Crystal Structure and Prediction

Tejender S. Thakur,[1] Ritesh Dubey,[2] and Gautam R. Desiraju[2]

[1]Molecular and Structural Biology Division, CSIR-Central Drug Research Institute, Lucknow 226 031, India

[2]Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560 012, India; email: desiraju@sscu.iisc.ernet.in

Annu. Rev. Phys. Chem. 2015. 66:21–42

First published online as a Review in Advance on November 19, 2014

The *Annual Review of Physical Chemistry* is online at physchem.annualreviews.org

This article's doi: 10.1146/annurev-physchem-040214-121452

Keywords

atom-atom potentials, crystal engineering, landscape, crystallization, polymorphism, supramolecular synthon, protein folding

Abstract

The notion of structure is central to the subject of chemistry. This review traces the development of the idea of crystal structure since the time when a crystal structure could be determined from a three-dimensional diffraction pattern and assesses the feasibility of computationally predicting an unknown crystal structure of a given molecule. Crystal structure prediction is of considerable fundamental and applied importance, and its successful execution is by no means a solved problem. The ease of crystal structure determination today has resulted in the availability of large numbers of crystal structures of higher-energy polymorphs and pseudopolymorphs. These structural libraries lead to the concept of a crystal structure landscape. A crystal structure of a compound may accordingly be taken as a data point in such a landscape.

"Well! I've often seen a cat without a grin," thought Alice; "but a grin without a cat! It's the most curious thing I ever saw in my life!"

Lewis Carroll, *Alice's Adventures in Wonderland*

Crystal structure: the arrangement of atoms, molecules, or ions that constitutes the internal structure of a crystal

Crystal engineering: the design of solids with desired properties using knowledge of intermolecular interactions in a crystal packing context

Pseudopolymorph: solvated form of a compound that has a different crystal structure and/or differs in the nature of the included solvent

CSP: crystal structure prediction

Trial-and-error method: real space method of solving a crystal structure by employing auxiliary information and packing arguments

1. INTRODUCTION: STRUCTURE AND CRYSTAL STRUCTURE

Structure is an immanent concept in scientific inquiry and deals with the relationships between objects and entities. In chemistry, these objects are atoms, ions, and molecules, and structure may be defined in terms of geometry, topology, or dimensionality. All chemistry may indeed be referred to as structural chemistry, or at least all chemistry may be said to lead from structural chemistry in that without the notion of structure, it would be quite difficult to conceive of synthesis and of course of dynamics, which is concerned with transformations between structures.

The molecular paradigm is of overwhelming importance in chemistry, and it is interesting to note that a molecule can be defined equally effectively in terms of geometry or energy. Kitaigorodskii (1) provided a geometrical definition of a molecule within a crystal as a group of atoms such that for every atom in the group, at least one interatomic distance within this group is "significantly smaller" than the shortest interatomic distance to an atom in another group. Energetic definitions are based on chemical considerations, but dividing the energy of a molecule in a crystal into bonded intramolecular interactions and nonbonded intermolecular interactions is, strictly speaking, not justified. In any event, the energetic criteria that may distinguish between intramolecular and intermolecular aspects of a molecule are examined and assessed with crystallography, spectroscopy, and computation to provide a modern framework for structural chemistry.

A combined consideration of intramolecular and intermolecular geometry leads to modern definitions of the term crystal structure. Crystallographically, this term is easy enough to define as the symmetry relationships among the (molecular) modular units that occupy the unit cell, which is defined in terms of a crystal system. Chemically speaking, a crystal structure is a wonderful template within which the relationships among molecules may be understood, predicted, modulated, and modified. The subject of crystal engineering attempts to do this so that new crystals with desired properties can be realized (2). A crystal structure is also a springboard to other crystal structures of the same or another molecule. A group of crystal structures of the same molecule (polymorphs) or of related systems (pseudopolymorphs) may also be said to constitute a larger platform, which may be termed the structural landscape of a molecule.

This review attempts to trace the development of the notion of crystal structure since the time when a crystal structure could be determined from a three-dimensional diffraction pattern and to assess the feasibility of predicting an unknown crystal structure of a given molecular solid. Our analysis therefore deals with two distinct topics, namely crystal structure and prediction, and the degree to which they are interrelated. In recent years, the term crystal structure prediction (CSP) has gained considerable popularity (Section 4) (3–5). CSP is the computational derivation of the unit cell and space group of a typically unknown crystal structure and of the positional parameters of all the atoms in it. CSP is of considerable fundamental and applied importance, and its successful execution is by no means a closed issue.

2. HISTORICAL PERSPECTIVE

The origins of the concepts of crystal structure and of CSP go back to the early twentieth century when W.H. Bragg developed a knowledge-based chemical route to identify probable molecular arrangements in crystals. This approach constitutes the beginnings of the so-called trial-and-error

RAMACHANDRAN PLOT

Ramachandran's work on the development of an analytical method for obtaining the allowed and disallowed configurations of polypeptide chains was also based on similar ideas of the exclusion of certain spatial regions in the close approach of two nonbonded atoms. Essentially, his approach resembles Kitaigorodskii's structure-seeking apparatus in the sense that it is based on hard-sphere potentials. This work helped in obtaining extremely useful information on the allowed and disallowed regions of peptide chain conformations in terms of φ-ψ angles, popularly known as the Ramachandran plot. This has become an indispensable tool in modern macromolecular crystallography.

methods for crystal structure determination (Section 3). Bragg (6) applied this method to crystals of condensed polycyclic hydrocarbons as well as hydrogen-bonded molecules. In the former, geometrical information could be obtained; for example, the dimensions of the benzene ring were estimated by assuming that "certain units of structure, like the benzene or naphthalene ring, having definite size and form, might be preserved with little or no alteration in passing from one crystalline derivative to another" (7, p. 161). By the late 1920s, Pauling (8) understood the importance of SiO_4 tetrahedra in the structures of silicates and was able to consider putative structures that contained these units. In the late 1930s, Robertson determined the crystal structures of 1,4-benzoquinone (9) and the resorcinol polymorphs using packing considerations (10, 11). Full crystal structures of molecular solids, rather than just probable molecular arrangements, were beginning to be reported (12).

With the advent of the Patterson (or heavy atom) method for phase determination, the trial-and-error method gradually fell into disuse but, even so, did not vanish entirely (12). Kitaigorodskii (1) made a notable contribution when he obtained the dimensions of a molecule from gas phase experiments (hydrogen atom positions could not be determined accurately with X-ray analysis) and, using this information, calculated fairly accurate nearest-neighbor nonbonded distances in crystals. With intermolecular radii so derived, he proposed a volume-based model for the crystal structure, which he cross-checked with a manually operated structure-seeking apparatus. Although Kitaigorodskii's approach was never competitive with respect to crystal structure determination (direct methods were becoming very important by the late 1950s), the thought process that underlies it was profitably harnessed toward the analysis of crystal packing and provides the foundation for the subject of this review (see the sidebar Ramachandran Plot).

