Predicting Crystalline Packing Arrangements of Molecules That Form Hydrogen-Bonded Tapes

Donovan N. Chin,*,† G. Tayhas R. Palmore,‡ and George M. Whitesides*,§

Contribution from Moldyn Inc., 955 Massachusetts Avenue, Cambridge, Massachusetts 02139, Department of Chemistry, University of California, Davis, Davis, California 95616, and Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received September 3, 1998

Abstract: A procedure based on simulated annealing Monte Carlo (SAMC) was used to predict the crystal structures of hydrogen-bonded organic molecules that form molecular tapes. This procedure was optimized to select structures with good hydrogen-bond geometries; it allowed for deformations in molecular conformation due to packing pressures. Biasing the SAMC procedure to select structures with good hydrogen bonds improves the efficiency of generating hydrogen-bonding motifs significantly. This procedure, written in the syntax of the CHARMM molecular modeling program, correctly predicted the crystalline structures of three test molecules derived from diketopiperazine: (C_5DKP = 3,6-(cyclotetramethylene)-2,5-diketopiperazine; Me_4DKP = 3,6-(tetramethyl)-2,5-diketopiperazine; $(Me_2C_6)_2$DKP = 3,6-(4,4-dimethylcyclohexane)-2,5-diketopiperazine). The computational and experimental results were compared using powder diffraction patterns, visualization, and values of C_k* (a measure of the crystalline packing efficiency).

Introduction

This paper describes the development and use of an optimized simulated annealing Monte Carlo (SAMC) procedure to predict the crystalline packing of organic molecules that form molecular tapes by connections through intermolecular hydrogen bonds (Scheme 1). The computational procedure is written in the syntax of CHARMM,[1] is easy to modify, and contains no major constraints on the crystallographic space groups that can be studied. It is therefore useful in studying the crystalline packing of certain classes of organic molecules. The procedure provides a useful approach to the analysis of the energetics of crystal packing of systems of low and intermediate complexity; the complexity of problems that can be studied is however limited by the computational power that is available.

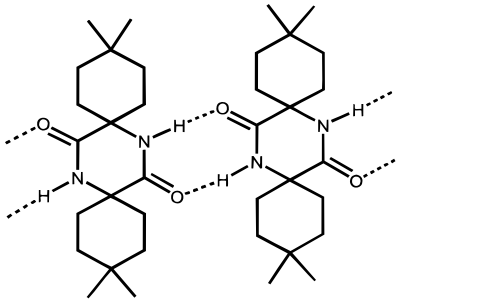
We (and others) are developing strategies for the design and synthesis of organic crystals that form well-defined aggregates and that pack in a predictable manner.[2−18] Among the goals of these efforts is the development of model systems that can be used to study crystal packing, and ultimately of a capability to *design* organic molecular crystals with specified structures and properties.[19,20] It is still difficult to understand and control the organization of organic crystals using computations for all but the simplest cases—for example, small rigid molecules—although substantial progress has been made using a number of different strategies: examples include research by Gavezzotti[21] and Ammon[22] in the packing of molecular clusters; Perlstein's work on the packing of molecules into one-dimensional aggregates for non-hydrogen bonding molecules using SAMC;[23,24] Hofmann and Lenguaer's approach of using fast, but approximate, statistically derived potential energy functions for finding low energy crystal structures;[25] Scheraga's application of the diffusion equation in predicting the crystal

† Moldyn Inc.
‡ University of California, Davis.
§ Harvard University.

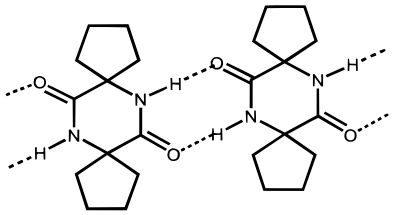
(1) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(2) Zerkowski, J. A.; Seto, C. T.; Whitesides, G. M. *J. Am. Chem. Soc.* **1992**, *114*, 5473−5475.
(3) Zerkowski, J. A.; Mathias, J. P.; Whitesides, G. M. *J. Am. Chem. Soc.* **1994**, *116*, 4305−4315.
(4) Zerkowski, J. A.; Whitesides, G. M. *J. Am. Chem. Soc.* **1994**, *116*, 4298−4304.
(5) Zerkowski, J. A.; MacDonald, J. C.; Seto, C. T.; Wierda, D. A.; Whitesides, G. M. *J. Am. Chem. Soc.* **1994**, *116*, 2382.
(6) Schwiebert, K. E.; Chin, D. N.; MacDonald, J. C.; Whitesides, G. M. *J. Am. Chem. Soc.* **1996**, *118*, 4018−4029.
(7) Palacin, S.; Chin, D. N.; Simanek, E. E.; MacDonald, J. C.; Whitesides, G. M.; McBride, M. T.; Palmore, G. T. R. *J. Am. Chem. Soc.* **1997**, *119*, 11807−11816.
(8) Russell, V. A.; Evans, C. C.; Li, W.; Ward, M. D. *Science* **1997**, *276* (5312), 575−579.
(9) Russell, V. A.; Ward, M. D. *J. Mater. Chem.* **1997**, *7*, 1123−1133.
(10) Russell, V.; Ward, M. D. *Chem. Mater.* **1996**, *8*, 1654−1666.
(11) Palmore, G. T. R.; McBride, M. T. *Chem. Commun.* **1998**, 145−146.
(12) Hollingsworth, M. D.; Ward, M. D. In *Structure and Chemistry of the Organic Solid State*; Hollingsworth, M. D., Ward, M. D., Eds.; American Chemical Society: Washington, DC, 1994; Vol. 6.
(13) Fan, E.; Vicent, C.; Geib, S. J.; Hamilton, A. D. *Chem. Mater.* **1994**, *6*, 1113−1117.
(14) Aakeroy, C. B.; Nieuwenhuyzen, M. *J. Am. Chem. Soc.* **1994**, *116*, 10983−10991.
(15) Aakeroy, C. B.; Hughes, D. P.; Nieuwenhuyzen, M. *J. Am. Chem. Soc.* **1996**, *118*, 10134−10140.
(16) MacGillivray, L. R.; Atwood, J. L. *J. Am. Chem. Soc.* **1997**, *119*, 6931−6932.
(17) Ghadiri, M. R.; Granja, J. R.; Milligan, R. A.; McRee, D. E.; Khazanovich, N. *Nature* **1993**, *366*, 324.
(18) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. *Acta Crystallogr., Sect. B: Struct. Sci.* **1998**, *B54*, 50−65.
(19) Bernstein, J. *J. Phys. D: Appl. Phys.* **1993**, *26*, B66−B76.
(20) Haleblian, J.; McCrone, W. *J. Pharm. Sci.* **1969**, *58*, 911−929.
(21) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622−4629.
(22) Holden, J. R.; Du, Z.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422−437.
(23) Perlstein, J. *J. Am. Chem. Soc.* **1992**, *114*, 1955−1963.
(24) Perlstein, J. *J. Am. Chem. Soc.* **1994**, *116*, 455−470.
(25) Hofmann, D. W. M.; Lengauer, T. *Acta Crystallogr.* **1997**, *A53*, 225−235.

