Molecular Self-Assemblies. 2. A Computational Method for the Prediction of the Structure of One-Dimensional Screw, Glide, and Inversion Molecular Aggregates and Implications for the Packing of Molecules in Monolayers and Crystals

Jerry Perlstein†

Contribution from the Center for Photoinduced Charge Transfer, Department of Chemistry, University of Rochester, Rochester, New York 14627

Received October 13, 1992®

Abstract: A computational method based on Monte Carlo cooling has been developed to quantitatively predict the geometric packing of rigid molecular units into screw, glide, and inversion aggregates using a force field containing only nonbonded and electrostatic terms. Of 60 aggregate structures selected at random from the Cambridge Structural Database (containing elements restricted to C, H, O, N, F, Cl, Br, I, S; no hydrogen bonds and only one molecule in the crystallographic asymmetric unit), 53 were found to lie at a local energy minimum which was 0.1–5.7 kcal above the global minimum but visually very different from it. The difference results from the existence of surface cavities in the global minimum structure created by the screw, glide, or inversion offset which are generally not fillable by parts of any other identical molecule. The electrostatic contribution to the total energy was found to be small, averaging 5% of the total energy and not structure determining. These results, taken together with our previous simulations for translation aggregates, confirm a generalized Aufbau principle implied by Kitaigrodskii, that the complex ordering of molecules in three dimensions can be broken down into substructures, each of which is in a local energy minimum. Implications of these results for the design of ordered molecular materials and for the quantitative predictions of monolayer packing and full 3-dimensional crystal packing are discussed.

Introduction

"*One of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical compositions.*" [1]

While this statement is still true today, it is nevertheless also true that the forces that hold molecules together are reasonably well understood. However, the complexity associated with the large number of possible point and spatial symmetries and the combinatorial problem associated with this has prevented the development of direct methods for solving the generalized crystal structure problem. Recent efforts by Karfunkel and Gdanitz,[2] Holden, Du, and Ammon,[3] and Gavezzotti[4] point to the extreme difficulty in finding suitable computational pathways to the solution of the 3-dimensional crystal structure of molecular systems in which only weak nonbonded and electrostatic forces are at play. It is the purpose of this work to show that the problem can be broken down into smaller components, the summed solution of which can lead to a predictive methodology not only for the packing of molecules into 3-dimensional crystals but also for the simpler problem of the packing of molecules into monolayers given only the connectivity of the atoms for a single molecule.

The ability of organic molecules to form ordered crystalline arrays is one of the most remarkable occurrences in nature. Regardless of the complexity of the molecular shape, there is always a space group which allows the molecule to assemble itself into a closed packed arrangement of low free energy. The relationship between the 3-dimensional geometric molecular structure and the various possible packing modes that a molecule can have was formulated and studied in some detail many years ago by Kitaigorodskii,[5] but it has been only recently, with the advent of high-speed affordable computers, that these ideas have been shown to be extremely fruitful for predicting the detailed packing structures of ordered arrays in a quantitative way.[6]

Kitaigorodskii's view of organic molecular crystals contains an Aufbau principle (KAP) which says the following: Molecular crystals are built up from 1-dimensional aggregates which are packed together to form 2-dimensional monolayers which in turn are packed together to form full 3-dimensional crystalline solids.[5]

What is most remarkable about KAP is the small number of 1-dimensional chains and 2-dimensional layers that actually occur. As complex as a crystal structure may seem at first blush, Scaringe and Perez[7] have shown that only four types of 1-dimensional aggregates actually occur in 92% of all crystals, and Scaringe[6] has shown that these can combine to form only seven types of layers.

We have examined in some detail the 1-dimensional aggregates that occur in close to 150 different crystal structures taken at random from the Cambridge Structural Database and have found only six in which a unique low-energy aggregate does not occur. In 96% of the structures, a unique low-energy aggregate is discernible.

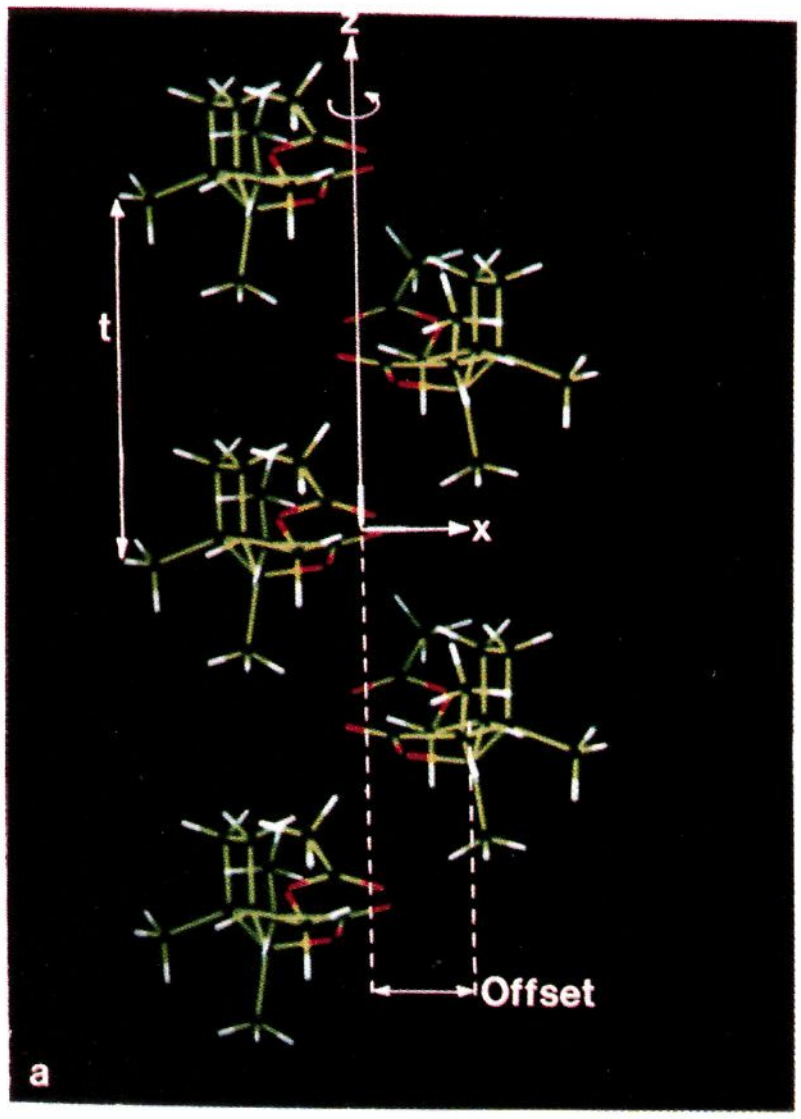
Given that molecules in ordered arrays play a significant role in various physicochemical properties and that from a molecular engineering standpoint it would be highly desirable to be able to predict the structural arrangements of the aggregates and layers that such molecules form, we have undertaken a program that will allow us to predict such arrangements with the minimum number of assumptions. The basis of the method is to employ KAP using a Monte Carlo cooling technique[8-10] combined with

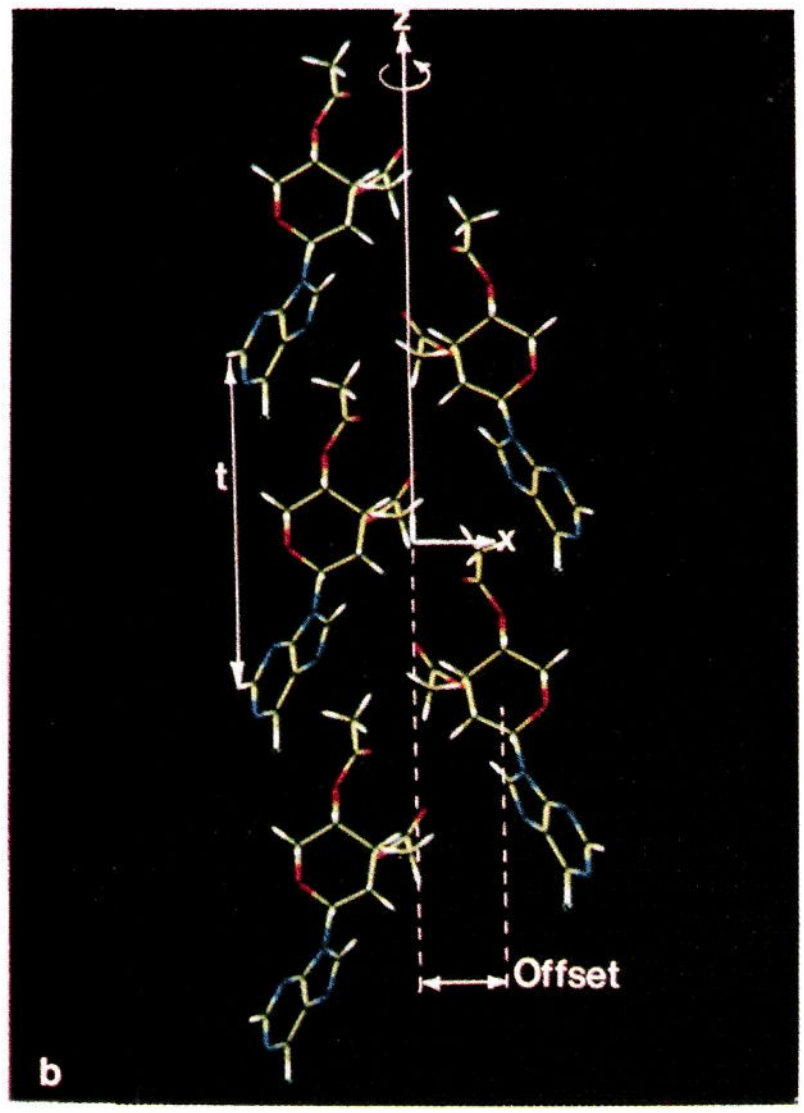
† Work done at Office Imaging Research and Technology Development, Eastman Kodak Co., Rochester, NY.

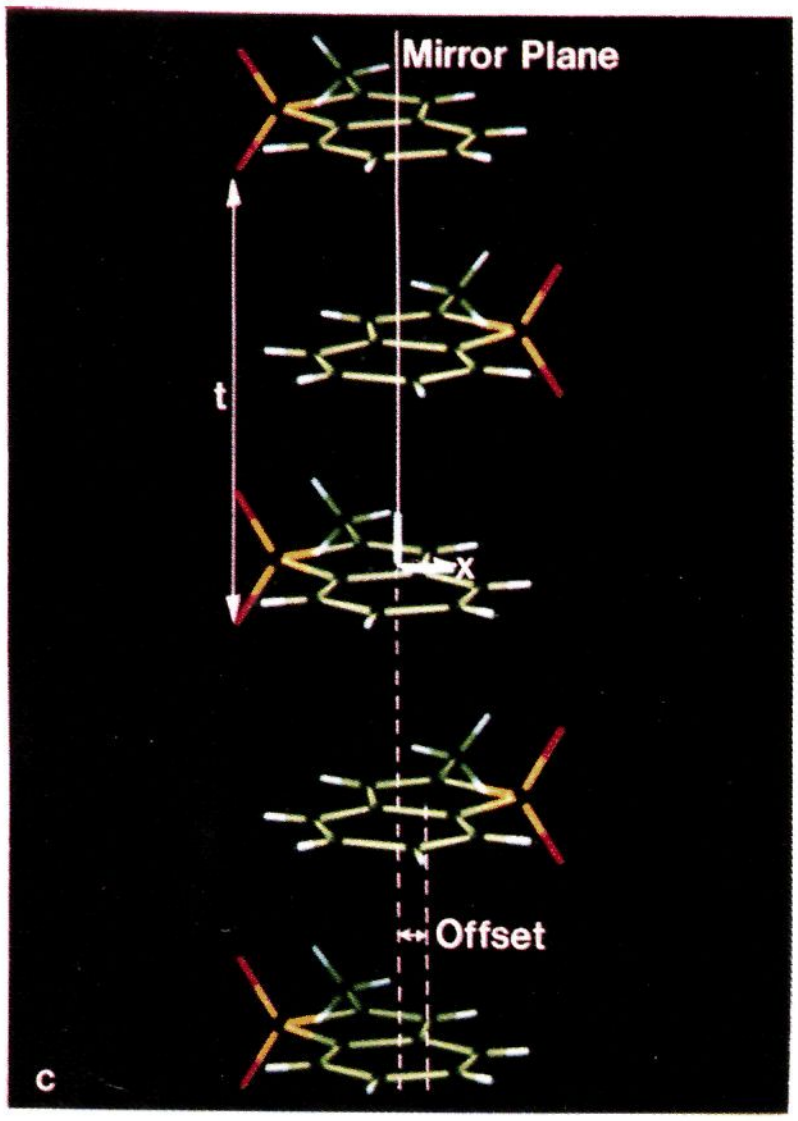
® Abstract published in *Advance ACS Abstracts*, December 15, 1993.

(1) Maddox, J. *Nature* **1988**, *335*, 201.
(2) Karfunkel, H. R.; Gdanitz, R. J. *J. Comput. Chem.* **1992**, *13*, 1171.
(3) Holden, J. R.; Du, Z.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422.
(4) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622.
(5) Kitaigorodskii, A. I. *Organic Chemical Crystallography*; Consultants Bureau: New York, 1961.
(6) Scaringe, R. P. In *Electron Crystallography of Organic Molecules*; Fryer, J. R., Dorset, D. L., Eds.; Kluwer: Dordrecht, 1990.
(7) Scaringe, R. P.; Perez, S. *J. Phys. Chem.* **1987**, *91*, 2394.
(8) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
(9) Szu, H.; Harley, R. *Phys. Lett. A* **1987**, *122*, 157.

