suitable for such exploitation, e.g., for selective oxidation of hydrocarbons.

Concluding Remarks

It should be encouraging, for organic chemists and chemical engineers alike, to know that enzymes can work not only in water but also in organic solvents. The still-existing prejudice against enzymes as practical catalysts will be lessened further by the forthcoming greater availability of a wider variety of enzymes at a lower cost, brought about by modern biotechnology. Moreover, protein engineering, particularly redesigning enzymes by site-directed mutagenesis,[90] has an exciting potential of altering the enzymatic properties at will, e.g., broadening substrate specificity,[91] as well as enhancing enzyme action in organic media.[92]

The ever intensifying search for novel biologically active compounds inevitably results in increasingly complex molecules, often containing stereocenters and existing as distinct structural isomers. Usually, one of the enantiomers or isomers has a higher biological activity than others. Consequently, there is mounting pressure on the manufacturers of pharmaceuticals and agricultural chemicals to produce only the desired stereoisomer, which would enhance the compound's benefit-to-harm ratio. As a result, the need for simple and scalable asymmetric conversions will grow. Hopefully, this Account demonstrates that enzymatic catalysis in organic solvents provides a useful methodology toward that aim.

I am grateful to David Volkin and Paul Burke for their help in the preparation of this manuscript and to the National Institutes of Health for financial support (Grant GM39794).

(90) Fersht, A. R.; et al. *Angew. Chem., Int. Ed. Engl.* **1984**, *23*, 467–473. Knowles, J. R. *Science* **1987**, *236*, 1252–1258.
(91) Bone, R.; Silen, J. L.; Agard, D. A. *Nature* **1989**, *339*, 191–195.
(92) Arnold, F. H. *Protein Eng.* **1988**, *2*, 21–25.

Encoding and Decoding Hydrogen-Bond Patterns of Organic Compounds

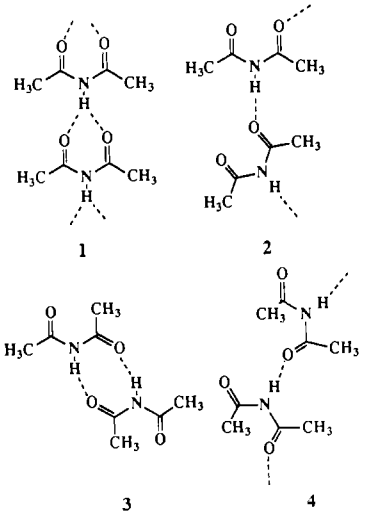
MARGARET C. ETTER

Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455

Received June 20, 1989 (Revised Manuscript Received January 31, 1990)

Are hydrogen-bond patterns predictable? Benzoic acids form cyclic dimers. Salicylic acid has an intramolecular hydrogen bond. These hydrogen-bond patterns seem obvious because they satisfy the chemical criteria of pairing a somewhat acidic hydrogen with an electronegative atom, and because spectroscopic and crystallographic studies have confirmed our chemical intuition about how and when hydrogen bonds should form in these and in similar frequently studied structures.[1] But what about other, less studied, yet simple compounds like diacetamide? Diacetamide has two proton acceptors, one proton donor, and three possible conformers each of which could form two- or three-centered hydrogen-bonded chains or dimers. Four of the hydrogen-bond possibilities are shown. Patterns 1 and 3 are found in the two polymorphs of diacetamide.[2]

Besides these patterns, diacetamide also cocrystallizes with many other hydrogen-bonding molecules.[3] The inherent complexities in making such predictions are apparent in the intricate hydrogen-bond pattern observed for diacetamide and acetamide,[4] shown below. It certainly is not obvious that the cyclic pattern shown



1 2 3 4

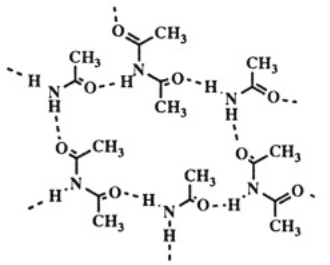
Margaret C. Etter, a native of Wilmington, DE, obtained her B.A. in chemistry from the University of Pennsylvania, an M.Sc. from the University of Delaware, and her Ph.D. in organic chemistry from the University of Minnesota in 1974 under the direction of J. Z. Gougoutas. After seven years as a solid-state chemist and crystallographer at the 3M Company in St. Paul, MN, she returned to the University of Minnesota to do postdoctoral work with R. G. Bryant in solid-state NMR spectroscopy. She is currently an Associate Professor of Chemistry at the University of Minnesota, where her work on solid-state organic chemistry involves interdisciplinary studies of structures and properties of small-molecule organic solids, with special interests in hydrogen-bond interactions and crystal growth. She was recently awarded an Alfred P. Sloan Fellowship.

(1) The idea that some hydrogen-bond patterns are obvious and can be used to prepare aggregate patterns with predictable structures has been demonstrated by several different groups. (a) Ducharme, Y.; Wuest, J. D. *J. Org. Chem.* **1988**, *53*, 5789–5791. (b) Chang, S. K.; Hamilton, A. D. *J. Am. Chem. Soc.* **1988**, *110*, 1318–1319. (c) Askew, B.; Ballester, P.; Buhr, C.; Jeong, K. S.; Jones, S.; Parris, K.; Williams, K.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1989**, *111*, 1082–1090, (d) Stadler, R. *Prog. Colloid Polym. Sci.* **1987**, *75*, 140–145.
(2) (a) Kuroda, Y.; Taira, Z.; Uno, T.; Osaki, K. *Cryst. Struct. Commun.* **1975**, *4*, 321–324. (b) Kuroda, Y.; Taira, Z.; Uno, T.; Osaki, K. *Ibid.* **1975**, *4*, 325–328.
(3) Etter, M. C.; Reutzel, S., unpublished results.