3. THE CRYSTAL STRUCTURE

The main focus of Kitaigorodskii's method was obtaining the crystal structure of a compound that satisfies the condition of close packing of molecules in a given unit cell. His model was successful in obtaining accurate information about molecular packing in many simple compounds. In some cases, however, the observed crystal structure showed noticeable deviations from the ideal close-packed structure. In iodoform, for example, close packing predicts that each iodine atom is in contact with four iodine atoms from three neighboring molecules rather than with the two iodine atoms from two adjacent molecules observed in the real structure (**Figure 1**) (13). The differences between the real crystal structure and the ideal crystal packing underline the role played by directed intermolecular interactions in understanding molecular association in solids (2). A thorough study

Direct methods: powerful methods that solve the phase problem in crystallography from the magnitudes of the structure factors alone

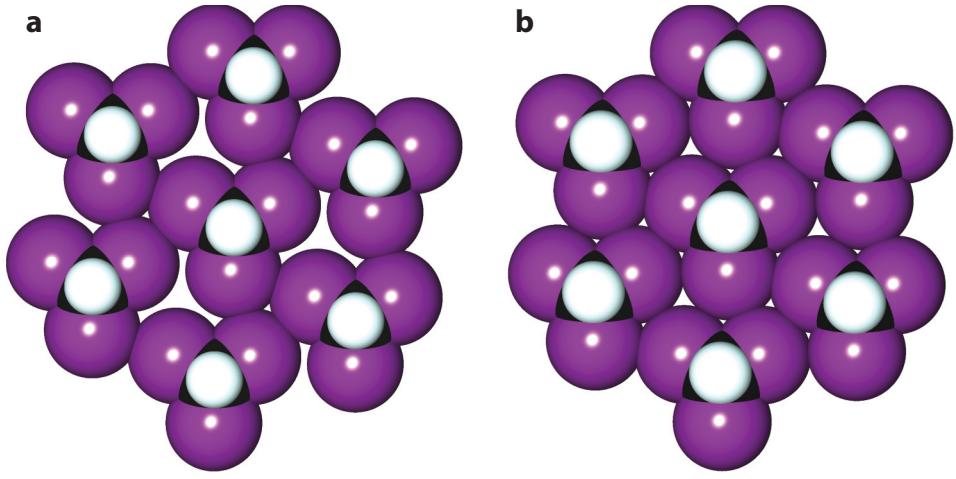


Figure 1

Molecular packing in iodoform: (*a*) experimental structure and (*b*) hypothetical close-packed structure. Note that the experimental structure contains some loosely packed regions.

of atom-atom potentials was the need of the hour to understand and to accurately reproduce the effects of these intermolecular interactions. It was believed at that time (the 1960s and 1970s) that the correct atom-atom potential parameters would reproduce the observed crystal structure. Even today, much research in CSP problems devolves around the modeling of accurate potentials (14–18).

Kitaigorodskii made seminal contributions toward the development of the atom-atom potential method for molecular crystals. According to this model, the potential energy (u_{ij}) of a nonbonded pair of atoms i and j separated by a distance r_{ij} is given by Equation 1. For general applicability, the constants A_{ij}, B_{ij}, and C_{ij} in the expression were assumed to depend only on the types of atoms involved in the interaction, irrespective of their valency and molecular and chemical environments. The derived potentials were also termed isotropic pair potentials as they assume that the interactions between pairs of atoms are radially symmetric. Additionally, the total potential energy (U) for a system of interacting molecules was assumed to be pair-wise additive such that

$$U = \frac{1}{2}\sum u_{ij} = \frac{1}{2}\sum\left[-A_{ij}r_{ij}^{-6} + B_{ij}\exp(-C_{ij}r_{ij})\right]. \quad (1)$$

Kitaigorodskii (19) used intermolecular distances obtained from the study of various crystal structures to propose a universal atom-atom interaction curve for various nonbonded atom-atom interactions. He showed that the experimental crystal structure can be reproduced computationally by performing lattice energy minimization with the help of these nonbonded potential parameters (20). Williams (21) extended Kitaigorodskii's ideas to include information obtained from known elastic constants and sublimation energies to calibrate these potentials to improve the accuracy of atom-atom potential parameters. Further improvements in the atom-atom potential models were proposed by many researchers to include electrostatic, polarization, and interaction anisotropy in the calculation (22–24). Although these methods became obsolete after direct methods became important for crystal structure solution, they did lead to progress in the understanding of molecular packing in crystals and thus toward the computational prediction of crystal structures.

Crystal structure solution: any method that solves the phase problem for diffracted intensities so that electron density information can be obtained

Polymorphism was still looked upon as a curiosity, a less prevalent phenomenon, although well-documented cases had started to appear by then (25, 26). The role of nucleation kinetics was also totally neglected in these studies. Differences between predicted and experimental crystal structures were ascribed to the empirical nature of the atom-atom potential method itself. It was believed that correct potentials should or would eventually lead to the experimentally observed crystal structure.

Polymorphism: phenomenon in which the same chemical compound exhibits different crystal forms

Halogen-containing compounds posed additional challenges to the atom-atom potential parameterization and attracted special attention. Unlike other simple diatomics (N_2, NO, CO) that crystallize with a close-packed structure in the cubic space group Pa3, the solid halogens Cl_2, Br_2, and I_2 take a lower-symmetry, layered orthorhombic (space group Cmca) structure. Atom-atom-based calculations with 6–12 potentials, however, compute the cubic Pa3 as the most stable structure (2). This discrepancy questions the applicability of the atom-atom potential method for halogen-containing compounds. New computational approaches were proposed to tackle this problem. Price & Stone (23) proposed the use of distributed multipole-based anisotropic potential terms for halogens and other aspherical molecules. Their method was the only successful method available for predicting the crystal structures of halogenated compounds before dispersion corrected density functional theory (DFT-D) methods came on to the scene.

However, there are several studies that indicate that DFT-D methods are not always reliable for computing the relative energies of polymorphs as they fail to provide a complete description of dispersion interactions and also show its inability to systematically increase accuracy (27, 28). In spite of these advances, the use of isotropic potentials was still preferred because of the expensive computations involved in the minimization of thousands of structures in each space group during the initial crystal structure search.

It was recently suggested that the quality of the isotropic model can be further improved by making these potentials more system specific with the use of tailor-made force fields (29). With the tremendous advances in computing power in recent years, it is not surprising to see that CSP methodologies today take a brute-force approach.

The success of CSP requires accurate lattice energy computation approaches reliable enough to obtain the correct stability order for polymorphs sometimes differing by less than 1 kJ/mol. Wave-function-based electronic structure methods provide alternatives to DFT in providing a better description of the dispersion energy. Second-order Møller-Plesset perturbation theory (MP2) is a widely used method for obtaining correlation energy contributions. However, the high computational cost involved in these calculations makes them unsuitable for crystals. Recent developments in the fragment-based electronic structure methods have also made these methods computationally feasible for molecular crystals. Several fragment-based methods have been proposed by many researchers using these wave-function-based electronic structure methods for large periodic systems (30–34).

