94	71.01	107.00	12.93	−4.76	3.21	−3.09	1.76	6.75	14.72
CEBIMA	20.89	94.98	110.23	9.09	−0.11	7.97	−0.54	−0.19	7.99	0.23
MSTAZO	46.51	97.03	44.35	5.10	−4.32	−8.22	1.18	0.30	9.37	0.84
AZBNON	73.43	98.61	18.79	6.91	−17.74	−1.33	8.78	0.40	19.84	3.56
DUPRIP	100.42	91.05	169.53	5.00	−12.83	−17.21	19.79	−0.96	29.22	0.00

[a] Angles in degrees. [b] Repeat distance in angstroms. [c] Deviations = computed value − observed value. [d] RSS is the square root of the sum of the squares of the deviations. [e] $\Delta(E)$ is how far in kcal the best structure lies above the global minimum.

Results

Small-molecule crystal structures containing translation aggregates were chosen at random from the Cambridge Structural Database with the following constraints:

(1) No hydrogen bonds or potential hydrogen bonds were chosen.

(2) No metal atoms—or any atoms for which we have no force field parameters were used.

(3) If the molecule was a salt, we followed S&P's suggestion of treating it as an ion-pair. The position of the cation relative to the anion was left fixed and the entire molecule was treated as if it were neutral with a charge distribution consistent with positive–negative charge separation on the ions.

(4) The relative atomic coordinates within the molecules were held fixed—viz. no rotations or bond bending were allowed during the Monte Carlo simulation of the aggregate. This restriction can be easily removed and will be considered in subsequent papers when we discuss layer aggregates.

Figure 3 shows 13 structures we have looked at to date.[19] They are identified by their Cambridge Structural Database reference code. They all form translation aggregates, many of them within space groups that can have only translation symmetry such as P1, or P-1 when the molecule sits on an inversion center, but in some cases the translation aggregates occur in other higher symmetry space groups. Some of the structures are neutral, and some are ionic. Some are small in size, and others are structurally very complex with the number of atoms/molecule varying from 25 to 62. The Monte Carlo simulation was run for each molecule, and the energy vs repeat distance was tabulated for some of the local minima.

Part of a typical output is shown in Table I for MORPCY. The total number of Monte Carlo cycles was typically several hundred thousand. This represents the number of times an attempt was made to construct a new aggregate from an old aggregate using the random variables. If an attempt was rejected by importance sampling it was because the new aggregate had an energy that was too high. An acceptance ratio of 50% is considered good. For MORPCY the acceptance ratio was 52.4%. Local minima at repeat distances between 3.47 and 8.91 Å were found by the simulation with the global minimum at −15.24 kcal and a repeat distance of 4.12 Å (structure **5**). Within 4 kT of this minimum are 4 other structures with repeat distances ranging from 3.90 to 4.50 Å. The table also shows that the coulomb contribution to the total energy is quite small, less than 3.5% for the global minimum structure. We find this to be generally true for nonionic structures, but for ionic structures the electrostatic term can sometimes be the largest term.

Figure 4 compares the Monte Carlo global minimum aggregate with the translation aggregate extracted from the X-ray data. The tilt of the molecules with respect to the z axis is quite evident. The direction cosines of the dummy atom vectors with respect to z are a measure of this tilt. If the average plane through the atoms of MORPCY was perpendicular to the z axis, then the direction cosines $\mathbf{i}'\cdot\mathbf{k}$ and $\mathbf{j}'\cdot\mathbf{k}$ would be 0 corresponding to 90° angles, viz., no tilt with respect to z. The deviation of these angles from 90° is thus a measure of the tilt angle. Table II shows the angles for $\mathbf{i}'\cdot\mathbf{k}$, $\mathbf{j}'\cdot\mathbf{k}$, and $\mathbf{k}'\cdot\mathbf{k}$ for the MORPCY structures in Table I and the