0001-4842/90/0123-0120$02.50/0

is the most likely aggregate pattern.

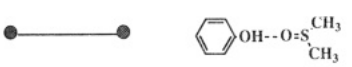
Diacetamide-Acetamide (1:1) Cocrystal



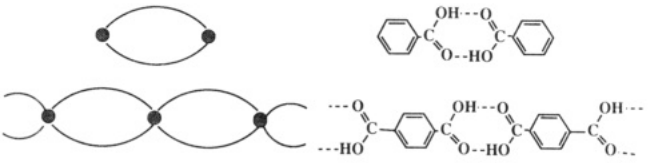
In the solid state, hydrogen-bond patterns are usually well defined and often involve infinite chains or arrays. Dimer patterns may be preferred in dilute solutions, and chains preferred in concentrated solutions or melts. Solvents with high dielectrics might favor more polar conformers. Such considerations pose additional difficulties for making empirical predictions of hydrogen-bond patterns. Theoretical methods have proven useful in addressing some of these problems.[5]

In this Account it is shown that empirical hydrogen-bond rules useful for determining preferred modes of hydrogen bonding can be developed by using the vast resource of data on intermolecular contacts found in the Cambridge Crystallographic Database.[6] Different functional-group classes show clear preferences for specific hydrogen-bond patterns in their crystal structures, despite the presence of other unpredictable and nonspecific lattice forces. These preferences can be framed as empirical "rules", which are most useful for predicting complete hydrogen-bond patterns of systems with a limited number of functional groups, or for predicting hydrogen-bond patterns of subsets of more complex systems. The rules are to be used as indicators of hydrogen-bond preferences of a particular functional group, whether in the solid state or in solution, in the absence of other competing forces.

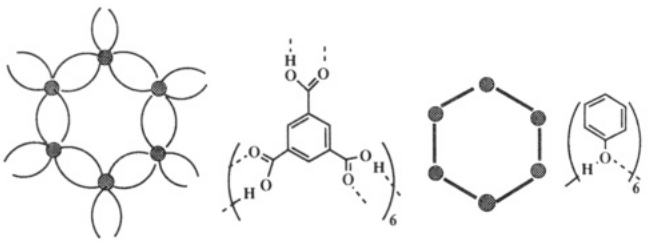
One of the challenges in determining hydrogen-bond preferences from crystal structures is to find a useful way to define a hydrogen-bond pattern. Frequently there are multiple intertwined hydrogen-bond patterns present in a single crystal structure, and similarities between complex patterns of chemically different systems may be completely lost without a consistent and analytically straightforward way to define the patterns. For example, a phenol–DMSO pair could be represented graphically by two points (the molecules) and one line (the hydrogen bond). The pattern is finite, and many other molecules can be imagined that fit this pattern:



Monocarboxylic acid dimers can be represented similarly, showing a finite dimeric graph with two hydrogen bonds per molecule. The infinite repeating dimer pattern is representative of the hydrogen-bond pattern of a dicarboxylic acid, such as terephthalic acid:



Other patterns can be constructed at will and molecules designed to fit the patterns. Conversely, similarities between subsets of hydrogen-bond patterns of known structures are readily recognized from their graphs. For example, trimesic acid (1,3,5–benzenetricarboxylic acid), left, and a cyclic hexameric phenol, right, would have a common hexagonal pattern. Other subsets of the trimesic acid structure match those of benzoic acid, and of terephthalic acid.



Assigning graph sets to particular hydrogen-bond patterns is the encoding process discussed in this Account. Development of empirical hydrogen-bond rules, based on the frequency of occurrence of specific graph sets within a functional group class, and based on observations about stereoelectronic hydrogen-bond preferences and hydrogen-bond selectivity, is the decoding process.

Hydrogen Bonds

The criteria for identifying hydrogen bonds must be established in order to define a hydrogen-bond pattern. Typically, criteria for defining hydrogen bonds are based on physical[7] and geometric properties.[8] Useful summaries of these criteria are given in recent reviews,[9] and comprehensive analyses are also available.[10] Most of these methods, however, focus on the three or four atoms involved in a hydrogen bond rather than on the consequences of such contacts for the rest of the atoms and molecules in the system. Now, with the luxury of a large data base of crystal structures of organic compounds,[6] the function of hydrogen bonds in aggregate structures can be evaluated, and the behavior of parts of a molecule or of many molecules in relation to their hydrogen bonds can be studied. The collective behavior of hydrogen bonds in many different related structures

(4) Matias, P. M.; Jeffrey, G. A.; Ruble, J. R. *Acta Crystallogr.* **1988**, *B44*, 516–522.