10.1021/ja983163k CCC: $18.00
Published on Web 02/25/1999

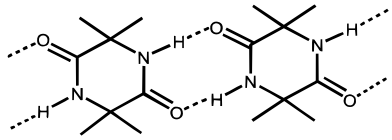
Scheme 1. Test Molecules



$(Me_2C_6)_2$DKP = 3,6-bis(4,4-dimethylcyclohexane)diketopiperazine



C_5DKP = 3,6-bis(cylcotetramethylene)-2,5-diketopiperazine



Me_4DKP = 3,6-bis(tetramethyl)-2,5-diketopiperazine

structure of S_6;[26] and Karfunkel's program for packing molecules in a crystalline environment (with periodic boundary conditions) using SAMC and starting from a single molecule.[27] Our method builds on that of Karfunkel by adding capabilities to search hydrogen-bonded structures efficiently and to find the local energy minimum of each predicted crystal structure.

The difficulties in predicting crystal structures have been eloquently described by others.[26,28−31] Empirical energy force fields are not yet accurate enough to distinguish polymorphs within a few kcals/mol; it is unreasonable to perform repetitive ab initio calculations on crystal structures of molecules of the size and complexity as those described in this paper. The many degrees of freedom for rigid body translations and rotations that are intrinsic to small molecules in solution, and the inclusion of any "soft" internal degrees of freedom (for example, torsions) contribute to the very large number of possible arrangements that must be sorted in order to find the one that is thermodynamically preferred in the crystalline state. Computational approaches that attempt to overpower the problem of sorting multiple minima by brute force are inadequate for the magnitude of the dimensionality associated with crystal engineering of most organic molecules. Currently, the most promising computational approach to the multiple minima problem is to apply one of the various optimization algorithms for finding the global minimum and as many low energy minima as possible.[32] A particular algorithm should be carefully matched and optimized to the problem at hand for efficiency (e.g., the "rate of cooling" in simulated annealing methods). The difficulty of estimating entropic contributions to the free energy of crystallization,[33] the problems of including interactions with solvent,[34] and kinetic effects on crystallization, reflected in the formation of polymorphs all compound the difficulty of the problem.[33,35−37] Theories that simplify the task of studying and understanding crystal growth are still being developed.[31,38,39]

Our approach to the problem of predicting the packing patterns of organic crystals starts with simplifying it to a level of complexity that is, we believe, tractable with current experimental and computational tools. The system that we are developing has three components: (i) it uses hydrogen-bonded *tapes* as a supramolecular architecture to limit the number of space groups that must be examined; (ii) it involves molecules with limited conformational flexibility to eliminate most complexities from conformational mobility; (iii) it uses an optimized SAMC sampling method to bias the computational search toward local minimum structures that have good hydrogen-bond geometries. We summarize each component in turn.

First, we have examined only organic molecules that form a geometrically well-defined type of aggregate, that is, a tape.[2,4,5,40] Molecules in these tapes are connected by intermolecular hydrogen bonds and form approximately flat, linear aggregates that are infinite in one dimension.[41] The use of molecules that can form planar, linear tapes greatly reduces the number of degrees of freedom of the system relative to that of weakly and nondirectionally interacting molecules of similar structural complexity. Second, through correct molecular design, the constituent molecules in the tapes can be made more or less rigid; rigidity minimizes the number of soft internal degrees of freedom in the crystal.[42] Third, computations involving SAMC can be used to search through the many packing possibilities of tapes.[43] We have optimized the SAMC method for crystal prediction by biasing the formation of hydrogen bonds and by minimizing the energy of all *intramolecular* and unit cell degrees of freedom. Biasing the SAMC procedure to generate crystals with hydrogen-bonded networks is necessary for this class of compounds because the hydrogen-bond energy is typically much smaller in magnitude than other electrostatic energies or the total crystal energy. In other words, the SAMC without this bias will tend to find low-energy structures with no guarantee that these structures will contain hydrogen bonds; the biasing strategy therefore ensures that predicted structures have hydrogen bonds. Although our method emphasizes the importance of hydrogen bonds, the resulting crystal structures are checked for efficient packing consistent with organic molecules. The two benefits of minimization are the following. First, it will find a local minimum for a predicted structure and therefore improve

(26) Wawak, R. J.; Gibson, K. D.; Liwo, A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1743−1746.
(27) Karfunkel, H. R.; Gdanitz, R. J. *J. Comput. Chem.* **1992**, *13*, 1171−1183.
(28) Maddox, J. *Nature* **1988**, *335*, 201.
(29) Desiraju, G. R. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2311−2327.
(30) Desiraju, G. R. *Crystal Engineering: The Design of Organic Solids*; Elsevier: New York, 1989.
(31) Gavezzotti, A. *Acc. Chem. Res.* **1994**, *27*, 309−314.
(32) Straub, J. E. In *Optimization Techniques with Applications to Proteins*; Elber, R., Ed.; World Scientific: Singapore, 1996.
(33) Dunitz, J. D.; Bernstein, J. *Acc. Chem. Res.* **1995**, *28*, 193.
(34) Khoshkhoo, S.; Anwar, J. *J. Phys. D: Appl. Phys* **1993**, *26*, B90−B93.
(35) Dunitz, J. D. *Acta Crystallogr.* **1995**, *B51*, 619−631.
(36) Bernstein, J. *J. Phys. D: Appl. Phys.* **1993**, *26*, B66−B76.
(37) McCrone, W. C. In *Polymorphism*; Fox, D., Labes, M. M., Weissberger, A., Eds.; John Wiley & Sons: New York, 1965; Vol. 2, pp 725−767.
(38) Sato, K. *J. Phys. D: Appl. Phys.* **1993**, B77−B84.
(39) Sangwal, K. *J. Cryst. Growth* **1989**, 393−405.
(40) Chin, D. N.; Zerkowski, J. A.; MacDonald, J. C.; Whitesides, G. M. In *Strategies for the Design and Assembly of Hydrogen-bonded Aggregates in the Solid State*; Whitesell, J. T., Ed.; John Wiley & Sons: London, 1998. In press.
(41) MacDonald, J. C.; Whitesides, G. M. *Chem. Rev.* **1994**, *94*, 383−420.
(42) For example, torsions where the barriers of rotation are on the order of *kT* (0.6 kcal/mol).
(43) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Clarendon Press: Oxford, 1987.

the *rate* of generating structures with favorable Boltzmann probabilities (explained more fully in the methodology). Second, it will allow conformational deformations—flexibility—due to packing pressures. By simplifying the problem of predicting the packing of organic crystals through choice of molecular types, supramolecular architecture, and computational methods, we greatly reduce the *magnitude* of the packing problem. The task of predicting crystalline structure from molecular structure still remains formidable, but we believe that with this design it is within reach of the computational power available through current workstations.

