# Crystal Packing, Hydrogen Bonding, and the Effect of Crystal Forces on Molecular Conformation

PNINA DAUBER and ARNOLD T. HAGLER*

*Chemical Physics Department, Weizmann Institute of Science, Rehovot, Israel*

*Received May 16, 1979*

The studies of crystal structures of organic molecules by X-ray and neutron diffraction provide valuable information about the spatial arrangement of the atoms in molecules and the packing of molecules in lattices. In many respects, however, this is just the beginning of the story, rather than the end. Why do these molecules pack in the observed space group? What are the intermolecular forces determining crystal structure? How do crystal forces influence the conformation of flexible molecules? These are some of the fundamental questions which we may now ask, having the wealth of crystal data at our disposal.

Over the past several years we have been developing methodology for answering some of these questions. Our studies involve a "marriage" of X-ray crystallography with theoretical techniques such as lattice energy calculations, conformational analysis, and ab initio molecular orbital calculations. These techniques allow us to analyze crystal structure in terms of energy as well as interatomic distances and angles and thus provide a deeper probe into packing phenomena. In addition to the studies of the questions posed above, the potentials and techniques discussed here are being applied to a wide range of problems, especially in studies of the conformational and dynamic properties of biological systems.[1a-e] A logical extension of this work is the prediction of crystal structure. Although this is an elusive goal, it remains a valuable objective. The ability to predict crystal structures would be of great importance, for example, in designing crystals with oriented reactive groups for solid-state reactions.[1f]

The use of crystal data to obtain interaction energies is a relatively recent practice, the first extensive work being Rice's,[2] for the Ar–Ar interaction. The study of organic crystals is more recent. The pioneering work in this area came in the 1960s with studies on hydrocarbon crystals,[3] which is a subject of continuing research.[4] Later, mainly sparked by the desire to better describe the energy surface of biological molecules, a large number of studies were carried out on polar systems.[5]

In order to carry out the studies outlined above, we must be able to express the energy of the crystal in terms of the structural parameters defining the interatomic distances. Because of the wide range of subjects for which interatomic functions are used[1] and limitations of computer time when large molecular systems are treated, the analytical form or "force field" used to represent this energy surface must be general, yet as simple as possible. These two, to some extent mutually exclusive, criteria shall be used to judge the utility of various analytical forms below.

In general the simplifying assumption is made that the total interaction between two molecules may be represented by the sum of all the (pairwise additive) atom–atom interactions between the two molecules.[6] For our studies of amides[7] we started from the common representation of the energy of interaction $V_{ab}$ between two atoms, a and b, as given in eq 1 in two equivalent

$$V_{ab} = A_{ab}/r^{12} - C_{ab}/r^6 + q_a q_b/r = \epsilon_{ab}[(r^*_{ab}/r)^{12} - 2(r^*_{ab}/r)^6] + q_a q_b/r \quad (1)$$

forms. The first term represents the exchange repulsion between a and b as they approach and their electron "clouds" begin to overlap. This is a steep repulsion and hence the inverse 12th power dependence on the distance, $r$. We have also considered an inverse 9th power dependence. The second term represents the dispersion interaction which exists even between nonpolar atoms such as argon. It is an attractive interaction known to go as the inverse 6th power of $r$ at large distances.[8] These two terms, commonly known as the 6–12 (or 6–9) Lennard-Jones (or Mie) potential,[9a,b] are depicted

Arnold T. Hagler is Associate Professor at the Weizmann Institute, Rehovot, Israel, where he has been since 1972. He was born in New York City and received both B.Ch.E. and Ph.D. degrees from Cornell University. His research interests include Monte Carlo and molecular dynamics simulation of biological systems, the structure and function of peptides and proteins, and the nature of the hydrogen bond and intermolecular forces. Dr. Hagler is currently visiting professor at University of California, San Diego.

Pnina Dauber was born in Romania in 1949. She studied at Bar-Ilan University, Ramat Gan, Israel, for her B.Sc. and M.Sc. degrees, and was an assistant there for 3 years before moving to her present position as chemist at the Weizmann Institute.

(1) (a) G. Nemethy and H. A. Scheraga, *Q. Rev. Biophys.*, **10**, 239 (1977); (b) B. Pullman and A. Pullman, *Adv. Protein Chem.*, **28**, 348 (1974); (c) A. T. Hagler, *Israel J. Chem.*, **10**, 202 (1977); (d) M. Karplus and J. A. McCammon, *Nature (London)*, **277**, 578 (1979); (e) A. T. Hagler and J. Moult, *ibid.*, **272**, 222 (1978); (f) G. M. J. Schmidt, *Pure Appl. Chem.*, **27**, 647 (1971).

(2) O. K. Rice, *J. Am. Chem. Soc.*, **63**, 3 (1941).

(3) (a) A. I. Kitaigorodskii, "Molecular Crystals and Molecules", Academic Press, New York, 1973, and references therein; (b) *Acta Crystallogr.*, **18**, 585 (1965); (c) D. E. Williams, *J. Chem. Phys.*, **43**, 4424 (1965); **47**, 4680 (1965).

(4) (a) D. E. Williams and T. L. Starr, *Comput. Chem.*, **1**, 173 (1977); (b) K. Mirsky, *Acta Crystallogr., Sect. A*, **32**, 199 (1976); (c) A. Warshel and S. Lifson, *J. Chem. Phys.*, **53**, 582 (1970).

(5) A. M. Liquori in "Symmetry and Function of Biological Systems at the Macromolecular Level", A. Engstrom and b. Strandberg, Eds., Wiley, New York, 1969, p 101; (b) F. A. Momany, L. M. Carruthers, R. F. McGuire, and H. A. Scheraga, *J. Phys. Chem.*, **78**, 1595 (1974); (c) L. L. Shipman, A. W. Burgess, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 543 (1975); (d) D. R. Ferro and J. Hermans in "Liquid Crystals and Ordered Fluids", J. F. Johnson and R. Porter, Eds., Plenum Press, New York, 1970; (e) J. Caillet and P. Claverie, *Acta Crystallogr., Sect. A*, **31**, 448 (1975); (f) J. L. Derissen and P. H. Smit, *ibid.*, **34**, 842 (1978).

(6) Kitaigorodskii[3a] has presented an extensive discussion of the history and applications of the atom–atom potential.