(5) Ab initio calculations: (a) Jorgensen, W. L. *Acc. Chem. Res.*, in press, (b) Dory, M.; Delhalle, J.; Fripiat, J. G.; Andre, J.-M. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *14*, 85–103. Semiempirical calculations: (c) Radom, L.; Riggs, N. V. *Aust. J. Chem.* **1980**, *33*, 2337–2342. (d) Vinson, L. K.; Dannenberg, J. J. *J. Am. Chem. Soc.* **1989**, *111*, 2777–2781. Molecular mechanics calculations: (e) Dauber, P.; Hagler, A. T. *Acc. Chem. Res.* **1980**, *13*, 105–112 and references therein.

(6) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. *Acta Crystallogr.* **1979**, *B35*, 2331–2339.

(7) (a) Arnett, E. M. *Prog. Phys. Org. Chem.* **1963**, *1*, 223–403. (b) Arnett, E. M.; Mitchell, E. J.; Murty, T. S. S. R. *J. Am. Chem. Soc.* **1974**, *96*, 3875–3891.

(8) Taylor, R.; Kennard, O. *Acc. Chem. Res.* **1984**, *17*, 320–326.

(9) (a) Kroon, J.; Kanters, J. A.; van Duijneveldt van de Rijdt, J. G. C. M.; Duijneveldt, F. B.; Vliegenthart, J. A. *J. Mol. Struct.* **1975**, *24*, 109–129. (b) Brown, I. D. *Acta Crystallogr.* **1976**, *A32*, 24–31. (c) Ceccarelli, C.; Jeffrey, G. A.; Taylor, R. *J. Mol. Struct.* **1981**, *70*, 255–271. (d) Taylor, R.; Kennard, O.; Versichel, W. *J. Am. Chem. Soc.* **1984**, *106*, 244–248. (e) Donohue, J. *Structural Chemistry and Molecular Biology*; Rich, A., Davidson, N., Eds.; W. H. Freeman: San Francisco, 1968; pp 443–465. (f) Joesten, M. D. *J. Chem. Educ.* **1982**, *59*, 362–366.

(10) Schuster, P., Zundel, G., Sandorfy, C., Eds. *The Hydrogen Bond*; North-Holland: Amsterdam, 1976; Vols. I–III.

is the focus of this study. Our operative definition of hydrogen bonds is taken from Pauling's early definition of chemical bonds, which emphasizes the *organizational consequences* of bonds:[11]

> A hydrogen bond is an interaction that directs the association of a covalently bound hydrogen atom with one or more other atoms, groups of atoms, or molecules into an aggregate structure that is sufficiently stable to make it convenient for the chemist to consider it as an independent chemical species.

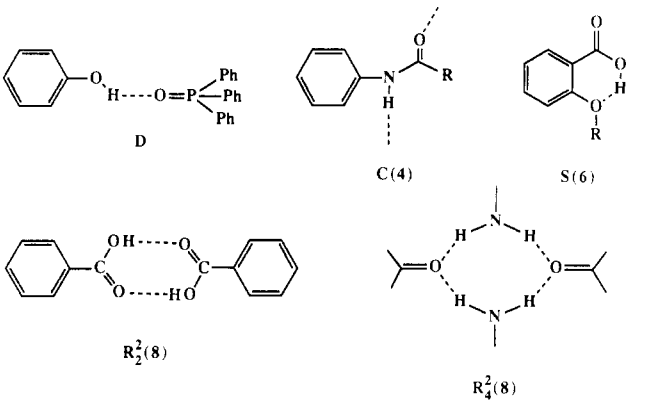
X-ray crystallography has made it convenient to consider hydrogen-bonded aggregates as independent species. The range of energies implied for hydrogen bonds defined in this way is yet to be determined. The range will probably extend down to less than 1 kcal/mol, but those very weak interactions will be included only when they have resulted in recognizable and recurring aggregate patterns.

Graph Sets

The purpose of graph-set assignments is to define the morphology of hydrogen-bonded arrays. Kuleshova and Zorkii[12] were the first to apply graph theory[13] to organic crystal structures, while Wells[14] and Hamilton and Ibers[15] had developed accounting schemes for hydrogen bonds, which were precursors of the graph-set concept used here. In our method, real molecules rather than points are used as nodes, and hydrogen bonds are differentiated by the type of donors and acceptors that are present rather than being represented as just abstract lines.[16]

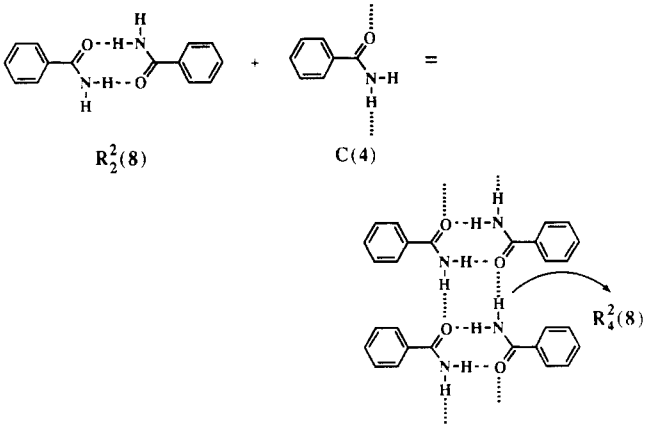
The process of assigning a graph set begins with identification of the number of different types of hydrogen bonds, as defined by the nature of the donors and acceptors in a hydrogen bond, that are present in the structure of interest. The set of molecules that are hydrogen bonded to one another by repetition of just one of these types of hydrogen bonds is called a motif and is characterized by one of four designators that indicate whether the motif is infinite or finite, and cyclic or not. For motifs generated from intermolecular hydrogen bonds, these designators are **C** (chain), **R** (ring), and **D** (dimer or other finite set), while **S** denotes an intramolecular hydrogen bond. The number of donors (**d**) and acceptors (**a**) used in each motif are assigned as subscripts and superscripts, respectively, and the size or degree of the motif (corresponding to the number of atoms in the repeat unit) is indicated in parentheses.