0002-7863/94/1516-0455$04.50/0







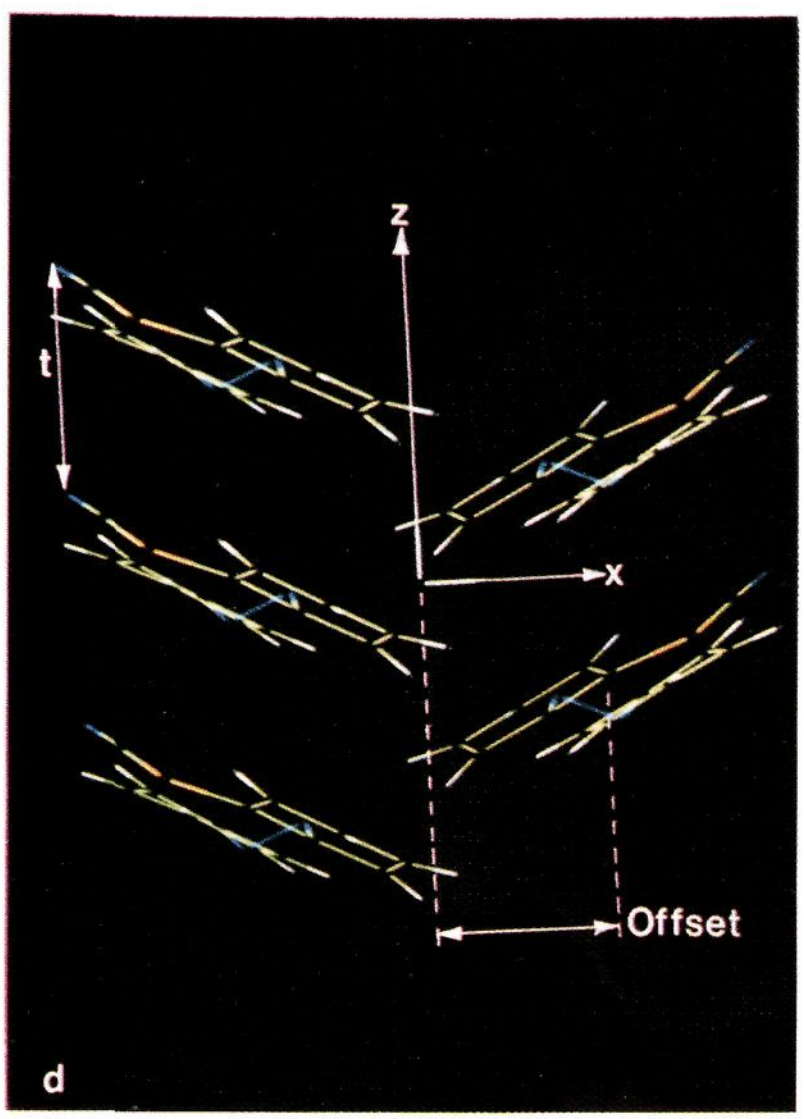


Figure 1. (a) ACMCOC screw aggregate. Molecules are related by 180° rotation about z with repeat t = 7.467 Å along z and offset 2.092 Å along the $\pm x$ axis. (b) CACRPU screw aggregate, t = 10.955 Å, offset = 2.736 Å. (c) GAKPEN glide aggregate. Molecules are related by reflection in the yz plane with repeat t = 7.643 Å along z and offset = 0.444 Å along the $\pm x$ axis. (d) ABSFCN looks like a screw aggregate with t = 3.79 Å, offset = 3.022 Å, but energetically it is two translation chains (see text).

a simple force field derived from MM2 which efficiently searches for local minima in the vicinity of the global minimum. We have presented part of this work earlier for the 1-dimensional translation aggregate,[11] and here we present our results for the remaining 1-dimensional aggregates, the screw, glide, and inversion types.

(10) Matsuba, I. *Phys. Rev. A* **1989**, *39*, 2635.

What remains is to show how to include torsional terms as variables in the development of the method and to demonstrate that 2- and 3-dimensional ordered arrays can be effectively simulated. This we will do in succeeding papers.

Types of One-Dimensional Aggregates

One-dimensional aggregates occur in four varieties: the translation aggregate, screw aggregate (2_1 screw), glide aggregate, and inversion aggregate.[7] The translation aggregate has already been discussed.[11]

A. Screw Aggregate. A 2_1 screw aggregate consists of a collection of molecules arranged along a unique axis in such a way that every molecule is related to its neighbor by a 180° rotation about the unique axis. Figures 1a and 1b show this for ACMCOC and CACRPU, two molecules used in the study (see Figures 3–5 for molecular structures).

As indicated by Scaringe and Perez, there are several important points to make about this structure which also apply to the glide aggregate. (a) The screw axis does not lie at the center of coordinates (the centroid) of any of the molecules; rather, the centroids are offset from the axis by a finite distance called the offset distance, as shown in Figure 1. Every molecule is thus not only rotated 180° with respect to its neighbors but also offset from the rotation axis by an amount which is equal and opposite to that of its neighbors. We take the screw direction to be the z axis and the offset direction to be the x axis. The y axis is then the cross product of z into x. (b) The repeat distance is the distance between the centroids of every other molecule. If we call this distance t, then a requirement for the screw aggregate is that the distance between the z component of neighboring centroids be $t/2$. Thus the molecules in the screw aggregate are equally spaced along the z axis at $t/2$ but offset from it in opposite direction along the x axis.

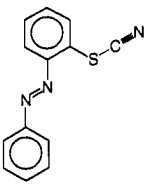
B. Glide Aggregate. The glide aggregate is similar, except instead of neighboring molecules being related by 180° rotation, they are related by reflection in a mirror plane whose normal is parallel to the offset direction. Figure 1c demonstrates this for GAKPEN. As with the screw aggregate, if the repeat distance is t, then neighboring mirror image molecules are at $t/2$ along the z axis but offset from it in equal and opposite directions along the x axis.

C. Inversion Aggregate. As its name indicates, the inversion aggregate consists of an array of inversion-related molecules aligned along an inversion axis but with their molecular centroids offset from this axis. Figure 2a shows an example. Between any two molecules there is an inversion point lying on the axis. Taken as origin, the inversion point has the property that the coordinates (x,y,z) of one molecule are related to those of its neighbor (and to those of any other molecule whose centroid is on the opposite side of the inversion axis) by change of sign $(-x,-y,-z)$.

While the inversion points are spaced $t/2$ apart along the inversion axis, where t is the repeat distance, the separation distance between the molecular centroids of inversion-related molecules need not be equally spaced, as indicated in Figure 2b for GABJOI. Molecules 3 and 4 are separated by 1.531 Å, whereas molecules 2 and 3 are separated by 6.45 Å. There is thus an additional translational degree of freedom which was not present in the screw and glide aggregates. This additional translation also introduces some complications in the computation of the lattice energy, which we will discuss below. As with the screw and glide aggregates, the inversion-related molecules are offset from the inversion axis in equal and opposite directions, as shown in Figure 2b. To measure this distance, draw a vector from the centroid of any molecule to its inversion-related neighbor. The length of the x component of this vector equals twice the offset distance.

(11) Perlstein, J. *J. Am. Chem. Soc.* **1992**, *114*, 1955. Equation 1 in this reference was incorrectly typed. See Appendix, eq A1 for the correct form.

We cannot overemphasize the importance of the offset distance in all three aggregate types (it is absent in the translation type). It can vary from 0 to 4 or 5 Å. Examine Figures 1a–d to see this tremendous variation. In Figure 1c, the offset distance is only 0.444 Å; it is 2.092 Å in Figure 1a and 2.736 Å in Figure 1b. In Figure 1d for ABSFCN (**I**), it is large enough compared to the molecular dimensions that instead of describing the structure as a single screw aggregate, we are forced to consider it as two translation aggregates side by side.



I

Aggregate Energetics

We use the simplest possible intermolecular potential to describe the potential energy, u_i of a molecule i, in these aggregates. It consists of an empirically parametrized nonbonded atom–atom potential E^{nb} and an electrostatic atom–atom potential E^{el} summed over all atoms in molecule i with all the other atoms in all the other molecules in the aggregate. For an N-molecule aggregate with the center molecule labeled 1, the potential energy of molecule 1 is given by

$$u_1 = \sum_{k=2}^{N} U_{1k} \tag{1}$$

where each of the U_{1k} values is given by

$$U_{1k} = E^{nb}_{1k} + E^{el}_{1k} \tag{2}$$

The form of E^{nb} is taken from the MM2 force field of Allinger[12] with empirical parameters from MACROMODEL.[13] For E^{el} we use a Coulomb potential with a distance-dependent dielectric constant and empirical partial charges of Gasteiger.[14] A detailed description of their use and values for the parameters are given in the Appendix. Interestingly enough, if eq 1 is the potential energy of a single molecule in the aggregate, then the 1-dimensional lattice energy, $U_{lattice}$, the energy released when all the molecules are brought together from infinity to form the aggregate, is given by

$$U_{lattice} = u_1/2 \tag{3}$$

namely half the single molecule potential.[15]

Identification of the Aggregate Types in Crystal Structures

Experimentally observed aggregates were identified from X-ray crystal structures for comparison with our simulation results. We used the above potential to determine the kinds of aggregates in crystal structures chosen at random from a subset of the Cambridge Structural Database with the following constraints: (a) the molecule contains only atoms for which we have force field parameters, (b) molecules with potential hydrogen bonds are excluded as we have not included a hydrogen bond term in the force field, (c) the crystal structure can be generated from a single molecule.

(12) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.
(13) 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.
(14) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
(15) Busing, W. R. *J. Phys. Chem. Solids* **1978**, *39*, 691.

Figures 3, 4, and 5 show the molecular structures which form screw,[16] glide,[17] and inversion aggregates,[18] respectively. A variety of structures, large and small, polar and nonpolar, with and without ring heteroatoms, are included.

A. Identification of Screw and Glide Types. The lowest energy aggregate type within a crystal structure can be determined as shown by the following example. In Figure 6, we show the projection of GAKPEN onto the *b* axis with the symmetry of the molecules referenced to molecule number 1 labeled by color. The interaction potentials between molecule 1 and each of the other molecules surrounding it in all directions were computed using eq 1 and are shown in Table 1. Molecule 1 and the two lowest energy molecules, namely 2 and 3, form a three-molecule aggregate whose symmetry is one of the four types. To determine which type, examine the colors in Figure 6. Molecule 1 is red and molecules 2 and 3 are green. Since the green molecules are related to the red ones by a glide plane, this type is thus a glide aggregate. The *z* axis is now unique for this aggregate, so all molecules lying along this axis, namely 9-12 (alternating red and green colors), are also part of this aggregate chain. The total energy for this seven-molecule aggregate is then one-half the sum of the individual energies (eq 3). Table 2 shows the energy for aggregates ranging in size from three to seven molecules.

Table 1. Molecule–Molecule Interaction Potential for GAKPKEN[a]

molecule no.[b]	E^{nb}	E^{el}	U_i[c]	symmetry (color)[d]
2	−7.54	−0.49	−8.03	glide (green)
3	−7.54	−0.49	−8.03	glide (green)
4	−2.74	−0.47	−3.21	inversion (blue)
5a	−2.21	−0.31	−2.52	screw (white)
5b	−2.21	−0.31	−2.52	screw (white)
6a	−1.93	−0.12	−2.05	screw (white)
6b	−1.93	−0.12	−2.05	screw (white)
7	−1.58	−0.18	−1.76	inversion (blue)
8	−1.35	−0.37	−1.72	inversion (blue)
9	−0.38	+0.14	−0.24	translation (red)
10	−0.38	+0.14	−0.24	translation (red)
11	−0.036	−0.016	−0.05	glide (green)
12	−0.036	−0.016	−0.05	glide (green)

[a] Interaction potential between molecule 1 and molecules 2–12, Figure 6a. *E* and *U* values are in kcal/mol. [b] Molecule numbering taken from Figure 6a. Molecules 5a and 6a are superimposed on 5b and 6b. [c] $U_i = E^{nb} + E^{el}$; E^{nb} computed from eq A1. E^{el} from eq A4. [d] Symmetry relation relative to molecule 1 (shown by the color indicated in Figure 6a).

Table 2. Aggregate Energy as a Function of Size[a] for GAKPEN[b]

aggregate size	*E* (kcal/mol)	% change
3 molecules (1, 2, 3)	−8.03	
5 molecules (1, 2, 3, 9, 10)	−8.27	2.99
7 molecules (1, 2, 3, 9, 10, 11, 12)	−8.32	0.60

[a] Computed from eq 3 using data from Table 1. [b] Molecules as numbered in Figure 6.

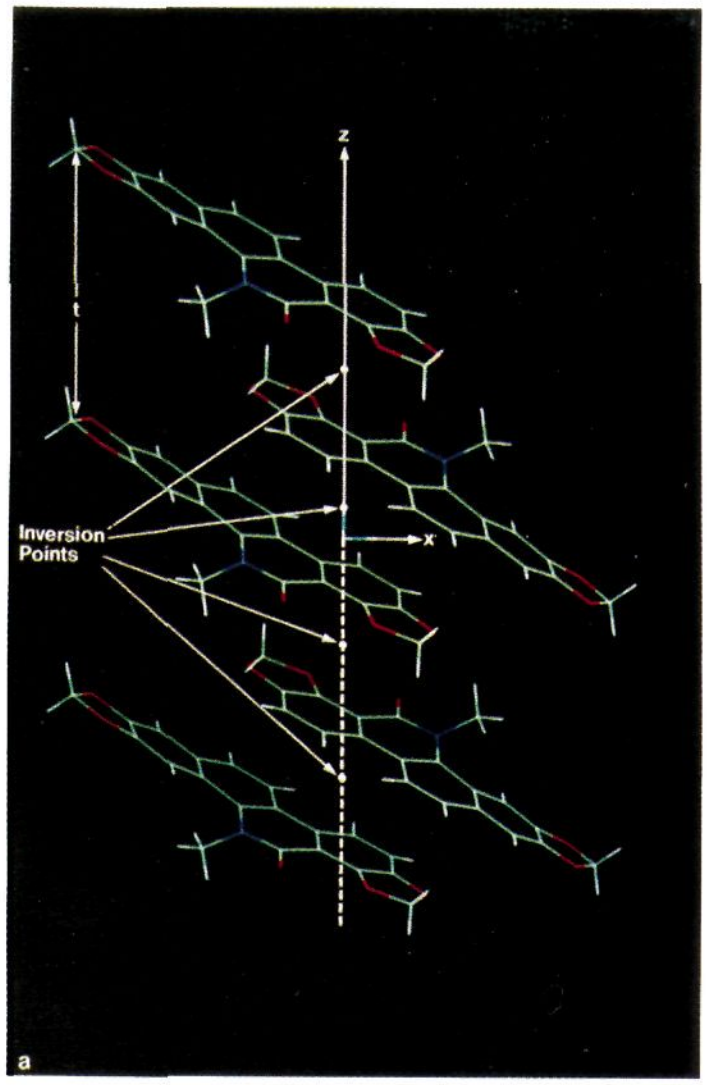
Since the decrease in lattice potential is minimal beyond five molecules, we use a five-molecule aggregate which includes nearest and next-nearest neighbors in most computations (but see discussion below for the inversion aggregates for some exceptions). The coordinates for this five-molecule structure are then extracted from the crystal and used as our experimental results for comparison with the Monte Carlo predictions described below.

