Molecular Self-Assemblies. 4. Using Kitaigorodskii's Aufbau Principle for Quantitatively Predicting the Packing Geometry of Semiflexible Organic Molecules in Translation Monolayer Aggregates

Jerry Perlstein

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

Received April 25, 1994[⊗]

Abstract: We make use of an aufbau principle, first suggested by Kitaigorodskii, to design a Monte Carlo cooling algorithm which can predict the local and apparent global energy minima of semiflexible molecules that are packed into translationally symmetric monolayer structures without any assumptions about the unit cell dimensions, molecular orientation, or exocyclic torsional conformation. We find the algorithm works effectively on molecules containing up to 12 exocyclic torsion bonds. Using the aufbau, the algorithm (a) packs molecules into 1-dimensional stacks generating a collection of local minima in stage 1, followed by (b) the packing of each of these minima into layers in stage 2. The only assumption is that the monolayer is made from a single molecular unit. The only additional information needed is the valence bond geometry of the molecule (viz. its atom connectivity, bond lengths, and bending angles, but not the exocyclic dihedral angles) and a suitable force field. We find, quite surprisingly, that the important features of the molecular orientation in the final monolayer packing geometry are already exhibited in stage 1 (but not the fine molecular conformational details), with the conformational details finally exhibiting themselves in stage 2. It is this expression of the orientational detail in stage 1 that makes the aufbau a practical quantitative tool for predicting the packing geometry of molecules with large numbers of single bonds. Coupled with a limited amount of experimental information, the aufbau can be used to determine which of the local minina in stage 2 are experimentally observable. The use of the aufbau for predicting full 3-dimensional crystal structures in a final stage 3 is discussed.

Introduction

The packing of organic molecules into regular assemblies is of considerable interest for the molecular engineering of Langmuir Blodgett films,[1] epitaxial grown films,[2] dyes aggregated in polymer films,[3] and other areas of materials science where a crystalline monolayer structure with known properties is desired. Experimental methods for determining these structures are oftentimes coupled with some molecular modeling to determine the packing geometry. For example, molecular modeling has been used to determine the packing of a triacontanoic acid monolayer from grazing incidence X-ray data.[4] Eckhardt and co-workers developed modeling techniques[5] in conjunction with their atomic force microscopy studies to predict the packing geometry of layers containing rigid norbornane molecules.[6] Möhwald and co-workers used modeling to determine the packing of cyanine dye monolayers from electron diffraction patterns.[7]

For 3-dimensional crystals of organic molecules, there have been a number of methods developed for predicting the packing mainly of rigid molecules. Holden, Du, and Ammon used an aufbau procedure coupled with a systematic search and energy optimization method.[8] Karfunkel and Gdanitz used a Monte Carlo simulated annealing method to obtain a collection of crude structures which were then independently energy optimized to find the best low-energy crystals for comparison to experiment.[9,10] Gavezzotti used a symmetry constrained small cluster approach to build nuclei of hydrocarbons from which the full crystal could be constructed.[11] Only rigid molecules were considered in these studies. Karfunkel and co-workers' method is amenable to inclusion of internal torsional degrees of freedom, and it has been tested on one molecule with two identical torsion bonds.[12] Extension of the method to molecules with more than two torsion bonds has not been reported. All these groups pointed out the considerable difficulty of including the torsion angles as variables.

It has recently been shown that *rigid* molecules (viz. molecules for which the exact conformation is known beforehand) can be close packed into 1-dimensional stacks[13,14] and

[⊗] Abstract published in *Advance ACS Abstracts,* November 15, 1994.

(1) Ulman, A. *An Introduction to Ultra Thin Organic Films*; Academic Press: Boston, 1991.

(2) (a) Patrick, D. L.; Beebe, T. P., Jr. *Langmuir* **1994**, *10*, 298-302. (b) Chau, L. K.; Arbour, C.; Collins, G. E.; Nebesny, K. W.; Lee, P. A.; England, C. D.; Armstrong, N. R.; Parkinson, B. A. *J. Phys. Chem.* **1993**, *97*, 2690-2698. (c) Chau, L. K.; England, C. D.; Chen, S. Y.; Armstrong, N. R. *J. Phys. Chem.* **1993**, *97*, 2699-2706.

(3) Perlstein, J. H. In *Electrical Properties of Polymers*; Seanor, D., Ed.; Academic Press: New York, 1982; pp 59-91.

(4) Leveiller, F.; Jacquemain, D.; Leiserowitz, L.; Kjaer, K.; Als-Nielsen, J. *J. Phys. Chem.* **1992**, *96*, 10380-10389.

(5) Eckhardt, C. J.; Swanson, D. R. *Chem. Phys. Lett.* **1992**, *194*, 370-374.

(6) Eckhardt, C. J.; Peachy, N. M.; Swanson, D. R.; Kim, J.-H.; Wang, J.; Uphaus, R. A.; Lutz, G. P.; Beak, P. *Langmuir* **1992**, *8*, 2591-2594.

(7) Bliznyuk, V. N.; Kirstein, S.; Möhwald, H. *J. Phys. Chem.* **1993**, *97*, 569-574.

(8) Holden, J. R.; Du, Z.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422-437.

(9) Karfunkel, H. R.; Gdanitz, R. J. *J. Comput. Chem.* **1992**, *13*, 1171-1183.

(10) Gdanitz, R. J. *Chem. Phys. Lett.* **1992**, *190*, 391-396.

(11) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622-4629.

(12) Karfunkel, H. R.; Rohde, B.; Leusen, F. J. J.; Gdanitz, R. J.; Rihs, G. *J. Comput. Chem.* **1993**, *14*, 1125-1135.

(13) Scaringe, R. P.; Perez, S. *J. Phys. Chem.* **1987**, *91*, 2394-2403.

(14) Perlstein, J. *J. Am. Chem. Soc.* **1992**, *114*, 1955-1963.

0002-7863/94/1516-11420$04.50/0

two-dimensional layers and that these low-dimensional packing arrangements represent local minima of real systems.[15,16] This work has prompted us to explore the question as to whether flexibility can be introduced at an early stage in the packing problem. The work presented here is an outgrowth of the pioneering efforts of Kitaigorodskii in the early 50's to determine the relationship between molecular geometry and the possible packing modes in one, two, and three dimensions.[17] The historical picture of this work and of other Russian and British workers has recently been put into perspective.[15]

Inherent in Kitaigorodskii's work is an aufbau principle which we have dubbed KAP. It can be used as a qualitative and a quantitative tool for both analyzing and predicting molecular packing geometry. KAP can be thought of as a staging process for the assembly of molecules in zero, one, two, and three dimensions. In stage 0 we have a single molecule or a finite collection of molecules not necessarily related by any symmetry. For example, KAP stage 0 could be one or more asymmetric molecular units that make up a full 3-dimensional crystal structure, or it could be a finite collection of symmetry related molecules such as a cyclic trimer, tetramer, or hexamer. The important point about stage 0 is that the number of molecules is finite. Since a number of studies have shown that the vast majority of crystal structures can be generated from a single molecular unit,[8,15] stage 0 will usually contain only a single molecule and we will focus our attention on this case.

KAP stage 1 is composed of the molecules in stage 0 stacked on top of one another with a single translational repeat distance. The important feature of stage 1 is that there is one and only one repeat distance but now the number of molecules is 1-dimensionally infinite. While the number of ways of stacking molecules in 1 dimension is large,[18] surprisingly, based on close packing arguments, there are only 4 which are of any significance;[13] the translation, glide, screw, and inversion stacks are shown schematically in Figure 1.

The geometric properties of these stacks and methods for generating and predicting their structure for *rigid* organic molecules have been discussed in great detail elsewhere.[13,14]

KAP stage 2 represents the bringing together of stage 1 stacks to form a 2-dimensional monolayer with two translational repeat distances as shown schematically in Figure 2. The first translation repeat is along the stage 1 stacks and the second is in a direction determined by the coupling of the stage 1 stacks. There are 80 symmetry types in which molecules can pack into monolayers all of which were categorized in the late 1920's. Wood[19] has presented these in terms of the more modern notation of the *International Tables of Crystallography*.[20] These layers can include structures containing higher order cyclic *n*-mers in stage 0 viz dimers, trimers, tetramers, and hexamers. Based on close packing arguments only seven types make up approximately 92% of all monolayers.[15] These types can all be constructed from various combinations of the stage 1 stacks of Figure 1 as shown by Kitaigorodskii[17] and more recently by Scaringe.[15]

Finally KAP stage 3 represents the full 3-dimensional crystal structure consisting of various combinations of stage 2 monolayers generating a third translation repeat.

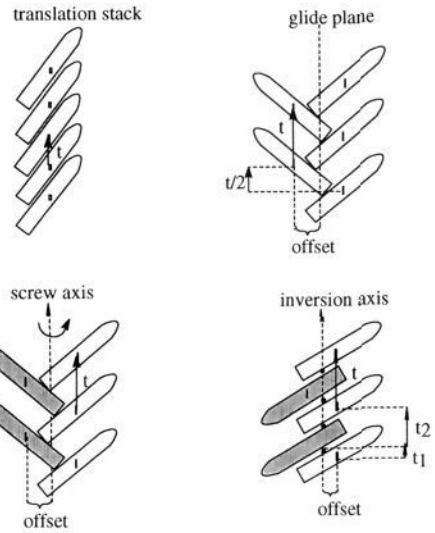

Figure 1. Schematic of the four most common types of Stage 1 molecular stacks with repeat distance, *t*, indicated by arrows. (Top Left) Simple translation stack: All molecules are related by simple translation. (Top Right) Glide Stack: Molecules on either side of the glide plane are related by reflecting in the plane and then translating *t*/2. Note that the centroids of the glide related molecules are offset from the plane. (Bottom Left) Screw stack: Like the glide stack but molecules are related by rotation of 180° about the screw axis before translation by *t*/2. (Bottom Right) Inversion stack: Molecules are related by changing sign of all coordinates (*xyz*) to (−*x*,−*y*,−*z*) across the inversion axis. Inversion points (●) are equally spaced at *t*/2. Note however that the molecular centroids in addition to being offset are not equally spaced along the inversion axis. The centroid spacing is such that $t_1 + t_2 = t$.

KAP as a Qualitative Tool. There are numerous discussions of using ideas like KAP particularly with regard to hydrogen bonding in solids. The ideas of H-bonded chains and layers were used extensively to categorize the structure of carboxylic acids[21] and amides.[22] More recently Whitesides and co-workers have used it for designing ribbon-like structures of melamines H-bonded to barbituric acids.[23] Lauher, Fowler, and co-workers used the symmetries of stage 1 and stage 2 to design ureadiamides which H-bond in two dimensions to form layers.[24] Ward and co-workers used similar ideas in elucidating the packing of molecules containing polyanions and polycations[25] and for designing monolayer motifs in 3-D crystals.[26] KAP is thus a useful way to view existing solid state structures and to offer direction for synthesizing new ones.