(19) X-ray structures for the listed reference codes are as follows: (a) BUCPTZ: Nelsen, S. F.; Landis, R. T.; Calabrese, J. C. *J. Org. Chem.* **1977**, *42*, 4192. (b) CEHVEQ: Massiot, G.; Lavaud, C.; Vercauteren, J.; Le Men-Olivier, L.; Levy, J.; Guilham, J.; Pascard, C. *Helv. Chim. Acta* **1983**, *66*, 2414. (c) APCHIB: Freer, A. A.; Gilmore, C. J.; Mont, D. M.; McCormick, J. *J. Chem. Soc., Chem. Commun.* **1977**, *296*. (d) BPPHTH10: Kalyani, V.; Manohar, H.; Mani, N. V. *Acta Crystallogr.* **1967**, *23*, 272. (e) ACQUIN: Baker, R. W.; Pauli, P. J. *J. Chem. Soc., Perkin Trans. 2* **1972**, *2340*. (f) MORPCY: Kulpe, S.; Schulz, B. *Krist. Tech.* **1976**, *11*, 707. (g) ADGSMF: Remy, G.; Cottier, L.; Descotes, G.; Foure, R.; Loiseleur, H.; Thomas-David, G. *Acta Crystallogr., Sect. B* **1980**, *36*, 873. (h) ZBEBSM: Rerat, B.; Rerat, C.; Julia, M.; Deprez, D. *J. Chim. Phys. Phys.-Chim. Biol.* **1979**, *76*, 253. (i) EHBSAZ10: Allmann, R. *Acta Crystallogr.* **1967**, *22*, 46. (j) CEBIMA: Smith, D. L.; Luss, H. R. *Acta Crystallogr., Sect. B* **1972**, *28*, 2793. (k) MSTAZO: Germain, G.; Paternatte, C.; Piret, P.; van Meerssche, M. J. *Chim. Phys. Phys.-Chim. Biol.* **1964**, *61*, 1059. (l) AZBNON: Nelsen, S. F.; Hollinsed, W. C.; Kessel, C. R.; Calabrese, J. C. *J. Am. Chem. Soc.* **1978**, *100*, 7876. (m) DUPRIP: Montgomery, L. K.; Huffman, J. C.; Jurczak, E. A.; Grendze, M. P. *J. Am. Chem. Soc.* **1986**, *108*, 6004.

Table IV. One-Dimensional/Three-Dimensional Energy Anisotropy (%)

	$100(E_{1d}/E_{3d})$		$100(E_{1d}/E_{3d})$
BUCPTZ	34.57	ZBEBSM	41.97
CEHVEQ	32.00	EHBSAZ10	24.99
APCHIB	44.00	CEBIMA	35.64
BPPHTH10	26.22	MSTAZO	31.48
ACQUIN	30.83	AZBNON	23.58
MORPCY	31.29	DUPRIP	29.70
ADGSMF	38.07		

deviation of these angles along with the repeat distance from the observed X-ray data. As a measure of the closeness to the observed data we also show the square root of the sum of the squares of these deviations (the root sum square, RSS).

The closest structure to that observed is structure **6** which is 0.56 kcal above the global minimum structure **5**. The difference between structure **5** and structure **6** is, nevertheless, rather small. We should point out that the above table represents only a small number of the local minima. There are more minima within the rotation angle space which we have not considered here. They differ little from the ones shown above but are easily accessible by the Monte Carlo method.

Table III shows the best Monte Carlo predictions for the 13 translation aggregates of Figure 3. The best structure was taken to be the one with the smallest root sum square deviation (RSS) of the angles and repeat distance. We report the direction cosine angles and repeat distance (columns 2–5) along with the deviations from the observed X-ray structure (columns 6–9) as well as the RSS (column 10) and how high in energy above the global minimum this best structure appears (column 11).

In most cases the best structure lies less than 2 kcal above the global minimum, and this structure has very little deviation from the observed as indicated by the RSS values. There are four poor structures either having a large RSS or lying significantly above the global minimum. ACQUIN, EHBSAZ10, and AZBNON all lie more than 2 kcal above the global minimum whereas DUPRIP's best structure is at the global minimum but deviates considerably from the observed. Visual comparison of the remaining structures with the global minimum shows very little difference. The agreement between the calculated and observed values is thus quite remarkable.

Discussion

It seems surprising that it is possible to predict the structure of an isolated translation aggregate which is buried in a three-dimensional crystal structure. The surrounding crystal field should significantly distort the aggregate structure, but the above results appear to indicate otherwise.

It is of interest to determine how much the one-dimensional aggregate energy E_{1d} contributes to the total three-dimensional binding energy E_{3d} of the crystal. In Table IV we have computed the ratio E_{1d}/E_{3d} for the above structures.

Most of the aggregates for which the Monte Carlo prediction agree with the observed have an anisotropy ratio of at least 30%. For an isotropic structure this ratio would be closer to 17%. But there are obvious exceptions. For BPPHTH10 with a ratio of 26.22% the agreement with observed is quit good, but for DUPRIP for which the ratio is 29% the agreement with observed is rather poor. Thus on energetic grounds the predicted aggregates have a considerable increase in their 1-d energies over what might be expected for an isotropic environment, but there is no clear break at which this occurs. The bulk of the crystal energy, 70%, is still tied up in two- and three-dimensional interactions.