Chart I
Graph-Set Assignments for Representative Hydrogen-Bond Motifs



Thus a benzoic acid dimer is a motif with graph set $\mathbf{R}_2^2(8)$, and a phenol hydrogen-bond chain has graph set $\mathbf{C}(2)$. Examples are given in Chart I.

One of the most useful aspects of this methodology involves differentiating between motifs containing only one type of hydrogen bond, and networks, $\mathbf{N}_i$, containing two or more types of hydrogen bonds. For example, a primary amide has two motifs corresponding to a cyclic dimer $\mathbf{R}_2^2(8)$ and a chain $\mathbf{C}(4)$. Combination of the two motifs generates a new $\mathbf{R}_4^2(8)$ pattern, shown below, containing two kinds of hydrogen bonds. In some cases the nature of an implied pattern like this one is important to the properties or structures of the compounds being studied, so it is helpful to specify it independently.



Graph-set assignments, thus, are done in stages. The first stage involves assigning all the motifs. Higher order networks resulting from combinations of the motifs may be assigned or not, as dictated by the chemical problem at hand. For very complex patterns, higher and higher order networks, involving three, four, or more different combinations of hydrogen bonds, can be identified systematically in this manner. Details of the process of making graph-set assignments have been published elsewhere.[16]

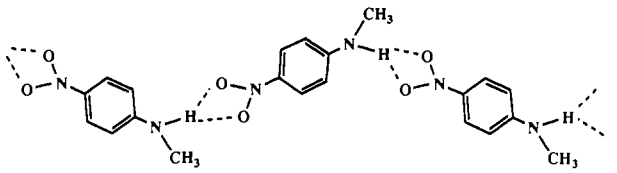
Using Graph Sets

By assigning graph sets to a series of crystal structures that contain one type of functional group, preferred patterns of hydrogen bonding may be seen. Leiserowitz found that most primary amides had the preferred hydrogen-bond motifs involving cyclic dimers and chains, $\mathbf{C}(4)\mathbf{R}_2^2(8)$.[17] Special relationships between

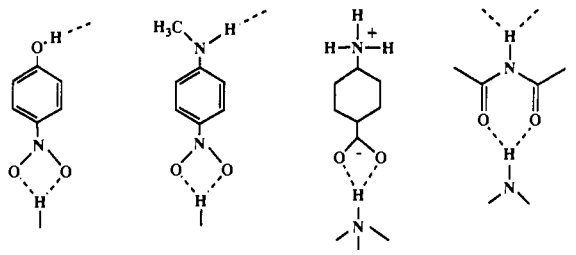
(11) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: New York, 1960; p 6.
(12) (a) Kuleshova, L. N.; Zorky, P. M. *Acta Crystallogr.* **1980**, *B36*, 2113–2115. (b) Zorky, P. M.; Kuleshova, L. N. *Zh. Strukt. Khim.* **1980**, *22*, 153–156.
(13) (a) Harary, F. *Graph Theory and Theoretical Physics*; Academic Press: New York, 1967. (b) Merrifield, R. E.; Simmons, H. E. *Topological Methods in Chemistry*; John Wiley and Sons, Inc.: New York, 1989.
(14) Wells, A. F. *Structural Inorganic Chemistry*; Clarendon Press: Oxford, 1962; pp 294–315.
(15) Hamilton, W. C.; Ibers, J. A. *Hydrogen Bonding in Solids*; W. A. Benjamin, Inc.: New York, 1968; pp 19–21.
(16) (a) Etter, M. C.; Bernstein, J. B.; MacDonald, J. M. *Acta Crystallogr., Sect. B*, in press. (b) Etter, M. C. *Isr. J. Chem.* **1985**, *25*, 312–319. (Graph-set notation was introduced in this paper. Some of the designations have been changed in the present Account to be more consistent with other work in the field.)

different amides and between amides and other kinds of molecules become evident upon comparison of their graph sets. The orthorhombic form of acetamide, for example, has an anomalous hydrogen-bond pattern, reflected in its unusual graph set, $C(4)R_1^1(12)$.[18] Most carboxylic acids have graph sets that match one of the primary amide motifs, $R_2^2(8)$, while secondary amides have the same graph set as the other primary amide motif, $C(4)$.

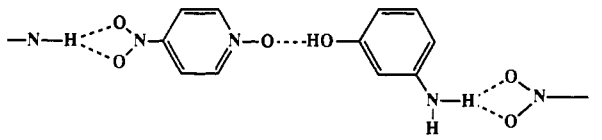
A study of all known crystal structures of small-molecule nitroaniline compounds showed that nitroanilines prefer a motif involving one amino proton associating with both oxygens of a nitro group,[19] as shown below. This pattern was observed even when bond lengths from nitro oxygens to the nearest amino hydrogens were longer than the sum of van der Waals radii. In several of the papers cited in the literature, it was stated that these interactions were not hydrogen bonds. Nevertheless, we found that the same aggregate pattern was present in all cases except those with unusual steric hindrance around the amino or nitro groups.