KAP as a Quantitative Tool. (a) Extracting Stacks and Layers from Known Crystal Structures. KAP's power comes not only from its qualitative significance but also from the quantitative information it can give about existing structures. We have recently shown that for rigid molecules stage 1 is a local energy minimum in the interaction potential of most organic molecules,[27] and stage 2 for rigid molecules has been

(15) Scaringe, R. P. In *Electron Crystallography of Organic Molecules*; Fryer, J. R., Dorset, D. L., Eds.; Dordrecht, 1990; pp 85−113.
(16) Scaringe, R. P. *Trans. Am. Cryst. Assoc.* **1992**, *28*, 11−21.
(17) Kitaigorodskii, A. I. *Organic Chemical Crystallography*; Consultants Bureau; New York, 1961; pp 65−112.
(18) Lauher, J. W.; Chang, Y.-L.; Fowler, F. W. *Mol. Cryst. Liq. Cryst.* **1992**, *211*, 99−109.
(19) Wood, E. A. *Bell Sys. Tech. J.* **1964**, *43*, 541−559 and references therein to the earlier catagorization of the layer groups.
(20) *International Tables for Crystallography*, Hahn, T., Ed.; Reidel: Dorderecht, 1983; Vol. A.
(21) Leiserowitz, L. *Acta Crystallogr. Sect. B* **1976**, *32*, 775−802.
(22) Leiserowitz, L.; Hagler, A. T. *Proc. R. Soc. London A* **1983**, *388*, 133−175.
(23) Zerkowski, J. A.; Seto, C. T.; Whitesides, E. M. *J. Am. Chem. Soc.* **1992**, *114*, 5473−5475.
(24) Chang, Y.-L.; West, M.-A.; Fowler, F. W.; Lauher, J. W. *J. Am. Chem. Soc.* **1993**, *115*, 5991−6000.
(25) (a) Ward, M. D. *Pure Appl. Chem.* **1992**, *64*, 1623−1627. (b) Ward, M. D.; Fagan, P. J. *Sci. Am.* **1992**, *267*, 48−54.
(26) Russel, V. A.; Etter, M. C.; Ward, M. D. *J. Am. Chem. Soc.* **1994**, *116*, 1941−1952.
(27) Perlstein, J. *J. Am. Chem. Soc.* **1994**, *116*, 455−470.

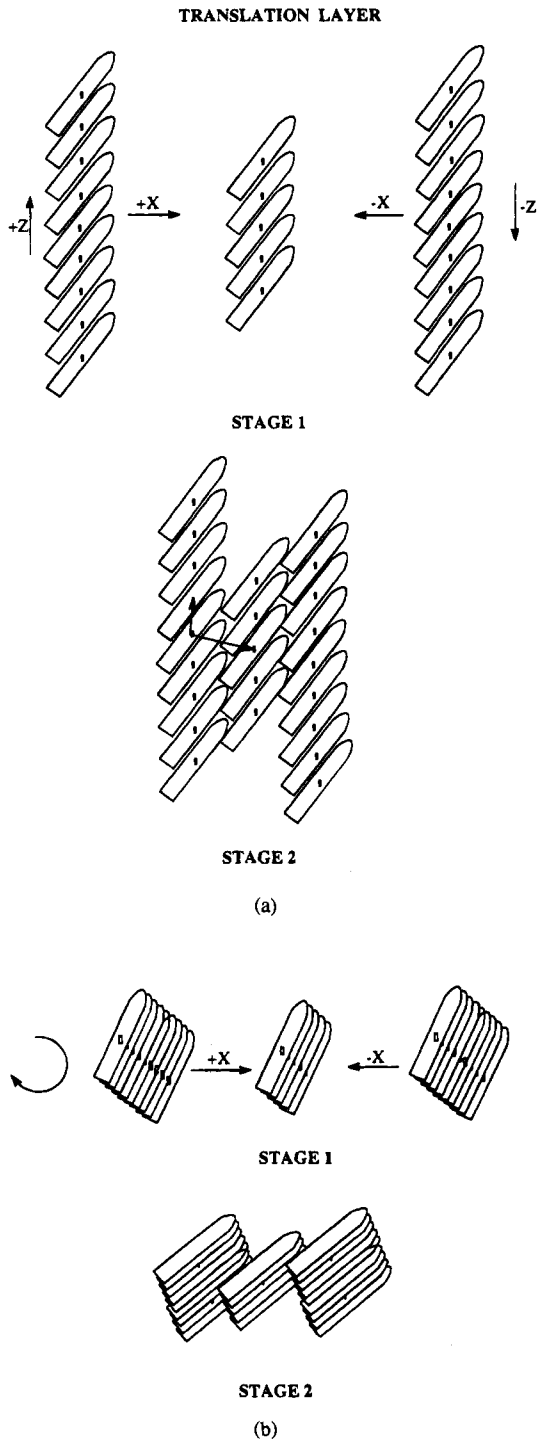

Figure 2. Construction of stage 2 monolayers from stage 1 stacks. (a) View looking along the repeat axis (z-axis) of three stage 1 stacks with a 5-molecule stack sandwiched between two 9-molecule stacks. Two of the three variables involved in making the stage 2 monolayer are shown. The two outer stacks are moved together in opposite directions along the x-axis and translated no more than $\pm t/2$ in opposite directions along the z-axis. (b) The third variable in the stage 2 monolayer formation. View looking down the repeat axis (slightly tilted) showing mutual rotation of all three stacks about the z-axis.

shown to be likewise.[15] These observations have a number of important consequences. Full 3-dimensional crystal structures can now be analyzed energetically in terms of the stage 1 and stage 2 substructures which make them up. By computing the interaction potential of a molecule with its nearest neighbors, next nearest neighbors, etc. the lowest energy stage 1 structures can be elucidated.[14,27] The stage 2 layers can then be extracted by computing the interaction potential of a molecule in a stage 1 stack with the molecules in neighboring stacks. The lowest energy layer in the crystal can be extracted this way (see below for details) and its physicochemical properties studied as a separate substructural entity.

(b) Predicting New Substructural Units. The stage 2 substructural layer of a 3-dimensional molecular crystal represents only one of many low-energy layers that can form. For a given molecule there can be many local energy minima belonging to the same layer type. The nature of this problem can be appreciated by looking at ECPRPR01[28] (see Figure 6 for molecular structures) with 10 internal rotational degrees of freedom. Even if there were only 3 possible conformations for each bond, this would amount to 3^{10} or approximately 59 000 possible conformations. Each of these conformations can form a 2-dimensional layer which itself can have more than 10^9 conformations associated with the orientational and translational variables of the layer. This results in a total conformation space for the layer of 5.9×10^{13} possible geometries. Nature has picked out only one or two of these to use as building blocks for stage 3. What we wish to demonstrate in this paper is that by applying KAP to semiflexible organic molecules and using a minimal number of assumptions, the low-energy stage 2 local minima (including the ones that nature likes) can be found by a careful application of a Monte Carlo cooling technique.

In what follows, we demonstrate for the first time the application of KAP for determining the packing of semiflexible organic molecules in the simplest of the seven-layer types, namely the translation layer defined as a layer with only translation symmetry elements. The results presented here represent the most extensive study to date on the inclusion of exocyclic torsion angles in the packing problem. The paper is divided into several parts as follows: (a) an overview of the method, (b) computational details for the Monte Carlo method applied to translation layers, (c) results for nine molecules with 0 to 12 torsion angles, and (d) summary and conclusions.

An Overview of the Method

Figure 3 displays a flow diagram for implementation of the method. Starting with the valence bond geometry of a single molecule in a random torsional conformation of low energy, a Monte Carlo cooling procedure TCHAIN[29] is implemented to generate 700 local minima for the translation stack. Each local minimum is generated by cooling from 4000 to 300 K and then repeating the process 700 times. At each temperature in the cooling process, a single Monte Carlo step consists of a random change of one of the external variables in the construction

(28) Allen, F. H.; Kennard, O.; Taylor, R. *Acc. Chem. Res.* **1983**, *16*, 146–153. Reference codes are from the Cambridge Structural Database, 12 Union Road, Cambridge CB2 1EZ, UK, and are as follows: (a) DOVCAS: Cordes, A. W.; Hojo, M.; Koenig, H.; Noble, M. C.; Oakley, R. T.; Pennington, W. T. *Inorg. Chem.* **1986**, *25*, 1137–1145. (b) ETIMQO12: Ito, T.; Sakurai, T. *Acta. Crystallogr., Sect. B* **1973**, *29*, 1594–1603. (c) BPOXSH: Cheng, P.-T.; Nyburg, S. C. *Acta Crystallogr., Sect. B* **1976**, *32*, 930–932. (d) DADDUH: Drendel, W. B.; Sundaralingam, M. *Acta Crystallogr., Sect. C (Cryst. Struct. Commun.)* **1985**, *41*, 950–953. (e) JIPBIT: Green, T. P.; Galinis, D. L.; Wiemer, D. F. *Phytochemistry* **1991**, *30*, 1649–1652. (f) DICNIM: Hadicke, E.; Graser, F. *Acta Crystallogr., Sect. C (Cryst. Struct. Commun.)* **1986**, *42*, 189–195. (g) ABMHFO: Pettus, J. A., Jr.; Wing, R. M.; Sims, J. J. *Tetrahedron Lett.* **1977**, 41–44. (h) ECPRPR01: Berkovitch-Yellin, Z. *Acta Crystallogr. Sect. B* **1980**, *36*, 2440–2442. (i) VARHUR: Mandel, J. B.; Douglas, B. E. *Inorg. Chim. Acta* **1989**, *155*, 55–69.

(29) Copies of the routines TCHAIN and TLAYER, supporting routines for the starting geometries and analysis for use with CHEM-X/CHEMLIB implemented on SGI's or IBM RS/6000's, are available from the author upon request. Contact the author at internet address lsaw00@risque.chem.rochester.edu or perlstein@chem.chem.rochester.edu.