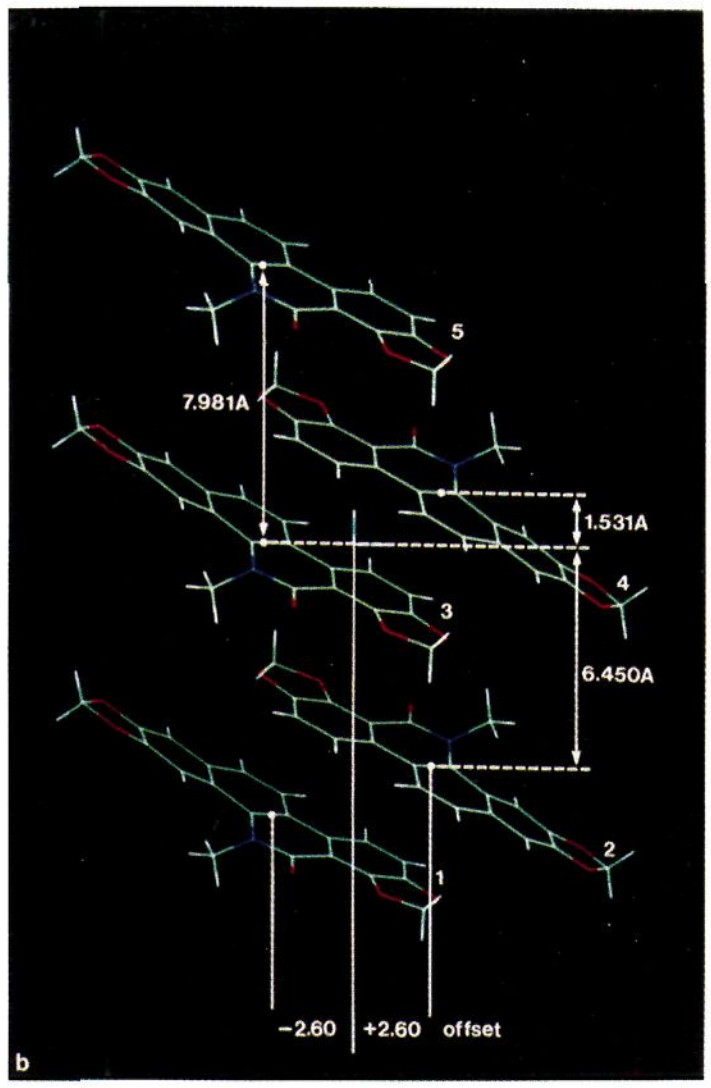
B. Identification of Inversion Types. For the screw and glide aggregates, the two lowest energy molecules relative to the test molecule are always part of the lowest energy aggregate structure extracted from the X-ray data. This is so because the two molecules have the same energy. The presence of two translations in the inversion aggregate, however, precludes this from being generally true. Thus, while the lowest energy molecule will always

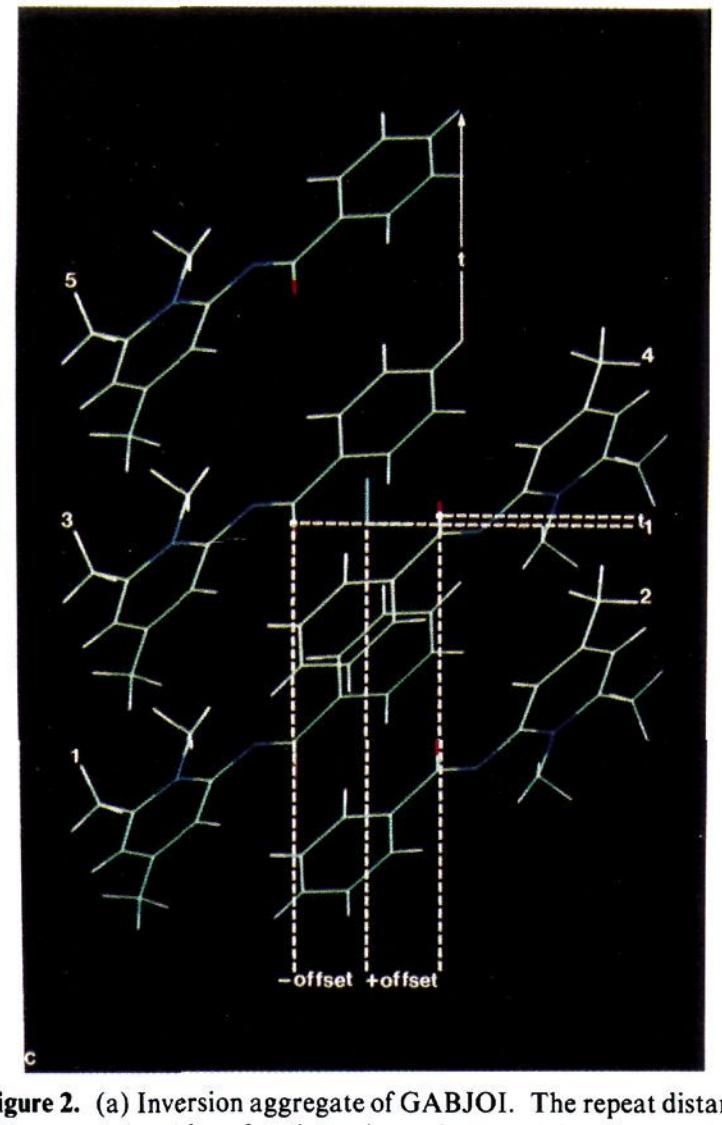
(16) X-ray structures were taken from the Cambridge Structural Database, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, U.K. The listed screw aggregates are found as follows: (a) ACALPB: Foces-Foces, C.; Alemany, A.; Bernabe, N.; Martin-Lomas, M. *J. Org. Chem.* **1980**, *45*, 3502. (b) CAMPTC10: McPhail, A. T.; Sim, G. A. *J. Chem. Soc. B* **1968**, 923. (c) MACAZC10: Preuss, J.; Gieren, A. *Acta Crystallogr., Sect. B* **1976**, *32*, 1299. (d) CACRED: Destro, R.; Ortoleva, E.; Simonetta, M.; Todeschini, R. *J. Chem. Soc., Perkin Trans. 2* **1983**, 1227. (e) ABAXES: Boeyens, J. C. A.; Bull, J. R.; Floor, J.; Tuinman, A. *J. Chem. Soc., Perkin Trans 1* **1978**, 808. (f) ACMCOC: Kitagawa, I.; Shibuya, H.; Fujioka, H.; Yamamoto, Y.; Kajiwara, A.; Kitamura, K.; Miya, A.; Hakoshima, T.; Tomita, K. *Tetrahedron Lett.* **1980**, *21*, 1963. (g) ARTEGA: Schmalle, H. W.; Klaska, K. H.; Jarchow, O. *Acta Crystallogr., Sect. B* **1977**, *33*, 2213. (h) ACESTB: Boeyens, J. C. A.; Bull, J. R.; van Rooyen, P. H. *S. Afr. J. Chem.* **1980**, *33*, 45. (i) BACHAO: Giguere, R. J.; Hoffmann, H. M. R.; Hursthouse, M. B.; Trotter, J. *J. Org. Chem.* **1981**, *46*, 2868. (j) BAMHEC: Hoffmann, R. W.; Ladner, W.; Steinbach, K.; Massa, W.; Schmidt, R.; Snatzke, G. *Chem. Ber.* **1981**, *114*, 2786. (k) CACRPU: Abraham, D. J.; Rosenstien, R. D.; Cochran, T. G.; Leutzinger, E. E.; Townsend, L. B. *Tetrahedron Lett.* **1971**, 2353. (l) ACSESO10: Weeks, C. M.; Rohrer, D. C.; Duax, W. L. *Cryst. Struct. Commun.* **1976**, *5*, 99. (m) BALSOW: Pfaendler, H. R.; Gosteli, J.; Woodward, R. B.; Rihs, G. *J. Am. Chem. Soc.* **1981**, *103*, 4526. (n) BAWNES: Ponnuswamy, M. N.; Parthasarathy, S. *Cryst. Struct. Commun.* **1981**, *10*, 1203. (o) BAHDGL: Ul-Haque, M.; Caughlan, C. N.; Emerson, M. T.; Geissman, T. A.; Matsueda, S. *J. Chem. Soc. B* **1970**, 598. (p) BAYNAQ: Woodward, R. B. *J. Am. Chem. Soc.* **1981**, *103*, 3210. (q) BAFJOH: Britton, D. *Cryst. Struct. Commun.* **1981**, *10*, 1061. (r) LUPANE; Doucerain, H.; Chiaroni, A.; Riche, C. *Acta Crystallogr., Sect. B* **1976**, *32*, 3213. (s) AAXTHP: Kojic-Prodic, B.; Rogio, V.; Ruzic-Toros, Z. *Acta Crystallogr., Sect. B* **1976**, *32*, 1833. (t) BACCHB10: Wagner, H.; Seitz, R.; Lotter, H.; Herz, W. *J. Org. Chem.* **1978**, *43*, 3339. (u) ABSFCN: Kakati, K. K.; Chaudhuri, B. *Acta Crystallogr., Sect. B* **1968**, *24*, 1645.

(17) X-ray structures for the listed glide aggregates are found as follows: (a) CAHVUC: Parfonry, A.; Tinant, B.; Declercq, J. P.; van Meerssche, M. *Bull. Soc. Chim. Belg.* **1983**, *92*, 437. (b) GAKPEN: El Amoudi El Faghi, M. S.; Genests, P.; Olive, J. L.; Rambaud, J.; Declercq, J.-P. *Acta Crystallogr., Sect. C (Cryst. Struct. Commun.)* **1988**, *44*, 498. (c) CMAPTX: Post, M. L.; Kennard, O.; Horn, A. S. *Acta Crystallogr., Sect. B* **1974**, *30*, 1644. (d) CBUMUR10: Flippen-Anderson, J. L.; Gilardi, R. *Acta Crystallogr., Sect. C (Cryst. Struct. Commun.)* **1984**, *40*, 1957. (e) FADVEL: Brink, K.; Mattes, R. *Acta Crystallogr., Sect. C (Cryst. Struct. Commun.)* **1986**, *42*, 1625. (f) BIFCAU: Atabaev, T.; Podberezskaya, N. V.; Gatilov, Yu. V.; Borixov, S. V.; Ashirov, A. *Zh. Strukt. Khim.* **1982**, *23*, 166. (g) CPTCET10: Delacy, T. P.; Kennard, C. H. L. *J. Chem. Soc., Perkin Trans. 2* **1972**, 2148. (h) SAFSAT: Chehna, M.; Pradere, J.-P.; Vicens, J.; Toupet, L.; Quiniou, H. *Bull. Soc. Chim. Fr.* **1988**, 897. (i) DEZXEL: Dillen, J. L. M.; Meth-Cohn, O. *S. Afr. J. Chem.* **1984**, *37*, 171. (j) BABDIR: Nirmala, K. A.; Gowda, D. S. S.; Watson, W. H. *Acta Crystallogr., Sect. B* **1981**, *37*, 1788. (k) BPACLA: Calabrese, J. C.; McPhail, A. T.; Sim, G. A. *J. Chem. Soc. B* **1970**, 285. (l) MACHYF10: Mo, F.; Sivertsen, B. K. *Acta Crystallogr., Sect. B* **1971**, *27*, 115. (m) MXNAMK: Schweizer, W. B.; Procter, G.; Kaftory, M.; Dunitz, J. D. *Helv. Chim. Acta* **1978**, *61*, 2783. (n) CARPIU: Davies, J. E.; Ham, P.; McKie, C. H.; Pearson, A. J. *Acta Crystallogr., Sect. C (Cryst. Struct. Commun.)* **1983**, *39*, 1573. (o) BOWZES: Inouye, Y. Bull. *Chem. Soc. Jpn.* **1983**, *56*, 244. (p) BIJSOC: Bened, A.; Durand, R.; Pioch, D.; Geneste, P.; Declercq, J. P.; Germain, G.; Rambaud, J.; Roques, R.; Guiman, C.; Guillouzo, G. P. *J. Org. Chem.* **1982**, *47*, 2461. (q) BAGYIR: Baker, R. J.; Sauter, E.; Fahmi, A. A.; Trefonas, L. M.; Griffin, G. W. *Cryst. Struct. Commun.* **1981**, *10*, 843. (r) KAMYEC: Dunn, P. J.; Rees, C. W.; Slawin, A. M. Z.; Williams, D. J. *J. Chem. Soc., Chem. Commun.* **1989**, 1134.

(18) X-ray structures were taken from the Cambridge Structural Database, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, U.K. The listed inversion aggregates are found as follows: (a) CIGLAF: Nakagawa, H.; Yamada, K.; Kawazura, H.; Miyame, H. *Acta Crystallogr., Sect. C* **1984**, *40*, 1039. (b) JABTUB: Iyengar, R.; Pina, R.; Grahmann, K.; Todaro, L. *J. Am. Chem. Soc.* **1988**, *110*, 2643. (c) MAZPAL: Brufani, M.; Casini, G.; Fedeli, W.; Mazza, F.; Vaciago, A. *Gazz. Chim. Ital.* **1971**, *101*, 322. (d) MBBPQC10: Acheson, R. M.; Procter, G.; Critchley, S. R. *Acta Crystallogr., Sect. B* **1977**, *33*, 916. (e) GAKPOX: Nalini, V.; Desiraju, G. R. *Acta Crystallogr., Sect. C* **1988**, *44*, 510. (f) CURFOK: Dauter, Z.; Greenhill, J. V.; Karaulov, A.; Reynolds, C. D. *Acta Crystallogr., Sect. C* **1985**, *41*, 630. (g) JABFOH: Banwell, M. G.; Herbert, K. A.; Buckleton, J. R.; Clark, G. R.; Rickard, C. E. F.; Lin, C. M.; Hamel, E. *J. Org. Chem.* **1988**, *53*, 4945. (h) GAHHEC: Sbit, M.; Dupont, L.; Dideberg, O.; Liegeois, J. F.; Delarge, J. *Acta Crystallogr., Sect. C* **1988**, *44*, 319. (i) MABYOZ: Lukac, J.; Bieri, J. H.; Heimgartner, H. *Helv. Chim. Acta* **1977**, *60*, 1657. (j) IMENIN10: Cook, R. E.; Glick, M. D. *Acta Crystallogr., Sect. B* **1970**, *26*, 2102. (k) JABFEX: Adam, W.; Kliem, U.; Mosandl, T.; Peters, E.-M.; Peters, K.; van Schnering, H. G. *J. Org. Chem.* **1988**, *53*, 4986. (l) HFPCDN: Golic, L.; Leban I. *Acta Crystallogr., Sect. B* **1980**, *36*, 1520. (m) HMNPHT: Boeckman, R. K., Jr.; Delton, M. H.; Dolak, T. M.; Watanabe, T.; Glick, M. D. *J. Org. Chem.* **1979**, *44*, 4396. (n) FADFEV: Rodier, N.; Gillo, M.-P.; Piessard, S.; Le Baut, G. *Acta Crystallogr., Sect. C.* **1986**, *42*, 1397. (o) MAMPHS10: Bandoli, G.; Clemente, D. A.; Tondello, E.; Dondoni, A. *J. Chem. Soc., Perkin Trans. 2* **1974**, 157. (p) GAFHEA: Braverman, S.; Freund, M.; Reisman, D.; Goldberg, I. *Tetrahedron Lett.* **1986**, *27*, 1297. (q) HPCCAM: Swanson, K. L.; Hope, H.; Landrum, P. F. *Acta Crystallogr., Sect. B* **1978**, *34*, 3411. (r) GABJOI: Horn, S. P.; Folting, K.; Huffman, J. C. *Proc. Indiana Acad. Sci.* **1986**, *95*, 177. (s) HCTDPT: Redhouse, A. D. *J. Chem. Soc., Perkin Trans. 2* **1974**, 1925. (t) GALHIK: Litvinov, I. A.; Struchkow, Yu. T.; Valitova, S. N.; Kataev, V. E.; Vereshchagin, A. N. *Izv. Akad. Nauk. SSSR, Ser. Khim.* **1986**, 2718. (u) GAJNIO: Butt, G. L.; Deady, L. W.; Mackay, M. F. *J. Heterocycl. Chem.* **1988**, *25*, 321. (v) GAKHIJ: Robins, D. J.; Sim, G. A. *J. Chem. Soc., Perkin Trans. 2* **1987**, 1379.