Our present work focuses on derivatives of 2,5-diketopiperazine (DKP) as a set of compounds against which we can test our ability to predict packing in crystals. As a class of molecules for crystal engineering, DKPs are attractive; they are easy to synthesize; they are more or less conformationally rigid (they are planar, or approximately planar); they form tapes reliably.[7,41] Scheme 1 shows the three molecules that we used to test our procedure for prediction of crystal structures (C_5-DKP = 3,6-(cyclotetramethylene)-2,5-diketopiperazine; Me_4-DKP = 3,6-(tetramethyl)-2,5-diketopiperazine; $(Me_2C_6)_2$DKP = 3,6-(4,4-dimethylcyclohexane)-2,5-diketopiperazine), all had crystallographic R values that were <0.09.[7] These molecules were chosen because they represent sizes and shapes that span the range of structures that we have been investigating and because the substituents on the DKP ring were alkyl groups. The absence of polar functions simplified the problem.[7]

Model and Methodology

Crystal Structures Are Predicted from a Single Molecule. Our procedure requires only the knowledge of the chemistry and geometry of a single molecule, and the identification of atoms that are hydrogen-bond donors and acceptors. In predicting crystal structures from a single molecule of a DKP, we considered only the five most populated space groups observed for organic molecules: that is, $P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $P2_1$, $C2/c$ (these five groups describe 80% of all organic molecules in the Cambridge Structural Database[44]).[45] When a molecule has a center of symmetry (centrosymmetric), these five groups can be reduced further to $P2_1/c$, $P\bar{1}$, and $C2/c$. Restricting the search to these space groups is not a limitation of our procedure, but it does improve its efficiency. By increasing the number of molecules in the unit cell (but at the expense of increasing the computational burden), and using pure translation (P1), higher symmetry elements would emerge if present.

The molecule was initially built and parametrized using Quanta 4.0,[46] and its energy minimized using CHARMM.[1] A molecular frame of reference was placed at the center-of-mass of the molecule to facilitate changes in molecular orientations.[47] A trial crystal structure was built in CHARMM for a chosen space group with randomly chosen values within limits of the unit cell (chosen intuitively) and molecular rotation window (Figure 1). These limits were the following: a 60° window for the angle of molecular rotation and a "loose" window for the lengths of the axes of the unit cell (for example, an arbitrary upper limit of 30% greater than the largest molecular dimension, and a lower limit 3% less the smallest dimension of the molecule). These loose limits ensured within reasonable time (seconds on a SGI R10000 machine) that the procedure initially found crystal structures that did not have energetically unfavorable van der Waals contacts. The SAMC method and minimization protocol, however, slowly increased the number of energetically favorable contacts between the molecule and its crystalline environment as the temperature was lowered. The crystal energy (CE) of this trial structure (and all subsequent structures) was determined using CHARMM with the noncovalent energy terms in eq 1.

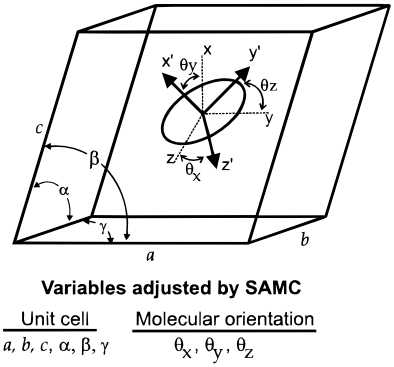


Figure 1. The variables adjusted by the SAMC procedure (some variables are fixed, or equal to others, depending on the space group).

$$\mathrm{CE} = \sum_{ij}\left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}\right] \quad (1)$$

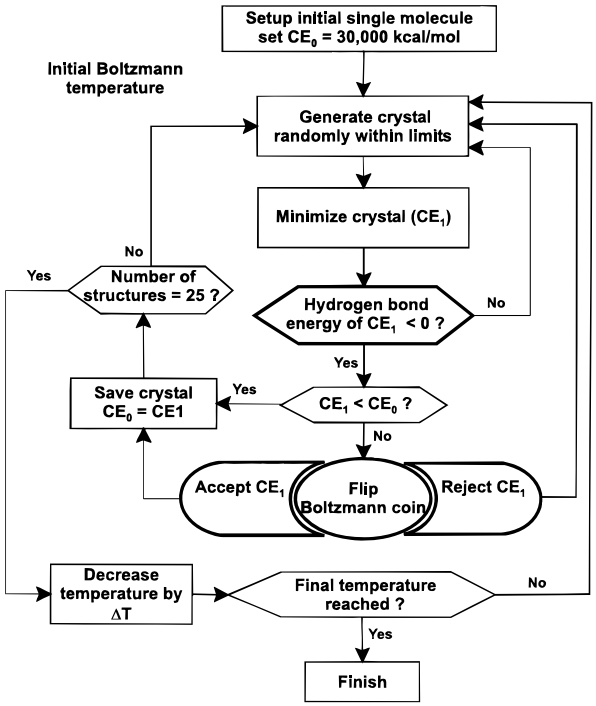
Values of CE were calculated by summing contributions between the i^{th} atom on the initial molecule and the j^{th} atom of the surrounding molecules in the crystal (for all pairs of atoms at distances r_{ij} up to 20 Å); A_{ij} and B_{ij} were the van der Waals parameters determined by, for example, $A_{ij} = (A_iA_j)^{1/2}$; q was the partial atomic charge; ϵ_0 was the permittivity of vacuum (8.854×10^{-12} C^2/Nm^2). Nonbonded interactions in the crystal were cut off at 20 Å in the computations; a switching potential between 16 and 19 Å was used for van der Waals and electrostatic terms (with the dielectric constant equal to 1). The value of CE for the current crystal was referred to as CE_0.