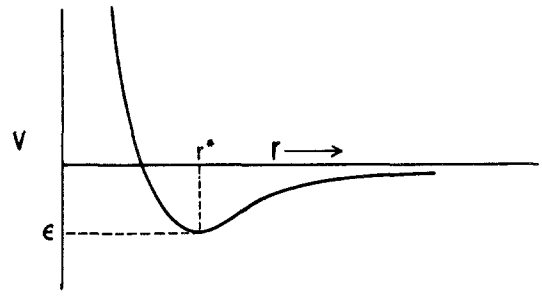
(7) (a) A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1974); (b) A. T. Hagler and S. Lifson, *ibid.*, **96**, 5327 (1974).

(8) See discussions of intermolecular forces in (a) J. O. Hirshfelder, C. F. Curtiss, and R. B. Bird, "Molecular Theory of Gases and Liquids", Wiley, New York, 1954; (b) E. A. Mason and L. Monchick, *Adv. Chem. Phys.*, **12**, 329 (1967); (c) A. D. Buckingham and B. D. Utting, *Annu. Rev. Phys. Chem.*, **21**, 287 (1970).

(9) (a) J. E. Lennard-Jones, *Proc. Phys. Soc., London*, **43**, 461 (1931); (b) G. Mie, *Ann. Phys. (Leipzig)*, **11**, 657 (1903).

0001-4842/80/0113-0105$01.00/0 

schematically in I. The values $r^*$ and $\epsilon$ indicated on




I

I correspond to the constants given in the second form of eq 1 and depend on the particular pair of atoms (as do the constants A and C). The last term represents the Coulomb interaction between the two atoms which carry partial charges $q_a$ and $q_b$. The latter term, when summed over all atoms in the two molecules, represents the dipole–dipole interaction between polar molecules. If we invoke the approximation of pairwise additivity, the lattice energy of a crystal may be written as the sum over all interatomic interactions between all molecules in the crystal (eq 2) where the sum over $i$ is over all

$$E_L = \tfrac{1}{2} \sum_i \sum_j A_{ij}/r_{ij}^{12} - C_{ij}/r_{ij}^{6} + q_i q_j/r_{ij} \qquad (2)$$

atoms of a reference molecule, while $j$ runs over the atoms in all other molecules.

It should be emphasized that eq 1 is a *model* of the interatomic energy. We are now faced with finding the best values of the constants $A$, $C$, and $q$ for each atom pair and testing how well this analytical representation reproduces the energy surface. We have undertaken this task as part of a study aimed at obtaining a set of energy functions for amides, acids, peptides, and other biological molecules and studying the hydrogen bond in these systems.[10] The goal is to understand the intermolecular forces in these systems, use them for further studies of crystals, and provide a simple well-tested set of functions for the growing, important field of computer simulation studies of biological systems. The strategy has been to use the information about the energy dependence inherent in the structures and sublimation energies of model compounds, such as the structure of urea[11a] shown in stereo[11b] in Figure 1.

This information is contained in the statement that the lattice energy is a minimum at the experimentally observed structure,[12] or by the $n$ equations,

$$\partial E_L/\partial a_i = 0, \qquad i = 1, \ldots n \qquad (3)$$

where the $a_i$ represent all degrees of freedom of the lattice. If crystal symmetry is maintained, these may be represented by the lattice constants (unit cell vectors and angles, $a$, $b$, $c$, $\alpha$, $\beta$, $\gamma$) and the translational and rotational degrees of freedom of the asymmetric unit

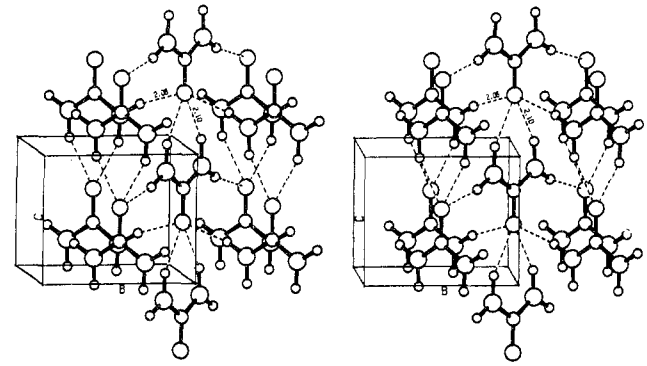


**Figure 1.** Crystal structure of urea[11] ($NH_2CONH_2$, $P\bar{4}2_1m$). Hydrogen bonds are indicated by dashed lines. The atoms H, C, O, and N are represented by circles of increasing size. Urea has an exceptional hydrogen-bonding network since it has only one oxygen but four hydrogens per molecule. This one oxygen is hydrogen bonded to four proton donors of adjacent molecules. Two of these hydrogen bonds are formed between the oxygen and two hydrogens from a molecule in the same plane as the acceptor molecule. The other two hydrogen bonds come from molecules on either side of the acceptor in a plane perpendicular to it.

($t_x$, $t_y$, $t_z$, $\theta_x$, $\theta_y$, $\theta_z$). Now if we know the laws describing the interactions between molecules (for example, in eq 1) we can calculate the lattice energy, $E_L$, of a given crystal structure and find the $a_i$'s corresponding to the minimum lattice energy for this structure. That is, we solve eq 3. We can in principle reverse the procedure and deduce the laws governing the intermolecular interactions by observing the crystal structure which resulted from the system "following these laws". For example, when urea "follows these laws", the structure in Figure 1 results.[13] Although deriving a crystal structure from the analytical form for the lattice energy, $E_L$, is a straightforward, unambiguous process, this is unfortunately not the case for the converse. One cannot map crystal structures directly into a functional form. This is basically why different analytical forms and potential constants have been proposed for the same interactions.