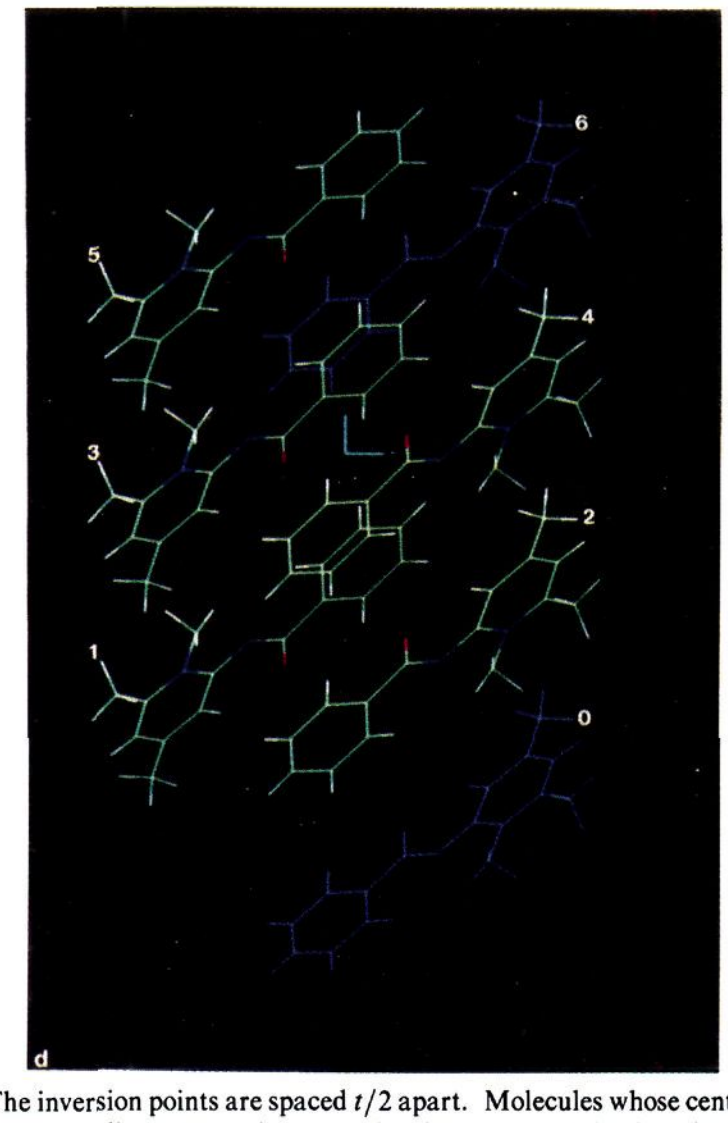


Figure 2. (a) Inversion aggregate of GABJOI. The repeat distance t = 7.981 Å. The inversion points are spaced $t/2$ apart. Molecules whose centroids are on opposite sides of an inversion point as origin are related by changing all atom coordinates x,y,z in one molecule to $-x,-y,-z$ in the other. (b) Inversion-related GABJOI molecules are not equally spaced along the z axis. There is one short distance, t_1 = 1.531 Å and one longer distance, t_2 = 6.45 Å. $t_1 + t_2$ is the repeat distance t. Inversion-related molecules also have an x offset of 2.60 Å. (c) A Monte Carlo constructed inversion aggregate for FADFEV. t_1 = 0.23 Å, t_2 = 5.06 Å, t = 5.29 Å, offset = 2.92 Å. Note that molecules 1 and 4 touch so that the computed interaction potential of molecule 3 with the other four molecules gives incorrect results for the total energy. (d) Addition of virtual molecules 0 and 6 to the energy computation corrects the energy since now molecule 3 touches molecule 6. The inclusion of molecule 0 prevents an incorrect computation when molecule 5 touches molecule 2.

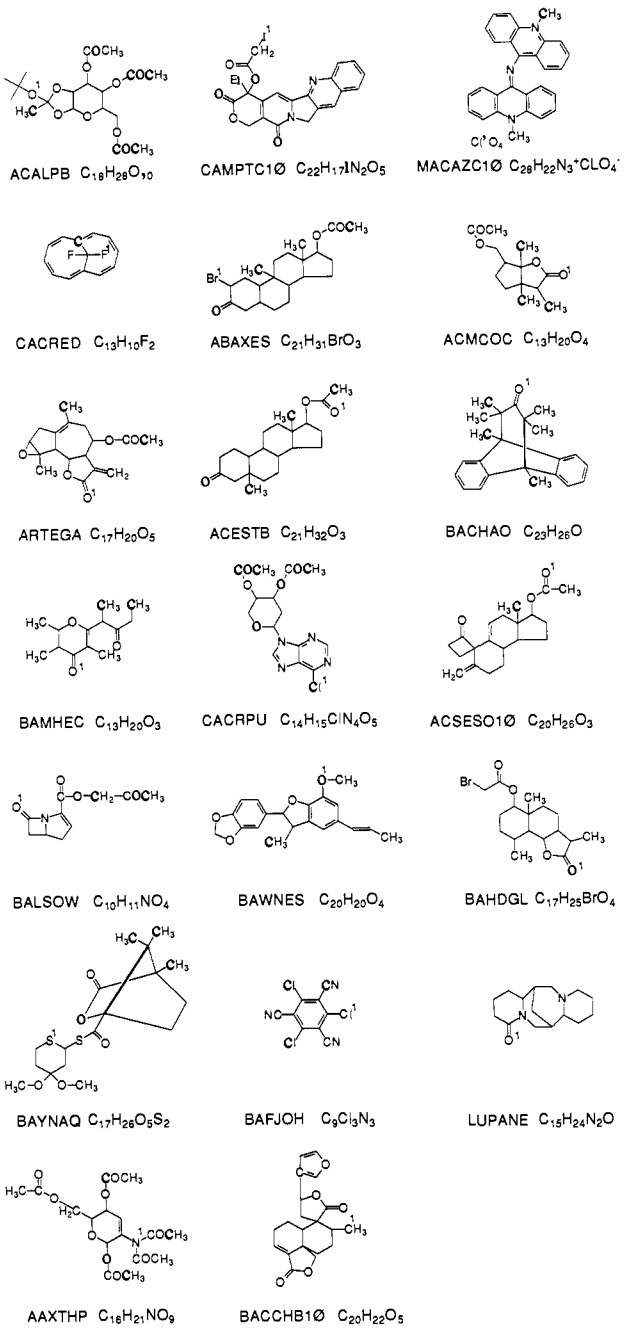


Figure 3. Molecular structures and formulas for molecules forming screw aggregates. Reference codes are from the Cambridge Structural Database.[16] Atoms numbered 1 are for reference in creating the Monte Carlo starting geometry and have coordinate positions given in the supplementary material.

be part of the lowest energy inversion aggregate, the next lowest energy molecule may not be part of this structure. Nevertheless, a visual examination of the lowest energy molecules easily allows for the extraction of the lowest energy inversion aggregate.

While there is no reason to expect that the aggregate type found this way will have a unique symmetry, as we mentioned earlier, of 150 structures we have examined, only six do not show a unique symmetry belonging to one of the four types.

Computational Methodology

In the following sections we detail the computational method for the prediction of the local energy minima of the various aggregate types using a Monte Carlo simulation technique. We first detail the method for constructing a random aggregate and then show how to use this construction procedure in the simulation process to find the important local and apparent global minima. We wish to stress here that the only information about the packing used in this prediction methodology is the

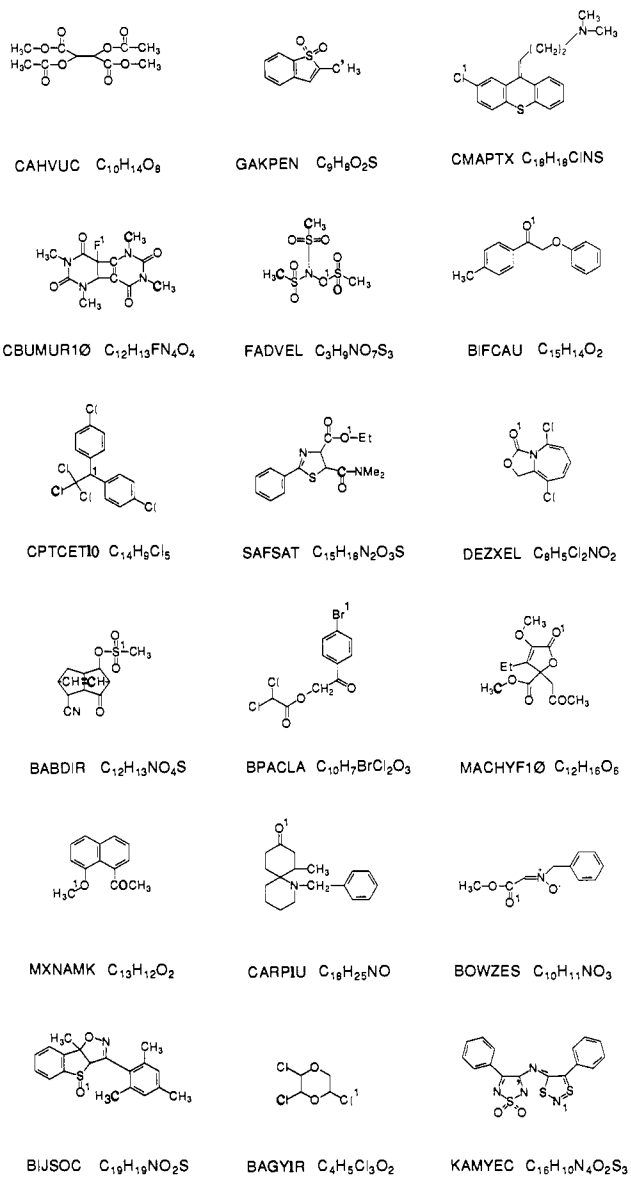


Figure 4. Molecular structures and formulas for molecules forming glide aggregates. Reference codes are from the Cambridge Structural Database.[17] Atoms numbered 1 have coordinates given in the supplementary material for the Monte Carlo starting geometry.

valence bond geometry of the starting molecule and the expected aggregate type. No information about the molecular orientation or spacing was used.

A. Screw and Glide Aggregate Construction Procedures. To construct a screw or glide aggregate, we use a variation of the method given by Scaringe and Perez.[7] (a) Start with a single molecule with its centroid at the origin of an orthogonal coordinate system with unit vectors **i**,**j**,**k** and a molecular reference frame with unit vectors **i**′,**j**′,**k**′ superimposed on **i**,**j**,**k** (see Figure 7a for GAKPEN). (Details for reproducing the starting geometries of this first molecule used in this work are described in the supplementary material.[37]) (b) Rotate the molecule by arbitrary angles θ_x, θ_y, and θ_z about the x, y, and z axes using the definition of the rotation matrix.[11,19] The dot products of the molecular unit vectors with the stationary unit vectors then define a direction cosine matrix which can be used for comparison with experimental values (Figure 7b). (c) Make a copy of this molecule. To do this for a screw aggregate, change all the x and y atom coordinates to $-x$ and $-y$ for the copy. For the glide, change only the x coordinates to $-x$ for the copy (Figure 7c). (d) Translate the copy a random offset distance along $+x$ and the original the same distance along $-x$ (Figure 7d). (e) Duplicate the original twice more, and translate the duplicates a random distance t along $+z$ and $-z$ (Figure 7e). (f) Duplicate the copy once more, and translate the copy and its duplicate a distance $t/2$ along $+z$ and $-z$, respectively (Figure 7f).

(19) Jeffreys, H.; Jeffreys, B. S. *Methods of Mathematical Physics*; Cambridge University Press: Cambridge, U.K., 1972; p 122.