Perhaps a better way to view this ability for molecules to form 1-d aggregates is to define a molecule–molecule anisotropy ratio for the interaction potential as follows: let $E_{onchain}$ be the largest energy of interaction of a single molecule with a neighbor in the aggregate and $E_{offchain}$ be the largest energy of interaction of this same molecule with a neighbor that is not in the chain. Then define the anisotropy ratio as $E_{onchain}/E_{offchain}$.[20] For a molecule which has isotropic interactions this ratio is 1.0. Values larger than 1.00 are then a measure of the one-dimensional molecular anisotropy. In Table V we have computed the molecular anisotropy for the 13 aggregates.

Table V. One-Dimensional Molecular Anisotropy

	$E_{onchain}/E_{offchain}$		$E_{onchain}/E_{offchain}$
BUCPTZ	1.60	ZBEBSM	2.67
CEHVEQ	2.30	EHBSAZ10	1.16
APCHIB	2.96	CEBIMA	1.57
BPPHTH10	1.53	MSTAZO	2.94
ACQUIN	1.06	AZBNON	1.10
MORPCY	1.51	DUPRIP	1.39
ADGSMF	2.46		

From the table, we can see that the least anisotropic structures are ACQUIN(1.06), EHBSAZ10(1.16), AZBNON(1.10), and DUPRIP(1.39) the same four whose Monte Carlo predictions deviate significantly from the observed. For the other molecules, this ratio is greater than 1.5.

Summary and Conclusions

A Monte Carlo technique has been developed which can predict the structure of one-dimensional molecular self-assemblies of rigid molecules in the solid state. A random sampling of crystal structures both ionic and nonionic containing one-dimensional translation aggregates indicates that these aggregates are fully predictable as isolated structures usually lying less than 2 kcal above the global minimum. Aside from choosing an appropriate force field, the only assumptions made are the following: (a) the molecule is rigid and (b) a symmetry type for the assembly. Assumption (a) could be relaxed by including torsion terms in the force field. Our preliminary studies indicate that at least for one-dimensional aggregates a very broad range of local minima occur indicative of many low-energy structures. We will reconsider inclusion of torsion terms when we examine the Monte Carlo predictions for two-dimensional layer structures. Assumption (b) could be relaxed by doing the Monte Carlo simulation in each of the four symmetry types and then picking out the one with the lowest energy. What one finds in fact is that the four symmetry aggregates differ in energy by only 1–2 kcal. Which particular type will occur then depends on the energetics of the interaction with the surroundings. CEBIMA for example is known to form three of the four symmetry types depending on the solvent used to crystallize it.

The Monte Carlo method we have used here clearly does not depend on the nature of the force field. Nevertheless, the quality of the MM2 force field leaves something to be desired. For example, it does a poor job of predicting heats of sublimation of alkanes[21] as well as the energy of the benzene dimer.[22] What effect does this have on predicting geometric molecular packing? Clearly not much; as we have shown here MM2 is perfectly adequate for predicting the geometric orientation of molecules in an extended one-dimensional lattice. Indeed other force fields, such as those of Momany et al.[23] and Williams and Starr[24] can and have been used for this purpose[1] as well as for the geometry of two-dimensional layer structures.[2]

The molecule–molecule energy anisotropy is typically 1.5 or more for those aggregates which are predictable, but this correlation is far from complete. Recent studies on the computer simulation of liquid crystalline structure show that molecular shape alone can account for the orientational ordering that occurs in these mesophases but not the specific details.[25] Various QSAR's based on molecular shape for the packing energy problem also have been formulated.[26] However, the a priori connection between

(20) The author thanks Yitzhak Shnidman for this suggestion.
(21) Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576.
(22) Pettersson, I.; Liljefors, T. *J. Comp. Chem.* **1987**, *8*, 1139.
(23) Momany, F. A.; Carruthers, L. M.; McGuire, R. F.; Scheraga, H. A. *J. Phys. Chem.* **1974**, *78*, 1595.
(24) Williams, D. E.; Starr, T. L. *Comput. Chem.* **1977**, *1*, 173.
(25) Allen, M. P.; Wilson, M. R. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 335.