Simulated Annealing Monte Carlo (SAMC). SAMC was used to avoid trapping crystal structures in local minima. Many discussions of SAMC exist in the literature,[23,24,43] we will therefore only highlight pertinent details here. Our SAMC method is similar to that of Karfunkel in that we include the dimensions of the unit cell as variables to be searched (in the presence of the crystalline environment),[27] but with the added constraint of selecting structures with good hydrogen bonds (as we will show), and with the improved rate of accepting structures by finding a local minimum in energy using minimization.[48] To improve the efficiency of the SAMC procedure, we only evaluated crystals that had reasonable intermolecular hydrogen-bond geometries (a procedure we refer to as hydrogen-bond biased simulated annealing Monte Carlo, or HBB-SAMC).[49−51]

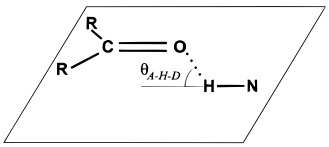
Our method is outlined in Scheme 2, and the details are the following. A trial crystal structure was generated by making random changes to the unit cell parameters appropriate for a given space group, and molecular rotations within the limits discussed previously (Figure 1), and then the symmetry operations were applied to the molecule. Each "crystal" was built by translations of the unit cell by a factor of 2 (that is, 124 translations). The energy of this crystal was minimized using adopted-basis Newton Raphson (ABNR) until the gradient in the energy was <0.2 kcal/(mol.Å^2); the final energy and crystal are referred to as CE_1. We included all the *internal* degrees of freedom (DOF) for the molecule along with the DOF for the unit cell in the minimization. Including the internal DOF of the molecule during minimization allowed for conformational deformations due to packing effects; we have observed previously that many of these DKPs adopt boat conformations in their energy minimized gas state, but planar conformations in their crystalline environment.[7] This minimization is not strictly rigorous

(44) Allen, F. H.; Kennard, O.; Taylor, R. *Acc. Chem. Res.* **1983**, *16*, 146−156.
(45) Brock, C. P.; Dunitz, J. D. *Chem. Mater.* **1994**, *6*, 1118−1127.
(46) MSI, *QUANTA Molecular Modeling Program*; MSI: San Diego, CA, 1995.
(47) The molecular frame of reference can be conveniently represented by placing "dummy" atoms on a Cartesian axis proximate to the molecule. These dummy atoms are not included in the calculations of the energy, but their presence conveniently faciliates the molecular rotation commands in CHARMM.
(48) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611−6615.
(49) Etter, M. C.; Huang, K. S. *Chem. Mater.* **1992**, *4*, 824.
(50) Etter, M. C.; Urfanczyk-Lipkowska, Z.; Zia-Ebrahimi, M.; Panunto, T. W. *J. Am. Chem. Soc.* **1990**, *112*, 8415−8426.
(51) Etter, M. C. *J. Phys. Chem.* **1991**, *95*, 4601−4610.

Scheme 2. Flow Diagram Describing the Prediction Scheme



Scheme 3. Definition of the Deviation from Linearity, θ_{A-H-D}, for NH···O Interactions



(more quantitative computation occurs at a later stage), but it greatly improves the rate of accepting reasonable structures as discussed by Scheraga.[48] That is, minimization will remove any unfavorable conformational, van der Waals, or electrostatic energies and direct structures to their nearest minimum in energy; therefore the probability of selecting these minimized structures via the Metropolis criteria (described later) is significantly greater than it is without minimization. We then screened the minimized crystal by evaluating the energy of the intermolecular hydrogen bonds using eq 2.[1] The value of HBE was only calculated if

$$\mathrm{HBE} = \sum_{AD}\left[\frac{A_{ij}}{r_{AD}{}^{6}} - \frac{B_{ij}}{r_{AD}{}^{4}}\right]\cos^{4}(\theta_{\mathrm{A-H-D}}) \qquad (2)$$

the angle, θ_{A-H-D} shown in Scheme 3, describing the deviation from linearity between the acceptor (A), hydrogen (H), and donor (D) atom was $<60°$,[52] and the distance between a hydrogen-bond donor and acceptor atom (r_{AD}) was <5 Å. If the value of HBE ≥ 0, the structure was rejected and a new crystal was generated from the previous step; if HBE < 0, then the energy of the minimized crystal is compared to that of the previous crystal. The QUANTA 4.1/CHARMM force field uses the Lennard-Jones and Coulombic terms in eq 1 to describe interactions that happen to be hydrogen bonds; it does not use the term in eq 2. We used eq 2 only to gauge the quality of hydrogen bonds. If CE_1 was less than CE_0, then the new crystal was accepted and became the current crystal ($CE_0 = CE_1$). If CE_1 was greater than CE_0, then the difference between the value of CE_1 and CE_0 was weighted by the Boltzmann expression in eq 3: W was the Boltzmann weight (values between 0 and 1), k_b was

$$W = e^{-(\mathrm{CE}_1 - \mathrm{CE}_0/k_b T)} \qquad (3)$$

the Boltzmann constant, and T was the "temperature" (initially set to 4300 K). At this point in the Monte Carlo procedure, a random number, R, between 1 and 0 was generated; if $R \geq W$, then this new crystal associated with CE_1 was rejected. If $R < W$, then the new crystal was accepted and became the current crystal ($CE_0 = CE_1$). The entire HBB-SAMC procedure was performed using a CHARMM script (with minor additions to the source code).

The procedure of generating and evaluating new crystals was repeated until 25 structures were accepted for a given temperature. The temperature was then lowered by 500 K, and the procedure for evaluating crystals was repeated starting from the 25th structure collected from the previous temperature. The annealing procedure was continued until the temperature reached 300 K; this resulted in a collection of 225 structures (that is, $(((T_{\mathrm{initial}} - T_{\mathrm{final}})/\Delta T) + 1) \times 25$, where $T_{\mathrm{initial}} = 4300$ K, $T_{\mathrm{final}} = 300$ K, and $\Delta T = 500$ K). We then ranked this collection of 225 structures from the lowest energy to the highest, and the lowest energy structure was further analyzed with the Cerius2 modeling program[53,54] and compared to the experimental structure.