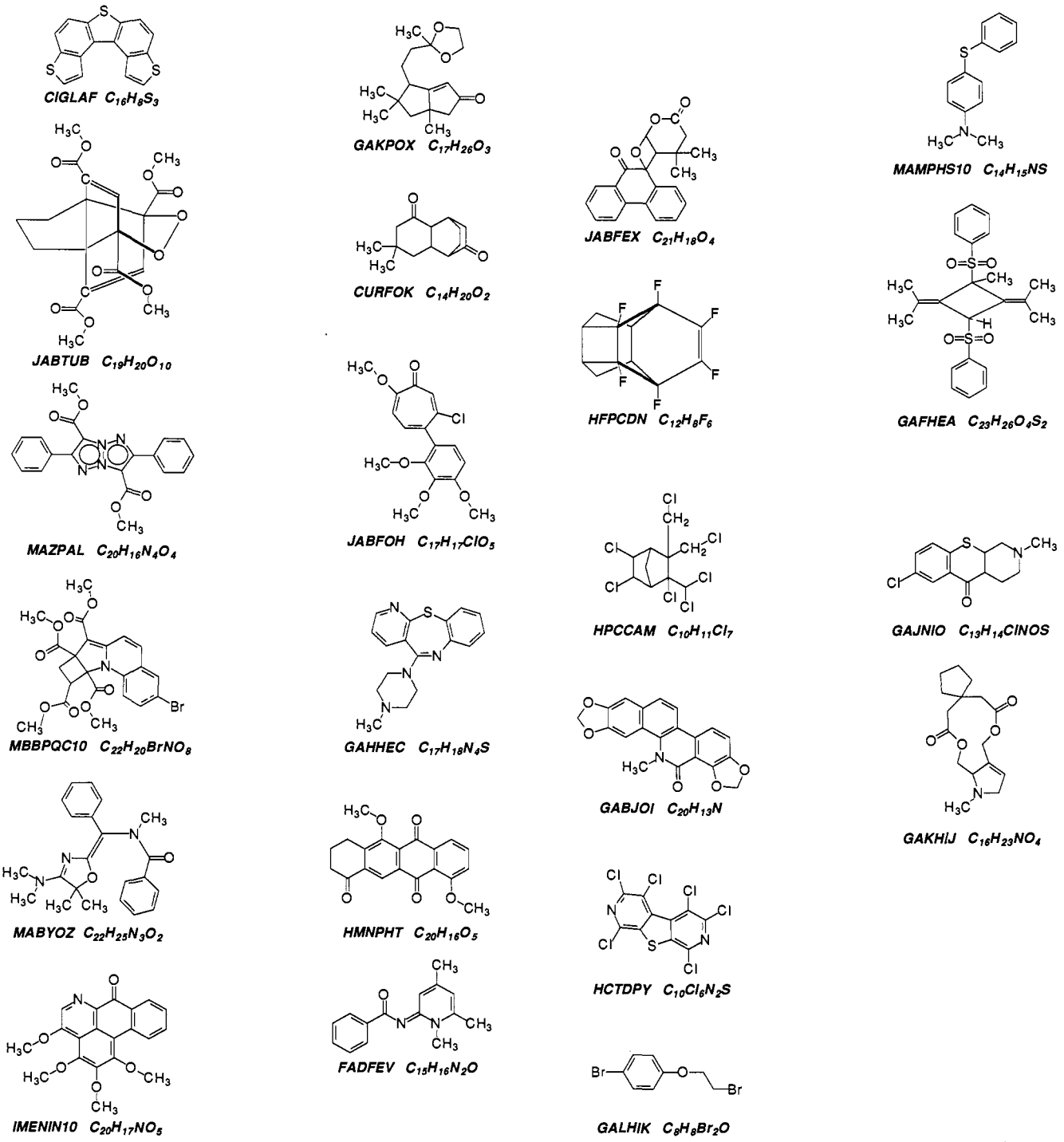
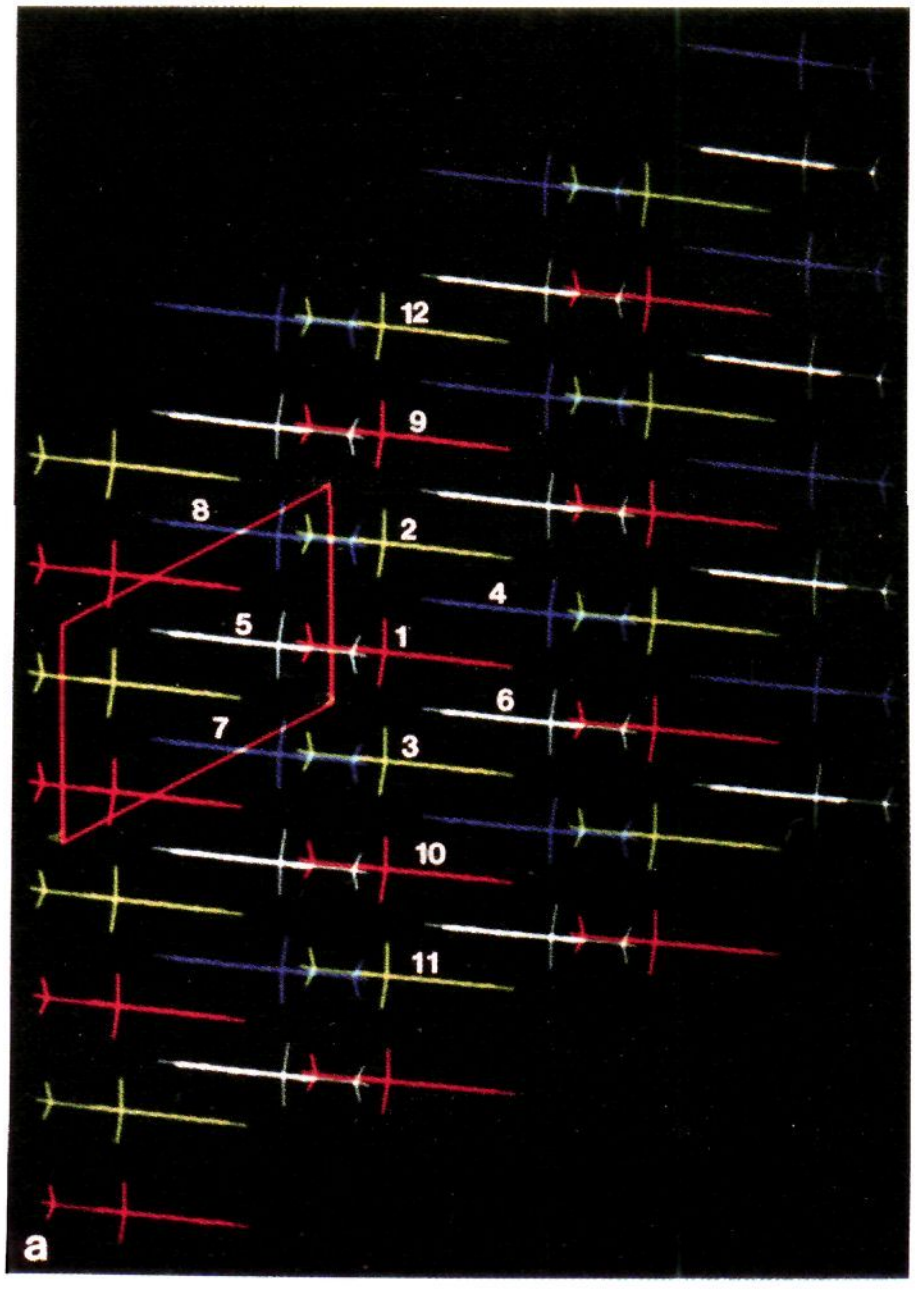


Figure 5. Molecular structures and formulas for 22 molecules known to form inversion aggregates. The molecules were chosen at random from a subset of the Cambridge Structural Database.[18]

This completes the construction of a five-molecule screw or glide aggregate. There are clearly five variables here, three rotation and two translation. The number of possible structures is enormous. If the angle variables θ_x and θ_z can take values of 0–360° and θ_y 0–180° in 5-deg increments and the translation variables can span 7–13 Å for the z repeat and 0–4 Å for the x axis offset in increments of 0.2 Å, then the total number of possible structures is $72 \times 72 \times 36 \times 30 \times 20 = 112 \times 10^6$.

B. Geometric Construction of the Inversion Aggregate. Construction of the inversion aggregate for the Monte Carlo simulation uses a modified method also described by Scaringe and Perez.[7] We demonstrate this for MAMPHS10 in Figure 8 as follows. (a) Start with a single molecule centered on an orthogonal coordinate system with unit vectors **i**,**j**,**k** and a molecular coordinate system **i**′,**j**′,**k**′ superimposed. (b) Rotate the molecule about x, y, and z. (c) Construct an inversion copy by changing all atom coordinates x,y,z to $-x,-y,-z$. (d) Set the offset distance by translation of the copy along $+x$ and translation of the original an equal amount along $-x$. (e) Duplicate the original twice, and translate the duplicates in opposite directions a distance t, the repeat distance along $\pm z$. (f) Translate the copy a distance t_1 along z such that $t_1 < t$. (g) Duplicate the inversion copy and translate this duplicate a distance $-t$ along z. This completes the construction of a five-molecule inversion aggregate. Note that there are two unequal separation distances between the molecular centroids along the z axis, t_1 and $t_2 = t - t_1$.

Most of these structures are of no interest since they lie far above the global minimum. What is needed is an efficient searching algorithm which will look for local energy minima around the global minimum. This is provided by Monte Carlo cooling techniques. Unlike other minimization techniques such as steepest descent searches which find different local minima depending on the starting point, Monte Carlo techniques will find local as well as global minima independent of the starting geometry. In this sense it has many of the advantages of a brute force systematic search and none of the disadvantages when the number of degrees of freedom grows large.



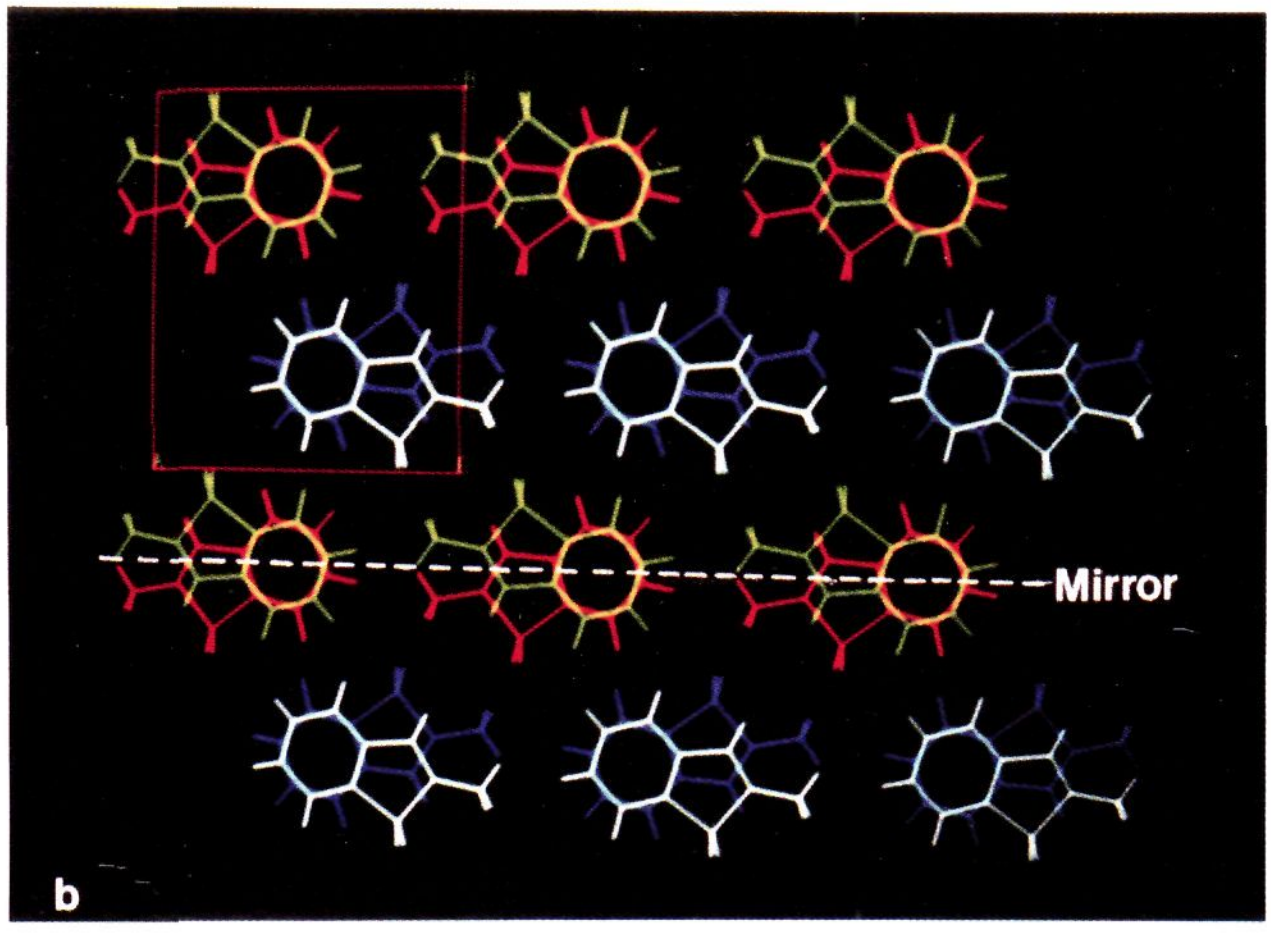


Figure 6. (a) Structure of GAKPEN projected onto the crystallographic *b* axis. The molecules are related by symmetry according to color as follows: red–blue or green–white, inversion; red–green or blue–white, glide; red–white or blue–green, screw. Energy of interaction of molecule 1 (red) with the surrounding molecules given in Table 1. The lowest energy occurs between the red–green molecules which form the basis of a glide aggregate with a repeat t = 7.643 Å along c. (b) GAKPEN projected onto the c axis with the mirror planes of the glide aggregates clearly discernible as the *ac* plane.

C. Monte Carlo Simulation with Cooling. The Monte Carlo simulation technique as it applies to the aggregate problem has been described previously for the translation aggregate.[11] We indicate here only those features unique to the screw, glide, and inversion aggregates.

C1. Monte Carlo Step. Coordinates of the starting geometries for the Monte Carlo cooling were initialized using the molecular modeling program CHEM-X[20] on a VAX8600 (see supplementary material[37]). The starting geometry for each aggregate consisted of five molecules

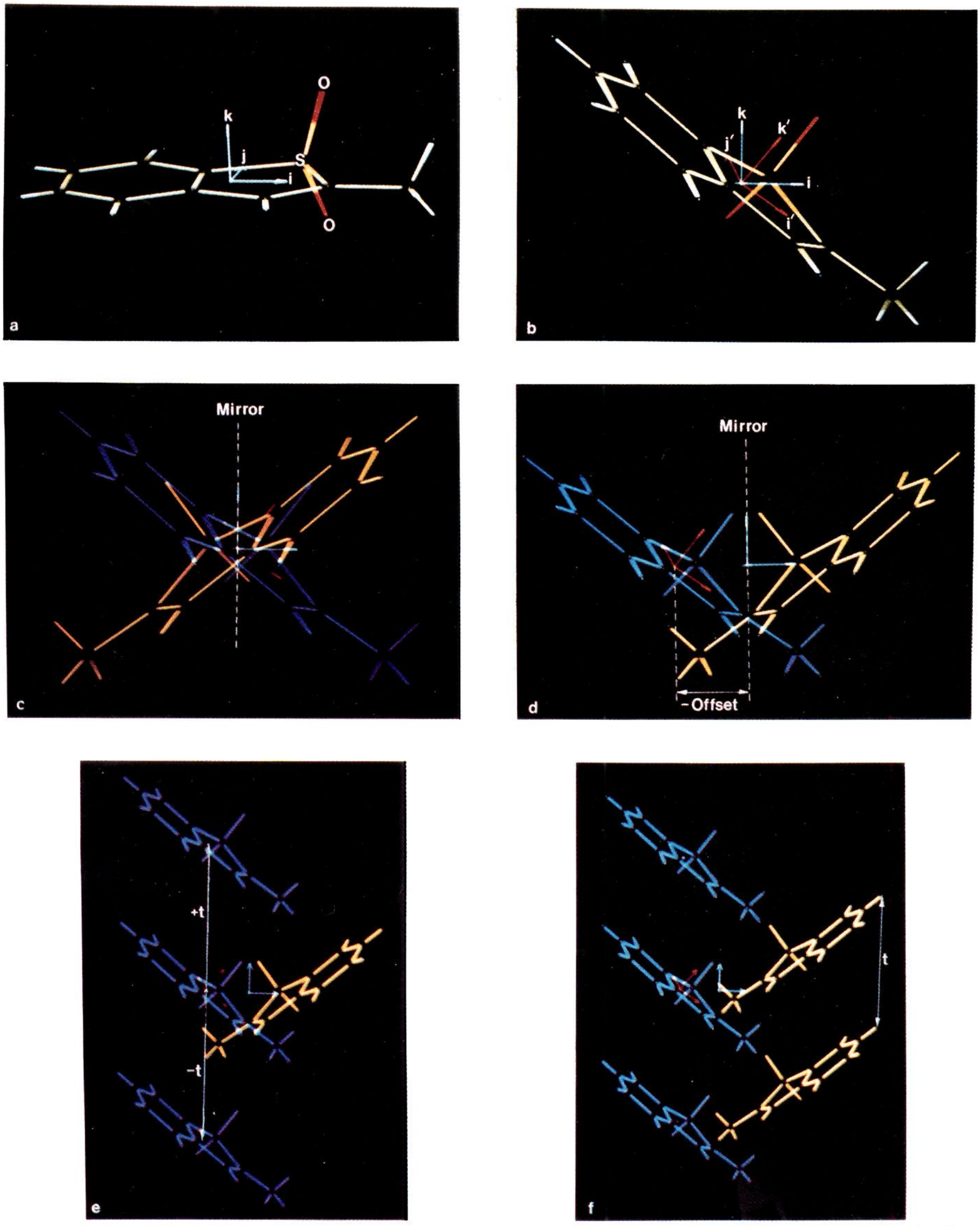


Figure 7. Random glide aggregate construction procedure for the Monte Carlo simulation as demonstrated on GAKPEN. (a) Single molecule with centroid at the origin of the orthogonal coordinate system with unit vectors **i**,**j**,**k** and molecular coordinate system **i′**,**j′**,**k′** superimposed (not shown). (b) GAKPEN rotated 30° about each axis. Note that the molecular coordinate system (in red) rotates with the molecule. (c) Mirror image copy (yellow) constructed from the original (blue) by changing all the atomic x coordinates to $-x$. (d) Offset distance created by moving the copy 1.5 Å along x, the original moved 1.5 Å along $-x$. (e) Two duplicates of the original created at the repeat distance $\pm t$ = 5.00 Å. (f) Duplicate of the copy created and both moved half the repeat length to $\pm t/2$ = 2.5 Å. Result is a five-molecule glide aggregate of GAKPEN.