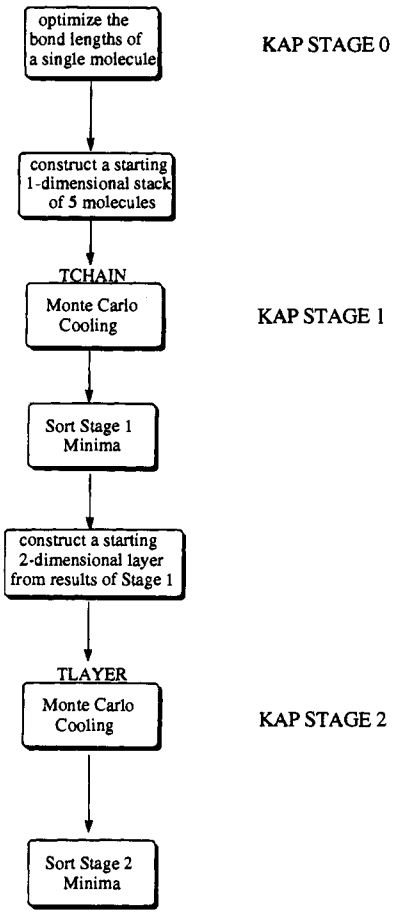

Figure 3. Flow chart for assembly of monolayers. Starting with a single molecule in stage 0, the monolayer is gradually built up by Monte Carlo cooling first to form translation stacks in stage 1 and then to form layers in stage 2.

procedure for the stack as described in detail previously[14] plus a random change in one of the internal torsion variables. The newly constructed stack is then accepted or rejected based on the Metropolis algorithm.[30] Twenty Monte Carlo steps are carried out at each temperature before the temperature is decremented by 10%. The choice of 20 steps and 700 local minima is purely heuristic, but was found to be adequate for molecules containing up to 12 exocyclic torsion bonds.

The most significant result of this stage 1 process is as follows: *Among the 700 local minima collected there is at least one structure whose external-variable geometry is close to the observed X-ray geometry even though the internal torsional geometry is still very far from the observed.* This result is rather remarkable in light of the large number of shapes that the molecule can have simply by rotating the single bonds. It implies that the external and internal variables are partially separable. There are many molecular conformations which have the same 1-dimensional packing pattern. The 1-dimensional packed molecules in stage 1 have a geometry which in essence allows for the ends of the molecules to "wave in the breeze". It is this observation which makes KAP a useful quantitative packing tool. In addition we find that for the translation layer the local minimum which is close to the observed is usually no more than 10 kcal above the apparent global minimum.

The 700 local minima output of TCHAIN are then passed on to TLAYER[29] containing a second round of Monte Carlo cooling for stage 2 with the following change in procedure: (a) at 4000 K, none of the stage 1 external stack variables are changed. Only the external layer variables and internal torsional variables are changed. The purpose of this first step is to select an acceptable starting point on the hypersurface and to get this starting layer moving energetically downhill without loss of information of the 1-dimensional stack geometry obtained in stage 1. Clearly if one of the stage 1 minima already has the correct packing geometry we do not want to disturb this geometry very much in stage 2. (b) After the 4000 K step, the layer is quenched to 400 K before continuing. This quickly brings the stage 1 stacks into contact. (c) Since the stack–stack interactions can be considerably weaker than the intrastack interactions, and since the torsion energy is considerably smaller than both of these, the final temperature for the cooling cycle is set to 40 K. (d) For each stage 1 minimum used in stage 2, the cooling cycle is repeated only 10 times.

Seven thousand local minima can be collected by TLAYER (10 minima for each of the 700 stage 1 stacks) which can then be compared with experiment. A considerable improvement in collection efficiency is obtained however if the 700 minima found by TCHAIN are first sorted by geometry. Not all the minima are unique. In most cases there are less than 200 unique stacks and of these there are usually fewer than 70 with an energy within 10 kcal of the apparent global minimum. A sorting method has been described elsewhere.[31] The 7000 layers can thus be conveniently reduced to 700 by first carrying out a geometry and energy presort on TCHAIN's output prior to starting TLAYER.

Below we present the complete computational details of the method and apply it to examples taken from the Cambridge Structural Database.

Computational Details

(a) Extracting Experimental Layers from the Cambridge Crystallographic Database. The Cambridge Structural Database[28] is a rich source of experimental layer geometries contained as substructures of full 3-dimensional crystal structures. The most likely crystals in which to find translation layers are those in space group $P1$ (or $P\bar{1}$ when the molecules sit on a center of symmetry), but translation layers can be found in any space group (the DICNIM layer is from space group $C2/c$ and DOVCAS is from Pc). To find these layers, the molecule is packed into as many unit-cell repeats as necessary so that at least one molecule in the lattice is surrounded by all possible nearest neighbors. If the initial asymmetric unit is not a full molecule, the asymmetric unit is packed until a full molecule is obtained and then this molecule is used as the asymmetric unit to pack the crystal. It is convenient to do this visually using any of a number of available software packages. Because of its open architecture which allows us to incorporate our own software, we use CHEM-X and CHEMLIB for this purpose.[32] The interaction energy (see below for details of the energy computation) of this one test molecule with all the others is then computed.[33] If the test molecule is in a translation layer, then the two lowest energy interactions are with molecules that form a translation stack. The repeat length along this stack forms one of the repeat lengths of the layer. As viewed down the stack axis, the next lowest energy interactions are with molecules in neighboring stacks related to the first one by simple translation. A vector from the test molecule to a molecule in the neighbor stack is the second repeat length. The two repeat lengths and the angle between them define the unit cell for the layer. It should be noted here that this unit-cell is not unique. There are many combinations of translation repeat lengths that could be used to define the translation layer. This complicates the final analysis (*vide infra*).

(30) Binder, K. *Topics in Current Physics*; Springer-Verlag: New York, 1984; Vol. 46.

(31) Perlstein, J. *Chem. Mater.* **1994**, *6*, 319–326.

(32) CHEM-X is a molecular modeling program developed and distributed by Chemical Design Ltd., Roundway House, Cromwell Park, Chipping Norton, Oxfordshire OX7 5SR, UK. CHEMLIB is a CHEM-X interface which allows the user to link his own subroutines to CHEM-X.

(33) FORTRAN 77 code for this computation for use with CHEM-X/CHEMLIB is available from the author upon request.

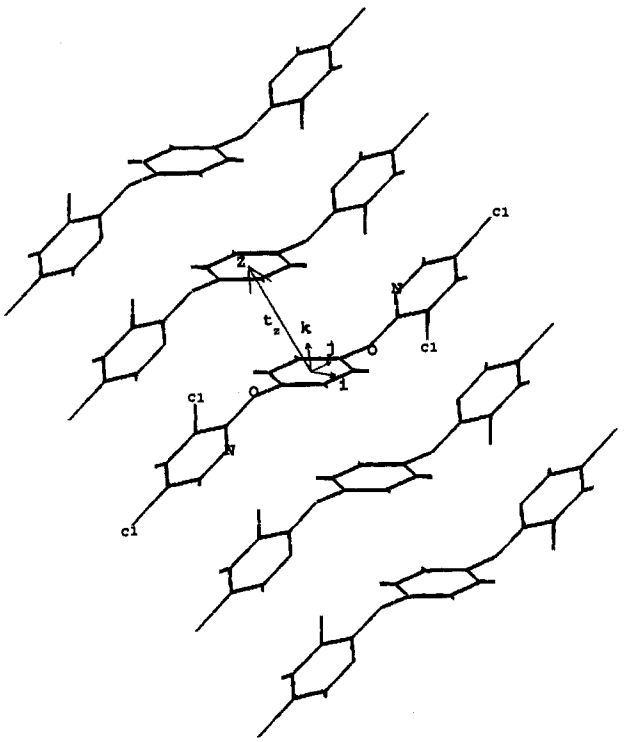

Figure 4. A representative translation stack of DADDUH. The repeat distance t_z lies along the z-axis. Orthogonally placed dummy atoms on the center molecule define the molecular reference frame **i**, **j**, **k** which rotates with the molecule. The value of t_z and the three angles formed between **i**, **j**, **k**, and **z** quantitatively define the packing geometry of the translation stack.

Once it has been established that the molecule in question forms a low-energy translation layer, this layer can be extracted from the structure and used as the experimentally observed result to compare with the Monte Carlo simulations described below.

(b) Starting Geometry KAP Stage 0. The starting geometry is a single molecule extracted from the experimental layer geometry described above with atom coordinates referenced to an orthogonal coordinate system with origin at the molecular centroid. We find it convenient to use CHEM-X for all graphic visualizations, initializations, and final analysis of results. All exocyclic single bond torsion angles are then randomized and optimized using CHEM-X's VDW optimizer prior to a full optimization of all the bond lengths (but not bond angles) using CHEM-X's level 1 optimizer.[34] The result is a relaxed stage 0 starting geometry in a convenient local minimum whose torsion angles are far from the experimentally observed ones. The purpose of this optimization step is twofold: (a) to provide a low-energy reference state for the internal torsional energy calculations during the Monte Carlo simulation and (b) to remove any forcefield "stresses" due to slight mismatches between the observed and forcefield computed bond distances. Without this optimization, rotations about the single bonds produce considerable noise in the internal energy computation.

(c) Starting Geometry KAP Stage 1. The starting geometry for the stage 1 Monte Carlo simulation consists of five stage 0 molecules oriented with their centroids (represented by dummy atoms) spaced 15 Å apart along a common z-axis. (The position of the centroid is computed from the coordinates of either all the atoms or a small group of atoms in the molecule. In the molecules studied here, the centroid was computed using the atoms connected by dark solid lines indicated in Figure 6.) Attached to the third molecule of the group are three additional dummy atoms arranged along the x-, y-, and z-axis. These three dummies form three orthogonal unit vectors as shown in Figure 4 which rotate with the molecule and serve as a molecular frame of reference for quantitative comparison with experimental geometries. The same three vectors are also placed in the same starting orientation in the experimental layer.

(d) Monte Carlo Cooling Stage 1. Routine TCHAIN for the Monte Carlo cooling was written in FORTRAN 77 and implemented on an IBM RS/6000 model 530 UNIX workstation. There are three external variables, two molecular rotations θ_x and θ_y about the x- and y-axes and one translation r_z along z. (Rotation about the z-axis produces no new structures for a translation stack.) The angle variables are grouped together so that for each Monte Carlo step either the angles are both changed or the translation is changed, the decision being made by a random number generator. In addition, during each step, a single torsion angle ω is changed. The maximum sizes of the changes depend on temperature and are as follows: external variables—for $T = 4000$ K, max $\Delta\theta_x$ = max $\Delta\theta_y = \pm 90°$ and r_z = any value in the range 3 to 15 Å, and for all other T, max $\Delta\theta_x$ = max $\Delta\theta_y = \pm 10°$ and r_z = any value in the range 3 to 15 Å; internal variables—for $T = 4000$ K, max $\Delta\omega = \pm 90°$, and for all other T, max $\Delta\omega = \pm 10°$.

Large changes in the angle variables were made at 4000 K in order to start the cooling simulation at different points on the hypersurface for each temperature cycle. Below 4000 K, the variable changes were kept small in order to search for the local minima in that region.