In order to proceed, we propose a trial functional form on the basis of a knowledge of interatomic forces of simpler systems such as Ar[8] and physical arguments such as those used to describe the various terms in eq 1. Some trial values for the constants are then taken, thus completely determining the interaction energy. Equation 3 is then solved and the calculated structural variables, $a_i$, are compared with the observed structure. If there are significant deviations (as there always are), the potential constants are changed so as to bring the calculated structure into best agreement with the observed. For example, in the case of the urea crystal (Figure 1), the calculated lattice constants should be $a = b = 5.66$ Å, $c = 4.71$ Å (the angles $\alpha = \beta = \gamma = 90°$ are determined by the tetragonal symmetry). Furthermore, the minimized orientation of the molecule should result in four amide hydrogens "hydrogen bonding" to the single carbonyl oxygen, two in the plane of the molecule with distances of ~2.10 Å and two approximately perpendicular to the molecule with distances of 2.07 Å.[11]

(10) A. T. Hagler, S. Lifson, and E. Huler in "Peptides, Polypeptides and Proteins", Wiley-Interscience, New York, 1974, p 35.

(11) (a) J. E. Worsham, Jr., H. A. Levy, and S. W. Peterson, *Acta Crystallogr.*, **10**, 319 (1957). (b) We strongly recommend that the reader use standard stereoglasses if necessary to view the figure given here and in the references. It is almost impossible to appreciate the relative spatial arrangements and hydrogen-bonding networks without seeing them in three dimensions.

(12) In fact the experimental structure corresponds to a free-energy minimum. All studies to date have invoked the approximation that the variation of the lattice entropy can be neglected, which results in the stated condition that the energy be a minimum.

(13) Deducing laws of intermolecular interactions in this way is often done qualitatively, perhaps without realizing it. For example, most of the notions of the "angular dependence of the hydrogen bond" until recently came from observing its geometry in crystals.

Table I
Experimental Hydrogen-Bond Geometry[a] and Packing Modes[b] of Amides and Carboxylic Acids

| amide | formula | ref[c] | O··H | ∠C=O··H | 180° − ∠X—H··O | space group | packing mode |
|---|---|---|---|---|---|---|---|
| oxamide | $NH_2COCONH_2$ | (1) | 2.02 | 155 | 29 | $P\bar{1}$ | planar layers |
| malonamide | $NH_2COCH_2CONH_2$ | (2) | 2.13 | 144 | 29 | $P2_1/c$ | two molecules/ asymmetric unit |
| succinamide | $NH_2CO(CH_2)_2CONH_2$ | (3) | 1.94 | 138 | 3 | $C2/c$ | planar layers |
| glutaramide | $NH_2CO(CH_2)_3CONH_2$ | (4) | 2.02 | 151 | 28 | $C2/c$ | nonparallel ribbons |
| adipamide | $NH_2CO(CH_2)_4CONH_2$ | (5) | 2.06 | 146 | 35 | $P2_1/c$ | layers without cyclic dimers |
| suberamide | $NH_2CO(CH_2)_6CONH_2$ | (6) | 1.98 | 150 | 29 | $C2/c$ | planar layers |
| urea | $NH_2CONH_2$ | (7) | 2.07 | 106 | 12 | $P42_1m$ | three-dimensional network |
| formamide | $HCONH_2$ | (8) | 1.91 | 125 | 19 | $P2_1/c$ | chains of cyclic dimers |
| diketopiperazine | $CONHCH_2CONHCH_2$ | (9) | 1.85 | 123 | 4 | $P2_1/a$ | nonparallel ribbons |
| L,L-dimethyl-diketopiperazine | $CONHCH(CH_3)NHCH(CH_3)$ | (10) | 1.91 | 120 | 11 | $P1$ | parallel ribbons |
| cyclopropanecarboxamide | $CH_2CH_2CHCONH_2$ | (11) | 1.99 | 136 | 19 | $P2_1/c$ | helical motif |
| *N*-methylacetamide | $CH_3NHCOCH_3$ | (12) | 1.81 | 139 | 6 | $Pnma$ | parallel ribbons |
| acid | | | | | | | |
| formic acid | $HCOOH$ | (13) | 1.58 | 123 | 7 | $Pna2_1$ | linear chains (catamers) |
| acetic acid | $CH_3COOH$ | (14) | 1.65 | 130 | 17 | $Pna2_1$ | linear chains (catamers) |
| propionic acid | $CH_3CH_2COOH$ | (15) | 1.63 | 124 | 2 | $P2_1/c$ | dimer rings |
| butyric acid | $CH_3(CH_2)_2COOH$ | (16) | 1.62 | 125 | 0 | $C2/m$ | dimer rings |
| valeric acid | $CH_3(CH_2)_3COOH$ | (17) | 1.65 | 126 | 12 | $P2_1/c$ | dimer rings |
| α-oxalic acid | $(COOH)_2$ | (18) | 1.80 | 123 | 33 | $Pcab$ | corrugated sheets |
| β-oxalic acid | $(COOH)_2$ | (18) | 1.67 | 120 | 5 | $P2_1/c$ | linear chains of cyclic dimers |
| malonic acid | $HOOCCH_2COOH$ | (19) | 1.71 | 114 | 8 | $P\bar{1}$ | twisted chains of cyclic dimers |
| methylmalonic acid[d] | $HOOCCH(CH_3)COOH$ | (20) | 1.74 | 124 | 11 | $P\bar{1}$ | twisted chains of cyclic dimers |
| succinic acid | $HOOC(CH_2)_2COOH$ | (21) | 1.61 | 125 | 0 | $P2_1/a$ | linear chains of cyclic dimers |
| glutaric acid | $HOOC(CH_2)_3COOH$ | (22) | 1.68 | 118 | 12 | $I2/a$ | linear chains of cyclic dimers |
| adipic acid | $HOOC(CH_2)_4COOH$ | (23) | 1.62 | 120 | 6 | $P2_1/c$ | linear chains of cyclic dimers |
| suberic acid | $HOOC(CH_2)_6COOH$ | (24) | 1.65 | 120 | 10 | $P2_1/c$ | linear chains of cyclic dimers |
| sebacic acid | $HOOC(CH_2)_8COOH$ | (25) | 1.64 | 118 | 10 | $P2_1/c$ | linear chains of cyclic dimers |