The fragment-based hybrid many-body interaction model developed by Beran and coworkers (30) provides a computationally affordable means of applying electronic structure wave-function methods to molecular crystals. It follows a combined quantum mechanical and molecular mechanics (QM/MM) approach that employs a highly accurate quantum mechanical treatment for molecules within the unit cell and a molecular mechanics approach for the computation of short-range pair-wise and long-range many-body dispersion (MBD) interactions. In a recent study, Chan and coworkers (32) demonstrated the successful application of fragment-based approaches for obtaining the lattice energy for the benzene crystal with sub-kJ/mol accuracy using the ab initio many-electron wave-function-based coupled cluster method. In another study, Hirata and coworkers (35) showed the utility of fragment-based ab initio electron-correlated methods in the computation of the phase diagram of solid carbon dioxide.

4. THE CCDC BLIND TESTS: A CONSORTIAL AND SYSTEMATIC EFFORT TO ADDRESS CRYSTAL STRUCTURE PREDICTION PROBLEMS IN ORGANIC COMPOUNDS

Phase problem: problem in which diffraction intensities contain information on the magnitudes of the structure factors but not on their phases

By the late 1960s, Kitaigorodskii (36) and Williams (37) had developed rudimentary atom-atom potential-based lattice energy–minimization programs (Section 3). However, their aim was (still) to provide a pool of trial structures as solutions to the phase problem. However, with the increasing reliability of direct methods for crystal structure solution of organic compounds, there was a change in emphasis in these energy-minimizing routines from crystal structure solution to crystal packing analysis and the design of new crystal structures—these were the early days of a new subject, crystal engineering (2). In 1980, Dauber & Hagler (38) set CSP as the next logical step emerging out of the current advances in X-ray crystal structure determination and computational techniques. The development of their crystal structure generation and lattice energy–minimization program became an important milestone toward the development of modern CSP methods. Leiserowitz & Hagler (39) proposed a new method for generating crystal structures of a compound through a consideration of energetic evaluation of possible hydrogen-bonded arrangements using atom-atom potentials. However, the quest for obtaining answers to the CSP problem remained a subject of limited scientific interest among researchers. In 1988, Maddox's (40) editorial in *Nature* helped draw the subject into focus and provoked considerable interest among the scientific community. In 1989, Desiraju's (2) book on crystal engineering highlighted this issue and pointed out that although organic crystal structures are predominantly governed by Kitaigorodskii's close-packing principle, the minor deviations from close packing, which result from chemical factors, are of the greatest importance because they lead to structures that can be systematically engineered. In the context of CSP, with the wisdom of hindsight, we can say that these deviations are the main reason for the difficulties encountered in the computational prediction of crystal structures.

The next decade saw a surge in various new CSP methodologies for organic compounds from researchers worldwide. Several novel approaches for crystal structure generation and for improving the accuracy of lattice energy computations were proposed during this period (41–55). However, many of these methodologies were restricted to small rigid molecules containing only C, H, N, and O atoms. Several novel approaches for crystal structure generation were proposed and validated, such as the grid search method, symmetry-guided buildup of molecular clusters, Monte Carlo simulated annealing, and genetic algorithm-based sampling of crystal structures. The Monte Carlo simulated annealing method became a popular choice for the fast and efficient generation of molecular packing in crystal structures and was preferred over systematic search methods; several CSP programs were developed based on this method. Some of these approaches are retained in contemporary CSP methodologies (14–18).

By the late 1990s, there was an urgent need to have a consortial and more organized approach to tackle the CSP problem. The Cambridge Crystallographic Data Centre (CCDC) took the lead in organizing CSP blind tests. These tests helped to bring CSP into focus and provided a common platform for researchers to work together toward the development and improvement of computational methodologies. The first of these tests was held in 1999, and 11 active research groups were invited to participate (14). Participation by invitation continued up to, and including, the third test held in 2004 (14–16). The blind tests were opened to all interested groups in the fourth and fifth tests, and they were widely publicized (17, 18). For the targets, unpublished, fully determined, high-quality structures with no disorder were held in confidence until the completion of the test. The molecular diagrams and crystallization solvents for the selected targets were provided to the participants. Participants were asked to give their three best choices for the crystal structures (unit cell parameters, positional parameters), and up to six months were given for the exercise.

In the second test, the experimental X-ray powder patterns were provided after the deadline had passed, and participants were given one more week to provide their best choices with this additional information. The targets occurred in categories of molecules that addressed different levels of computational complexity (**Figure 2**). These categories ranged from simple rigid molecules containing the elements carbon, hydrogen, nitrogen, and oxygen only with the total number of nonhydrogen atoms less than 25 (targets I, IV, VII, VIII, XI, XII, and XVI), small, rigid molecules with less common elements or functional groups that present a challenge for modeling methods with the total number of nonhydrogen atoms less than 40 (targets II, V, IX, XIII, and XVII), flexible molecules with several rotatable bonds (targets III, VI, X, XIV, XVIII, and XX), to complex multicomponent systems such as salts/cocrystals (targets XV and XIX). Molecules exhibiting polymorphism that fall into any of these categories (target XXI) were also considered as targets.

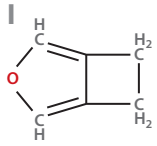
Disappearing polymorph: informal terminology for a crystal form that seemingly fails to appear after initial isolation

For 3,4-cyclobutylfuran, one of the CSP1999 targets (target I), two polymorphic forms exist. A metastable Pbca polymorph, which was discovered first, constituted the target structure. Subsequently, and after the blind test concluded, a more stable $P2_1/c$ polymorph was discovered (14). The metastable polymorph never appeared again, and it was termed a disappearing polymorph. There were four successful predictions for the Pbca form; however, none of the participants was able to predict the more stable $P2_1/c$ polymorph among their top three submissions. To complicate the issue further, very accurate hybrid DFT-based calculations, done in 2009, showed the Pbca structure as actually being more stable than the $P2_1/c$ polymorph by 0.5 kJ/mol (56). However, the authors pointed out that the method they employed did not account for zero-point vibrational energies and entropic contributions. The entire episode is thought provoking. Which is the correct answer? Which is incorrect? How different would the assessment of the CSP results have been if this molecule had been provided as a target after 2009? Would better predictions have been obtained if specific attention were paid to the weak C–H···O hydrogen bonds that are found in the $P2_1/c$ structure? Does the monoclinic polymorph appear more regularly because of this structural feature? C–H···O hydrogen bonds also feature in the experimental structure of target molecule X, as well as in a large number of related structures of other 1,2-dinitro-substituted aromatics (57). The overwhelming weight of evidence is in favor of these interactions as structure-defining elements in crystal structures (58). If one were to ignore these interactions, whether in CSP or in computer-aided drug design, it would lead to additional problems (59).