Table 1. Weighting Scheme for the Monte Carlo Variables in KAP Stage 2[a]

temp, K	stack rotation (z-axis)	stack translation (x-axis)	stack translation (z-axis)	each torsion angle	all molecule rotations	z-repeat
4000	3	3	3	1	0	0
400−40	3	3	3	1	1	1

[a] Numbers for each variable represent the number of Monte Carlo steps performed for that variable at each temperature.

(e) Sorting the Stage 1 Local Minima. The 700 local minima collected in stage 1 are sorted by similarity in packing geometry. A sorting method has been described elsewhere[27,31] but modified by the possible pseudo-2-fold rotation symmetry as described in the analysis section below. Briefly, the 700 local minima are first ordered by energy. The lowest energy structure is taken as the apparent global minimum of rank order = 0. All other structures are compared to this one as reference and those with an RSS < 10 (see analysis section for definition of RSS) are considered identical to the apparent global minimum and removed from the list. The next lowest energy structure remaining is rank order = 1. All remaining structures are compared to it and the similar ones removed from the list. This process is repeated to find structures of rank order = 2, 3, etc. until the list is exhausted. Of the 700 minima found in stage 1 only the 70 unique lowest energy structures (rank order 0 to 69) are used in KAP stage 2. These 70 are usually within 10 kcal of the apparent global minimum. Note that this sorting routine is internal to the 700 minima and in no way makes use of any experimental packing information.

(f) Starting Geometry KAP Stage 2. The starting geometry for the layer in stage 2 consists of three stage 1 molecular stacks placed side by side spaced 35 Å apart along the x-axis as shown schematically in Figure 2. The center stack contains five molecules and the two outer stacks contain nine molecules each. There are three degrees of freedom for the layer simulation as follows: (a) mutual rotation of the stacks about the z-axis, (b) mutual displacement of the outer stacks in opposite directions along the x-axis, and (c) mutual displacement of the outer stacks in opposite directions along the z-axis.

(g) Monte Carlo Cooling Stage 2. As for stage 1 the simulation for stage 2 is again started at 4000 K allowing for large changes in the layer variables and the internal torsion variables, 135° for all rotations and 6 Å for the x displacement. The z-displacement is limited by the repeat distance in the stack. To preserve the stage 1 stack geometry at 4000 K, no changes in the 1-dimensional stack variables are allowed. A Monte Carlo step consists of a random change in any one of the layer variables or one torsion variable. The number of steps at each temperature is determined by a weighting scheme. We use a weighting scheme shown in Table 1 to determine how many times a variable is allowed to change at each temperature. This weighting scheme was found by trial and error and applies to all molecules containing up to 12 torsions. The scheme shown is minimal. Larger numbers can be used if desired at the expense of longer cpu times. The weights shown represent the fewest number of Monte Carlo steps that can be done at

(34) The Level 1 optimizer in CHEM-X allows for geometry optimization by change of bond lengths keeping all bond angles fixed.

each temperature without seriously affecting the outcome. Note that at 4000 K, the intrastack weights for the external variables are 0. The intrastack variables are not changed at this temperature. Also note that the intrastack variables are turned on below 400 K. After the 4000 K step, the layer is quenched to 400 K where all variables (including one step for each of the stack variables) are allowed to change by small amounts (10° for all angles and 0.6 Å for all displacements including the stage 1 stack variables). This allows for the small intrastack adjustments that have to occur as the 1-dimensional stacks begin to interpenetrate to form the layer. Thus for a molecule with say five rotatable bonds, the number of Monte Carlo steps would be 14 at 4000 K (3 steps each for the x and z stack translations, and z rotation, 1 step for each torsion) and 16 at 400 K and below (adding one step for the molecular rotation about x and y (treated as a group), and one step for the z-repeat translation). This process of three Monte Carlo steps for the layer variables and one step for the stack variables and torsions is continued, lowering the temperature by 10% after each set of steps until the temperature reaches 40 K at which point the lowest energy layer is saved and the temperature cycle repeated nine more times. The choice of 40 K for the final temperature is predicated on the fact that for small torsional energies (of order 1 to 2 kcal) the rotational conformations will only "freeze out" at temperatures well below room temperature. For the nine molecules, the average number of Monte Carlo steps in stage 2 was 326 000 with an acceptance ratio of 47%.

(h) Sorting of Stage 2 Minima. As with stage 1 minima, not all of the 700 minima collected in stage 2 are unique. They can be internally sorted by geometry in the same way as described above for stage 1, paying particular attention to the fact that in comparing two minima, all possible orientations of the layers as well as pseudosymmetry comparisons have to be considered as described in the analysis section below. Two layers are considered identical if their root sum square deviation (RSS) < 20.

(i) Energy Computations. We compute the energy, E_t, for packing molecules into a 1-dimensional stack or a 2-dimensional layer by the force field eq 1

$$E_t = E^{nb} + E^{el} + (E^{int} - E^0) \tag{1}$$

where E^{nb}, the nonbonded energy between molecules in the layer, is given by the nonbonded term of the MM2 force field equation of Allinger[35]

$$E^{nb} = \frac{1}{2}\sum_{i,j} A_{ij}\left[2.90 \times 10^5 \exp\left(\frac{-12.50 r_{ij}}{B_{ij}}\right) - 2.25\left(\frac{B_{ij}}{r_{ij}}\right)^6\right] \tag{2}$$

and E^{el} is a Coulomb term with a distance dependent dielectric constant, $\epsilon(r)$, with $\epsilon_0 = 1.0$

$$E^{el} = \frac{1}{2}\sum_{i,j} \frac{q_i q_j}{\epsilon(r) r_{ij}} \qquad \epsilon(r) = \epsilon_0 r_{ij} \tag{3}$$

In eqs 2 and 3 the sum is taken over all atoms i in a single molecule with all other atoms j in all the other molecules of the stack or layer (four molecules in stage 1 and 22 molecules in stage 2) and the $^1/_2$ is inserted to prevent double counting. The single molecule containing the i atoms is conveniently chosen to be the third molecule in the five molecule stack of stage 1 or the third molecule in the middle stack of stage 2. The parameters A and B are taken from MACROMODEL[36] (which uses a modified MM2 force field as one of its choices) and the charges q are empirical charges of Gasteiger.[37] E^{int} is the internal energy of a single molecule relative to the energy E^0 of the molecule in the starting geometry and is given by

$$E^{int} = E^{nb} + E^{el} + E^{tor} \tag{4}$$

In eq 4, E^{nb} and E^{el} are given by eqs 2 and 3 but with the summations over all atoms in a single molecule that are separated by at least three bonds (viz. at least 1−4 interactions) and E^{tor} is the MM2 torsion energy given by

$$E^{tor} = \sum_{\substack{\text{all bond}\\ \text{dihedrals}}} \left\{\frac{V_1}{2}(1 + \cos\omega) + \frac{V_2}{2}(1 - \cos 2\omega) + \frac{V_3}{2}(1 + \cos 3\omega)\right\} \tag{5}$$

In eq 5 the summation is over all dihedral angles, ω, for each bond numbered in Figure 6 and the V_1, V_2, and V_3 parameters are those given in MACROMODEL.[36]

The choice of molecular starting geometry has been described above and represents a convenient stage 0 local minimum. E^0 is computed for this molecule using eq 4. This need not be the apparent global minimum but should be chosen such that eq 1 satisfies the relation

$$\text{abs}(E^{int} - E^0) \ll \text{abs}(E^{nb} + E^{el})$$

This can be determined very quickly at the start of the stage 1 simulation and guarantees that the internal energy changes will not dominate the packing calculation.

(j) Analysis of Stage 1 and Stage 2 Results and Comparison with Experiment. The resulting low-energy structures found in stage 2 simulation are compared to the experimental X-ray structures. The analysis is complicated by three factors: (1) the unit-cell for the translation layer is not unique, (2) the Monte Carlo layer orientation for stage 2 or stack orientation for stage 1 and the X-ray orientation can differ by a 180° rotation about x, y, or z, and (3) if the molecule possesses a pseudo-2-fold rotation axis, the Monte Carlo molecular frame and the molecular frame in the X-ray structure may differ by 180° rotation about the pseudoaxis (see Figure 5 for an example). Below is a complete deescription of the analysis method.

(1) The Non-unique Unit-Cell Problem. Since the experimental unit-cell is not unique, all possible unit-cells for the stage 2 simulation have to be computed and compared to the X-ray structure as follows: For each local minimum define all possible unit cell vectors **a** and **b** for every combination of non-collinear translation vector. For each of these vector pairs, define an orthogonal coordinate system as follows:

$$\mathbf{z} = \mathbf{b}$$

$$\mathbf{y} = \mathbf{b} \times \mathbf{a}$$

$$\mathbf{x} = \mathbf{y} \times \mathbf{z}$$

consisting of a z-axis parallel to the unit-cell b-axis, a y-axis perpendicular to the unit-cell, and an x-axis perpendicular to y and z. The tilt angles α_{iz}, α_{jz}, α_{kz} of the molecules are then computed from the direction cosine matrix of the molecular coordinate system **i**, **j**, **k**(using the dummy atoms attached to the third molecule in the middle stack) with respect to the z-axis. For every possible unit-cell, the root sum square deviation (RSS) of the unit-cell vector lengths, a and b, the unit-cell angle γ, the molecular tilt angles, and the stack rotation angle, ϕ_z, with respect to the starting geometry were computed and compared to the observed X-ray layer defining RSS as follows:

$$\text{RSS} = \{(\alpha_{iz} - \alpha_{iz0})^2 + (\alpha_{jz} - \alpha_{jz0})^2 + (\alpha_{kz} - \alpha_{kz0})^2 + (\varphi_z - \varphi_{z0})^2 + (18a - 18a_0)^2 + (18b - 18b_0)^2 + (\gamma - \gamma_0)^2\}^{1/2} \tag{6}$$

where α_{iz0}, α_{jz0}, α_{kz0}, φ_{z0}, a_0, b_0, and γ_0 are the corresponding tilt angles, stack z-axis-rotation angle, and unit-cell dimensions for the experimentally observed X-ray structure. The unit-cell deviations are weighted by an arbitrary factor of 18 so that a 0.3 Å length deviation corresponds to a 5.4° angle deviation.

A similar equation is defined for the RSS of stage 1 stacks

$$\text{RSS} = \{(\alpha_{iz} - \alpha_{iz0})^2 + (\alpha_{jz} - \alpha_{jz0})^2 + (\alpha_{kz} - \alpha_{kz0})^2 + (18t_z - 18t_{z0})^2\}^{1/2} \tag{7}$$

where the α's are as defined above and t_z and t_{z0} are the translation repeat distance for the simulation and the observed X-ray stack, respectively.

(35) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127−8134.
(36) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440−467.
(37) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219−3228.