Comparing Predicted Structures with Those from Experiment. We used three methods of comparing predicted and experimental crystal structures—visualization, packing efficiency (C_k*)[5,55−57] and *calculated* X-ray powder diffraction patterns (XPD)—the *experimental* XPD pattern was not used. Comparing unit cell dimensions was not reliable because they can be assigned arbitrarily; comparing unit cell dimensions that were reduced to a common frame was also problematic.[58] Analyses of powder diffraction patterns are particularly useful; these patterns provide a way of comparing crystalline packing arrangements of molecules determined using different methods (experimental or computational).[58,59] They are also a convenient and practical method for organic chemists to use.[60] Comparing XPD patterns, however, required a qualitative measure of similarity. Two commonly used analyses for comparing XPD patterns are profile R-factors (Rp) and XPD difference plots. These analyses, however, are very sensitive measures of the characteristics of the data being compared: Rp is appropriate as a quantitative measure of fit between two XPD patterns after refinement (such as in Rietveld methods), while difference plots depend sensitively on the line widths used, and can suggest enormous errors when in fact the structures are very close.[61] We therefore computed the similarity index (SI) shown in eq 4 to compare two XPD traces with an equal number of points, N.[58] Here, $\boldsymbol{d}$ is a vector that describes the difference ($\boldsymbol{d} = s_1 - s_2$)

$$\mathrm{SI} = \boldsymbol{d}\cdot\boldsymbol{F}\cdot\boldsymbol{d} \qquad (4)$$

between two normalized sets of points in the patterns (s_n), and $\boldsymbol{F}$ is an

(52) A value of 90° was chosen to allow some flexibility in the connecting hydrogen bonds.

(53) MSI, *Cerius 2 Molecular Modeling Program*; MSI: San Diego, CA, 1995.

(54) Cerius2 was used because it produced better hydrogen-bond geometries during its minimizations than CHARMM/QUANTA4.1.

(55) The packing fraction measures the amount of filled space in the crystal and is calculated as $C_k^* = N(V_m/V_c)$, where N is the number of molecules in the unit cell, V_m is the volume of the molecule, and V_c is the volume of the unit cell. V_m was calculated using van der Waals radii from Cerius 2; we believe that this way of calculating packing fractions is more accurate than that of Kitaigorodskii because his method uses pretabulated values of V_m for some functional groups and does not provide a sufficiently extensive range for general use among many organic compounds.

(56) Kitaigorodskii, A. I. *Molecular Crystals and Molecules*; Academic Press: New York, 1973.

(57) Spek, A. L. In *Platon 92*; Spek, A. L., Ed.; University of Utrecht: Utrecht, Netherlands, 1992.

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

(59) Cullity, B. D. *Elements of X-ray Diffraction*; Addison-Wesley: Reading, MA, 1978.

(60) Powder diffraction patterns were calculated using Cerius2 from Molecular Simulations/Biosym.

(61) E.g., the difference plot for the C5DKP structure before and after energy minimization shows enormous differences, the Rp was 62%, yet the structures are visually very close (shown in the Supporting Information).

Table 1. Structural and Energetic Values for Various Crystal Structures

	space group	Z^d	a (Å)	b (Å)	c (Å)	α (deg)	β (deg)	γ (deg)	C_k* [e]	cryst energy (kcal/mol)	vdW (kcal/mol)[f]	elec (kcal/mol)
Me_4DKP												
obsd[a]	$P\bar{1}$	1	5.65	5.87	8.36	69.89	113.04	116.03	0.681			
min[b]	$P1$	1	5.84	5.65	9.32	65.03	120.51	107.31	0.680	−32.19	−21.06	−11.13
pred[c]	$P1$	1	5.33	6.26	11.31	51.91	87.46	123.97	0.683	−32.55	−21.21	−11.34
	$P2_1$	2	5.49	8.20	5.38	90	109.49	90	0.620	−29.45	−17.17	−12.28
C_5DKP												
obsd[a]	$P\bar{1}$	1	6.21	8.93	5.79	104.97	116.84	78.18	0.716			
min[b]	$P1$	1	6.26	8.90	5.75	104.62	118.29	77.8	0.727	−40.40	−30.31	−10.09
pred[c]	$P1$	1	9.44	11.95	6.43	60.78	82.78	66.13	0.722	−40.40	−30.38	−10.02
	$P2_1$	2	5.55	8.84	6.24	90	62.81	90	0.623	−36.76	−25.26	−11.50
$(Me_2C_6)_2$DKP												
obsd[a]	$P\bar{1}$	1	6.09	6.32	11.57	90.09	101.15	109.83	0.709			
min[b]	$P1$	1	6.04	6.36	11.80	88.72	100.50	112.72	0.709	−47.09	−37.71	−9.38
pred[c]	$P1$	1	6.30	11.43	6.14	91.10	62.60	101.45	0.706	−47.08	−36.23	−10.85
	$P2_1$	2	6.28	11.17	6.49	90	122.46	90	0.641	−44.75	−35.10	−9.65

[a] The experimentally observed values. [b] The values from the experimental structure after minimization. Minimization was done to remove any high-energy deformations in the molecule and to judge the quality of the force field as discussed in the text. The inversion operation is removed because we are using whole molecules (as described in the text). [c] The predicted values for the two space groups studied. To validate our method, we chose the known space group to compare with an alternative (but well populated by organic molecules) space group. [d] Z is the number of molecules in the unit cell. [e] C_k* is the packing coefficient defined as $C_k^* = Z(V_m/V_c)$, where V_m is the volume of the molecule and V_c is the volume of the unit cell. [f] vdW = van der Waals.

$N \times N$ matrix called the "fold matrix"[58] that has the elements (i, j) in eq 5; α and β have been calibrated to the empirical values of $\alpha = 1.0 \times 10^{-8}$ and $\beta = 4$ by Karfunkel.[58] Quantitatively, eq 4 compares a

$$Fij = 1/(1+ \alpha(i - j)^{\beta}) \quad (5)$$

given point in one array with a set of points in its environment in the other array; this comparison is dictated by the functional form of the fold matrix in eq 5. The form of this equation gives decreasing weight to points further away from the given point; the maximum weight occurs along the diagonal ($i = j$). Karfunkel has shown the advantage of computing the similarity of nonoverlapping spectra or patterns in this way relative to simpler methods, such as Rp,[62] that compute only a sum of the mean-square-difference between two spectra (in other words, $Fij = 1$ for $i = j$, and $Fij = 0$ for $i \neq j$, or $\boldsymbol{d}\cdot\boldsymbol{d}$). Two identical arrays would have an SI value of 0; the more dissimilar the arrays, the larger the value of SI.