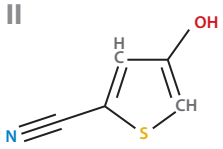
In CSP2001, participants searched for additional polymorphs for target VI (6-amino-2-phenylsulfonylimino-1,2-dihydropyridine) after the completion of the test because the given experimental polymorph of target VI (form I) was not predicted correctly. The second polymorph (form II) was obtained by Desiraju and coworkers (60) after an extensive polymorph screening; these workers claimed that the polymorph given for the test is a kinetic form. Because form I contains a two-point synthon (**Figure 3*b***), which facilitates one-dimensional growth, it was proposed that it is kinetically favored over form II, which contains a discrete four-point synthon (**Figure 3*a***). This four-point synthon occurred in some of the computational solutions in the CSP, although the exact structure of form II was not submitted as a solution. A third polymorph (form III) was later discovered using a polymer-induced heteronucleation strategy (61). Based on differential scanning calorimetry melting enthalpies and conversion of forms II and III to form I in slurrying experiments, it was concluded that form I is the most stable polymorph, followed by forms III and II. The observed stability order was also confirmed in CSP performed later with the hybrid DFT-D method (62). Is this a simple failure of the earlier CSP methodologies? Or are there more forms? Which is the most stable form? Does it matter? Does the occurrence of a crystal form depend on the stability of a structure or the ease of crystal growth?

The CSP2004 blind-test molecule 2,9-bis(iodo)anthrone, target IX, presented additional challenges. There were difficulties in modeling the interaction potentials accurately owing to the

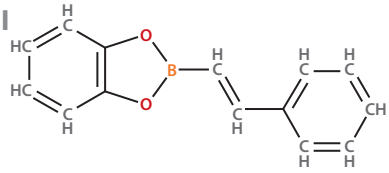
I



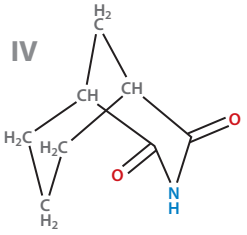
II



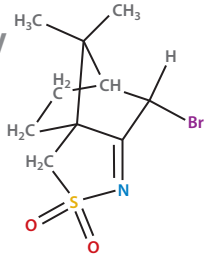
III



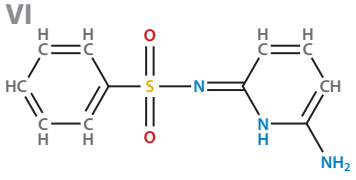
IV



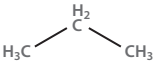
V



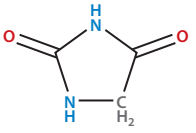
VI



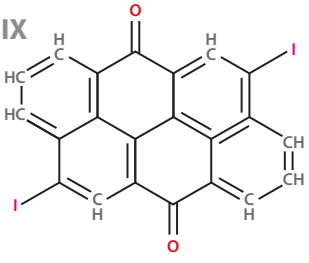
VII



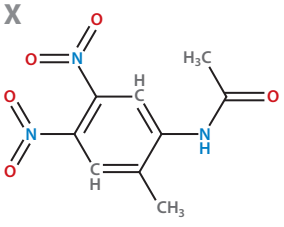
VIII



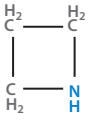
IX



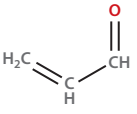
X



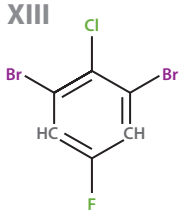
XI



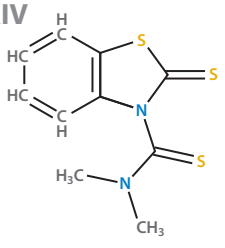
XII



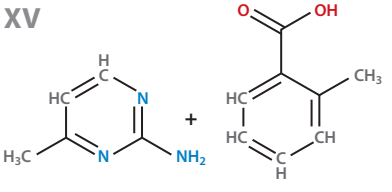
XIII



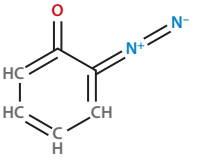
XIV



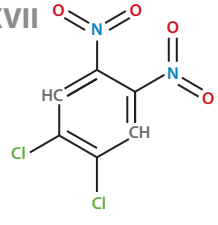
XV



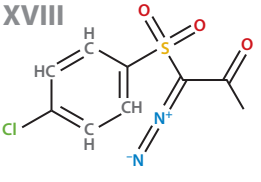
XVI



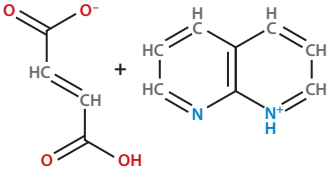
XVII



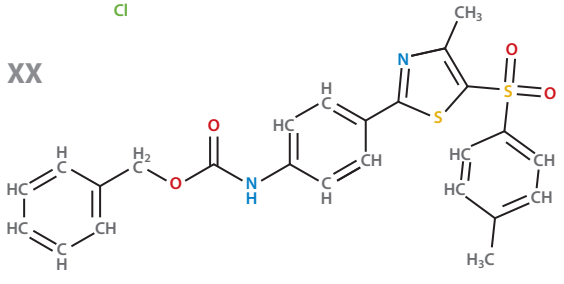
XVIII



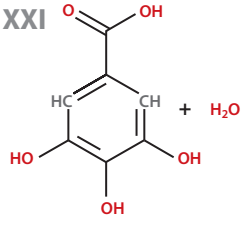
XIX



XX



XXI



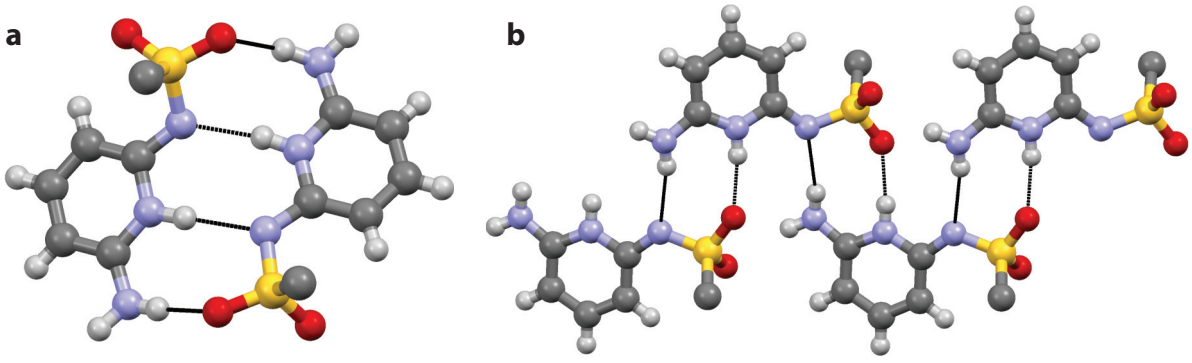


Figure 3

Supramolecular synthons in the polymorphs of target VI: (*a*) a discrete four-point synthon in form II and (*b*) a two-point synthon in forms I and III. Phenyl substituents on the sulfur atoms are omitted for clarity.

anisotropic nature of the iodine atom; that is, the electronic distribution is aspherical and results in short van der Waals radii along the C–X bond when compared to the perpendicular direction (a polar flattening effect). This anisotropy is so significant that all attempts at CSP using isotropic potentials were largely unsuccessful. The experimental structure was predicted correctly by only one participant as the first choice with the help of specifically developed, anisotropic atom-atom repulsion and electrostatic models that explicitly model the iodine anisotropy. However, there were eight other participants who found the experimental structure in their extended list as a high-energy structure. This example raises different questions. Could one use isotropic potentials and remain satisfied as long as the experimental structure occurred somewhere in the list of predicted structures? Or would the extra computational effort in custom fitting the structure to a tailor-made anisotropic force field be worthwhile? Once again, should CSP be cast in terms of a unique correct solution or in terms of many equivalent and comparable choices?