[a] The hydrogen atom positions are refined by minimization of intramolecular energy of these molecules with respect to hydrogen positions.[7] Lengths are in Å and angles in degrees. [b] Most amides form hydrogen-bonded cyclic dimers, and these dimers are interlinked by additional hydrogen bonds to form a two- or three-dimensional array. The values given in this table refer to the interdimer hydrogen bonds. [c] In order to save space, only literature citation is given—see also ref 7 and 17: (1) *Acta Crystallogr.*, **7**, 588 (1954); (2) *J. Chem. Soc., A*, *179* (1970); (3) *Acta Crystallogr.*, **9**, 334 (1956); (4) *ibid.*, **21**, 413 (1966); (5) *ibid.*, **20**, 626 (1966); (6) *ibid.*, **20**, 368 (1966); (7) *ibid.*, **10**, 319 (1957); (8) *ibid.*, **7**, 559 (1954); (9) *ibid.*, **12**, 1007 (1959); (10) *Biopolymers*, **7**, 751 (1969); (11) *Acta Crystallogr., Sect. B*, **25**, 2083 (1969); (12) *Acta Crystallogr.*, **13**, 624 (1960); (13) *ibid.*, **6**, 127 (1953); *Acta Crystallogr., Sect. B*, **34**, 315 (1978); (14) *ibid.*, **27**, 893 (1971); (15) *Acta Crystallogr.*, **15**, 1233 (1962); (16) *ibid.*, **15**, 1240 (1962); (17) *ibid.*, **15**, 1244 (1962); (18) *Acta Crystallogr., Sect. B*, **30**, 2240 (1974); (19) *Acta Crystallogr.*, **10**, 125 (1975); (20) *Acta Crystallogr., Sect. B*, **26**, 901 (1970); (21) *Proc. R. Soc. London, Ser. A*, **251**, 441 (1959); (22) *J. Chem. Soc.*, 1001 (1949); (23) *Acta Crystallogr.*, **18**, 693 (1965); (24) *ibid.*, **18**, 753 (1965); (25) *ibid.*, **20**, 325 (1966). [d] Only one of the two types of the hydrogen bonds is given.

In practice, it is not sufficient to treat a single crystal, both because there is not enough information to determine all the constants in eq 1 and because fitting the properties of even several crystals is not a sufficient test of the validity of the model. Thus a large set of crystals containing different packing characteristics should be used. Several algorithms have been developed[4a,14] to derive these parameters from the crystal data, with the stress being on optimizing the efficiency of this inherently time-consuming process. (For example, on an IBM 370/165 the minimization of the lattice energy of a typical crystal takes about 10 min.)

**Intermolecular Forces in Amides and Acids.** Following these considerations, we used the experimental data given in Table I to derive the potential constants for amides and acids for both the 6–9 and 6–12 potentials given in eq 1. These crystals represent a wide range of packing modes, hydrogen-bond geom-

(14) (a) A. T. Hagler and S. Lifson, *Acta Crystallogr., Sect. B*, **30**, 1336 (1974). (b) Techniques to improve the efficiency of the minimization procedure by using an accelerated convergance Fourier transform method are also available; see, for example, D. E. Williams, *Acta Crystallogr., Sect. A*, **27**, 452 (1971).

etries, and secondary and tertiary structures.[15,16] In the amide study, 9 parameters were optimized over 81 observables including lattice constants, sublimation energies, and dipole moments.[7] Once the force field had been derived, the equilibrium geometries were calculated by minimizing the lattice energy with respect to all degrees of freedom, including the positions and orientations of all molecules in the unit cell. Comparison with the experimental structures led to several conclusions concerning the nature of intermolecular forces in amides.[7] (1) The hydrogen-bonding interaction in these crystals is accounted for reasonably by the electrostatic (and van der Waals) forces. No explicit hydrogen-bonding term is needed to account for this interaction.[10] (2) One of the main features of the amide hydrogen bond is the negligible van der Waals repulsion of the amide hydrogen, which allows a short N···O distance. (3) Analysis of remaining deviations, especially in the case of formamide,[7] indicates that the explicit representation of the lone-pair electron density in the carbonyl oxygen would further improve the analytical representation. (4) The electrostatic energy accounts for ~50–75% of the energy in the primary amides and ~33% in the secondary amides. Thus *both* electrostatic and van der Waals interactions must be considered when analyzing the crystal packing.

Following this study, we decided to investigate the extent to which this simple model could account for packing phenomena and hydrogen bonding, and thus we continued to use it in extending the work to carboxylic acids.[17] There were several reasons to maintain the simple model. First and most importantly, we ultimately want to evaluate the degree to which different potentials represent the energy surface. Each improvement in the model will involve additional terms or interactions with additional constants to be evaluated empirically. The "first generation" Lennard-Jones Coulomb model is the simplest possible model and will provide a basis for judging the validity and utility of additional terms. Secondly, it seemed wise to investigate another family of molecules with different characteristics which might bring out other deficiencies in the simple model and thus provide a broader base of information for developing a "second generation" model. Finally, we wanted to investigate transferrability and the extent to which potential functions derived to describe atomic interactions in amides could account for the corresponding interactions in acids.

**Carboxylic Acids.** In this study we modified the strategy we used in the amide study in one respect. We incorporated information from ab initio molecular orbital studies in which we analyzed the electron distribution of acids, amides, and peptides.[18] Comparison of the partial atomic charges in different compounds obtained from population analysis as well as the spatial electron density (see Figure 2) was used in the modeling of the charge distribution. They also provided information as to the transferrability of sets of charges from one family to another.[18] This analysis led to several initial working hypotheses: (1) the parameters for the carboxyl carbonyl group were transferrable from the amides; (2) the hydroxylic oxygen could be described by the same charge and nonbonded parameters as the carbonyl oxygen; (3) a reasonable charge distribution could be obtained by assuming that the COOH group

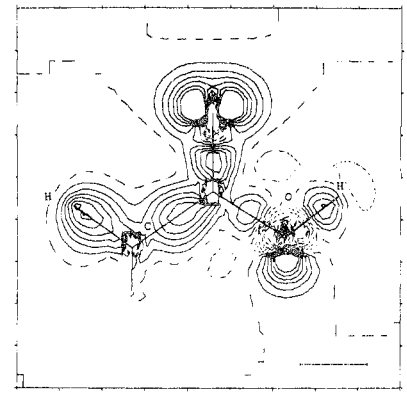


**Figure 2.** Difference electron density map of acetic acid in the plane of the carboxyl group.[18] Solid, dashed, and dotted lines represent positive zero and negative densities, respectively. Consideration of a similar plot for *N*-methylacetamide shows that (1) the difference densities around the carbonyl and alkyl group atoms are very similar in the acid and the amide. (2) The formation of the O–H and N–H bonds involves migration of electrons into the bond which results in a reduced "volume" of these hydrogens. (3) The lone-pair density on the carbonyl oxygen in the acids and the amides reflects the anisotropy of these atoms.