The hydrogen-bonding relationships derived for nitroanilines are useful for understanding the hydrogen-bonding properties of other molecules as well. For example, the OH group and the NH group in *p*-nitrophenol and *N*-methyl-*p*-nitroaniline both have a single proton donor that associates with a nitro group in a graph set $R_1^2(4)$. Chemically different functional groups having the same graph set are isographic. Recognition of isographic relationships between functional groups is a very useful and time-saving feature of this method since it allows the graph sets for one class of compounds to be transferred to another. Isographic functional groups usually have the same number of proton donors and acceptors, as shown.



The concept of isographic relationships also provides a powerful design tool for constructing cocrystals. In the nitroaniline examples above, the proton donors and acceptors were on the same molecule, but they should hydrogen bond to one another in the same way even if they are on different molecules. This principle is illustrated in the cocrystal structure of *p*-nitropyridine *N*-oxide and *m*-aminophenol, where the *N*-oxide and OH groups form a strong hydrogen bond to one another. The dimer is analogous to *m*-nitroaniline since the pair of molecules has a nitro group and an amino group oriented meta to one another. In this cocrystal, acentric chains form as a result of the expected intermolecular nitroaniline hydrogen bonding. The entire bulk cocrystal is acentric,[20] as is *m*-nitroaniline.[21]



Graph sets can also be used to understand polymorphic relationships. Iminodiacetic acid, $^-O_2CCH_2NH_2^+CH_2CO_2H$, is zwitterionic and exists in three polymorphic modifications. Two previous attempts had been made in the literature to compare and analyze the hydrogen-bond patterns in these structures.[22] The analyses involved comparing bond lengths and bond angles, symmetry relationships between neighboring hydrogen-bonded molecules, and identification of some of the ring patterns. Nevertheless, little insight was gained into similarities that existed between these structures or into the relative stabilities of the crystals. One of the previous authors, J. Bernstein, recently used the graph-set method to reanalyze these patterns. Similarities in the motifs were observed immediately. Five motifs occurred in different combinations in the three polymorphs. Two of the polymorphs had the same first-order network but differed in their second-order network, as shown.[23]

polymorph 1: $N_1 = C(5)R_2^2(10)C(8)$ $N_2 = R_2^2(14)$

polymorph 2: $N_1 = C(5)R_2^2(10)C(8)$ $N_2 = R_4^2(8)$

polymorph 3: $N_1 = C(5)C(5)C(8)$

Hydrogen-Bond Rules

Empirical hydrogen-bond rules (Table I) are a way to use correlations that exist between functional groups and hydrogen-bond patterns, with graph sets being one type of correlation. Graph sets deal with the connectivity or configuration of a set of hydrogen-bonded molecules, analogous to the primary structure of a macromolecule. Other stereoelectronic or structural factors affecting the primary, secondary, and even tertiary structures provide the other hydrogen-bond rules.[24]

The first three rules listed in Table I apply in general to functional groups in neutral organic molecules. To apply the first rule, which was originally put forth by

(17) (a) Carboxylic acids: Leiserowitz, L. *Acta Crystallogr.* **1976**, *B32*, 775–802. (b) Primary amides: Leiserowitz, L.; Schmidt, G. M. J. *J. Chem. Soc. A* **1969**, 2372–2382. (c) Secondary amides: Leiserowitz, L.; Tuval, M. *Acta Crystallogr.* **1978**, *B34*, 1230–1247. (d) From refs 17a–c, four simple hydrogen-bond rules are evident: carboxylic acids prefer to form cyclic hydrogen-bonded dimers ($R_2^2(8)$); a few carboxylic acids form hydrogen-bonded chains ($C(4)$); acyclic secondary amides form only chains ($C(4)$); primary amides form cyclic dimers which are hydrogen-bonded together to form chains ($C(4)R_2^2(8)$). Additional specific rules may be further deduced from these papers.

(18) Hamilton, W. C. *Acta Crystallogr.* **1965**, *18*, 866–870.

(19) Panunto, T. W.; Urbańczyk-Lipkowska, Z.; Johnson, R. B.; Etter, M. C. *J. Am. Chem. Soc.* **1987**, *109*, 7786–7797.

(20) Lechat, J. R.; de Almeida Santos, R. H.; Bueno, W. A. *Acta Crystallogr.* **1981**, *B37*, 1468–1470.

(21) Shapski, A. C.; Stevenson, J. L. *J. Chem. Soc., Perkin Trans. 2* **1973**, 1197–1200.

(22) (a) Bernstein, J. *Acta Crystallogr.* **1979**, *B35*, 360–366. (b) Boman, C.-E. Herbertsson, H.; Oskarsson, A. *Acta Crystallogr.* **1974**, *B30*, 378–382.

(23) Bernstein, J.; Etter, M. C.; MacDonald, J. M. *J. Chem. Soc., Perkin Trans. 2*, in press.

(24) The idea of secondary structure in sets of small molecules is discussed in the following: (a) Leiserowitz, L.; Hagler, A. T. *Proc. R. Soc. London* **1983**, *A388*, 133–175. Stereoelectronic effects in hydrogen bonding are discussed in the following: (b) Glusker, J. P.; Murray-Rust, P. *J. Am. Chem. Soc.* **1984**, *106*, 1018–1025.