From the mixed successes and failures of the first three blind tests, it was realized that routine force field–based methods would not offer significant further gains. None of the approaches employed till then was indeed consistent in its success. The CSP2007 blind test saw significant progress with a new method. The group of Neumann, Leusen, and Kendrick was successfully able to predict crystal structures of the four target compounds under various categories as their first choice. The introduction of highly accurate, but computationally intensive, DFT-D methods into the lattice energy computations was the key to their success (29). The use of tailor-made force fields helped them to efficiently sample putative crystal structures. This made their exercise robust and also computationally viable. The successes of this group gave rise to much hope in their method. However, they failed to maintain their exemplary position in CSP2010, although many of their ranks were still respectably high. The very high time demands of this method still keep it out of reach for many researchers. Interestingly, one of their good predictions was for target XIX, 1,8-naphthyridinium fumarate, and it was chosen based on its similarity to the packing found in a

Figure 2

Molecular diagrams of the targets given in the crystal structure prediction blind tests: CSP1999 (I–III and VII), CSP2001 (IV–VI), CSP2004 (VIII–XI), CSP2007 (XII–XV), and CSP2010 (XVI–XXI).

related experimental structure (CSD Refcode RABYID). This solution appeared as their twentieth ranked DFT-D structure, and it corresponds to the experimental structure.

CSD Refcode: six-letter identification code given to an entry in the Cambridge Structural Database of organic and organometallic crystal structures

However, there is still room for further improvement in the accuracy of DFT-based approaches by including MBD effects. For instance, a DFT-D method that accounts only for the empirical two-body dispersion corrections fails to predict the correct stability of the two aspirin polymorphs (27). Furthermore, improvements in DFT approaches to include MBD terms have also been proposed (63, 64). Grimme et al. (63) proposed the DFT-D3 dispersion correction energy term to include three-body effects by fitting interaction energy data computed at the CCSD(T)/CBS level of theory. In a recent study, Reilly & Tkatchenko (65) showed that the correct stability order for an aspirin polymorph can be obtained from DFT methods by including higher-order MBD terms in the calculations. The free-energy difference they obtained including the MBD and zero-point vibrational energy corrections shows form I to be more stable than form II in accordance with experiment. They also noticed significant changes in the phonon modes in the low-frequency region upon the inclusion of MBD terms, which emphasizes the importance of these neglected terms, not only in obtaining accurate lattice energies but also in calculating the phonon density of states and other crystal properties.

In CSP2011, target XX, benzyl-(4-(4-methyl-5-(p-tolylsulfonyl)-1,3-thiazol-2-yl)phenyl) carbamate, which contains eight rotatable bonds, was given under the newly added category of highly flexible molecules. This target poses difficulties in terms of completeness of search during the structural sampling. Out of the 15 participants, only 10 even attempted predictions of this molecule, and of these, only three were successful in sampling the experimental structure; two impressively predicted it as the first choice in their submission. The Neumann-Leusen-Kendrick group found the experimental crystal structure ranked seventh in their extended list. Target XXI, gallic acid monohydrate, was found to be the most challenging among all blind-test targets with an unknown positioning of the water molecule. There are two known polymorphs, forms I and II (CSD Refcodes KONTIQ and KONTIQ01). The crystal structure of a new polymorph (form III) was given for CSP. After the test, it was found that there is an alternative positioning of hydrogen-bonded hydrogen atoms for the same positioning of nonhydrogen atoms in the target structure. These alternatives have different energies and are actually distinct structures. There were no successful predictions of the experimental structure of form III in the top three submissions of any participant group. However, the structure with the alternative packing was found among the top three choices of two groups. Neither form I nor form II was found among the top three least energy structures of any group (18). The Neumann-Leusen-Kendrick group obtained forms I, II, and III in the 27th, 49th, and 81st rank in their extended list, respectively. They attributed the poor performance of their method to kinetic factors, as inferred from the large differences in the relative stability (1.4, 2.4, and 2.6 kcal/mol above the global minimum, respectively) observed for the three polymorphs in the post-blind-test analysis. Other disordered forms of gallic acid monohydrate have subsequently been reported (66). The importance of kinetic and entropic factors during crystallization may be enunciated repeatedly in all these examples. In general, the occurrence of polymorphism should always be considered as widely prevalent, and hence, the existence of just one thermodynamically stable crystal structure cannot be used as a guide to perform CSP. The criterion of assessing the results on the basis of finding the experimental structure among the top three submissions is also not really justified.

With these equivocal successes in the fifth blind test, the CCDC initiated discussion on the future of the CSP blind tests. Comments and suggestions were sought from all participants in the fifth blind test through a web-based survey on these issues. A hit in the top 100 rather than the top three submissions as a means of assessing the success of the method was proposed, keeping kinetic factors in mind. It was realized that restricting the test to the so-called three best structures is also

not fully justified because currently there is no a priori way of predicting that a thermodynamically stable or a kinetically driven metastable polymorphic form will be obtained in the experiment, even if experimental conditions are provided with the chemical diagrams. Participants also supported an exhaustive polymorph screening of a target molecule to validate the efficiency of various prediction methods in locating various possible minima. A sixth blind test will commence in September 2014, keeping with the above suggestions from the participants. This test will be of one-year duration.

Supramolecular synthon: structural unit within a crystal that can be assembled by known or conceivable synthetic operations involving intermolecular interactions

Following a completely deterministic approach and neglecting the role of kinetics, researchers reduced the CSP problem to the exhaustive sampling of probable crystal structures. Much effort was made toward achieving completeness of search and obtaining accurate lattice energies; these issues remained the main challenge of the first four tests. Several approaches had been tested thoroughly during the blind tests, but no single method emerged as a clear winner. This may be attributed to the complete neglect of kinetic factors in CSP. Desiraju & Sarma (67) presented an alternative view that the CSP problem may be better understood by identifying supramolecular precursors (synthons) from retrosynthetic analysis of known crystal structures. Target IV was attempted and knowledge-based strategies were used to rank upward and select structures that contain extended N–H···O chains rather than closed N–H···O dimers, which were identified as being kinetically disfavored. The kinetically favored prediction was close to the experimental structure, and other choices based on lowest-energy dimer structures were never found. The synthon approach to CSP mimics the molecular aggregation process, which is the assembling of molecules to a crystal via supramolecular synthon-directed recognition. A reranking of crystal structures was proposed by giving a preference to commonly occurring synthons in crystal structures of similar compounds (68). The incorporation of synthon information in the reranking of generated structures provides an indirect means of incorporating kinetic factors because these synthons represent the deviations between ideally close-packed structures and real structures (Section 3). Using the best known method for obtaining accurate lattice energies, one would be able to sample a kinetic crystal as a stationary point on the potential energy hypersurface within the top few hundred structures. The direct inclusion of kinetic effects into the CSP methodology is still a challenge and will be the main focus of next-generation CSP methods.