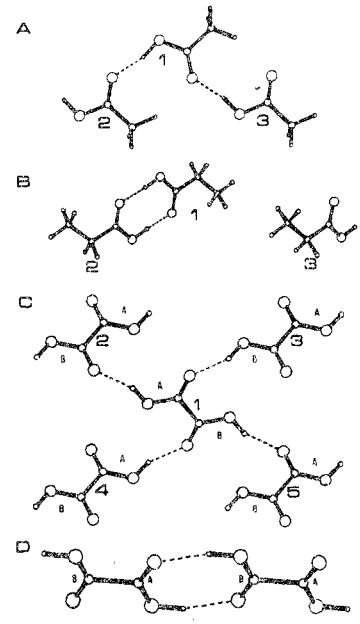



**Figure 3.** Secondary structure of carboxylic acids (for reference, see Table I). (A) Catamer motif in acetic acid crystal. The hydrogen-bonded molecules form a chain along the diagonals of the *bc* plane. Each molecule (1) in this chain is related to the neighboring molecules (2 and 3) by an *n* glide. (B) Cyclic dimer motif in propionic acid crystal. Pairs of molecules 1 and 2 are interlinked by two hydrogen bonds to form centrosymmetric rings. The neighboring molecule (3) is related to (1) by another center of symmetry, between the alkyl group. (C) Catamer motif in $\alpha$-oxalic acid crystal. Each molecule (1) is hydrogen bonded to four other molecules (2–5) via a glide plane, creating corrugated sheets. (D) Linear chains of cyclic dimers in $\beta$-oxalic acid crystal. While in the monocarboxylic acids the dimer motif forms linear arrays of distinct dimers, the existence of carboxyl groups at both ends of the dicarboxylic acids result in hydrogen-bonded ribbons.

(15) A. T. Hagler and L. Leiserowitz, *J. Am. Chem. Soc.*, 100, 5879 (1978).

(16) By secondary structure we refer to the topology of the hydrogen-bonded network while by tertiary structure we refer to the way these networks are packed together in the crystal, in analogy with the terminology used to describe levels of protein structure.

(17) S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.*, 101, 5111 (1979).

(18) A. T. Hagler and A. Lapiccirella, *Biopolymers*, 15, 1167 (1976).

Table II
Potential Parameters for Amides and Acids

| atom | 9-6-1[a] | | | 12-6-1[a] | | | MCMS[b] | |
|---|---|---|---|---|---|---|---|---|
| | $r^*$ | $\epsilon$ | $q^c$ | $r^*$ | $\epsilon$ | $q^c$ | $r^*$ | $\epsilon$ |
| $H_C$ | 3.54 | 0.0025 | 0.11 | 2.75 | 0.038 | 0.1 | 2.92 | 0.037 |
| $H_N$ | 0 | 0 | n | 0 | 0 | n | 2.68 | 0.062 |
| $H_O$ | 0 | 0 | 0.41 | 0 | 0 | 0.35 | 2.83 | 0.044 |
| C | 3.62 | 0.184 | n | 4.35 | 0.039 | n | 4.12 | 0.038 |
| C′ | 3.75 | 0.042 | n | 4.06 | 0.148 | n | 3.74 | 0.141 |
| O | 3.65 | 0.198 | −0.46 | 3.21 | 0.228 | −0.38 | 3.24 | 0.094 |
| O′ | 3.65 | 0.198 | −0.46 | 3.21 | 0.228 | −0.38 | 3.12 | 0.200 |
| N primary | 4.01 | 0.161 | −0.82 | 3.93 | 0.167 | −0.83 | 3.99 | 0.045 |
| N secondary | 4.01 | 0.161 | −0.26[d] | 3.93 | 0.167 | −0.28[d] | 3.99 | 0.045 |

[a] References 7a and 17. [b] Reference 5c (partial charges in this potential are obtained from CNDO calculations of each molecule). [c] An "n" indicates charges obtained from neutrality approximations. The partial charge of the alkyl carbon in CH, $CH_2$, and $CH_3$ is obtained from $q_H$ by assuming these groups are neutral. In addition, in the carboxylic acids, the COOH group is assumed neutral, while in the amides the CO, NH and $NH_2$ groups are constrained to neutrality. [d] The charge distribution in primary and secondary amides has been discussed in ref 18.

is neutral; (4) the van der Waals repulsion of the hydroxylic proton is negligible. This left *only one parameter* to optimize, the charge on the hydroxylic hydrogen, in order to fit 99 structural observables, 7 sublimation energies, and 4 dipole moments. The typical secondary structures[19] included in the least squares are represented schematically in Figure 3.

It was found that this single parameter along with the parameters transferred from the amide force field (see Table II) could account for the properties of the carboxylic acids. These included the crystal structure and sublimation energies of several crystals, not included in the original data set. It also accounted for the relative stabilities of $\alpha$- and $\beta$-oxalic acid and the dimerization energies and structures of formic and acetic acid dimers in the gas phase (see Table III and accompanying discussion).[20]

**Nature of the Intermolecular Interactions. Comparison with Molecular Orbital Calculations.** The legitimacy of using the simple electrostatic, dispersion, and exchange repulsion model of the hydrogen bond in these crystals follows from several observations. First, the carboxylic acid hydrogen-bond length of ~2.65 Å is reproduced within ~0.05 Å[20] by the parameters *derived to account* for amide crystal properties. In the latter systems, the N···O hydrogen-bond distance is significantly larger, ~2.9 Å. Furthermore, the fact that the dimerization energies and structures, which were not included in the optimization, were accounted for as well as sublimation energies indicates that the balance between short- and long-range forces is approximately correct. Also the charge distribution, which yields reasonable sublimation energies and structures, results in good dipole moments, indicating that the electrostatic term in the potential represents this physical interaction and is not an artifact or substituting for another significant term.