Table I
Hydrogen-Bond Rules for Organic Compounds

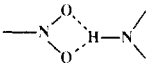
A. General Rules

1. All good proton donors and acceptors are used in hydrogen bonding.
2. Six-membered-ring intramolecular hydrogen bonds form in preference to intermolecular hydrogen bonds.
3. The best proton donors and acceptors remaining after intramolecular hydrogen-bond formation form intermolecular hydrogen bonds to one another.

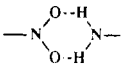
B. Additional Rules for Specific Classes of Functional Groups

Nitroanilines[19]

4. Amino protons will hydrogen bond to nitro groups.
5. One or more intermolecular amino–nitro hydrogen bonds will form.
6. The aggregate patterns formed from intermolecular hydrogen bonds between substituents in meta and para positions will be acentric.
7. The amino–nitro interaction is usually a three-center hydrogen bond, $R_1^2(4)$.

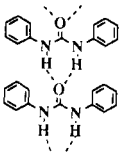


8. Ortho-substituted primary nitroanilines usually form two-center intermolecular hydrogen bonds, rather than three-center, with graph set $R_2^2(6)$.

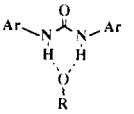


Diarylureas[28b]

4. The NH hydrogens prefer to adopt an anti relationship to the carbonyl group and to form three-center bonds to urea carbonyl groups, $C(4)[R_2^1(6)]$.



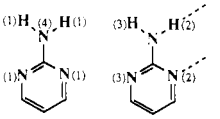
5. Cocrystals form in $R_2^1(6)$ patterns when there are strong meta-substituted electron-withdrawing substituents, like NO_2 groups, on the aryl rings, and when the guest molecules have acceptor groups that are stronger than the internally hydrogen bonded urea carbonyl oxygen.



6. When cocrystals form, the NH protons form three-center bonds to acceptor groups, $R_2^1(6)$.
7. Nitro groups of *m*-nitro-substituted diarylureas are not usually used as hydrogen-bond acceptors for urea NH hydrogens in the presence of guest molecules with good hydrogen-bond acceptors.

Carboxylic Acid Cocrystals with 2-Aminopyrimidine[27]

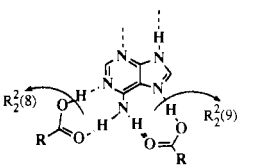
4. Both NH protons and both ring nitrogens are used in hydrogen bonds.
5. 2-Aminopyrimidine (2AP) prefers to form hydrogen bonds to acids rather than to itself.
6. 2AP forms cyclic $R_2^2(8)$ patterns with acids and with itself.
7. The two NH protons need not form hydrogen bonds to identical groups. Likewise for the ring nitrogens.
8. A ranking of proton-accepting ability consistent with these structures is N1 > N3 > acid carbonyl > N2, or N4, where the numbered pyrimidine atoms are those indicated in the monomeric and hydrogen-bonded structures.



9. Proton donor ranking is acid OH > NH(1) > NH(3) > NH(2).

Nucleotide Base (NB) Cocrystals[38]

4. Adenine (A) and cytosine (C) form cocrystals with many acidic organic compounds, but thymine and uracil do not.
5. Cocrystals of A or of C with carboxylic acids give patterns with $R_2^2(8)$ or $R_2^2(9)$ graph sets.



6. Neutral *N*-acylamino acids complex with A or C in $R_2^2(8)$ patterns.

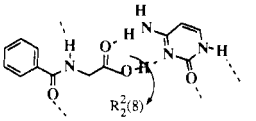
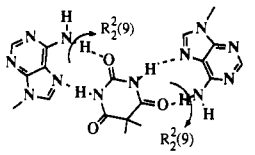
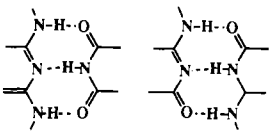


Table I (Continued)

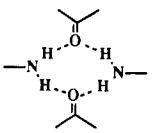
7. Cyclic alternant -C(O)-NH- groups complex with A and C with preferred graph set $R_2^2(9)R_2^2(9)$.



8. Three-centered hydrogen-bond contacts are frequently found with $N_1 = R_2^2(8)$ and $N_2 = R_2^2(12)$.

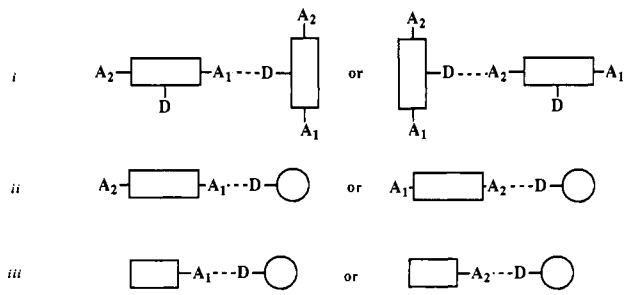


9. A common motif for NB complexes involves two primary amines and two carbonyl groups in an $R_4^2(8)$ or $C_2^1(4)$ pattern.