A perusal of the literature over the years indicates that researchers in CSP were certainly aware of the limitations of energy-minimization procedures based on simple two-body isotropic potentials. The difficulties encountered in CSP arise from the basic fact that the interactions in organic crystals are weak, numerous, and anisotropic. Many procedures employed in CSP are necessarily approximate because of the need to decrease the computer time required. The question is whether they are so simple that they tend toward the simplistic. One could argue that two-body potentials provide far too coarse a treatment, and therefore, one should aim for a multibody analysis (69). One could argue that atom-atom interactions are too simplistic, and therefore, one should analyze a structure in terms of molecule-molecule interactions (57). These types of arguments, in the end, do not lead to a resolution of the problem because the deviations between ideal structures and real structures have chemical origins. Because the modeling of some chemical interactions is still a matter of conjecture and debate, other approaches, such as synthon theory, have been advocated (70). Invoking the supramolecular synthon is a means of circumventing what is a very difficult computational problem—one assumes that certain modular units are favored in crystal packing, and one selects only the structures that contain these units (67). It has been said that qualitative analysis is just bad quantitative analysis. We believe it is an alternative approach. When the nature of the system is more complex than what can be revealed by the presently available quantitative treatments, the researcher may well be tempted to take more qualitative routes. Chemistry is finally a judicious blend of the qualitative and the quantitative (71).

5. FROM "THE CRYSTAL STRUCTURE" TO "A CRYSTAL STRUCTURE"

Z′: number of molecules in a crystal asymmetric unit

The predominant emphasis in structural crystallography for the first 70 or 80 years following the discovery of X-ray diffraction was the determination of what may generally be referred to as the crystal structure of a molecule. Tacit in this phraseology is that any given molecule had just one crystal structure associated with it, which was called the crystal structure. Perhaps this emphasis arose implicitly because crystallography was originally used to establish the constitution, structure, and stereochemistry of a molecule rather than to study crystal packing. Perhaps crystal structures were just too difficult to determine for anyone to want to explore crystal polymorphism in any detail (72, 73). In addition to polymorphism, there were other more indirect notions in the literature about the existence of virtual crystal forms that were beyond unique structural characterization from diffraction patterns, but the characterization techniques were in their infancy, and these structures were discussed as part of a transient molecular behavior of the compound in the crystalline state (74). All in all, a one-to-one correspondence between molecular structure and crystal structure permeated into the consciousness of chemists, so much so that even up to and during the early days of CSP, in which the practitioners were all seasoned specialists, the notion of the one "correct" structure, as opposed to the numerous "incorrect" solutions that were submitted, continued to persist.

The idea of multiple forms was partly considered when Kitaigorodskii's (1) close-packing model was unable to rationalize some real crystal structures, but effort was still focused on a single possibility for the crystal structure. The growing popularity of computational techniques provoked an interest in high-energy minima in the landscape and prompted the community to look into these possibilities more seriously with computational as well as diffraction-based techniques, and also into the existence of multiple crystal forms. Polymorphs are obtained because of the dichotomy between kinetics and thermodynamics during the crystallization process, and the dependence of crystal growth on experimental conditions, which subtly alter the free energy differences among polymorphs (75). These energy differences may be attained by various characteristic phenomena, such as the existence of multiple symmetry-independent molecules (*Z′*) and pseudopolymorphism (solvation) (see the sidebar Crystallization and Protein Folding). Beyond the close-packing principle, and to understand the origin of polymorphism as well as to provide a rationalization of the role of kinetics in crystallization, one can use the supramolecular synthon model as a template that may provide a reasonable understanding of the crystallization phenomenon. The complexity of

CRYSTALLIZATION AND PROTEIN FOLDING

The dynamic and progressive behavior of molecules during crystallization and protein folding shows marked similarities. These complex phenomena decode the emergence of multiple metastable forms before the final structures are attained. These intermediate kinetically driven species may be high-energy polymorphs and pseudopolymorphs of the compound in question or semicompact random globules for proteins. Understanding the role of these species in their respective processes is of critical importance in elucidating mechanisms.

An interesting structural feature of organic molecules that needs to be considered in crystallization is their carbon content. Somewhat similar to Levinthal's paradox in protein folding, the frequency of occurrence of polymorphism does not vary significantly until C_{80}, suggesting that the presence of multiple hydrophobic recognition units does not change basic crystallization behavior. However, polymorphism in small molecules is more common than are multiple folding arrangements in macromolecules. Maybe this means that protein folding is more discriminating and efficient than crystallization of organic molecules.

this phenomenon is trivialized in part by the simplistic McCrone (76, p. 725) dictum that "every compound has different polymorphic forms, and that, in general, the number of forms known for a given compound is proportional to the time and money spent in research on that compound." This dictum gives an (actually erroneous) feeling of pervasive behavior, or even the purportedly mysterious nature of disappearing polymorphs, a discussion of which (77) was cast against an elusive and dramatic backdrop that tended to obfuscate the challenges that had to be overcome to understand this so-called phenomenon. One must not confuse the metaphor with the real thing. No polymorph can ever disappear. Once it has been isolated experimentally, any failure to obtain it again must arise from the fact that the same experimental conditions were not employed in subsequent efforts. The problem is that some of these conditions are inherent to a particular experiment and need not be included specifically by the experimentalist. Therefore, they are hard to identify and reproduce.

In 1999, Sarma & Desiraju (78) discussed the propensity of polymorphism in organic compounds using statistical studies and highlighted some little-known aspects. The important message is that polymorphism is normal and real, even if incompletely understood. The conformational flexibility and presence of multiple hydrogen-bonding functionalities are important features that may play an important role in the existence of polymorphism in organic compounds. Synthon polymorphism was identified as such and has been subsequently observed in many well-documented cases (79, 80). Simple compounds such as phenyl 2-pyridyl ketone azine (81) and pyrazinamide (82) exemplify some of these generalizations. The polymorph scenario becomes more complex experimentally and computationally when both conformational flexibility and hydrogen-bonding functionality features are simultaneously available in an organic compound, for example, in 6-amino-2-phenylsulfonylimino-1,2-dihydropyridine (three crystal structures; discussed in Section 4) and flufenamic acid (eight crystal structures) (83). The former example contains several complicating features: possible C–N geometrical isomerism, conformational flexibility along the C–S single bond, and significant hydrogen-bond functionalities, which together may provide very diverse crystallization behavior. Flufenamic acid exists in eight polymorphic forms, which exhibit a variability of conformation as well as hydrogen bonding.