Results and Discussion

Table 1 summarizes the structural and energetic results. In this table, the energy of the observed structure (obsd) was minimized (min) to remove any distortions in the coordinates incurred during structure refinement, and this minimized structure was compared to the predicted structures (pred) from HBB-SAMC for each space group studied. The first measure of how appropriate a force field is for crystal energetics is that the differences in lattice parameters between the minimized and observed structures should be small, allowing only for thermal expansion in the observed structure. To this end, Table 1 shows that there are minor changes in C_k* and in the unit cell dimensions suggesting that the force field is adequate in this respect. We have also included in this table the number of molecules in the unit cells (Z). This number is related to the packing coefficient C_k*: that is, $C_k^* = Z(V_m/V_c)$, where V_m = the molecular volume and V_c is the volume of the unit cell. The structures that we considered here all have internal symmetry (one-half molecule as the asymmetric unit), and therefore the inversion operator in the observed space group $P\bar{1}$ serves to build the complete molecule.[63] Our program only works with complete molecules, and there is no need to use an inversion operator to build the molecules. We therefore have chosen the appropriate space groups for our test cases without the inversion operator (e.g., $P1$ for $P\bar{1}$, and $P2_1$ for $P2_1/c$). Our goal in this paper was to validate the method; we therefore chose the known space group for each molecule ($P\bar{1}$) and an alternative, but well populated by organic molecules, space group ($P2_1/c$) for this comparison.

Results for C_5DKP. The predicted crystal with the lowest (most favorable) value of the CE is that from the P1 space group (Table 1). In the $P1$ crystal, the predicted packing arrangement for C_5DKP is very close to that from experiment: two different views, and the calculated XPDs, of the predicted and experimental structure of C_5DKP are shown in Figures 2 and 3. The view of the predicted crystal (Figure 2) is particularly striking in that it correctly shows the subtleties of intertape packing. The calculated XPDs (Figure 3) are also similar (SI = 1.4) and indicate that the experimental and calculated arrangement of molecules in the crystals are almost identical: the major structural features of the crystal—the tape and its packing arrangement—are clearly revealed. The value of C_k* of the predicted $P1$ crystal structure—a value that measures the packing efficiency of the molecule—is within 1% of that from experiment (Table 1). The value of C_k* for the predicted $P2_1$ crystal in Table 1 is less favorable than that of $P1$; it was, in fact, outside the limit (0.67) normally observed for organic crystals.[56] This result is important because it shows that predicted crystals biased by hydrogen bonds must still pack efficiently. Carefully calculated hydrogen-bonded structures *that do not pack efficiently* may indicate that nonintuitive intermolecular interactions should be explored.

Results for Me_4DKP. The lowest predicted energy crystal is that from the P1 space group (Table 1). Again, the predicted packing is similar to that from experiment (Figure 4); the powder patterns also establish structural similarities between the respec-

(62) Giacovazzo, C. In *Fundamentals of Crystallography*; Giacovazzo, C., Ed.; Oxford University Press: Oxford, 1992; pp 109−111.

(63) For example, $Z = 1$ for the space group $P\bar{1}$. $P\bar{1}$ has two symmetry operators—translation and inversion—which implies that there is one-half of a molecule for the asymmetric unit (number of operatators × molecule in the unit cell = Z). In our prediction program, $Z = 1$ for the space group $P\bar{1}$; that is, one symmetry operator × one molecule for the asymmetric unit = Z.

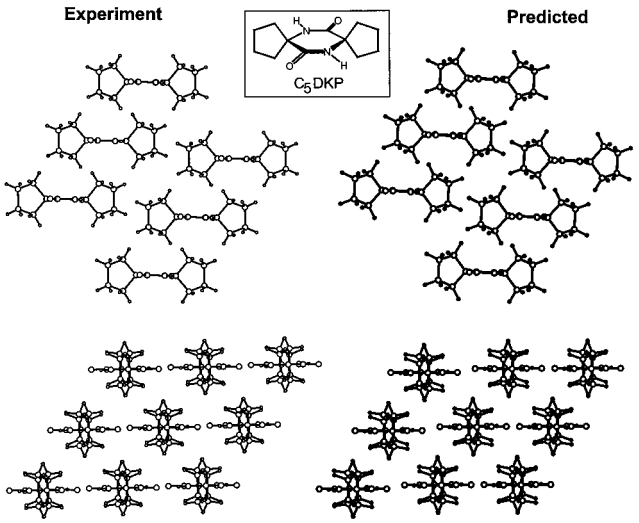


Figure 2. Two different views of the molecular arrangements for the experimental and predicted *P*1 crystal structure of C_5DKP.

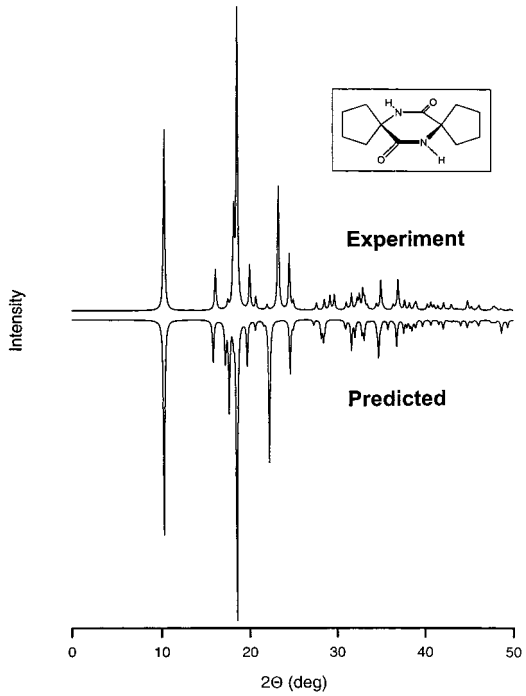


Figure 3. The *calculated* powder patterns of the experimental and predicted *P*1 crystal structures for C_5DKP.

tive atomic packing arrangements (Figure 5; SI = 2.4). The value of C_k* for the predicted *P*1 structure is 0.683, which is close to the experimental value of 0.680; the value for the space group $P2_1$ is well below the 0.67 limit for organic molecules (Table 1).

Results for $(Me_2C_6)_2$DKP. The calculated structure of *P*1 compares slightly less well with that from experiment than for C_5DKP and Me_4DKP, but still reveals the major structural elements of the crystal clearly. The values of C_k* are close for experimental and calculated *P*1 structures (Table 1); the packing arrangements are also close (Figure 6); the powder patterns are similar (Figure 7; SI = 5.0). Among the three test molecules, the similarity of the observed and calculated powder patterns for $(Me_2C_6)_2$DKP was the poorest. The poorer comparison with experiment is likely due to, but may not be limited to, the following two issues. First, $(Me_2C_6)_2$DKP has more DOF than the other two molecules, and the program may have sampled

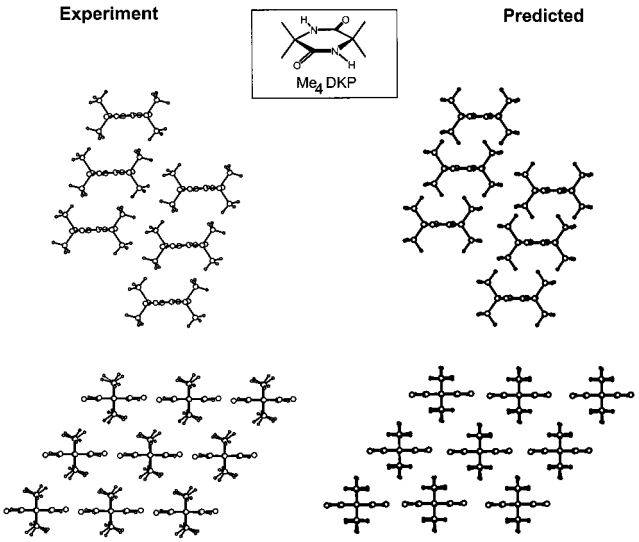


Figure 4. Two different views of the molecular arrangements for the experimental and predicted crystal structure of Me_4DKP.