arbitrarily oriented along a common z axis spaced 15 Å apart. Coordinates were transferred to an IBM RISC SYSTEM/6000 Model 530 workstation for the Monte Carlo cooling. Monte Carlo code was written in FORTRAN 77. Random numbers for the code were generated using the IMSL routine GGUBS.[21]

Beginning with this arbitrary starting geometry and energy as computed

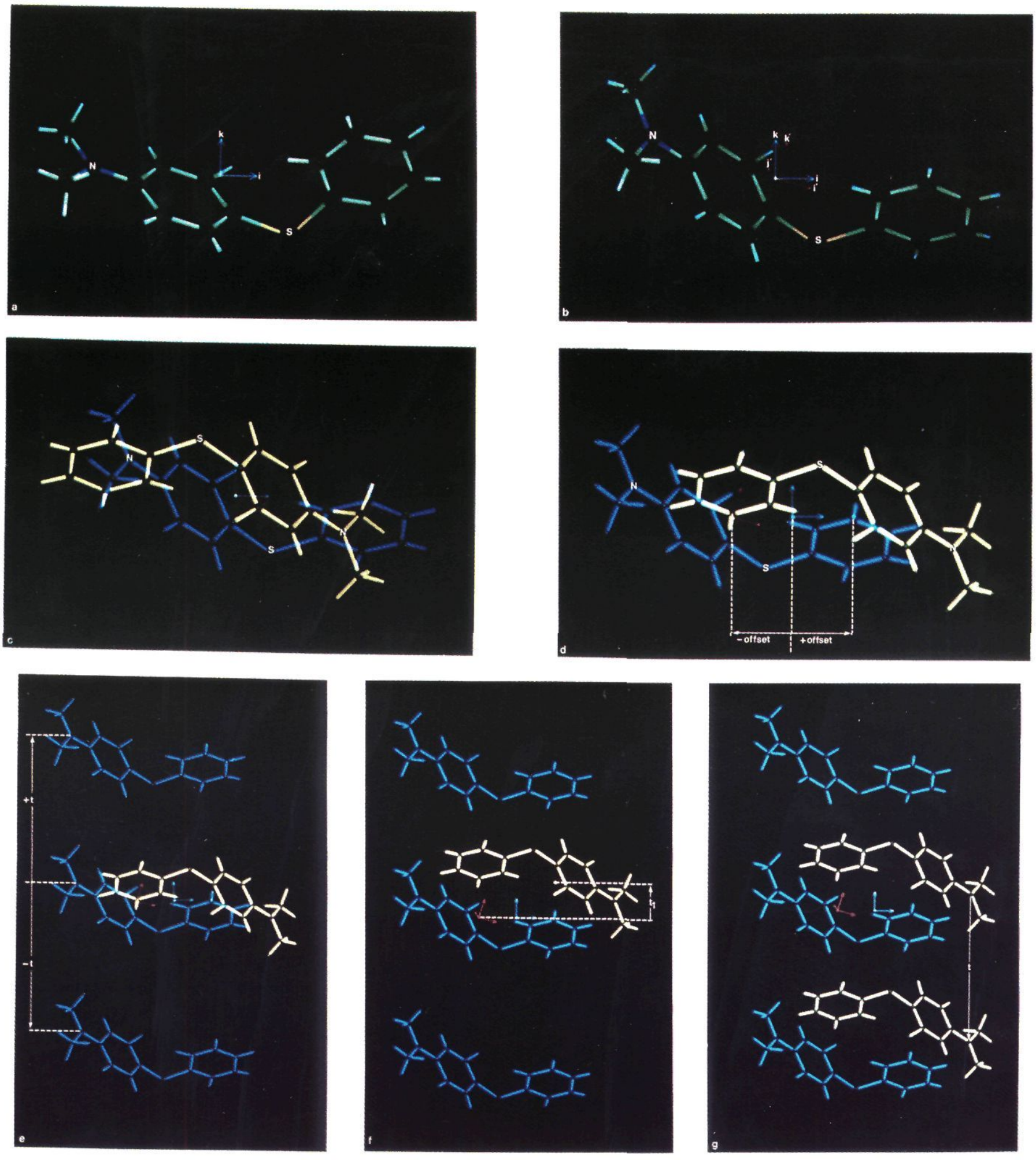


Figure 8. Construction of an inversion aggregate of MAMPHS10. (a) Single molecule with centroid at the origin of a coordinate system with unit vectors **i**,**j**,**k**. A molecular coordinate system with unit vectors **i**′,**j**′,**k**′ is superimposed (not shown). (b) Molecule is rotated 15° about the x, y, and z axes, respectively. Note that the molecular coordinate system (in red) rotates with the molecule. (c) Inversion-related molecule (yellow) constructed from original (blue) by changing all atomic coordinates x,y,z to $-x,-y,-z$. Both molecules still have their centroids at the origin. (d) Offset created by moving original (blue) to $x = -2.00$ Å and inversion (yellow) to $x = +2.00$ Å. (e) Two copies of the original constructed at $\pm t = 8.00$ Å. (f) Inversion molecule (yellow) moved to $t_1 = 2.00$ Å. (g) Duplicate of inversion molecule constructed and moved a distance $t = -8.00$ Å along z completing a five-molecule aggregate.

using eq 3 for the five-molecule aggregate, a Monte Carlo trial involves choosing random values for the variables (five for screw and glide, six for inversion), constructing a new aggregate as described above, and accepting or rejecting the new state depending upon whether the energy moves downhill or uphill compared to the starting structure. In order to enhance the acceptance ratio for the simulation, not all the variables were changed simultaneously for each trial. Rather they were grouped, and only those in a single group were changed. For the screw and glide aggregates there were three groups, the three rotation angles in one group and the repeat translation, t, and the offset each in a separate group. The inversion aggregate had four groups, the three rotation angles in one

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

(21) IMSL Problem Solving Software Systems, 2500 City West Blvd., Houston Texas 77042-3020. The newest version of GGUBS is now called RNUN.

group and the translations, t_1 and t, and the offset each in a separate group. If the energy heads downhill, the state is accepted. Otherwise Metropolis importance sampling is done to decide whether or not to accept an uphill move.[22] This procedure is then repeated with the new state as the starting state (or the old state if the last state was not energetically acceptable).

C2. Cooling Procedure. Monte Carlo cooling considerably shortens the time it takes to find the important local minima.[8-10] The cooling schedule is similar to that described previously.[11] The initial temperature for the simulations was 4000 K and the final temperature 300 K. Sixty of the above Monte Carlo steps were performed before the temperature was decremented by 10% so that the variables in each group were changed at least 20 times for the screw and glide aggregates and 15 times for the inversion aggregates at every temperature. This process was continued until the final temperature was reached. The lowest energy state found during the cooling process was saved. The temperature was then raised back to 4000 K and the cooling procedure repeated. Approximately one-half million Monte Carlo steps were performed for the screw and glide aggregates, resulting in 350 local minima found, and 1 million steps for the inversion aggregates, with 700 local minima found with a reasonably high acceptance ratio averaging 40%.

Problems Encountered in the Simulation and Their Solution

A. Offset Variable. The offset variable can pose a problem. It is possible for the aggregate to break apart into two translation chains, as depicted in Figure 1d, whose total energy might be lower than that of any possible screw or glide aggregate. The signature for the onset of this occurring is the molecular interaction potential for next-nearest neighbors growing larger than that for nearest neighbors. To prevent this, any state is rejected if the next-nearest-neighbor interaction potential is larger in absolute magnitude than the nearest-neighbor potential. For example, in Figure 1d for ABSFCN, the interaction potential between the translationally equivalent molecules is larger than that between the screw-related molecules. The offset is too large, and we would reject this state as a possible screw aggregate state.

B. Energy of the Constructed Inversion Aggregate. Because of the two different translations present in the inversion aggregate, a problem arises in computing the lattice energy using eq 3. In Figure 2c, for example, we show a constructed aggregate of FADFEV. Recall that the lattice energy for this structure is computed by summing the interaction potential of molecule 3 with molecules 1, 2, 4, and 5 and then dividing by 2. The lattice energy computed by eq 3 is −30.74 kcal, but this is incorrect as molecules 1 and 4 are almost touching and should contribute a very large positive potential to the total. To prevent this problem, the interaction potential of molecule 3 with two additional molecules labeled 0 and 6 needs to be included, as shown in Figure 2d. It is not necessary to actually construct these two molecules since the interaction energy of molecule 3 with 0 and 6 is identical to those for the interactions of 1 with 4 and 5 with 2. The total lattice energy is thus computed by adding these two interactions to the sum.

Saving the Results

All the quantitative information about the molecular packing geometry in each aggregate local minimum is contained in the direction cosine matrix, the x axis offset distance, and the z axis translations. For each local minimum, this information was saved in a file and transferred back to the VAX8600 for reconstruction of the minima in CHEM-X and visual observation on an Evans and Sutherland PS390 3-D graphics terminal. The lowest energy structure in the file was taken to be the apparent global minimum. Depending on molecular size and the number of energy computations, the entire simulation would run from 4 to 30 h on the IBM Model 530 workstation. Typically, a 45-atom molecule required 24 h of CPU time to collect 700 inversion minima.

That the states collected by this procedure were indeed local minima was ascertained by comparing the energy of the state with those of nearby states. In all cases of interest, states within 5° for any angle and 0.1 Å for the translations and offset were uphill in energy from the Monte Carlo state. A more exact treatment of this problem would be to compute the Hessian second-derivative matrix to determine if it is positive.

(22) Binder, K. In *Topics in Current Physics*, Springer-Verlag, New York 1984; Vol. 46, p 1.

Table 3. Deviations of Monte Carlo Predictions from Observed Screw Structures

ref code (space group)[a]	RSS[b] global	RSS[b] local	ΔE[c] (kcal)	anisotropy ratio[d]
ACALPB ($P2_12_12_1$)	6.08	1.82	0.16	3.01
CAMPTC10 ($P2_12_12_1$)	25.48	2.34	0.95	2.37
MACAZC10 ($P2_12_12_1$)	6.12	2.13	0.03	2.11
CACRED ($P2_12_12_1$)	5.68	2.50	0.18	2.06
ABAXES ($P2_1$)	131.30	3.64	0.96	1.84
ACMCOC ($P2_1$)	5.85	2.22	0.27	1.77
ARTEGA ($P2_12_12$)	140.54	8.87	1.86	1.65
ACESTB ($P2_12_12_1$)	120.99	3.35	1.46	1.65
BACHAO ($P2_1$)	20.53	10.28	1.31	1.57
BAMHEC ($P2_1$)	129.38	4.02	1.92	1.41
ACSESO10 ($P2_1$)	120.97	2.41	1.15	1.2
BALSOW ($P2_1$)	95.12	9.56	3.23	1.2
BAWNES ($P2_1$)	126.39	10.18	1.23	1.18
BAHDGL ($P2_12_12_1$)	95.41	9.32	2.14	1.16
BAYNAQ ($P2_1$)	99.12	19.83	5.69	1.09
BAFJOH ($P2_12_12_1$)	125.66	17.93	1.96	1.09
LUPANE ($P2_1/c$)	35.30	6.83	1.94	1.04
CACRPU ($P2_12_12_1$)	186.36	12.30	3.56	1.03
AAXTHP ($P2_12_12_1$)	121.83	2.88	0.70	1.03
BACCHB10 ($P2_12_12_1$)	9.78	6.83	0.33	1.02

[a] Reference codes and space groups from Cambridge Structural Data Base (see Figure 3). [b] RSS, root sum square divations (see discussion in Appendix). [c] ΔE, energy of local minimum above global minimum. [d] Ratio of largest on-chain to off-chain molecular potential.

Quantitative Comparison with Observed X-ray Data

The structures of the local minima generated by the above simulation were compared to the experimental aggregates by computing the tilt angles from the direction cosine matrix for each local minima and comparing these with the tilt angles from the experimental structures extracted from the X-ray crystal data. These tilt angles along with the translation distances and offset distance make up the complete quantitative information about the geometry of the aggregate. The root sum square deviation (RSS) of the tilt angles and the distances was computed for each local minimum. The distances were weighted by a factor of 18 so that a 0.3-Å deviation would have the same weight in the RSS as a 5.4° angular deviation. For the inversion aggregate, these deviations would amount to an RSS of 13.2 (see Appendix for details). Care must be taken in making this comparison as the local minimum may be incorrectly oriented by 180° rotations about x, y, or z. Each structure was thus compared to the experimental structure four times, once in its original orientation and three more times after 180° rotations about each axis. The best fit and their deviations are presented in Tables 3 and 4 for the screw and glide aggregates and in Tables 5 and 6 for the inversion aggregates.

Discussion of Results

A. Apparent Global Minima. Examination of the RSS values for the global minima in Tables 3, 4, and 5 clearly indicates that the apparent global minimum structures for most of the aggregates are far from the observed X-ray structures. For the screw aggregate (Table 3), 15 of the 20 have a global minimum that differs significantly (RSS > 20) from the observed structure. For the glide aggregates (Table 4), 14 of the 18 global minima have RSS > 20. For the inversion aggregate (Table 5), 19 of the 22 global minima structures have RSS > 20.