Donohue,[25] "good proton donors" and "good proton acceptors" have to be defined. We have found, not unexpectedly, that proton donors such as those in carboxylic acids, amides, ureas, anilines, imides, and phenols are nearly always used in hydrogen bonding in a crystal structure. Less acidic protons such as those in acetylenes, aldehydes, or activated aromatic and aliphatic compounds may be used in hydrogen bonding when there are extra proton acceptors available after all the more acidic protons have found an acceptor. Proton acceptors are typically acid and amide carbonyl groups, sulfoxides, phosphoryls, nitroxides, and amine nitrogens.

Solid-state acceptor and donor properties of molecules can be ranked by using the schemes shown below. Each scheme presents a different experimental technique for testing the competition between formation of $\mathbf{A_1}\cdots\mathbf{D}$ vs $\mathbf{A_2}\cdots\mathbf{D}$. In the first case, the two acceptor groups and the donor are on the same molecule. Two different crystal structures result from the two different hydrogen-bond choices. In the second case, the donor is on a guest molecule and the two possible hydrogen-bond contacts give rise to two different cocrystal structures. In the third case, the donor guest molecule selectively hydrogen bonds to one of the two acceptor molecules in a mixed solution. The chemical composition of cocrystals formed in scheme iii will actually be different. By changing the **A** groups in these experiments, relative acceptor strengths can be ranked.



The second rule derives from competition studies where we find that an intramolecular hydrogen bond is more difficult to break than a comparable intermolecular hydrogen bond formed between similar donors and acceptors. We have also observed that six-membered-ring intramolecular hydrogen bonds usually occur in crystal structures in preference to intermolecular hydrogen bonds when donor and acceptor groups are positioned to allow such an interaction. There are exceptions to this rule, but they are rare.[26] Intramolecular hydrogen bonds in five- and seven-membered rings are also common, but we do not have data at this point about competition between these bonds and intermolecular hydrogen bonds.

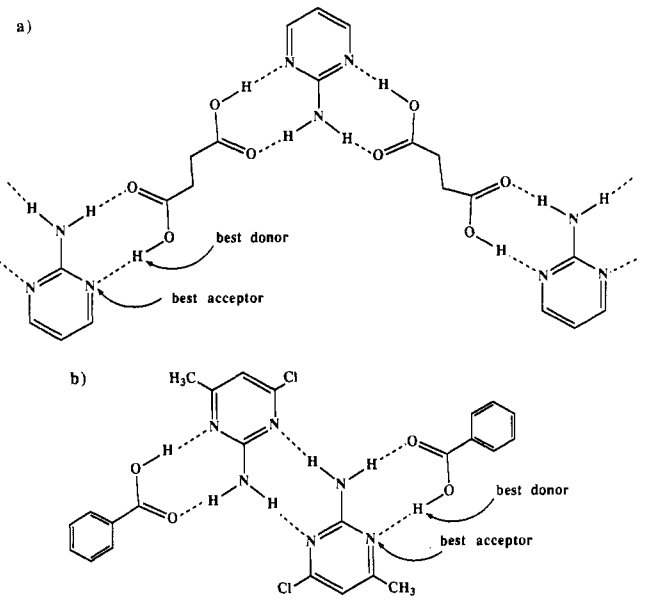


Figure 1. Hydrogen-bond patterns of 2-aminopyrimidine (2AP) and carboxylic acids. (a) This pattern shows the use of all good proton donors and acceptors, selectivity of the best proton donor (carboxylic acid proton) for the best proton acceptor (pyrimidine ring nitrogen), and formation of $\mathbf{R_2^2(8)}$ ring motifs.[27] (b) Cocrystals (1:1) of the 4-chloro-6-methyl derivative of 2AP and benzoic acid have been made, but the crystal structure is unknown.[39] A proposed hydrogen-bond pattern, consistent with the known stoichiometry, and based on the rules from Table I, is shown.

The third rule is illustrated by 2-aminopyrimidine–carboxylic acid cocrystals. In these structures (Figure 1), the best donors (acid OH) are paired with the best acceptors (ring N).[27] When substituents are present

(25) Donohue, J. *J. Phys. Chem.* **1952**, *56*, 502–510.

(26) Etter, M. C.; Urbańczyk-Lipkowska, Z.; Fish, P. A.; Panunto, T. W.; Baures, P. W. *J. Crystallogr. Spectrosc. Res.* **1988**, *18*, 311–325.

(27) Etter, M. C.; Adsmond, D. A. *J. Chem. Soc., Chem. Commun.*, in press.

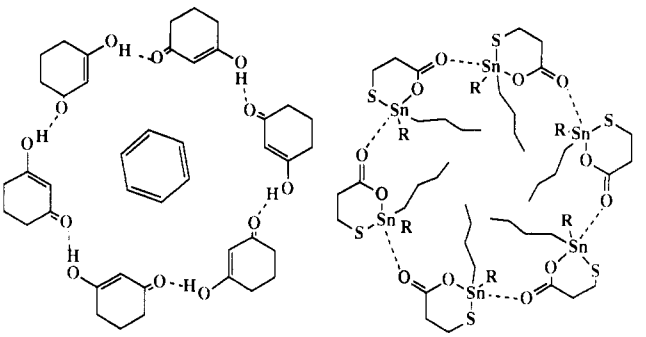
on the pyrimidine ring, the relative accepting strengths of the two ring nitrogens change in a predictable way, changing the aggregate hydrogen-bond patterns but still reflecting association of the best acceptor and the best donor. Examples of this rule have also been demonstrated for carboxylic acid hetero dimers, urea cocrystals, and triphenylphosphine oxide complexes.[28]

Specific hydrogen-bond rules, in addition to the three general rules given above, can be derived for other functional-group classes. Some examples are given in Table I. Others are available in the literature.[17d,29] Graph-set frequency for the functional group of interest is monitored from sets of crystal structures to see if preferences are apparent, while perturbations of hydrogen-bond patterns by competing functional groups give information about selectivity. These observations constitute the specific hydrogen-bonding rules. The rules should evolve as new structures become available, but we have found that the most important relationships are usually seen from the first 10 or so structures in a series.