It is worthwhile to compare these conclusions with those drawn from extensive quantum mechanical studies which have explored this interaction from another direction.[21] The description given above is in accord with recent studies which have emphasized the importance of electrostatic potentials (e.g., see discussion and citations in ref 21f). This is not to say that charge-transfer effects are completely negligible, but their importance may have been overestimated in the past.[22] In this regard it should be noted that we have *not* obtained a perfect fit to the crystal or dimer properties by any means. The residual deviation in structure and energy may be due to the omission of the charge-transfer and polarization interactions as well as to other factors such as the omission of anisotropy.[7,21] However, these recent results seem to bring the quantum mechanical and the empirical descriptons of the nature of the hydrogen bond much closer.

**A Benchmark. Validity of Different Analytical Representations.** The guidelines discussed above still leave significant latitude for the choice of the trial analytical function.[10] These functions, even when derived for the same functional group, tend to use different systems for their data base, different methods for their derivation, and different criteria to evaluate their validity. We felt it important to establish a "benchmark" to be used for an objective evaluation of various potential forms.[20] Such a benchmark is required by those who would like to apply the functions to various systems in order to judge which is best, what sort of deviations to expect, and what sacrifice in accuracy need be made in using a "computationally inexpensive" potential. In addition, such a comparison can be used to further investigate the physical meaningfulness and utility of various terms and approximations in different force fields.

The application of the benchmark to amides and acid crystals (Table III) resulted in several observations.[23] Overall, a reasonable fit to most properties is obtained,[20] especially with the 9–6–1 potential. The fit of amide properties is better than that for the acids. The better fit to amide properties was attributed to the larger

(19) L. Leiserowitz, *Acta Crystallogr., Sect. B*, **32**, 775 (1976).
(20) A. T. Hagler, S. Lifson, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5122 (1979).
(21) (a) M. Dreyfus and A. Pullman, *Theor. Chim. Acta*, **19**, 20 (1970); (b) K. Morokuma, *J. Chem. Phys.*, **55**, 1236 (1971); (c) P. A. Kollman and L .C. Allen, *Chem. Rev.*, **72**, 283 (1972); (d) L. C. Allen, *J. Am. Chem. Soc.*, **17**, 6921 (1975); (e) P. Schuster in "The Hydrogen Bond", P. Schuster, G. Zundell, and G. Sandorfy, Ed., North Holland, Amsterdam, 1976; (f) P. A. Kollman, *J. Am. Chem. Soc.*, **99**, 4874 (1977); (g) H. Umeyama and K. Morokuma, *ibid.*, **99**, 4874 (1977); (h) see the extensive series of papers by Del Bene and co-workers on hydrogen bonding, *ibid.*, **100**, 1387, 1395 (1978), and references therein.
(22) Smaller basis sets which have often been used due to the large size of the systems treated tend to overestimate the charge-transfer contribution.[21g] For example, early results from the partitioning of a water dimer[21b] yielded −7 kcal total energy; 9.9 kcal exchange repulsion; −8.0 kcal Coulomb; −0.7 kcal polarization, and −8.2 kcal charge transfer. More recent results with significantly improved basis sets yield slightly lower total energies with significantly smaller charge transfer contributions (e.g., −1.7 kcal/mol).[21g]
(23) A complete breakdown of results is given in ref 20.

Table III
Root-Mean-Square Deviations of Properties Calculated for Carboxylic Acids and Amides by the Various Force Fields (kcal/mol, Å, and Degree)

| property | no. of terms | 12-6-1 rms dev | 9-6-1 rms dev | MCMS rms dev |
|---|---|---|---|---|
| | | Acids | | |
| energy | 12 | 2.468 | 2.053 | 2.118 |
| UCV length | 42 | 0.489 | 0.307 | 0.604 |
| UCV angle | 17 | 3.456 | 2.856 | 4.465 |
| volume | 14 | 15.911 | 16.772 | 18.876 |
| $d < 4$ | 14 | 0.247 | 0.190 | 0.322 |
| H··O distance | 16 | 0.062 | 0.072 | 0.058 |
| O···O distance | 16 | 0.047 | 0.071 | 0.041 |
| C—O··O angle | 16 | 11.071 | 9.881 | 14.048 |
| O··O=C angle | 16 | 7.843 | 7.760 | 11.786 |
| H··O=C angle | 16 | 12.362 | 12.144 | 17.985 |
| 180° − O—H··O | 16 | 8.491 | 7.732 | 11.710 |
| | | Amides | | |
| energy | 6 | 1.574 | 1.930 | 8.446 |
| UCV length | 36 | 0.208 | 0.235 | 0.261 |
| UCV angle | 14 | 1.824 | 1.261 | 2.385 |
| volume | 12 | 7.057 | 17.797 | 13.951 |
| $d < 4$ | 12 | 0.145 | 0.145 | 0.164 |
| H··O distance | 30 | 0.049 | 0.059 | 0.056 |
| N··O distance | 30 | 0.055 | 0.055 | 0.076 |
| C—N··O angle | 22 | 3.337 | 3.575 | 4.071 |
| N··O=C angle | 22 | 5.931 | 5.502 | 9.257 |
| H··O=C angle | 30 | 5.830 | 5.609 | 7.329 |
| 180° − NH···O | 30 | 4.396 | 3.894 | 4.093 |

number of hydrogen bonds per amide group 4, compared to the carboxyl group 2. The hydrogen-bond distances as seen in Table III are fit significantly better than other interatomic distances, thus serving as "constraints".

A further analysis of the results shows that, in general, if one potential has problems accounting for a property of a given crystal, all three do. This indicates that although the functions were derived completely independently and the MCMS[5b] has an explicit 10–12 hydrogen-bond term, they have common deficiencies. It may be, for example, that an improved fit of the formamide structure and the lattice energies and structures of oxalic acid require the anisotropic nature of the electron distributions to be represented[7,21h,24] (see also Figure 2). Finally, this comparative analysis is useful in showing that the additional parameters used in the explicit 10–12 hydrogen-bond potential[5b] do not yield a better fit to the crystal properties. Thus, it does not appear that the remaining deviations in these crystals can be improved by adjusting the representation of this interaction in terms of an isotropic atom–atom potential.