Table VI. Coordinates for Atoms 1 and 2 in the Starting Geometries

structure	A1 *x, y, z*	A2 *x, y, z*
BUCPTZ	(1.72, 0.11, 0.00)	(0.45, −0.43, 0.00)
CEHVEQ	(2.12, 0.93, −0.45)	(−0.07, 1.66, 0.66)
APCHIB	(1.30, 0.06, −0.05)	(−0.81, 1.54, 0.01)
BPPHTH10	(0.79, 0.00, 0.00)	(−1.22, 1.22, 0.68)
ACQUIN	(1.27, 0.11, 0.00)	(−0.83, 0.67, 1.22)
MORPCY	(1.45, 0.28, 0.18)	(0.16, 0.63, 0.07)
ADGSMF	(1.40, −0.07, 0.11)	(−0.60, 1.30, 0.22)
ZBEBSM	(0.40, 0.08, 0.22)	(0.40, 0.08, 0.22)
EHBSAZ10	(1.65, −0.50, 0.00)	(0.53, 0.34, 0.01)
CEBIMA	(1.21, −0.35, −0.02)	(0.01, 0.33, 0.03)
MSTAZO	(1.00, −0.07, −0.01)	(−1.37, 1.38, 0.00)
AZBNON	(0.75, 0.00, 0.00)	(−0.71, 0.97, 2.32)
DUPRIP	(1.46, 0.00, 0.00)	(−0.67, 1.22, 0.00)

Table VII. Nonbonded Energy Parameters[a]

atom	A_i, kcal	B_i, Å	atom	A_i, kcal	B_i, Å
C(sp3)[b]	0.044	1.90	N	0.055	1.82
C(sp2)	0.0440	1.90	Cl	0.240	2.03
H	0.047	1.50	Br	0.320	2.18
O(sp3)	0.050	1.73	I	0.424	2.32
O(sp2)	0.066	1.74	S	0.202	2.11

[a] See eqs 1–3 for definitions and use. [b] For C–H bond $A_{CH} = 0.046$, $B_{CH} = 3.34$.

the molecular geometry and the various packing modes that a molecule can have still remains to be elucidated. Further work on the glide, screw, and inversion chains is in progress and will be presented shortly.

Acknowledgment. The author would like to thank Ray Scaringe for many discussions on the energetics of the crystal packing problem. The author is also very grateful to Alan Cameron, Jim Eilers, John Hamilton, John Mckelvey, Dennis Perchak, Yitzhak Shnidman, and Ian Watson from the Computational Science Laboratory of Eastman Kodak Company for their help, numerous discussions, and particularly their patience associated with the many frustrating hours required for the computer hardware and software implementation of the Monte Carlo method.

Appendix

Starting Geometries. For each of the molecules used in this study the starting geometries can be constructed in CHEM-X using the information provided in Figure 3 and Table VI as follows:

Obtain from the Cambridge Structural Database the coordinates relative to the unit cell using the reference code given in Figure 3. Transform the coordinates to an orthogonal coordinate system with unit vectors and construct the centroid as origin using the atoms numbered in Figure 3. In order to orient the structure properly, use CHEM-X's vector facility to construct the following four vectors:

A1 the vector from the centroid to atom 1

A2 the vector from the centroid to atom 2

A1* the vector from the centroid to the A1 coordinates in Table VI

A2* the vector from the centroid to the A2 coordinates in Table VII

The two vector pairs (**A1**, **A2**) and (**A1***, **A2***) define two planes with normals **N** = **A1** × **A2** and **N*** = **A1*** × **A2***. Rotate the structure so that **N** is superimposed on **N*** and then rotate the structure again by the angle given by the dot product **A1**·**A1***. This superimposes the two vector pairs. The resulting orientation will then be the starting configuration used in the Monte Carlo computations.

Final Geometries. The final Monte Carlo geometries can be obtained by using the information given in Table III for the direction cosine angles and repeat distance. The terms in the last row of the $\mathbf{R}_y\mathbf{R}_x$ rotation matrix are equivalent to the direction cosines. For MORPCY for example,

$$-\sin\theta_y = \cos(62.58)$$

$$\sin\theta_x \cos\theta_y = \cos(91.08)$$

$$\cos\theta_x \cos\theta_y = \cos(27.44)$$

which yield two solutions

$$\theta_x = -1.21,\ \theta_y = -27.42$$

and

$$\theta_x = 178.79,\ \theta_y = 207.42$$

The two solutions differ by 180° rotation about z and are therefore identical. To construct MORPCY, rotate the starting structure first about the x axis by −1.21° and then about the y axis by −27.42°. Complete the assembly by making additional copies of this structure at multiples of $t = 4.30$ Å along the z axis.

Energy Parameters. The nonbonded energy parameters used in eq 1 come from the MM2 force field as implemented in MACROMODEL version 2.5. They are listed in Table VII.

For the case of C (sp3) bonded to H the energy parameter A is set to 0.046 and the sum of the radii B is set to 3.34 in eq 1 rather than the computed table values of 0.045 and 3.40. No other special interactions were considered nor any interactions of lone pair electrons as separate entities.

(26) Gavezzotti, A. *Acta Crystallogr., Sect. B* **1990**, *46*, 275.