Other Weak Intermolecular Interactions and Competition with Hydrogen Bonds

Hydrogen bonds are not the only type of intermolecular interactions that are useful for directing molecular self-assembly. Lithium–oxygen chelation is a well-known example of such an interaction.[30] Other intermolecular interactions that occur in organic systems may also be candidates for "intermolecular syntheses" when they are sufficiently strong and sufficiently directional to give rise to recognizable and reproducible molecular assemblies.

Two directed intermolecular interactions that are about as strong as hydrogen bonds are S···O[31] and Sn···O bonds. A striking example of the use of the latter interaction as a hydrogen-bond surrogate involves the two structures shown below.[32,33]



Ionic interactions are considerably stronger than hydrogen bonds, and they usually dominate a packing pattern;[34] nevertheless, hydrogen bonds may still be useful as organizing forces even in the presence of ionic interactions.[35] Very weak electrostatic attractions sometimes give predictable association patterns, as shown by Burgi and Dunitz.[36] Van der Waals interactions, charge-transfer interactions, and halogen–halogen interactions have also been used.[37]

Graph-set analysis and the protocol presented here for developing hydrogen-bond rules provide a framework for studying hydrogen-bond-directed aggregation, with the long-range goal of learning how hydrogen bonds compete with and cooperate with other forces that determine aggregate structures. Hopefully in the future these results can be coupled with theoretical studies in order to gain a better understanding of why particular patterns are preferred.

I am particularly grateful to Prof. Joel Bernstein, Ben Gurion University of the Negev, for his enthusiasm and helpful comments about the work and the manuscript. I also thank the students in my research group for the many hours spent developing and testing graph sets. Financial support from the NIH (GM 42148-01), NSF (CHE8600383-02), and Office of Naval Research is gratefully acknowledged.

(28) (a) Etter, M. C.; Baures, P. W. *J. Am. Chem. Soc.* **1988**, *110*, 639–640. (b) Etter, M. C.; Panunto, T. W. *J. Am. Chem. Soc.* **1988**, *110*, 5896–5897. (c) Etter, M. C.; Frankenbach, G. M. *Chem. Mater.* **1989**, *1*, 10–11.

(29) (a) Carbohydrates: Jeffrey, G. A.; Takagi, S. *Acc. Chem. Res.* **1978**, *11*, 264–270. (b) Amino acids: Jeffrey, G. A. In *Structure and Biological Activity*; Griffin, J. F., Duax, W. L., Eds.; Elsevier Science Publ.: New York, 1982; pp 135–150.

(30) Setzer, W. N.; Schleyer, P. v. R. *Adv. Organomet. Chem.* **1985**, *24*, 353–451.

(31) (a) Kâlmân, A.; Parkanyi, L. *Acta Crystallogr., Sect. B* **1980**, *36*, 2372–2378. (b) Rosenfield, R. E., Jr.; Parthasarathy, R. J.; Dunitz, J. D. *J. Am. Chem. Soc.* **1977**, *99*, 4860–4862.

(32) Etter, M. C.; Urbañczyk-Lipkowska, Z.; Jahn, D. A.; Frye, J. S. *J. Am. Chem. Soc.* **1986**, *108*, 5871–5876.

(33) Lockhart, T. *Organometallics* **1988**, *7*, 1438–1443.

(34) Fagan, P. J.; Ward, M. D.; Calabrese, J. C. *J. Am. Chem. Soc.* **1989**, *111*, 1698–1719.

(35) Day, V. W.; Klemperer, W. G.; Maltbie, D. J. *J. Am. Chem. Soc.* **1987**, *109*, 2991–3002.

(36) Burgi, H. B.; Dunitz, J. D. *Acc. Chem. Res.* **1983**, *16*, 153–161.

(37) (a) Charge transfer: Torrance, J. B. *Acc. Chem. Res.* **1979**, *12*, 79–86. (b) Halogen–halogen: Desiraju, G.; Sarma, J. A. R. P. *Acc. Chem. Res.* **1986**, *19*, 222. (c) Van der Waals: Salemme, F. R.; Genieser, L.; Finzel, B. C.; Helmer, R. M.; Wendoloski, J. J. *J. Cryst. Growth* **1988**, *90*, 273–282. (d) Griffin, J. F.; Smith, G. D. *Structure and Expression*, Vol. I from *From Proteins to Ribosomes*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: New York, 1988; pp 129–147.

(38) Structural data taken from the following: Tamura, C.; Sato, S.; Hata, T. *Bull. Chem. Soc. Jpn.* **1973**, *46*, 2388–2394 and references therein.

(39) Etter, M. C.; Adsmond, D. A., unpublished results.