**Applications to the Study of Crystal-Packing Modes.** At this stage we turn to the way in which these potentials can be used as a tool to help us understand crystal packing. One of the major advantages of computer simulation over experiment is that the energy and its variation with structure may be investigated and partitioned into molecular and submolecular contributions. For example, the question arises as to why formic and acetic acids exhibit the catamer motif while most of the remaining acids form cyclic dimers (Figure 3). We first "removed" the hydrogen-bonded molecules forming the motif for each of the secondary structures from the crystal in order to study the intrinsic stabilities of the isolated hydrogen-bonding mode (Figure 3). The

(24) A. T. Hagler, P. Dauber, and S. Lifson, *J. Am. Chem. Soc.*, 101, 5131 (1979).

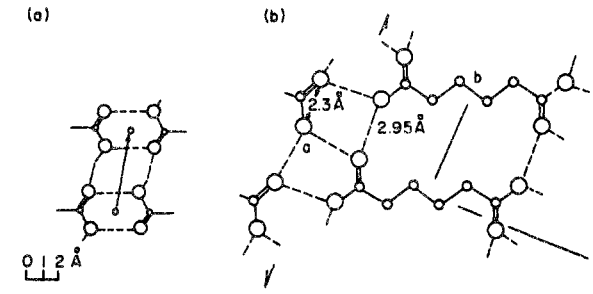



**Figure 4.** Secondary structure motifs of amides (for reference, see Table I). (a) Packing of amide dimers along a translational axis. Almost all primary amides form hydrogen-bonded cyclic dimers in the crystal. These can be interlinked by hydrogen bonds along a translational axis, as depicted here, or by a glide plane or twofold screw axis. (b) The exceptional packing arrangement in adipamide crystal. The amide group hydrogen bonds to two molecules along a twofold screw axis. The resulting secondary structure is distorted owing to the incompatability in intra- and intermolecular geometry as depicted in the figure. The intramolecular N··O distance is 2.3 Å, while the translation-produced hydrogen-bonded N··O distance is 2.95 Å, leading to NH··O hydrogen bonds across the twofold axis which are not parallel and splayed open.

first observation made was that the catamer motif is slightly more stable than the cyclic dimer with all potentials used. This is consistent with the fact that both secondary structures are available to formic and acetic acids where the catamer motif obtains. The higher homologues do not have a choice since a steric interaction occurs between the larger substituent and the carbonyl group[19] (Figure 3). In the catamer motif, the hydrogen bond itself, i.e., the O—H··O=C interaction, is less stable than in the cyclic dimer but this is more than compensated for by the electrostatic interactions between C=O··C=O and O—H··O—H groups, which reflect the favorable orientation of their respective dipoles in the catamer.[25]

**Why Does a Crystal Pack in Its Observed Packing Mode? Hypothetical Crystals.** The second application we address ourselves to is the understanding of the factors contributing to the stability of a given packing mode. Experimental analysis is restricted to the comparison of observed polymorphic crystals (e.g., Figure 3C,D); however, computer simulation offers an additional and perhaps more powerful degree of freedom. This is "crystallizing" hypothetical crystals, comparing their stability relative to the observed structure, and asking what factors are responsible for the stability of the latter.

This technique was employed recently in a study of the stability of the anomalous packing mode of adipamide (see discussion in legend of Figure 4).[15,26] Note that this technique must be used in conjunction with experiment in order to recognize anomalous packing and to build "relevant" hypothetical crystals.[15,27] To understand why adipamide packs in this apparently strained motif, 16 hypothetical crystals were constructed. All contained the cyclic hydrogen-bonded

(25) Further analyses of the packing modes of the carboxylic acids including the relative stabilities of $\alpha$- and $\beta$-oxalic acids and the source of calculated deviations in several of the structures is given in ref 24.

(26) A similar study has been carried out to compare two hypothetical crystals of acetic acid containing cyclic dimers packed with the observed structure and space group of trifluoroacetic acid: J. L. Derissen and P. H. Smit, *Acta Crystallogr., Sect. A*, 33, 230 (1977).

(27) (a) L. Leiserowitz and G. M. J. Schmidt, *J. Chem. Soc. A*, 2372 (1969). (b) L. Leiserowitz and A. T. Hagler, work in progress.

secondary structure in space groups ranging from triclinic to orthorhombic and packing modes typical of amides.[15] The stability of the observed structure over all 16 hypothetical structures was accounted for, despite the fact that the cyclic hydrogen-bonded secondary structure was *more favorable* than the distorted motif observed in adipamide. Energy partitioning revealed that by foregoing the most favorable hydrogen-bonding motif, adipamide achieved a better *inter*layer packing which more than compensated for the "poorer" hydrogen bonding.

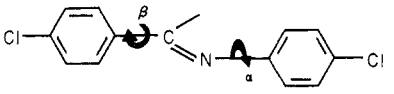
**Effect of Crystal Forces on Molecular Conformation.** Although X-ray crystal studies provide precise molecular geometry, structural determinations of flexible molecules yield the solid-state conformation which may deviate considerably from that in solution. This problem is of crucial importance in biological compounds where activity is related to conformation.[28]

A preliminary study of this effect was carried out in an attempt to estimate the magnitude of the crystal forces on the torsion angles about N–$C^{\alpha}$ ($\phi$) and $C^{\alpha}$–C′ ($\psi$) in *N*-methylacetamide.[29] It was found that the variation with conformation of the intermolecular and intramolecular energies (calculated by both ab initio and empirical techniques) was of the same magnitude. This is in accord with the experimental observations that in different crystals these angles tend to cluster about the most favored intramolecular values but with variations of up to 30° depending on the particular crystal.[29]

A similar study was performed on adrenaline, 5-methoxy-*N*,*N*-dimethyltryptonine, and serotonin by Caillet et al.[30] The molecules, in conformations corresponding to intramolecular local minima as obtained from the semiempirical PCILO method, were inserted into the observed crystal lattice and the lattice energy was minimized. It was found that the lattice containing the observed conformation was more stable than the lattices of any of the local intramolecular minimum energy conformations. The authors point out a fundamental problem with this approach, however, as they noted that the minimized hypothetical structures do not always correspond to the lowest lattice energy *in this space group* for the particular conformation. Furthermore it is possible and even probable (see below) that the different conformations would crystallize in lattices with completely different symmetry and packing. This was not checked since in general it is extremely difficult to know what packing modes are available to conformational isomers.

A different approach to the problem was formulated recently by Bernstein and Hagler.[31] The methodology is centered around the phenomena of "conformational polymorphism", in which a given molecule is observed to adopt significantly different conformations in different crystalline polymorphs. Thus, we avoid the problem of finding the most stable structure of the "conformational isomer" since it is experimentally observed.