Beyond routine polymorphism, there are examples that show multiple domains in the same crystal structure, highlighting other facets of polymorphism in molecular crystals. Many years ago, Boonstra & Herbstein (84) reported the composite nature of crystalline hexabromobenzene, in which two different domains of molecular arrangement are related to each other via rotation about one of the crystal axes. Along a similar line of thought, Bond et al. (85, 86) recently discussed the existence of two polymorphic domains in aspirin, which were distinguished on the basis of C–H···O dimers (form I) and catemers (form II). The primary O–H···O recognition layer unit is conserved in both domains, and the layers are arranged using alternative arrangements of weak C–H···O hydrogen bonds to obtain the three-dimensional structures. In omeprazole, there is a statistical distribution of two tautomeric forms of the molecule, and this distribution varies in the different crystal forms, which also have different physical and chemical properties (87). There are not many reports in the literature about these exotic types of polymorphism, but the examples above give an idea of the diversity of the phenomenon [see the sidebar Polymorphism in Ranitidine Hydrochloride (ZANTAC®)].

The transition from considerations of "the crystal structure" to "a crystal structure" was greatly facilitated by the power of the computational methods employed in CSP. Examples such as 6-amino-2-phenylsulfonylimino-1,2-dihydropyridine, in which a computer-predicted structure was subsequently realized experimentally, provided much reassurance that force fields were finally reaching a stage of acceptable reliability and that a computer-generated structure also had a physical reality that might eventually be realized experimentally. This also strengthened the notion that a given molecule could be associated with a large number of crystal structures, many of which

POLYMORPHISM IN RANITIDINE HYDROCHLORIDE (ZANTAC®)

Ranitidine hydrochloride is a blockbuster drug developed in the 1970s by GlaxoSmithKline (GSK) and became patented in 1977 [US patent 4,128,658 ('658 patent)]. GSK later discovered a new polymorph, which was also patented [US patent 4,521,431 ('431 patent)]. The '658 patent expired in 1995, and the '431 patent was set to expire in 2002 when Novopharm, a generic company now known as Teva Canada, filed an abbreviated new drug application (ANDA) to market form 2, claiming that '431 was invalid because form 2 is inherent in the expired '658. GSK contested this and sued Novopharm for infringement of '431. GSK argued that Novopharm's preparations of form 1 were contaminated by form 2 seeds. The court ruled in favor of GSK. In 1994, however, Novopharm filed another ANDA to market pure form 1, but which permitted up to 1% impurities. The district court now ruled in favor of Novopharm on multiple counts, and noted that 1% of form 2 in form 1 could only be an "independent component or impurity," and "not as the basis for some improvement or equivalent." The courts rejected Glaxo's theory that the scope of approval sought by Novopharm (which permitted 1% impurities) should be presumed infringing; and held Glaxo did not prove all form 2 peaks were in Novopharm's product, notwithstanding detection of what Glaxo called a form 2 peak within Novopharm's samples.

Erratum

had chemical credence, whether experimentally or in silico. Furthermore, the computational prediction of the crystal structure of an organic compound results in several choices, and it is possible that a collection of some of these forms a pattern that mimics the course of the crystallization process, very much in the manner that structure correlation mimics covalent bond breaking and making. Of course, such modeling of the crystallization process may not be accurate because the structures generated in CSP are heavily dependent on the input model. However, with this caveat, one assumes that some high-energy structures obtained in CSP, such as high-Z' structures and polymorphs and those of related pseudopolymorphs and solvates, may constitute an approximation of the late stages of crystallization. With all these developments, we are truly at the stage today at which any experimental or computed crystal structure is just that, "a crystal structure" of the molecule in question, and it is part of a complex and dynamic landscape that may include part of the supramolecular reaction trajectory for crystallization itself.

A discussion of the so-called high-Z' structures is also in order. These are crystallographic realities, and it is preferable to not dismiss them as mostly arising from inaccurate X-ray analyses or poor-quality crystals or data (88). Many high-Z' structures are of high accuracy, and there is no doubt that these structures are real. The symmetry-independent molecules in the asymmetric unit are sometimes related by a pseudoelement of symmetry located nearly at a special position or by one located on a general position (89), and sometimes the molecules are not related by any pseudosymmetry at all. Sometimes, the high-Z' structure can be more stable than the corresponding lower-Z' one, and sometimes it is the other way around (90). Sometimes, the high-Z' structure can be placed in a reaction trajectory for crystallization as a fossil relic, but at other times, it cannot (91, 92). It is sobering to note that on the few occasions when a target molecule for CSP takes a high-Z' structure, no CSP participant could ever obtain the experimental structure in their top three predictions. It is our view that these high-Z' structures play a valuable role in analyzing crystal landscapes and crystallization mechanisms (93–95).

6. CRYSTAL STRUCTURE PREDICTION IN THE CONTEXT OF THE CRYSTAL STRUCTURE LANDSCAPE

CSP is of fundamental importance in the context of understanding the crystallization process. All energy-based computational methods of CSP address this problem by scanning the

multidimensional energy hypersurface. This is performed by computing lattice energy changes with respect to parameters such as unit cell dimensions, space group symmetry, and the positional coordinates of atoms in the asymmetric unit. The aim is to seek all possible minima. These computations are generally performed at 0 K and generally do not include zero-point vibrational energy correction terms. Most importantly, they completely ignore entropic and kinetic contributions. The energy minima obtained in this exercise provide a scattered data set of putative crystal structures for a given compound. The distribution of these virtual crystal structures and their relative energies constitutes the crystal energy landscape (96).

However, it is not always possible to draw a direct relationship among these virtual minima because of the emergent nature of the crystal structure (71). A comparison of energies of any two crystal forms, although based on highly accurate state-of-the-art wave-function or DFT-based calculations, cannot provide a conclusive idea of the most probable experimental structure because of the missing kinetic features. Moreover, the experimental validation of many of these virtual structures is extremely difficult as it requires very precise control of experimental conditions. An alternative free energy–based metadynamics approach has been proposed that involves a relatively more efficient exploration of the free energy surface to study phase transformations over a range of temperatures and pressures to identify the most probable crystal structures of a compound (97). This approach helps reduce the larger number of computed structures obtained by lattice energy–based CSP methods to a smaller number of more realistic minima on the free energy landscape. Parrinello and coworkers (98) successfully employed the metadynamics approach for the prediction of all experimentally known thermodynamically stable crystalline phases for benzene. However, the applications of this method are restricted to the study of phase transitions. In another example, a metadynamics phase transition study of 5-fluorouracil characterizes the most stable form II structure as thermally unstable at room temperature (computed as a shallow minimum susceptible to phase transformations) (99). Such examples reaffirm the idea that a crystal structure is an emergent property of a molecule that depends on thermodynamic and kinetic factors in the molecule → crystal progression. There may be many crystal forms predicted in a CSP run that can be easily transformed from one to another in silico by moving across the crystal landscape. However, it is not always necessary that all such transformations be experimentally realized without undergoing melting and renucleation of new crystalline phases.