A visual picture of what is happening is quite clear. Examination of the Connolly[23] surfaces for the global minima indicates that there are cavities within the surface which cannot be filled by parts of any other similar molecule no matter how it is oriented. Figure 9 demonstrates this for the glide aggregate of CMAPTX

Table 4. Deviations of Monte Carlo Predictions from Observed Glide Structures

ref code (space group)[a]	RSS[b] global	RSS[b] local	ΔE[c] (kcal)	anisotropy E[c] ratio[d]
CAHVUC ($Pca2_1$)	114.03	7.97	1.10	2.55
GAKPEN ($P2_1/c$)	19.24	8.36	1.73	2.50
CMAPTX ($Pca2_1$)	92.37	5.45	1.92	2.41
CBUMUR10 (P_c)	4.77	3.13	0.31	2.21
FADVEL ($P2_1/c$)	148.58	7.02	1.55	1.98
BIFCAU ($Pca2_1$)	153.12	8.25	2.46	1.86
CPTCET10 ($Pca2_1$)	6.30	2.08	0.09	1.86
SAFSAT (P_c)	101.58	9.35	1.99	1.74
DEZXEL ($Pca2_1$)	85.91	27.99	1.29	1.41
BABDIR ($Pca2_1$)	116.93	35.68	2.79	1.28
BPACLA (P_c)	5.34	1.25	0.57	1.27
MACHYF10 ($P2_1/c$)	58.95	4.25	1.27	1.26
MXNAMK (P_c)	122.66	4.68	1.11	1.19
CARPIU ($Pca2_1$)	129.20	1.59	1.88	1.17
BOWZES (P_c)	97.16	15.84	2.49	1.15
BIJSOC ($Pca2_1$)	148.15	5.24	1.72	1.11
BAGYIR ($Pca2_1$)	128.11	43.72	2.26	1.03
KAMYEC (P_c)	159.89	3.05	5.73	1.03

[a] Reference codes and space groups from Cambridge Structural Data Base (see Figure 4). [b] RSS, root sum square deviations (see discussion in Appendix). [c] ΔE, energy of local minimum above global minimum. [d] Ratio of largest on-chain to off-chain molecular potential.

Table 5. Inversion Aggregates: Apparent Global Minima of Energy and Root Sum Square Deviations from Observed Structures

ref code (space group)[a]	energy (kcal) VDW[b]	Coulomb[c]	total[d]	RSS[e]
CIGLAF ($P2/c$)	−13.46	−0.18	−13.64	3.16
JABTUB ($P\bar{1}$)	−14.78	−0.56	−15.34	126.84
MAZPAL ($P2_1/c$)	−15.72	−0.68	−16.40	117.16
MBBPQC10 ($P2_1/c$)	−16.79	−0.16	−16.95	123.19
GAKPOX ($P\bar{1}$)	−12.93	−1.28	−14.21	8.18
CURFOK ($P2/c$)	−11.40	+0.01	−11.39	74.94
JABFOH ($P\bar{1}$)	−14.18	−1.53	−15.71	44.92
GAHHEC ($P\bar{1}$)	−14.48	−0.07	−14.55	166.31
MABYOZ ($P2_1/c$)	−17.32	+0.08	−17.24	84.60
IMENIN10 ($P\bar{1}$)	−16.85	−0.76	−17.61	62.34
JABFEX ($P\bar{1}$)	−14.83	−1.02	−15.85	138.25
HFPCDN ($P\bar{1}$)	−07.52	−1.68	−09.20	17.09
HMNPHT ($P\bar{1}$)	−16.05	−1.61	−17.66	91.46
FADFEV ($P2_1/c$)	−15.36	−0.05	−15.41	129.37
MAMPHS10 ($P2_1/c$)	−12.86	−0.15	−13.01	114.45
GAFHEA ($P2_1/c$)	−13.42	−0.84	−14.26	155.40
HPCCAM ($P\bar{1}$)	−12.64	−0.02	−12.66	135.34
GABJOI ($P\bar{1}$)	−16.40	−0.10	−16.50	83.69
HCTDPY ($P\bar{1}$)	−13.58	−0.19	−13.77	94.32
GALHIK ($P2_1/c$)	−11.22	−0.34	−11.56	137.51
GAJNIO ($P2_1/c$)	−14.84	−0.55	−15.38	133.38
GAKHIJ ($P2_1/c$)	−16.32	−0.29	−16.61	143.20

[a] Reference codes and space groups from Cambridge Structural Data Base (see Figure 5). [b] Nonbonded van der Waals energy. [c] Coulomb energy using partial Gasteiger charges. [d] Total energy = VDW + Coulomb. [e] Root sum square deviations from observed X-ray data with translations weighted by a factor of 18 (see Appendix for details).

and the screw aggregate of ABAXES. There is not room for a single carbon atom to fit into some of the global minimum cavities. When the aggregates pack together to form layers, there cannot be any spaces left as large as those shown. The appearances of these cavities are in large measure due to the fact that when the molecules are packed along the repeat axis (z axis) they are also offset in opposite directions along the x axis. The offset distance can be quite large, leaving spaces between the next-nearest neighbors.

In contrast to this, the translation aggregates have zero offset. As a consequence, the global minimum very often turns out to be the observed structure, as we have discussed previously.[11]

(23) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548. The program is available from the Quantum Chemical Program Exchange, program no. 429.

Table 6. Best Local Minima Energy and Deviations of Inversion Aggregates

ref code[a]	energy (kcal) VDW[b]	Coulomb[c]	total[d]	ratio[e]	ΔE[f] (kcal)	RSS[g]
CIGLAF	−13.44	−0.18	−13.62	0.014	0.02	1.56
JABTUB	−12.63	−0.68	−13.31	0.051	2.03	1.85
MAZPAL	−14.16	−1.14	−15.30	0.075	1.10	2.56
MBBPQC10	−12.48	−0.38	−12.86	0.030	4.09	2.60
GAKPOX	−12.38	−1.41	−13.79	0.102	0.420	3.48
CURFOK	−09.90	−0.83	−10.73	0.077	0.660	4.51
JABFOH	−11.22	−2.24	−13.46	0.166	2.25	5.63
GAHHEC	−12.14	−0.26	−12.40	0.021	2.15	6.15
MABYOZ	−13.60	+0.12	−13.48	0.009	3.76	6.27
IMENIN10	−13.83	−0.58	−14.41	0.040	3.20	6.55
JABFEX	−12.06	−0.07	−12.13	0.006	3.72	7.19
HFPCDN	−05.91	−1.70	−07.61	0.223	1.59	7.70
HMNPHT	−15.18	−0.62	−15.80	0.039	1.86	8.47
FADFEV	−14.64	−0.10	−14.74	0.007	0.67	9.36
MAMPHS10	−09.02	−0.23	−09.25	0.025	3.76	9.74
GAFHEA	−10.53	−0.64	−11.17	0.057	3.09	12.27
HPCCAM	−08.14	−0.48	−08.62	0.056	4.04	12.66
GABJOI	−13.82	−0.20	−14.02	0.014	2.48	14.47
HCTDPY	−10.34	−0.04	−10.38	0.004	3.39	27.92
GALHIK	−07.64	−0.06	−07.70	0.008	3.85	38.81[h]
GAJNIO	−09.02	−0.46	−09.48	0.048	5.90	39.65[h]
GAKHIJ	−11.54	−0.22	−11.76	0.019	4.84	44.29[h]

[a] Reference codes from Cambridge Structural Data Base (see Figure 5). [b] Nonbonded van der Waals energy. [c] Coulomb energy using partial Gasteiger charges. [d] Total energy = VDW + Coulomb. [e] Ratio of Coulomb energy to the total energy. [f] Deviation of local minimum energy from global energy (from Table 5). [g] Root sum square deviations from observed X-ray data with translations weighted by a factor of 18 (see Appendix for details). [h] GALHIK, local minimum is very broad for the repeat length. GAJNIO, a very flat local minimum for the t_1 variable. GAKHIJ, no minima for t_2 variable found.

B. Local Minima. The RSS values for local minima closest to the observed X-ray structures are displayed in column 3 of Tables 3 and 4 for the screw and glide aggregates and in column 7 of Table 6 for the inversion aggregates. The Monte Carlo cooling procedure found a local minimum which is close to the observed X-ray structure for 53 of the 60 molecules investigated (RSS < 20). These local minima are no more than 5.7 kcal above the global minimum (see the ΔE values in column 4 of Tables 3 and 4 and column 6 of Table 6), and the average angular deviations are less than 3.7° from the observed. Only seven deviate significantly from the observed (glide types DEZEXEL (RSS = 27.99), BABDIR (RSS = 35.68), BAGYIR (RSS = 43.72) and inversion types HCTDPY (RSS = 27.92), GALHIK (RSS = 38.81), GAJNIO (RSS = 39.65), and GAKHIJ (RSS = 44.29)). We discuss these deviations below.

C. Molecule–Molecule Energy Anisotropy. Included in Tables 3 and 4, column 5, for the screw and glide aggregates are the molecule–molecule energy anisotropy ratios, defined as the ratio of the largest molecule–molecule interaction in the aggregate to the largest molecule–molecule interaction off the aggregate (viz. with another molecule in the crystal). In our initial work on the translation aggregates, we thought there might be a connection between this anisotropy and the occurrence of the global minimum being close to the observed X-ray structure. While it is true that the global minimum structure occurs more frequently for translation chains, there appears to be no correlation of this ratio with either the RSS values for the global minimum or the energy difference between the best local minimum and the global minimum for the screw, glide, or inversion aggregates. Some of the largest anisotropies have very large RSS values for the global minimum (ABAXES, 131.30; CAHVUC, 114.03), yet there is a nearby local minimum close to the observed (ABAXES, 0.96 kcal, CAHVUC, 1.10 kcal). This is also true for some of the smallest anisotropies. AAXTHP and BIJSOC are almost isotropic in their molecular potential, yet the screw aggregate for AAXTHP is only 0.70 kcal above the global minimum and the glide aggregate for BIJSOC is only 1.72 kcal above its global

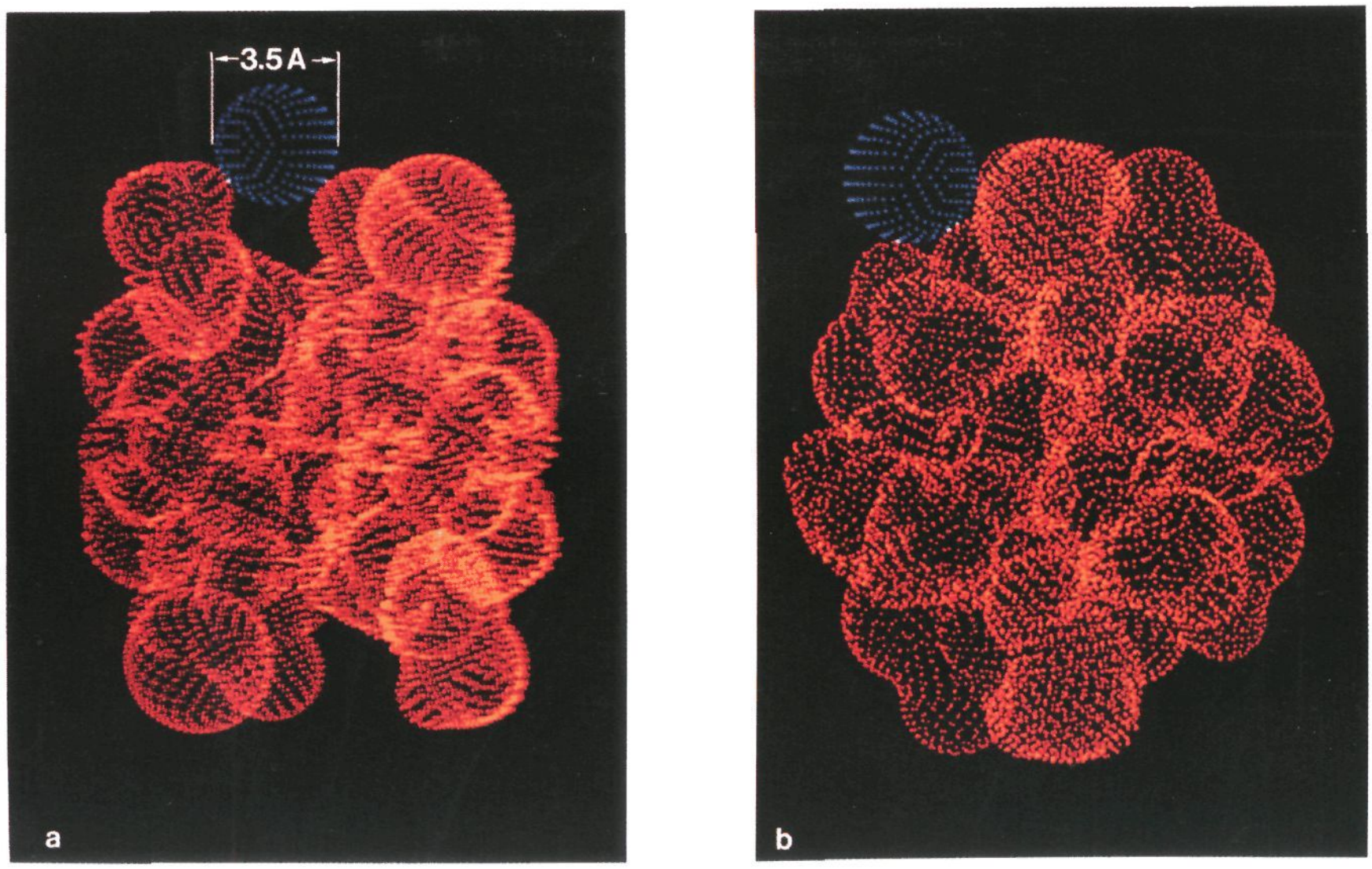


Figure 9. (Top) Connolly dot surface (red) of the CMAPTX glide aggregate as viewed down the repeat axis. In a, the global minimum structure, a carbon atom (blue), cannot fill the largest cavity, whereas in b the observed minimum, a carbon atom, easily fills the cavities. (Bottom) Connolly dot surface of the ABAXES screw aggregate for (a) the global minimum and (b) the observed. The view is along the screw axis. In a, a carbon atom cannot fill the largest cavity, but it easily fills even the smallest cavity in b.

minimum; both aggregates have global minimum structures which deviate considerably from the observed structures (BIJSOC, RSS = 148.15; AAXTHP, RSS = 121.83).

D. How Significant Is the Electrostatic Term? There is considerable interest in the role of the electrostatic term in contributing to the overall packing geometry of molecules. For

Table 7. Rank Ordering of the Local Minima for Glide Aggregates

ref code[a]	position of the best local minimum[b]	ref code[a]	position of the best local minimum[b]
CAHVUC	4	BABDIR	19
GAKPEN	14	BPACLA	0
CMAPTX	1	MACHYF10	4
CBUMUR10	0	MXNAMK	10
FADVEL	42	CARPIU	9
BIFCAU	24	BOWZES	30
CPTCET10	0	BIJSOC	5
SAFSAT	7	BAGYIR	59
DEZXEL	9	KAMYEC	41

[a] Reference codes are from the Cambridge Structural Database.[17] [b] The number of uniquely different structures with energy less than the best local minimum.

Table 8. Rank Ordering of the Local Minima for Screw Aggregates

ref code[a]	position of the best lcoal minimum[b]	ref code[a]	position of the best local minimum[b]
ACALPB	0	CACRPU	22
CAMPTC10	1	ACSESO10	1
MACAZC10	0	BALSOW	44
CACRED	0	BAWNES	28
ABAXES	2	BAHDGL	6
ACMCOC	0	BAYNAQ	58
ARTEGA	2	BAFJOH	23
ACESTB	8	LUPANE	6
BACHAO	0	AAXTHP	2
BAMHEC	10	BACCHB10	0

[a] Reference codes are from the Cambridge Structural Database.[16] [b] The number of uniquely different structures with energy less than the best local minimum.