A detailed investigation was carried out on a model system which included chloro and methyl derivatives of benzylideneaniline. The *p*-dichloro derivative (II)




II, dichlorobenzylideneaniline (BACl)

crystallizes in two polymorphic forms. In the triclinic form ($P\bar{1}$) the molecule is in the less stable planar form ($\alpha = \beta = 0°$), while in the orthorhombic form (*Pccn*) the molecule is twisted with the two rings rotated about the exocyclic bonds by ±24.8°.[32]

The analysis of the inter- and intramolecular energetics was developed in several stages. Lattice energy calculations of the triclinic and orthorhombic forms with three different potential functions all yielded more negative energies for the triclinic lattice in agreement with its observed (meta) stability. Ab initio molecular orbital calculations of the *intra*molecular energy were used next and showed that the planar form was less stable than the nonplanar molecule by approximately the difference in the lattice energies between the orthorhombic and triclinic forms. The total lattice energy was partitioned into "partial atomic energies", i.e., the contribution of each atom to the total lattice energy, to understand the mechanism by which the higher energy molecular conformation is stabilized in the triclinic form. Several interesting observations on the comparative energetics of the triclinic and orthorhombic lattices emerged from the analysis of these partial atomic energies. The chlorine atom, which yielded the single largest contribution to the total lattice energy, has essentially the same *energetic* environment in the two forms and does not contribute significantly to the difference in lattice energies. The bridge region and the phenyl rings both contribute to the stabilization of the triclinic form over the orthorhombic. Both are in energetically more favorable environments in the triclinic structure, to roughly equal extents. No one atom, or group, dominates the contribution to the stabilization of the triclinic form.

A second study aimed at answering the question as to why this molecule (II), does not pack in a structure containing the ordered, lowest energy molecular conformation ($\alpha \sim 0°$, $\beta \sim 45°$)[31b] followed. In general this would be an extremely difficult problem to consider, as no information is available as to the "nonexistent" packing of the low-energy molecular form. In this system, however, one of the polymorphic forms of the dimethyl analogue (space group $P2_1$) is observed to contain the low-energy form of the isolated benzylideneaniline molecule.[32] By replacing the methyl groups with Cl computationally, the problem becomes amenable to attack.

Comparing the lattice energies of the two observed crystal forms with that of BACl in this $P2_1$ structure shows the former to be more stable, compensating for the lower intramolecular energy. Analysis using the energy partitioning pinpointed the relatively unfavorable energetic environment of the aniline ring, and the

(28) (a) C. M. Deber, V. Madison, and E. R. Blout, *Acc. Chem. Res.*, **9**, 106 (1976); (b) F. A. Bovey in "Peptides, Polypeptides and Proteins", Proceedings of the Rehovot Symposium, Wiley, New York, 1974, p 248.

(29) A. T. Hagler, L. Leiserowtiz, and M. Tuval, *J. Am. Chem. Soc.*, **98**, 4600 (1976).

(30) (a) J. Caillet, P. Claverie, and B. Pullman, *Acta Crystallogr., Sect. B*, **32**, 2740 (1976); (b) *Acta Crystallogr., Sect. A*, **33**, 885 (1977).

(31) (a) J. Bernstein and A. T. Hagler, *J. Am. Chem. Soc.*, **100**, 673 (1978); (b) A. T. Hagler and J. Bernstein, *ibid.*, **100**, 6349 (1978); (c) J. Bernstein and A. T. Hagler, *Mol. Cryst. Liquid Cryst.*, in press.

(32) I. Bar and J. Bernstein, *Acta Crystallogr., Sect. B*, **33**, 1738 (1977).

chlorine on this ring, in this packing mode as the underlying cause for the lack of stability.

**Future Directions.** A start has been made in systematically extracting and evaluating analytical representations of intermolecular forces from crystal data. These potentials have been used, in conjunction with a knowledge of the packing modes available to a given functional group, to analyze the basis of the energetic preference of a given molecule for its packing motif. The phenomena of conformational polymorphism gives us the space group corresponding to the most stable packing of conformational isomers, enabling us to study the effect of crystal forces on molecular conformation. Where do we go from here? There are two general directions with much room for future progress. The first is the application of the existing methodology to additional systems. It is desirable to extend the derivation of a simple force field to such families as sugars, nucleotides, thio groups, and alcohols. Here the objective is twofold: to obtain a reasonable representation of the intermolecular forces in these systems, and, more importantly perhaps, to gain further insight into the intermolecular forces and the deficiencies in this simple analytical form. Related to this is the extension of the benchmark to all recent analytical forms proposed to account for acids and amides and to expand it to include the additional families discussed above. The same general comments apply to the analysis of the underlying basis for crystal packing and the effect of crystal forces on molecular conformation; it is worthwhile to apply these studies to still additional systems, both to further substantiate the validity of the methodolgy and to gain further understanding of the crystal state.

The second direction is the extension and improvement of the methodology itself. In the case of the analytical forms this might take the form of improvement of the models to account explicitly for such effects as polarization and anisotropy of the electron distribution about atoms. Here it might be worthwhile to incorporate both ab initio energy surfaces and spatial electron densities into the methodology as improvements in both software and computing power are making these techniques extremely powerful. Again, the existence of a benchmark should prove invaluable. It also seems worthwhile to attempt to generalize the building of systematic realistic, hypothetical crystal structures to the analysis of the effect of crystal forces on molecular conformation, relaxing the need for conformational polymorphs in the attempt to find the most stable packing motif for the various conformational isomers. Further developments in this field might also include the generalization of this analysis to understand the difference between solvent effects and crystal forces on the conformational behavior of flexible molecules using the rapidly developing techniques of Monte Carlo and molecular dynamics simulations to study the former. A well-known statement of the late Winston Churchill aptly summarizes the current status of this field: "Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning."

*This work was supported by grants from the U.S.-Israel Binational Science Foundation, Jerusalem, and a grant from the Israel Academy of Science. It is a pleasure to acknowledge Drs. J. Bernstein and L. Leiserowitz for sharing their insights into the problems of crystal packing with us. It is also a pleasure to thank Professor S. Lifson, on the occasion of his 65th birthday, for his continuing interest and support in all aspects of this work.*