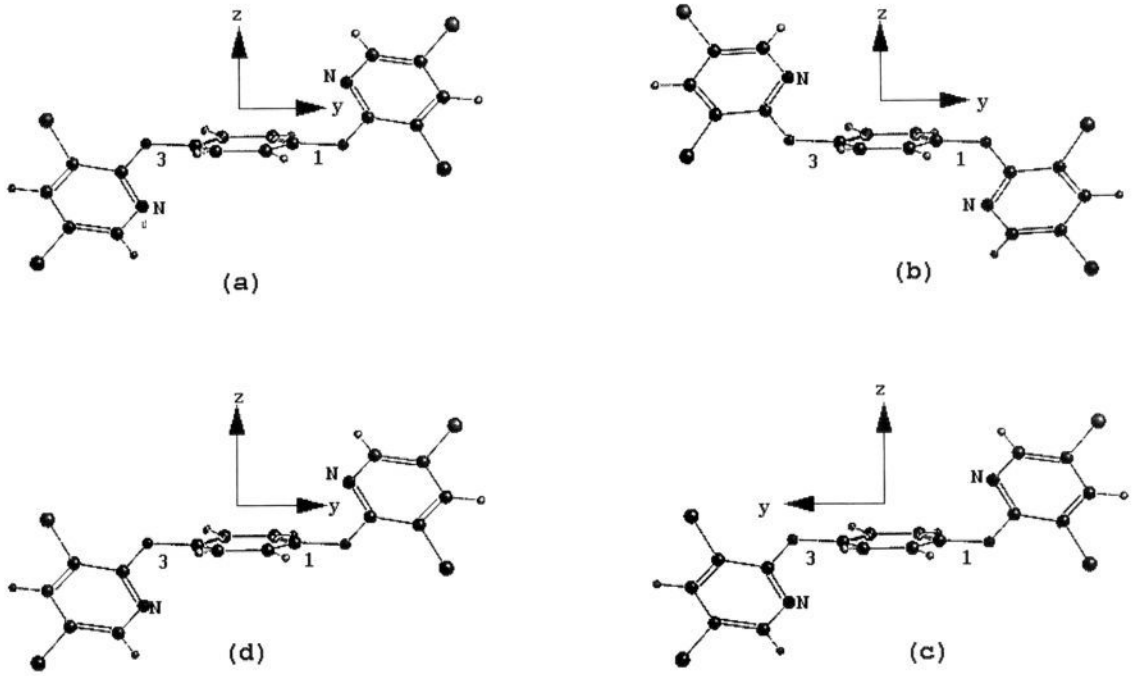

Figure 5. The problem encounted in the analysis of DADDUH which possesses pseudo-2-fold rotation axes (as also occurs for ETIMQO12, BPOXSH, and DICNIM which sit on a center of symmetry). In part a the molecule is shown with its molecular frame of reference. During the course of the simulation bonds 1 and 3 can rotate by 180° as in part b. The resulting conformation is similar to that in part a which can be seen by rotating the whole molecule about the z-axis as in part c except that the molecular frame is not oriented the same way. To superimpose the two molecules including the frame, the frame in part c has to be rotated 180° about the z-axis as in part d.

(2) The Orientation Problem. The orientation problem is solved simply by recomputing RSS three more times for layer or stack reorientations of 180° rotations about the **x**, **y**, and **z** axis.

(3) The Pseudo-2-Fold Axis Problem. If the molecule possesses a possible pseudo-2-fold rotation axis as shown in Figure 5, the entire analysis described above is repeated with the molecular frame first rotated about these axes.

(k) Comparison of the Dihedral Angles. For both stage 1 and stage 2 local minima, one dihedral angle for each rotatable single bond was compared with the X-ray structure and the root mean square deviation (RMSD) computed from eq 8

$$\mathrm{RMSD} = \sqrt{\sum_i (\omega_i - \omega_0)^2 / N} \qquad (8)$$

where $\omega_i - \omega_0$ is the difference between the Monte Carlo prediction and the X-ray data and N is the number of angles. The rotatable bonds are numbered in Figure 6. The four atoms used to compute each of the dihedral angles for comparison with the observed X-ray conformation are indicated by a dot (●). Where there is ambiguity in choice, the next appropriate atom is indicated first by an asterisk (*) and then by a diamond (♦).

Results and Discussion

Nine molecules that form translation layers as the lowest energy layer type within their crystal structures were randomly picked from the Cambridge Structural Database and are indicated by reference code[28] in Figure 6. These molecules have from 0 to 12 exocyclic torsion bonds as well as various kinds of heteroatoms, nonplanar rings, and molecular symmetries. In carrying out the Monte Carlo simulations, no assumptions were made concerning molecular symmetry, orientation, or exocyclic torsion angle conformation. Every effort was made to remove any preconceived prejudice about what the final packing geometry and torsion conformation should look like. We took particular care to use a starting exocyclic torsional geometry that was far from the observed X-ray structure. This is indicated in Table 6 by the large RMSD values for the starting geometries as computed using eq 8. For the external variables the RSS values for stage 1 stacks were computed using eq 7 and for stage 2 layers using eq 6. For the stage-2-layer deviation calculation an RSS of 20 for example would correspond to an average angular error of 7.6° and a unit cell length deviation of 0.42 Å.

To give some idea of the range of layers explored, we show in Table 2 the energy of the nine layers studied and the energy of crystals from which they come as computed from eq 1 with the internal torsional energy set to 0 ($E^{\mathrm{int}} - E^0 = 0$). For an isotropic lattice containing 12 nearest neighbors in the 3-dimensional crystal structure and 6 nearest neighbors in the layer the ratio of the layer energy to the crystal energy would be 50%. The layers in this study came from crystals covering a wide range of anisotropy ranging from almost isotropic (ABMHFO) to almost lamella (DICNIM). Nevertheless we always find a local minimum in stage 2 simulations which corresponds to the observed layer.

Results for KAP Stage 1. Table 3 presents the stage 1 simulation results for the nine molecules shown in Figure 6. For each molecule the energy of the best local minimum structures (as defined to be the one with the smallest RSS) relative to the apparent global minimum is given in column 3. The RSS and RMSD deviations of the apparent global minimum and the best local minimum are presented in columns 4 through 7 and the rank order of the best local minimum is given in column 8. The RSS and RMSD values for the best local minima are highlighted.

The RSS values for the best local minima are all small in the range from 1.75 to 10.69, whereas the RMSD values are very large for these same structures. These results indicate that the molecular shape is not all that important for the packing of molecules in 1-dimensional stacks. *There are many shapes which can adopt the packing pattern corresponding to the ordered experimental one.* The implications of this are quite

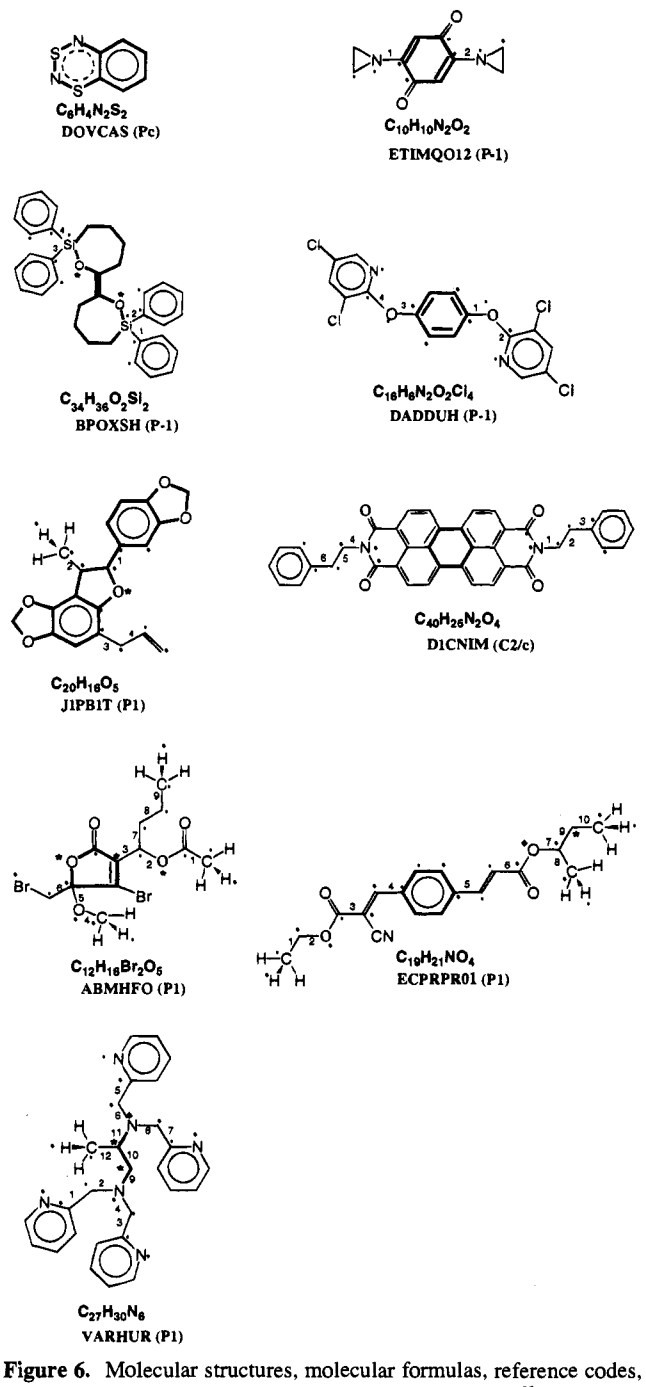

Figure 6. Molecular structures, molecular formulas, reference codes, and space groups from the Cambridge Structural Database.[28] Connected atoms used to compute the centroid are indicated by darker solid lines. Torsion bonds used in the simulation are numbered. For each bond, the dihedral angle used to compute the RMSD (in eq 8) is computed using the atoms indicated by a dot. An asterisk and then a diamond are used where there is ambiguity.

important for the next stages of the simulation. We do not have to explore the entire hyperspace for the molecular shape in order to find the few internal conformations which will pack into the observed 1-dimensional stacks. There are in fact many of them and we need only find one. This is what KAP stage 1 does. It finds at least one molecular conformation which is not necessarily near the observed experimental conformation but whose shape is sufficiently close to the experimental one to allow for molecular packing which is close to the observed conformation. Having found this one, the hyperspace for the layer packing in KAP stage 2 can then be explored but with considerable

Table 2. 2-Dimensional Layer Energy, 3-Dimensional Crystal Energy, and Energy Anisotropy