Table 9. Energy vs Rotation About the x, y, and z Axes for BAGYIR

rotation angle	energy (kcal)		
	x rotation	y rotation	z rotation
−20.00	−6.37	−1.86	−6.33
−15.00	−6.49	−4.77	−6.41
−10.00	−6.60	−6.11	−6.47
−5.00	−6.66	−6.61	−6.51
0.00	−6.59	−6.59	−6.59
+5.00	−6.26	−6.17	−6.72
+10.00	−5.50	−5.32	−6.87
+15.00	−4.19	−3.97	−7.00
+20.00	−2.34	−2.06	−7.08

instance, Hunter and Sanders[24] model the packing of porphyrins almost exclusively in terms of an electrostatic model. Lee and Subbiah[25] describe the packing of amino acid side chains in proteins by neglecting the electrostatic term completely, as do Scaringe and Perez[7] in modeling the packing of a highly ionic thiapyrylium dye.

While there is no doubt that the Coulomb term contributes to the total packing energy, our results support the idea that the major determiner of packing geometry for closed-shell molecules is the short-range nonbonded interaction potential (for open-shell molecules in which strong intermolecular exchange forces exist, as in organic conductors and organic magnetic materials, electrostatic forces maybe more important).

For example, in Table 6 for the local minima of the inversion aggregates, we show the contribution of the nonbonded energy (column 2), the Coulomb energy (column 3), the total energy (column 4), and the ratio of the Coulomb to the total energies (column 5). There are many highly anisotropic, highly polar structures, many of them with significant π conjugation shown in Figure 5 for the inversion aggregate type, yet the average electrostatic contribution to the total energy is less than 5%, with the largest ratio being 22.3% for HFPCDN.

E. How Sparse Are the Important Local Minima? We have collected up to 700 local minima for each structure, but not all of these minima are unique. In order to determine the number of different structures, we will assume that two structures are the same if the RSS between them is less than 20.00. This means that their tilt angles and offset and translation repeat on average do not differ by more than 7° and 0.38 Å, respectively. To find these unique structures we did the following. (1) The structures were ordered by energy with the global minimum at the top of the list. This is the zeroth-order structure. (2) Starting with the global minimum as the first unique structure, we compared the remaining structures by computing the RSS with respect to the global. All of those that had RSS < 20.0 were taken to be similar to the global. We removed them from the list. (3) The structure now at the top of the list became the next unique structure. This is the first-order structure. The remaining structures were compared to it. Those with RSS < 20.00 were removed from the list. (4) Step 3 was repeated, finding the second-, third-, etc. order structures, until the list was exhausted.

We can now ask the question, in which order does the best local minimum lie for each structure? Table 7 and 8 shows the answers for the glide and screw aggregates, respectively. The local minima are reasonably sparse for the majority of the structures. For example, for ACESTB in Table 8, there are only eight local minima whose energy are lower (viz. more negative) than that of the observed X-ray structure. More than two-thirds of the structures have a rank ordering of 10 or less. There are some much higher, such as BAYNAQ with 58 and BAGYIR with 59, yet even these numbers are not daunting.

F. Outliers. There are seven aggregates for which the Monte Carlo predictions do not find a local minimum close to the observed (DEZXEL, BABDIR, BAGYIR, HCTDPY, GALHIK, GAJNIO, and GAKHIJ). A careful examination of the observed structures from the X-ray data indicate that for all of these, the aggregate is either not at a local minimum or the local minimum is extremely broad or shallow. For example, BAGYIR is not at or near a local minimum for rotation about the z axis. This was determined as follows. First the energy of the aggregate for the experimental X-ray structure was computed using eq 3. Next the center molecule was rotated in increments of ±5° about the x axis, a new aggregate structure was constructed keeping all the other construction variables constant, and the energy was recomputed. This was repeated for rotations about the y and z axes as well. The results are shown in Table 9. The first column shows the rotational increments ranging from −20° to +20°, and the successive columns show the computed energies for rotations about the x, y, and z axes, respectively. The results are quite clear. For rotation about x or y, the energy is at a minimum for a rotation of −5° away from the X-ray result, whereas for a z rotation, the energy has not reached a minimum value even for a +20° rotation. The local minimum structure found by the Monte Carlo cooling will thus deviate from the X-ray result by at least 20°, which accounts for the large RSS we find.

In the case of BABDIR, we find that the local minimum is very shallow, less than kT at 300 K for changes in the repeat distance of 0.5 Å. While the majority of the structures we have examined lie at or very near a well-defined local minimum, it is clear that there are examples where this is not the case.

Summary and Conclusions

We have developed a Monte Carlo cooling technique to quantitatively predict the structures of 1-dimensional screw, glide, and inversion aggregates. These self-assemblies occur naturally not only as part of full 3-dimensional crystal structures but also in lower dimensional monolayer structures such as Langmuir–Blodgett films[26] as aggregate states in solution (the so-called Scheibe aggregates[27]), in dye layers within polymer matrices,[28] as epitaxial layers for example on silver halide grains,[29] in smectic

(24) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525.
(25) Lee, C.; Subbiah, S. *J. Mol. Biol.* **1991**, *217*, 373.

liquid crystals,[30] and in lipid bilayers.[31] Our results suggest the first part of Kitaigorodskii's Aufbau principle (KAP), namely that 2- and 3-dimensional molecular packing geometries are made up of simpler 1-dimensional aggregates.

Most of the 60 structures examined here have Monte Carlo local minima which are close in energy to the observed structures, generally less than 6 kcal above the global minimum and sparsely populated.

There appears to be no correlation of the global or best local minimum structure with any particular energetic anisotropy of the full 3-dimensional crystal. Rather, the local minimum which is observed appears to be the one which allows for the best space-filling by other similar molecules of the cavities occurring in the 1-dimensional aggregate.

That the isolated 1-dimensional aggregate is itself in a local minimum is quite surprising. When we started this work we expected the Monte Carlo procedure not to find any structures close to that observed for the 1-dimensional case. What has been generally believed is that crystal packing forces tend to drive isolated structures uphill in energy in order to allow for close packing. While this is clearly so, the uphill drive is not arbitrary but toward another local minimum, one in which cavities can be filled by parts of other similar molecules. The isolated aggregates themselves appear to be thermodynamically stable entities trapped in a local well. Energetically, the uphill movement is not large, typically only a few percent of the total crystal energy.

Implications of KAP

A. Designing Novel Structures. Ideas similar to Kitaigorodskii's Aufbau principle have been used by others in attempts to systematize ways of understanding crystal structure packing. Lauher, Chang, and Fowler,[32] for example, used the concept of rods and layers combined with known preferred symmetries of amides and carboxylic acids to predict layer structures. Gavezzotti[4] used small 1- and 2-dimensional clusters combined with appropriate symmetry elements and quantitative structure property relationships as starting points to predict crystal structures of aromatic hydrocarbons. Ward[33] and Fagan and Ward[34] used 1- and 2-dimensional motifs coupled with electrostatic arguments to explain the prenucleation geometries of polycations and polyanions.

What we have shown here is that KAP can be made quantitative. Subunits of crystal structures are local energy minima. Given the symmetry type, the shape of the most important subunits which are the global minimum and nearby local minima can be determined unambiguously. There are several important implications of KAP. It should be possible to express the other local minima structures (viz. to force the appearance of local minima other than the one that nature likes) under the right conditions if the cavities could be filled by something else, such as small solvent molecules or other molecules different than the one in the aggregate itself. Knowing the packing geometry of the important local minima and thus the shape of the unfilled cavities should aid in the design of the types of molecules that will fit. For example, there are many descriptions of solvate structures in which the aggregate structure depends on the solvent used to crystallize it.[35] This design problem is very much akin to the lock and key problem in enzyme–substrate interactions.

B. Prediction of Monolayer and Full Crystal Packing. As we indicated in our discussion of the translation aggregates,[11] there is no way to predict which particular symmetry type the 1-dimensional aggregate will prefer since the difference in energies between the four types is not large (typically less than 2 kcal). Nevertheless, the tendency for 1-dimensional aggregates to be in local minima suggests simple procedures for predicting the structure of monolayers and crystal packing using KAP. Start by finding all the important translation, screw, glide, and inversion local minima for the 1-dimensional aggregate using Monte Carlo cooling as described here. Each of these minima can then be used as a building block in a Monte Carlo cooling procedure for the construction of one of the seven monolayer types describe by Scaringe.[6] The result will be a collection of monolayer local minima, each of which could then be used as a building block for obtaining the full 3-dimensional crystal structure. An idea similar to this was used by Pincus, Klausner, and Scheraga for the protein folding problem. They used an Aufbau principle to first find local minima of small peptides which were then linked to find the important local minima for a larger protein.[36]

Scaringe has clearly demonstrated that 2-dimensional layer structures of rigid molecules sit in local minima. That result has given us further impetus to develop methods which allow for the predictions of layer structures in which the molecules are not rigid but contain flexible units. We have used KAP to develop such methods and will present these methods in future papers.

Acknowledgment. The author would like to acknowledge the assistance of Kyle Werner for help in obtaining the experimental aggregate geometries from the crystal structure data and for running many of the Monte Carlo simulations. The author is also very grateful for many helpful discussions with Ray Scaringe on the symmetry aspects of this problem and discussions with Jim Eilers, Dennis Perchak, and John Hamilton on the nature of the molecular packing problem.

Appendix

Force Field. The force field used for the aggregate energy computations and the Monte Carlo simulations is the MM2 atom–atom nonbonded potential of Allinger[12] plus a Coulomb electrostatic term. The nonbonded potential for the interaction of two molecules labeled 1 and 2 is given by

$$E_{12}^{nb} = \sum_{i,j} A_{ij}\left[2.90 \times 10^5 \exp\left(\frac{-12.50 r_{ij}}{B_{ij}}\right) - 2.25\left(\frac{B_{ij}}{r_{ij}}\right)^6\right] \quad \text{(A1)}$$

where r_{ij} is the distance between atom i in molecule 1 and atom j in molecule 2, and the parameters A_{ij} and B_{ij} are given by

$$A_{ij} = (A_i A_j)^{1/2} \quad \text{(A2)}$$

(26) Ulman, A. *Introduction to Thin Organic Films*; Academic Press: New York, 1991.
(27) Herz, A. H. *Adv. Colloid Interface Sci.* **1977**, *8*, 237.
(28) Perlstein, J. H. In *Electrical Properties of Polymers*; Seanor, D., Ed.; Academic Press: New York, 1982; p 59.
(29) Maskasky, J. E. *Langmuir* **1991**, *7*, 407.
(30) Leadbetter, A. J. In *Thermotropic Liquid Crystals*; Gray, G. W., Ed.; John Wiley & Sons: New York, 1987; p 1.
(31) Shimomura, M.; Ando, R.; Kunitake, T. *Ber. Bunsen-Ges. Phys. Chem.* **1983**, *87*, 1134.
(32) Lauer, J. W.; Chang, Y.-L.; Fowler, F. W. *Mol. Cryst. Liq. Cryst.* **1992**, *211*, 99.
(33) Ward, M. D. *Pure Appl. Chem.* **1992**, *64*, 1623.
(34) Fagan, P. J.; Ward, M. D. *Sci. Am.* **1992**, 48.

(35) Unusual examples of this are the structures formed by 5,5′,6,6′-tetrachloro-1,1′,3,3′-tetraethylbenzimidazolocarbocyanine iodide, which forms no less than five crystal structures and aggregates in three of the four possible symmetries depending on the solvent of crystallization (ref codes CEBIMA, CEBIMM, DYEDCM, DYEETS, DYEMES).
(36) Pincus, M.; Klausner, R. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5107.
(37) In order to reproduce the results shown in this paper, the reader needs to know the starting geometry of one molecule for each of the aggregates. In addition, visual observation of the global minima, as well as the best local minima found by the Monte Carlo simulations, is extremely useful. The CHEM-X software allows for simple construction of crystal structures using unit cell, space group, and coordinate data from the Cambridge Structural Database file. In addition, the software has a powerful vector facility which allows the construction of vectors and their manipulation using vector algebra, and, via its user interface CHEMLIB, it gives the user the flexibility to write his own software routines to interface with the graphics manipulation and visualization. Accordingly, since CHEM-X as well as the Cambridge Structural Database is now generally available for most workstations, we have detailed in the supplementary material a CHEM-X description for obtaining the starting geometries as well as visualizing the various minima presented in this paper.

Table 10. Nonbonded Energy Parameters from MACROMODEL

atom	A_1 (kcal)	B_1 (Å)	atom	A_1 (kcal)	B_1 (Å)
C (sp^3)[a]	0.044	1.90	F	0.078	1.65
C (sp^2)	0.044	1.90	Cl	0.240	2.03
H	0.047	1.50	Br	0.320	2.18
O (sp^3)	0.050	1.73	I	0.424	2.32
O (sp^2)	0.066	1.74	S	0.202	2.11
N	0.055	1.82			

[a] For C–H bond $A_{CH} = 0.046$, $B_{CH} = 3.34$; all other A and B values are computed using eqs A2 and A3.

$$B_{ij} = B_i + B_j \tag{A3}$$

The parameters A_i and B_i for the atoms are from the molecular modeling program MACROMODEL,[13] which uses the MM2 force field, and are listed in Table 10. For a C–H bond, MM2 first shrinks r_{ij} for the bond by 8.5% before computing E^{nb}. The above equations are used as described in the text for computing the molecule–molecule potential for each aggregate type.

The electrostatic interaction between molecule 1 and 2 is given by

$$E_{12}^{el} = \sum_{i,j} \frac{q_i q_j}{\epsilon r_{ij}} \tag{A4}$$

where again the summation is over all atoms i in molecule 1 and atoms j in molecule 2 with empirical Gasteiger[14] charges q_i and a distance-dependent dielectric constant ϵ of the form

$$\epsilon = \epsilon_0 r_{ij} \tag{A5}$$

with $\epsilon_0 = 1.0$.

Best Fit Structures. Comparison of the local minima with the observed structures was done by computing the root sum square deviations (RSS) of each minima from the observed. Although the direction cosine matrix is orthogonal so that only three of the nine components are independent, we use the first and third rows of this matrix plus the offset and translations each weighted by a factor of 18 to compute RSS. For the local minimum of GAKPEN, for example (Table B2 in the supplementary material), the angles from the first and third rows of the matrix are as follows: row 1, 100.2, 169.5, and 88.0, row 3, 83.4, 89.1, and 6.67. For the observed the angles are as follows: row 1, 103.4, 166.6, and 90.8, row 3, 83.8, 92.2, and 6.56. The local minimum and observed offsets are 0.3 and 0.44 Å and the repeat distances are 7.93 and 7.643 Å, respectively. The deviations from the observed are then (3.2, −2.9, 2.8) and (0.4, 3.1, −0.11) for the angles, 2.52 for the weighted offset, and −5.22 for the weighted repeat distance, yielding RSS of 8.36, the result shown in Table 4.

Supplementary Material Available: A CHEM-X description for obtaining the starting geometries of the aggregates and visualizing the various minima described herein; Tables B1 and B2, giving the results for the screw and glide aggregates, respectively, in terms of the direction cosine matrix, offset, and repeat distances for the experimental data and the Monte Carlo data for the best best local minimum and the global minimum (12 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.