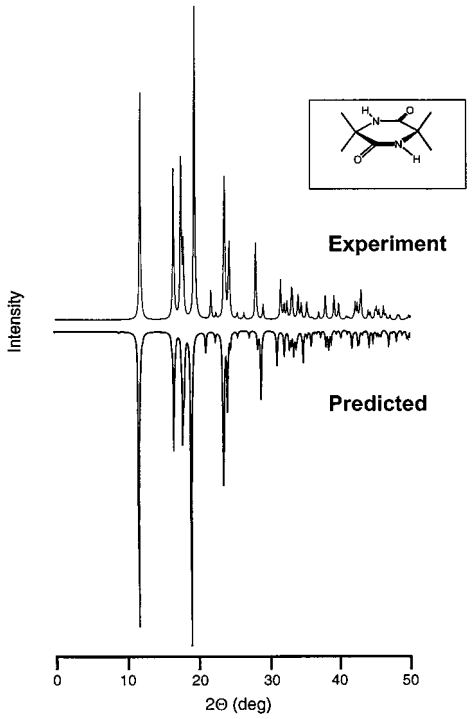


Figure 5. The *calculated* powder patterns of the experimental and predicted *P*1 crystal structure for Me_4DKP.

these DOF incompletely. Second, the partial charge distribution may be inappropriate for this molecule. Table 1 suggests that the electrostatic energy for the predicted $(Me_2C_6)_2$DKP may be slightly more favorable (and, to compensate, the van der Waals energy slightly less favorable) than that for the experimental structure, and this increased electrostatic energy may be the cause of minor distortions in the molecule.

Constraining the Search to Aggregates with Good Hydrogen Bonds Improves the Efficiency of the Procedure. Evaluating only those structures that had favorable values of HBE (with reasonable hydrogen-bond geometries) improved the efficiency of the procedure by counteracting limitations of incomplete sampling by SAMC: the collected 225 minima all had negative (favorable) values for HBE. When this bias for favorable hydrogen-bond geometry was removed, structures that were much poorer in quality were collected (for example,

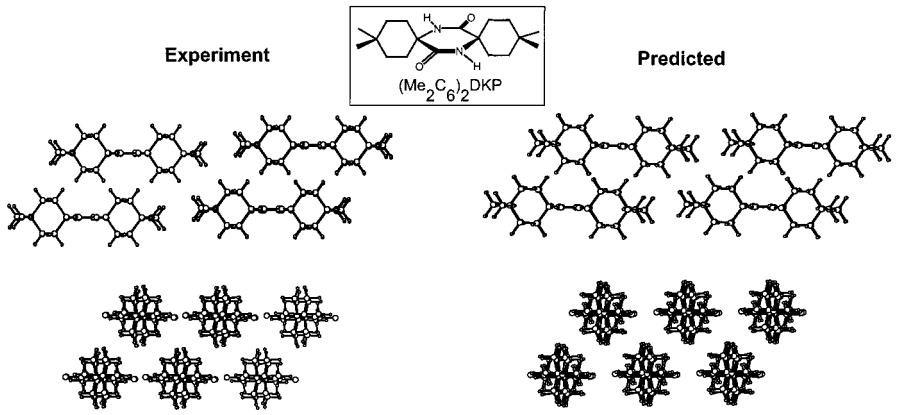


Figure 6. Two different views of the molecular arrangements for the experimental and predicted *P*1 crystal structure of $(Me_2C_6)_2DKP$.

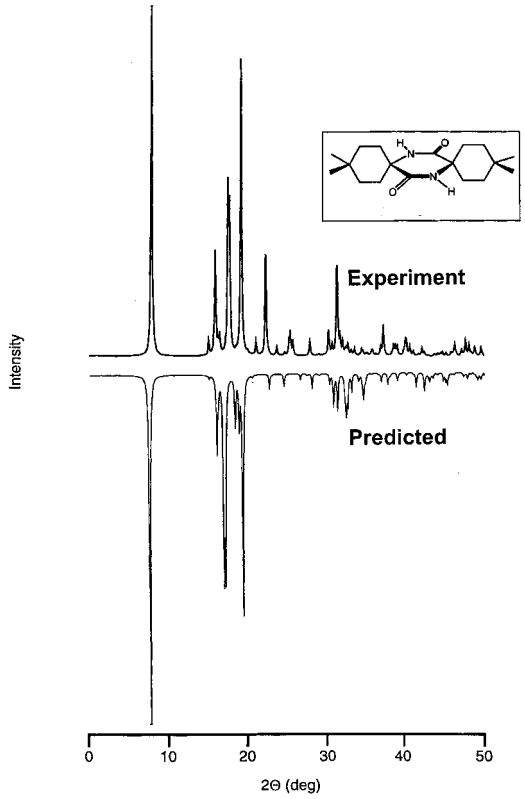


Figure 7. The *calculated* powder patterns of the experimental and predicted *P*1 crystal structure for $(Me_2C_6)_2DKP$.

structures with unsatisfied hydrogen bonds). The effectiveness of screening with and without HBE for Me_4DKP is shown dramatically in Figure 8.

Constraining the search through hydrogen bonds results in a modest increase in the number of trial crystal structures that must be generated, but minimization of these structures improves the acceptance ratio and therefore offsets the added computational expense.

Polymorphism Was Not Prevalent in These DKPs. Of the seven DKPs studied experimentally by us so far (with substituents in the 3,6 positions representing a substantial range of sizes and shapes; the three included in this paper span this range), only one structure showed evidence of a polymorph by XPD (IndDKP= 3,6-(2-indane)-2,5-diketopiperazine);[7] of the 40 DKPs identified from the CSD, one was polymorphic.[41] We were, however, unable to obtain useful single-crystal data from IndDKP, and thus cannot test it with this computational approach. Such a test, however, would place the emphasis on a more accurate force field and would be very useful nevertheless.