	Energy, kcal/mol		
REFC[a]	2-D layer	3-D crystal	ratio (2D/3D), %
ABMHFO	−26.62	−47.99	55.5
VARHUR	−49.55	−83.63	59.3
BPOXSH	−68.24	−112.3	60.8
DOVCAS	−20.16	−31.47	64.1
ETIMQO12	−29.46	−45.91	64.2
JIPBIT	−46.13	−65.14	70.8
DADDUH	−50.26	−66.39	75.7
ECPRPR01	−47.07	−61.92	76.0
DICNIM	−94.28	−115.8	81.4

[a] Reference codes are from the Cambridge Structural Database.[28]

Table 3. Monte Carlo Results for Stage 1 (1-Dimensional Stacks)

		local	RSS[d]		RMSD[e]		rank
REFC[a]	torsions[b]	energy[c]	global	**local**	global	**local**	order[f]
DOVCAS	0	1.31	38.86	**10.06**	0	**0**	41
ETIMQO12	2	0.66	25.71	**10.69**	6.12	**10.29**	4
BPOXSH	4	8.54	80.83	**2.37**	45.92	**77.67**	10
DADDUH	4	2.54	5.85	**1.75**	76.85	**34.90**	1
JIPBIT	4	7.49	16.41	**2.08**	152.3	**145.5**	1
DICNIM	6	3.36	12.66	**2.13**	37.35	**28.08**	3
ABMHFO	9	5.96	20.03	**5.61**	192.9	**195.3**	56
ECPRPR01	10	10.34	85.12	**9.86**	82.81	**81.05**	44
VARHUR	12	8.37	30.03	**8.21**	61.84	**75.45**	65

[a] Reference codes are from the Cambridge Structural Database.[28] [b] The number of torsion bonds used in the simulation. [c] Energy difference in kcal between the best local minimum (as defined by the smallest RSS) and the apparent global minimum. [d] The root sum square deviation from the observed X-ray structure for the apparant global energy minimum and the best local minimum as computed from eq 7. [e] The root mean square deviation of the dihedral angles computed from eq 8 for the torsion bonds numbered in Figure 6. [f] The rank order for the best local minimum defined to be the number of unique structures whose energy is less than the best local minimum.

constraint imposed by the stage 1 results. Note also in Table 3 that the best local minima lie in a 10 kcal window above the apparent global minimum (column 3) and that the worst ranking is 65 for VARHUR (column 8). In KAP stage 2 we can then limit our search to the first 65 unique structures found in stage 1 or to those structures which are within about 10 kcal above the apparent global minimum energy.

Results for KAP Stage 2. Results for stage 2 are presented in Tables 4−6. In Table 4 the energy of the best local minimum structure (as defined to be the one with the smallest RSS) relative to the apparent global minimum is given in column 3. The RSS deviation for the apparent global minimum and the structure closest to the observed conformation is given in columns 4 and 5 with the rank order in column 6. The best structures are all within 3 kcal of the apparent global minimum. The rank orders are mostly 1 or 2 but with some larger such as ABMHFO. The RSS of the best local minimum structures are all low with the worst being ECPRPR01 at 26.11 which is also the apparent global minimum. These layer structures thus lie energetically close to the apparent global minima but are structurally very different from them as indicated by the considerably larger RSS values for the apparent global minima (column 4). Figure 7 through 10 display some of these differences.

In Figure 7 for DOVCAS, the structural difference between the apparent global minimum and the best local minimum is illusory. The two layers are mirror images of one another (and not superimposable). Even though the molecule is achiral, the translation layer is chiral with both chiral layers occurring in the full 3-dimensional crystal structure.

Table 4. Monte Carlo Results for Stage 2 (Layer Structures)

REFC[a]	torsions[b]	best local energy[c]	RSS[d] global	RSS[d] local	rank order[e]
DOVCAS	0	0.24	50.84	11.10	1
ETIMQO12	2	2.97	64.31	23.95	13
BPOXSH	4	2.0	29.52	5.81	2
DADDUH	4	1.05	42.45	5.98	2
JIPBIT	4	0.38	30.23	8.28	1
DICNIM	6	2.88	48.07	17.75	2
ABMHFO	9	2.68	35.45	4.71	17
ECPRPR01	10	0	26.11	26.11	0
VARHUR	12	2.71	82.94	16.56	9

[a] Reference codes are from the Cambridge Structural Database.[28] [b] The number of torsion bonds used in the simulation. [c] Energy difference in kcal between the best local minimum (as defined by the smallest RSS) and the apparent global minimum. [d] The root sum square deviation from the observed X-ray structure for the apparent global energy minimum and the best local minimum as computed from eq 6. [e] The rank order for the best local minimum defined to be the number of unique structures whose energy is less than the best local minimum.

Table 5. Comparison of Unit-Cell Constants between the Predicted Stage 2 Local Minima and the Observed X-ray Structures

REFC[a]	torsions[b]	structure[c]	unit cell constants[d] a	b	γ
DOVCAS	0	local	5.94	3.82	86.47
		X-ray	5.616	3.896	90.00
		deviation	0.32	−0.08	−3.53
ETIMQO12	2	local	7.17	3.81	105.73
		X-ray	6.810	3.863	99.35
		deviation	0.36	−0.05	6.38
BPOXSH	4	local	8.92	9.12	113.71
		X-ray	9.226	9.169	114.63
		deviation	−0.31	−0.05	−0.92
DADDUH	4	local	7.27	3.86	106.17
		X-ray	7.268	3.964	104.68
		deviation	0	−0.10	1.49
JIPBIT	4	local	10.46	4.48	91.63
		X-ray	10.174	4.686	93.47
		deviation	0.29	−0.21	−1.84
DICNIM	6	local	9.67	4.60	97.90
		X-ray	10.33	4.75	103.31
		deviation	−0.66	−0.15	−5.41
ABMHFO	9	local	7.70	7.57	110.86
		X-ray	7.755	7.429	111.25
		deviation	−0.06	0.14	−0.39
ECPRPR01	10	local	7.22	4.61	110.73
		X-ray	6.941	5.263	103.11
		deviation	0.28	−0.65	7.62
VARHUR	12	local	9.91	6.05	98.58
		X-ray	9.570	6.258	105.78
		deviation	0.34	−0.21	−7.20

[a] Reference codes are from the Cambridge Structural Database.[28] [b] The number of torsion bonds used in the simulation. [c] Local: the best local minimum whose structure has the lowest RSS value (from eq. 6). X-ray: experimental unit cell constants from ref 28. Deviation: unit cell difference between the best local minimum structure and the X-ray structure. [d] Unit cell constants: a and b in Å, γ in deg.

Figure 9 shows the apparent global minimum and the best local minimum for DADDUH. The two structures are energetically only 1.05 kcal apart (Table 4) but are structurally very different (RSS = 42.45 compared to 5.98). What is the difference? Visual examination indicates that the two layers differ by a 109° rotation about the stacking axis. Note the large cavity in the local minimum structure which is absent in the apparent global minimum structure. This cavity has the correct complimentary shape to accommodate the chloropyridine ring of the next layer in the full 3-dimensional crystal and to lower the total energy more than the packing of layers with the global minimum structure.

In contrast to this look at the difference between the apparent global and best local minima of JIPBIT (Figure 10) (RSS = 30.23 compared to 8.28). The major distinguishing feature is the difference in internal rotation angles of bonds 1 and 4. In the apparent global minimum these bonds are rotated such that the methylenedioxo group and the ethylene group stick out the top and bottom of the layer creating large gaping cavities in the structure which cannot be filled by another layer to make a crystal. These cavities are absent in the best local minimum.

Unit-Cell Predictions. Table 5 presents the unit-cell predictions for the best local minima whose RSS's are given in Table 4 and compares them to the experimental values. The method for making this comparison for each of the 700 local minima is described above in the analysis section. Since the choice of the layer cell is arbitrary the experimental unit-cell dimensions reported in Table 5 are not necessarily those reported in the literature. We chose them so that the b-axis was along the lowest energy stage 1 stack and the angle between a and b was $\geq 90°$.

In making this comparison we wanted to emphasize the deficiencies of the forcefield so that no attempt was made to optimize the experimental structure. The results show that the deviations from experiment are not large for most of the examples, typically less than 0.35 Å for one cell dimension although there are some unexplained larger values for the a-axis of DICNIM and the b-axis of ECPRPR01. The direction of these deviations can be positive or negative. In general the deviations are typical of what has been found by other workers packing rigid molecules.[8,9]

Prediction of Dihedral Angles. Comparison of the dihedral angles with experiment is shown in Table 6. For each rotatable bond we show the initial value for one dihedral angle in the starting geometry, followed by the simulation results for the apparent global minimum, the best local minimum, the experimental result, and the deviation of the best local minimum from the experimental. In the last column is the RMSD computed from eq 8. Bond numbering and the specific angle are indicated in Figure 6 for each molecule. The large RMSD values for the starting geometry (greater than 60° in all cases) indicates how far from the observed conformation the simulation was started. In setting up the starting geometry we manually rotated all the bonds 180° (subject to steric constraints), 90° (for bonds with 2-fold symmetry, e.g. phenyl rings), or 60° (for 3-fold symmetry, e.g. methyl groups) from their observed configurations prior to initial optimization. For example, if two bonds were syn in the experimental structure, we set them anti in the starting structure and vice versa. We wanted to preclude any possibility of starting the simulation close to the observed conformation. Results of stage 2 simulation for each structure are described below.

DOVCAS. There are no exocyclic torsion bonds for this molecule.

ETIMQO12 (2 Torsions). The dihedral angle comparison is very good with an RMSD of only 1.96. Note that the predicted dihedral angles are approximately equal and opposite in sign as would be expected for a centrosymmetric structure although this was not assumed in the simulation. Comparison of the apparent global and local dihedral angle geometry is shown in Figure 8. Part of the 2.97 kcal difference in energy between these two conformations (Table 4) comes from a 1.94 kcal conformation energy difference in favor of the global conformation. This is a simple example where 3-dimensional packing requirements forces the molecule to take up both a higher internal energy conformation and a higher layer energy conformation. This is very common in most of the layers we

Table 6. Comparison of Computed and Observed Dihedral Angles (deg) for Stage 2 Layer Structures

	bond no.												
	1	2	3	4	5	6	7	8	9	10	11	12	RMSD[a]
DOVCAS	no exocyclic dihedrals												
ETIMQO12													
initial[b]	−58.75	−62.43											91.53
global[c]	−141.02	+138.45											72.98
local[d]	−66.01	+64.11											**1.96**
X-ray[e]	−66.77	+66.77											
deviation[f]	0.76	2.66											
BPOXSH													
initial[b]	−47.69	+6.91	−30.07	−32.37									61.46
global[c]	−4.60	−28.59	+28.93	+4.11									36.44
local[d]	−56.23	−31.49	+54.63	+31.35									**5.38**
X-ray[e]	−60.88	−36.66	+60.88	+36.66									
deviation[f]	4.65	5.17	6.25	5.31									
DADDUH													
initial[b]	−87.65	−66.84	−86.20	−64.18									88.76
global[c]	−125.56	+2.08	+122.91	+6.14									9.86
local[d]	−129.80	+15.81	+125.85	8.98									**10.16**
X-ray[e]	−129.03	+10.29	+129.03	−10.29									
deviation[f]	0.77	5.52	3.18	19.27									
JIPBIT													
initial[b]	+136.06	−179.25	−96.14	+60.42									151.22
global[c]	+142.09	−179.88	+113.74	+103.81									109.05
local[d]	−40.10	−177.15	+68.08	−114.73									**7.92**
X-ray[e]	−52.41	180.00	+76.60	−119.06									
deviation[f]	12.31	2.85	8.52	4.33									
DICNIM													
initial[b]	+85.90	+52.64	−92.92	+84.50	+56.51	+88.98							73.55
global[c]	+97.36	+170.47	−82.11	−100.69	−172.54	+85.60							7.45
local[d]	+99.31	+171.10	−87.78	−97.37	−166.17	+86.37							**9.25**
X-ray[e]	+94.65	+171.51	−72.00	−94.65	−171.51	+72.00							
deviation[f]	4.66	0.41	15.78	2.72	5.34	14.37							
ABMHFO													
initial[b]	−179.52	+162.71	−166.92	+178.48	−142.15	−52.74	+64.31	+109.36	+59.88				86.15
global[c]	+166.82	+71.44	+63.21	−172.11	−48.22	+64.34	−58.82	−168.96	+53.45				39.76
local[d]	+162.00	+77.47	+61.32	+170.31	−54.21	+68.56	180.00	+175.80	+56.30				**7.70**
X-ray[e]	180.00	+74.93	+59.04	180.00	−60.01	+72.51	−174.46	+172.54	+60.00				
deviation[f]	18.00	2.54	2.28	9.69	5.80	3.95	5.54	3.26	3.70				
ECPRPR01													
initial[b]	−177.52	−78.53	+90.13	−32.30	−159.96	+29.31	+57.83	−152.13	+150.77	−48.64			103.10
global[c]	+179.19	−178.38	−179.58	−2.19	−174.77	+178.83	−152.22	+175.58	+68.23	−171.27			14.07
local[d]	+179.19	−178.38	−179.58	−2.19	−174.77	+178.83	−152.22	+175.58	+68.23	−171.27			**14.07**
X-ray[e]	180.00	+145.74	−177.26	−7.20	−171.35	−176.72	−129.07	180.00	+68.14	180.00			
deviation[f]	0.81	35.88	2.32	5.01	3.42	4.45	23.15	4.42	0.09	8.64			
VARHUR													
initial[b]	−79.46	+80.46	+74.56	+69.52	+76.93	−162.75	−127.02	+54.91	+76.89	−176.83	+64.77	+165.31	85.32
global[c]	−66.95	+173.92	−132.55	−162.23	+113.51	+171.90	−42.85	−54.98	−61.49	−38.22	−56.46	−172.44	93.14
local[d]	−147.55	+72.46	+36.51	−162.83	−82.10	+161.99	+108.14	−62.02	−164.03	+173.11	−78.43	+170.12	**41.77**
X-ray[e]	−129.88	+73.68	+50.68	−164.16	+136.77	+162.72	+123.13	−73.68	−161.85	−178.76	−76.69	+178.38	
deviation[f]	17.67	1.22	14.17	1.33	141.13	0.73	14.99	11.66	2.18	5.65	1.74	8.26	

[a] Root mean square deviation of the dihedral angles from the observed X-ray data (as computed from eq 8). [b] Dihedral angles of the initial optimized starting geometry Stage 0. Specific angles are for bonds numbered in Figure 6. [c] Predicted dihedral angles for the apparent global minimum. [d] Predicted dihedral angles for the best local minimum structure. [e] Observed dihedral angles from X-ray data ref 28. [f] Deviation of dihedral angles for the best local minimum structure.

have studied. The apparent global minimum monolayer is rarely the observed structure.

BPOXSH, DADDUH, JIPBIT (4 Torsions). BPOXSH and DADDUH are both centrosymmetric. Except for bond 4 of DADDUH, the simulation predicts the expected magnitude and sign alternation of the dihedral angles with a small RMSD.

BPOXSH is another example of how a small energy difference in internal energy can result in a dramatically different packing geometry. The apparent global and local minima while differing only by 2 kcal in total energy (Table 4) are nevertheless structurally very different as a result of the difference in torsion angles for the phenyl rings in the two layers. The contribution of the internal energy of the molecule to this difference amounts to only 0.92 kcal.

In the case of DADDUH, the dihedral conformation of the apparent global and local minima are similar (RMSD = 9.86 compared to 10.16) yet the two layers are structurally very different (see Table 4, Figure 9, and discussion above).

JIPBIT is the first of four molecules in the list which is not centrosymmetric. The predicted dihedrals are close to the observed conformation with the largest deviation being 12.31° for bond 1.

DICNIM (6 Torsions). DICNIM is a large centrosymmetric insoluble perylene dicarboximide pigment molecule of interest as an electron accepting component in electrophotographic films.[38] There are at least 18 known perylene dicarboximide crystal structures with various substituents on the chromophore.[28f] Many of them form translation layers. A study of the stage 1 one-dimensional stacking behavior for all of them as rigid molecules has been described.[31] The simulation correctly

(38) Law, K.-Y. *Chem. Rev.* **1993**, *93*, 449−486.

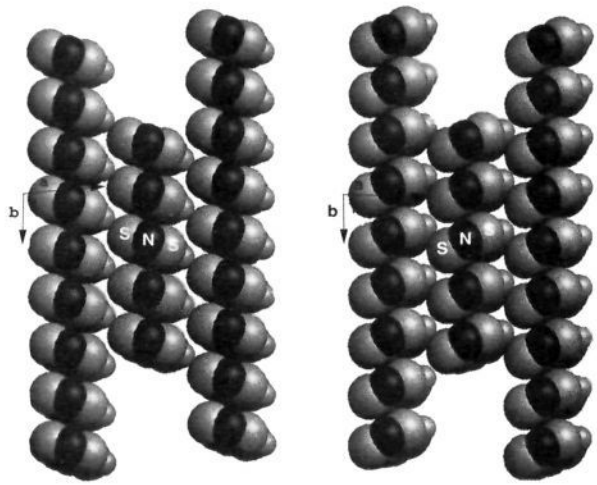

Figure 7. The stage 2 layer packing geometry for the apparent global minimum of DOVCAS on the left and the best local minimum structure on the right? Also shown are typical unit cell vectors **a** and **b** and angle γ with values for the best local minimum structure given in Table 5. The layer plane is parallel to the paper with the two S and one N atoms labeled (the other N atom is hidden from view). The two layers are related by a mirror plane perpendicular to the paper.

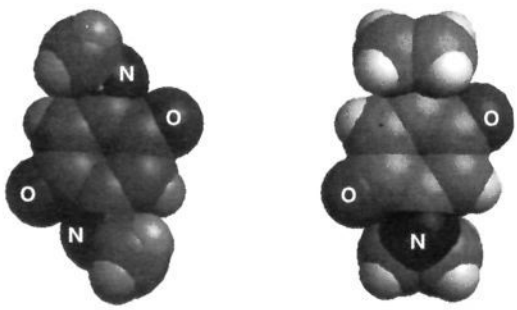

Figure 8. Apparent global minimum molecular conformation (left) and best local minimum molecular conformation (right) for ETIMQO12 with the heteroatoms labeled. In both conformations the ethyleneimine rings are anti. In the apparent global minimum conformation the three-membered rings are perpendicular to the quinone plane whereas in the local minimum conformation the rings lie above and below the quinone plane.

predicts the magnitude and sign alternation of the angles expected for a molecule with a center of symmetry. The largest deviations are for bonds 3 and 6 which by being close to the layer surface are floppy.

ABMHFO (9 Torsions). The chiral molecule ABMHFO crystallizes in space group $P1$ with nine rotatable single bonds. The predicted dihedral angles for the local minimum are quite close to the observed conformation with an RMSD of only 7.7°.

ECPRPR01 (10 Torsions). This molecule is one of a class of asymmetrically substituted dienes of interest in solid-state photopolymerization reactions with a centrosymmetric backbone. ECPRPR01 crystallizes in space group $P1$ with ten rotatable single bonds (the two ester linkages are considered partially double and not rotated). The dihedrals for the local minimum (which is also the apparent global minimum) have RMSD = 14.07°. Examination of the angles and visual examination of the layer structure indicates that bonds 2 and 7 deviating 35.88° and 23.15° respectively from the observed are floppy since they lie close to the layer surface. These two bonds account for most of the RMSD.

VARHUR (12 Torsions). This molecule contains the largest number of torsions we have considered for a translation layer.

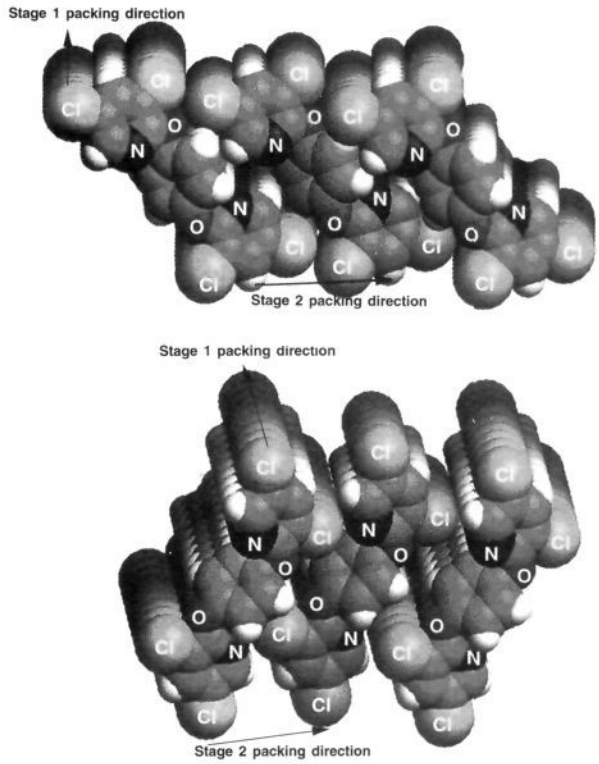

Figure 9. Cross-sectional view (slightly tilted to show the stage 1 stacking) of the layer packing for DADDUH. Both the apparent global minimum structure (top) and best local minimum structure (bottom) have the pyridine rings in the anti configuration with the same dihedral geometry. The two layers however differ by 109° rotation of the stage 1 stacks about the stacking axis.