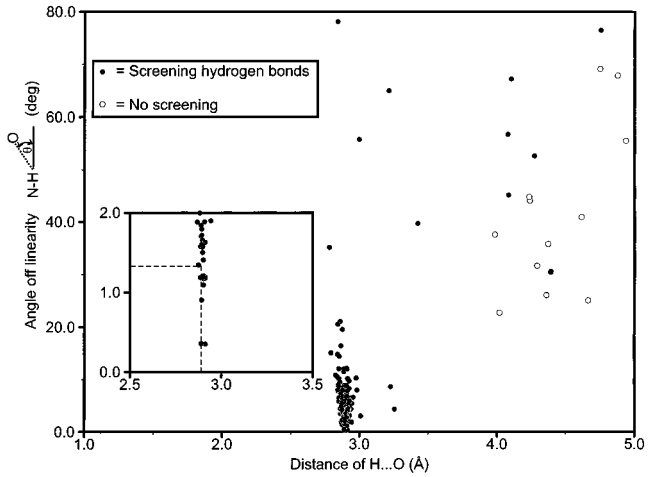


Figure 8. An example of the improved quality of the predicted crystal structures that results from constraining the search to find structures with good hydrogen bonds. These data are for Me_4DKP, and the hydrogen-bonding distances (<5.0 Å) and angles off linearity (<80°) for the predicted structures with and without screening are shown. The insert is an expansion of the values near the experimental structure (dashed line).

Conclusions

A new and optimized SAMC procedure for predicting crystal structures of hydrogen-bonded organic molecules, written in the syntax of CHARMM, correctly predicted the crystalline structures for C_5DKP, Me_4DKP, and $(Me_2C_6)_2DKP$. There were three strategic points that led to the successful predictions using SAMC. First, we used molecules that, through their topography, were predisposed to form molecular tapes; this constraint greatly reduced the dimensionality of the problem.[51,64,65] Since hydrogen-bonded tapes derived from DKPs are a type of structure that should allow usefully broad variation in solid-state architecture, we believe that this approach will allow us to test a range of predictions in crystal engineering. Second, we biased trial structures in the SAMC procedure to have reasonable hydrogen-bond geometries (that is, within the limits described in the section on eq 2) and check that these structures have packed efficiently. Third, we allowed for conformational deformations of the molecule due to packing pressures by minimizing the internal DOF of the molecule and the DOF of the unit cell. Minimization helps the HBB-SAMC procedure to generate predicted structures that represent a local minimum in energy.[48]

(64) Lehn, J. M. *Supramolecular Chemistry*; VCH: Weinheim, 1995.
(65) Perlstein, J.; Steppe, K.; Vaday, S.; Ndip, E. M. N. *J. Am. Chem. Soc.* **1996**, *118*, 8433−8443.

An additional test of the HBB-SAMC method was successful on another molecule, 4,5-(dimethyl)-2-benzimidazolone, that is very different (and more challenging to predict) than DKPs.[66]

The primary goal of our paper was to show that the hydrogen-bonded screening strategy is useful in restricting the search to structures with good hydrogen-bond networks and that this strategy increases the efficiency of the search. This strategy, however, does not completely compensate for limitations in empirical force fields (as it did in Figure 8). The large electrostatic energy in the $(Me_2C_6)_2$DKP case resulted in a slightly distorted structure (but a structure that still showed the major packing features). The problem of high quality force fields in the solid state remains largely unsolved. The development of force fields with improved descriptions of intermolecular interactions in the solid state will therefore be needed if they are to aid in useful predictions for crystal engineering.[67] Screening strategies in concert with optimization methods, nevertheless, constitute a convenient alternative approach (albeit one limited to specific types of structures) to the development of high quality force fields. Screening other interactions (such as CH···O and halogen−halogen interactions) may broaden the application and usefulness of this approach. The reliability of screening strategies improves with the availability of experimental structures (such as those in the Cambridge Structural Database).

Three methods were used to compare the predicted and experimental structures. Powder diffraction and visualization methods provided convenient and unambiguous ways of comparing the arrangements of molecules in crystals; values of C_k* were useful as a final check of the accuracy of the predicted crystals. We found the values of SI to be a useful metric with which to compare and rank different calculated powder patterns with nonzero line widths. Difference plots and values of Rp are too sensitive for the goals of this paper (and in fact may not be useful) but are more appropriate for cases where predicted structures are used as models for further refinement such as in Rietveld methods. Interpreting powder patterns requires care, however, as the problems of reducing a 3-dimensional structure to a 2-dimensional representation (that is, a powder pattern) are well-known.[59]

To our knowledge, this is the first crystal prediction program that is written entirely in a scripting language (CHARMM) for studies in the field of organic solid state.[68] This point is an important one because this approach eliminates the need for detailed expertise and knowledge of lower-level programming languages (e.g., FORTRAN, C); programming in the much simpler environment of scripting languages is a paradigm that should encourage the use and development of such prediction programs by noncomputational experts in the field of crystal engineering.[69]

It remains difficult to predict the structures of organic molecules in general.[31] We have shown here, however, that using HBB-SAMC methods on molecules having restricted packing arrangements due to specific hydrogen-bond interactions is a promising approach toward the goals of crystal engineering.[40,41]

Acknowledgment. This work was supported by the NSF (NSF CHE 91-2231). We are grateful to John C. MacDonald at N.A.U. for discussions on crystals and crystallography.

Supporting Information Available: Listings of the parameters used for all molecules. This material is available free of charge via the Internet at http://pubs.acs.org. All CHARMM scripts for running the program can be obtained from the authors. The CHARMM program can be obtained from Martin Karplus at Harvard University, or from MSI. Cerius2 is available from MSI.

JA983163K

(66) Chin, D. N. In *Improving the Efficiency Of Predicting Hydrogen-Bonded Organic Molecules*; Zaworotko, M., Rodgers, R., Eds.; MRS Transactions: Washington, DC. In press.

(67) Price, S. L. *Intermolecular Forces-from the Molecular Charge Distribution to the Molecular Packing*; Gavezzotti, Ed.; John Wiley & Sons: Chichester, 1997; Vol. 1, pp 31−60.

(68) Gavezzotti, G. *Theoretical Aspects and Computer Modeling of the Molecular Solid State*; Gavezzotti, G., Ed.; John Wiley & Sons: New York, 1997; Vol. 1.

(69) Beazley, D. M.; Lomdahl, P. S. *Comput. Phys.* **1997**, *11*, 231−237.