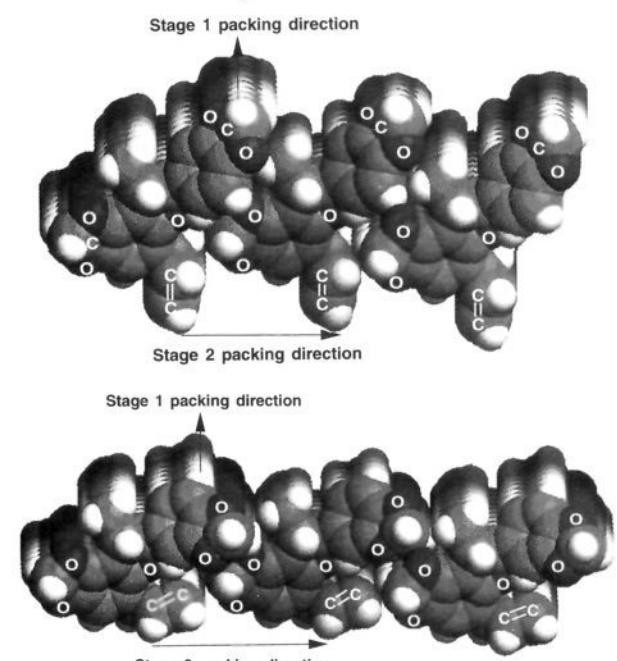

Figure 10. Cross-sectional view (slightly tilted to show the stage 1 stacking) of the layer packing for JIPBIT with the O and ethylene C atoms labeled. In the global packing (top), the methylenedioxo group sticks out of the top of the layer surface, the ethylene group sticks out of the bottom. In the best local minimum (bottom), 180° rotation about bond 1 and 142° rotation about bond 4 forces these groups to point into the layer. (See Figure 6 for bond numbering.)

The RSS of 16.56 for the best local minimum is small (Table 4) as are the unit-cell deviations (Table 5), but surprisingly the

RMSD of the dihedrals is very large (RMSD = 41.77). A look at the individual angles indicates that all except bond 5 are close to the observed conformation. Bond 5 differs from the observed conformation by 141.13°, viz. the nitrogen of the pyridine ring on the wrong side is an indication that the forcefield is not distinguishing clearly the difference between N and C−H (they have about the same van der Waals radius).

Summary

We have demonstrated that Kitaigorodskii's aufbau principle can be used as a quantitative tool for the analysis of packing geometries in molecular self-assemblies. It allows for the extraction of the lowest energy 1-dimensional stacks and 2-dimensional layers from the 3-dimensional crystal structures of organic molecules. In addition when coupled with a Monte Carlo cooling algorithm, it can be used to find the most important local minima of stacks and layers.

Using KAP, we have demonstrated that the important local minima of monolayers with simple translational symmetry containing molecules with up to 12 torsional degrees of freedom can be predicted, and that one of these minima occurs as the lowest energy monolayer in the full 3-dimensional crystal structure regardless of the anisotropy of the lattice energy.

Conclusions

There are several important conclusions that can be made about these results:

(1) An arbitrarily shaped organic molecule has many possible monolayer geometries represented by the local minima found by the Monte Carlo simulation. The minima we have found here belong to the translation layer type. There are of course other layer types which are possible and we are in the processes of writing the Monte Carlo code to generate these. Which of the possible monolayer minima will express themselves depends on the local environment. The particular local minimum that occurs for example in the three-dimensional crystal structure will be the one that allows for the best packing of the layers in stage 3 of KAP. This need not be the same layer that occurs in another environment such as in Langmuir films at the air−water interface, or on a metallic surface, or in a polymer matrix. There is already clear recognition of this by the fact that the same molecule with different conformational geometry can crystallize in different crystal structures[39] containing different monolayer geometries. Monolayer polymorphs at the air−water interface are also beginning to be recognized.[40] *Nevertheless, the collection of local minima that one can generate by the Monte Carlo technique should contain all the important ones of interest.*

(2) if the topology of the surface layer is such that the apparent global minimum (or any other minimum for that matter) has cavities or hollows which cannot be filled by other molecules, then it is unlikely that such a layer will form as it cannot be close packed. It seems reasonable to use that it should be possible to discard local minima as realizable possibilities by some quantitative measure such as the local surface curvature.[41,42] A preponderance of values greater than some cutoff value (say that for a hydrogen or carbon atom) could be eliminated.

(3) It should be possible to sift through the Monte Carlo local minima using limited experimental information to pin down experimental packing geometries more precisely. For example, molecular surface area measurement in Langmuir−Blodgett experiments can be compared with the predicted surface areas from the simulations to focus on one or two of the possible minima. We have done this for squarylium dye monolayers to pinpoint their structures.[43] Other kinds of experimental information can be used such as molecular tilt angle determinations from polarized absorption measurements,[44] and aggregate spectral shifts,[45] periodicities from atomic force microscopy,[46] scanning tunneling microscopy,[47] or grazing incidence X-ray diffraction[48] can each be used to sift through the local minima.

Table 7. Cpu Times for Some Monte Carlo Simulations of KAP Stages 1 and 2 (IBM RS/6000 Model 530-AIX 3.1 Code Written in FORTRAN 77)

		cpu time[b]		
structure[a]	atoms/molecule	stage 1	stage 2	total time[b]
DOVCAS	14	0.5	1	1.5
ETIMQO12	24	1.5	4.5	6.0
DADDUH	32	2.75	2.5	5.25
ABMHFO	35	4	17	21
JIPBIT	43	5	19.75	24.75
ECPRPR01	45	5	20	25
VARHUR	63	12	69.5	81.5
BPOXSH	74	13.75	70	83.75

[a] Reference codes are from the Cambridge Structural Database.[28] [b] Cpu time in hours.

(4) How many torsional variables can our simulation algorithm handle? In its present form we believe 12−13 internal variables can be confidently handled for the translation layer. We routinely do 8−10 torsions in molecules containing up to 80 atoms without difficulty. CPU time starts to become a factor as shown in Table 7. As the atom count approaches 100 atoms, the cpu time can become excessive. However improvements in our present coding technique such as introducing van der Waals or electrostatic cutoff distances[49] should alleviate this problem. Some preliminary results for glide layers on molecules containing up to 17 torsions indicates no specific problems with the present technique except long cpu times.[50] It seems reasonable to us at this stage in the development of the algorithm that the number of torsional variables that can be handled will be determined mainly by the size of E^{int}. The routine should remain robust as long as $abs(E^{int} - E^{0}) \ll abs(E^{nb} + E^{el})$. This condition could be seriously violated for example if the molecule folded up onto itself during the course of the simulation.

(5) The inclusion of endocyclic torsions should be possible. Various methods for doing endocyclic torsions have been described[51,52] including Monte Carlo methods.[53,54]

(6) The inclusion of hydrogen bonding should also be possible provided the force field is properly parametrized for H-bonds. H-bonded systems offer an interesting case to both analyze and predict using KAP. Since the H-bonded interaction is mainly electrostatic,[55] it will be of interest to see if KAP stage 1 and stage 2 packing geometries are realizable in which the electro-

(39) Bernstein, J. In *Organic Solid State Chemistry*; Desiraju, G. R., Ed.; Elsevier: Amsterdam, 1987; pp 471−518.
(40) Fujimoto, Y.; Ozaki, Y.; Kato, T.; Matsumoto, N.; Iriyama, K. *Chem. Phys. Lett.* **1992**, *196*, 347−352.
(41) Olson, A. H.; Duncan, B. S. *Biopolymers* **1993**, *33*, 219−229.
(42) Olson, A. J.; Duncan, B. S. *Biopolymers* **1993**, *33*, 231−238.
(43) Chen, H.; Herkstroeter, W. G.; Law, K.-Y.; Perlstein, J.; Whitten, D. B. *J. Phys. Chem.* **1994**, *98*, 5138.
(44) Kawai, T.; Umemura, J.; Takenaka, T. *Langmuir* **1989**, *5*, 1378−1383.
(45) Czekkely, V.; Forsterling, H. D.; Kuhn, H. *Chem. Phys. Lett.* **1970**, *6*, 207−210.
(46) Schwartz, D. K.; Viswanathan, R.; Garnaes, J.; Zasadzinski, J. A. *J. Am. Chem. Soc.* **1993**, *115*, 7374−7380.
(47) Roder, H.; Hahna, E.; Brune, H.; Bucher, J.-P.; Kern, K. *Nature* **1993**, *366*, 141−143.
(48) Jacquemain, D.; Wolf, S. G.; Leveiller, F.; Deutsch, M.; Kjaer, K.; Als-Nielsen, J.; Lahav, M.; Leiserowitz, L. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 130−152.
(49) Brooks, C. L., III; Pettitt, B. M.; Karplus, M. *J. Chem. Phys.* **1985**, *83*, 5897−5908.

static term now becomes important. The results we have shown here and elsewhere[27,31] and the results of Scaringe[15] and Perez[13] indicate that in the absence of H-bonds or other exchange forces, the electrostatic contribution to the packing geometry is not overly significant.

(7) It should be possible to predict the packing geometry of *semiflexible* molecules in their full 3-dimensional crystal structure using the technique described here in the next stage of KAP. As we pointed out in the introduction, several methods have already been developed for doing this,[8,9,11] but none has been demonstrated that can pack molecules containing more than two torsion bonds as variables. As we have demonstrated, except for some bonds close to the layer surface, most of the dihedral angles have already been determined in stage 2. The monolayer local minima found in KAP stage 2 can be used as starting points for a Monte Carlo simulation of the packing of these layers in KAP stage 3. Any floppy bonds close to the layer surface should resolve themselves in this last stage.

(8) The design of organic solids requires control of the packing arrangement of the molecules.[56] Based on the results presented here, molecular engineering of monolayer self-assemblies using semiflexible molecules should no longer be a major obstacle. Until now there has been a tendency to limit the design of monolayers to molecular systems which were either rigid, contained large rigid parts, or were isomorphous with pre-existing structures and for which the packing was either simple or easy to calculate. This is no longer necessary. Complex molecular shapes with non-obvious packing arrangements can now be treated in a reasonably simple way using KAP and the Monte Carlo cooling technique described here.

(50) Perlstein, J. To be submitted for publication.
(51) Goto, H.; Osawa, E. *J. Chem. Soc., Perkin Trans. 2* **1993**, 187–198.
(52) Goodman, J. M.; Still, W. C. *J. Comput. Chem.* **1991**, *12*, 1110–1117.
(53) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.
(54) Ferguson, D. M.; Raber, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 4371–4378.
(55) Shields, G. C.; Jurema, M. W. *J. Comput. Chem.* **1993**, *14*, 89–104.
(56) Desiraju, G. R. *Crystal Engineering, The Design of Organic Solids*; Elsevier: Amsterdam, 1989.