During crystallization, there are several possible nucleation pathways—only some of these actually lead to nucleation and growth of distinct crystal forms (75). These crystal forms may include both polymorphs and pseudopolymorphs. Each of these experimentally observed structures represents a data point on a crystallization pathway predefined at the nucleation stages. The collection of all these experimentally observed data points constitutes the crystal structure landscape (100, 101). The existence of multiple polymorphs in general provides experimental evidence for the existence of many crystallization pathways. The supramolecular synthon approach can help in the identification and classification of some of these pathways and in establishing inter-relationships among these structural data points (102). The crystal structure landscape deals with experimentally viable crystal structures of a compound and may also include structures from similar compounds that establish the existence of other probable crystallization pathways for a given system. The existence of several molecular aggregates in the supersaturated solution, which resemble the final crystalline phases of the polymorphs, has also been proposed (74). By using suitable auxiliary molecules as templates, metastable forms may be accessed (103).

The study of the crystal structure landscape includes the identification of well-characterized crystallization routes through the analysis of various crystal forms. Blagden & Davey (104) pointed out three important aspects: (*a*) A thorough understanding of the structural similarities and dissimilarities of all known polymorphs is essential (i.e., the characterization of crystallization

routes/pathways). (*b*) The role played by both thermodynamics and kinetic factors in deciding the outcome of crystallization processes cannot be neglected. (*c*) The identification of the growth unit or the supramolecular synthons for a given crystal form is important. A synthon or combination of synthons acts as a primary growth unit for a given crystallization pathway. Most importantly, the assessment of the thermodynamic/kinetic feasibility of various crystal forms/pathways is obtained by studying the outcome of the crystallization of structurally related compounds. This is also an important and distinctive feature between the crystal energy landscape and crystal structure landscape. The former does not take into account kinetic factors even as it generates a large number of energy-ranked virtual structures, whereas, in the latter, all minima belong to real space, and the experimental search of these structural end points (minima) need not be restricted to a single compound.

A recent study by Desiraju and coworkers (100) on the crystal structure landscape of orcinol, 5-methyl-1,3-dihydroxybenzene, emphasized the importance of supramolecular synthons as structural modules of emergent crystal structures of nearly 30 orcinol polymorphs, pseudopolymorphs, and cocrystals. The study identifies 12 distinct crystallization pathways for orcinol crystallization with mono and bi-N-acceptor coformers. Unlike the computed crystal energy landscape obtained from a CSP study, these pathways represent the most probable and experimentally established, robust crystallization routes for the orcinol cocrystal system. This group also extended the idea of exploring a specific crystallization route for the orcinol–4,4′-bipyridine cocrystal system through the isolation of five distinct polymorphs based on O–H···N supramolecular synthons (105). The scope and application of this approach can also be extrapolated to other 1,3-dihydroxybenzenes that are widely used in industrial and analytical applications.

Desiraju and coworkers (106) recently exploited the small difference in the atomic volume, near-inert supramolecular behavior, and reasonably different chemical features of the fluorine atom in fluoro-substituted benzoic acids to characterize distinct crystallization pathways for both single-component and cocrystal systems. The fluoro-substitution method of studying the crystal structure landscape provides an alternative strategy to sample the inaccessible high-energy regions of the landscape of the nonfluorinated analog under normal experimental conditions. This strategy was found successful in exploring the structural landscapes of both single-component and multicomponent systems (107). The fluoro-substitution strategy effectively raises the energy of the native (unsubstituted) structure slightly above the energy minimum. Accordingly, one can explore the high-energy regions by moving vertically across the energy hypersurfaces. This strategy provides a subtle balance between chemical and geometrical features for exploring alterations in molecular packing in compounds and their fluoro-substituted analogs.

We may finally consider the ways in which certain core ideas in physical organic chemistry may be applied to crystallization, which is a supramolecular reaction. Hammond's postulate in chemical kinetics states that, "if two states, as, for example, a transition state and an unstable intermediate, occur consecutively during a reaction process and have nearly the same energy content, their interconversion will involve only a small reorganization of the molecular structures" (108, p. 334). The Curtin-Hammett principle states that "the distribution of products in a reaction that has many pathways need bear no relation to the relative stability of those products" (109, p. 111). If we apply these ideas to crystallization, one finds that crystal nuclei that lead to kinetic products resemble early transition states, whereas those that lead to thermodynamic products resemble the final crystal structures themselves; in other words, they are like late-transition states. Again, some kinetic crystals may be obtained quite easily and may persist under certain preferred crystallization conditions. Concomitancy is also not hard to understand, and crystal forms may arise

from independent pathways. These analogies between molecular and supramolecular processes could act as guides for the interpretation of crystallization events.

7. OUTLOOK

Over the past few decades, CSP methods have evolved to a stage at which we can obtain lattice energies of crystal structures very accurately. However, it is still a challenge to predict the most probable crystal structure of a compound under given conditions. In other words, the prediction of optimal experimental conditions for obtaining a predicted virtual polymorph of a compound is hopelessly difficult. However, novel experimental strategies such as fluoro substitution have provided a means of exploring these uncharted regions of crystal structural or energy landscapes. The focus should now be on deciphering the role of kinetics in crystal nucleation and growth to obtain a molecular-level understanding of the mechanism of crystallization. The whole crystalline phase space of a compound is a vast sea of crystal structures that includes polymorphs, pseudopolymorphs, and other multicomponent forms that originate from a common mother phase. The interrelationship between these phases has been brought out by the idea of the crystal structure landscape in terms of growth units, nuclei, and crystallization pathways. In summary, emphasis should be given to combining experimental and computational approaches for the study of crystallization pathways and to improve the success rate of CSP.

SUMMARY POINTS

1. The subject of CSP has evolved from the earlier endeavors that attempted to provide trial solutions for the phase problem in X-ray crystallography.
2. The beginning of the CSP blind tests in 1999 helped to bring structure prediction into focus and to provide a common platform for researchers to work together toward the development of better computational methodologies.
3. Frequent failures in the prediction of experimental crystal structures in the blind tests hinted at the importance of kinetic and entropic factors that are generally ignored in the computation.
4. Polymorphism should always be considered as being widely prevalent, and hence the existence of just one thermodynamically stable crystal structure cannot be used as a guide to perform CSP.
5. A collection of virtual structures obtained from CSP can be useful in the understanding of crystallization routes. However, it is not always possible to draw a direct relationship among these virtual minima because of the emergent nature of the crystal structure. The collection of all these experimentally observed data points (structures) constitutes the crystal structure landscape.
6. The assessment of the thermodynamic/kinetic feasibility of various crystal forms/pathways obtained from the study of the crystal structure landscape of a compound is greatly assisted by the supramolecular synthon approach.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

T.S.T. thanks the Department of Science and Technology (DST), Government of India, for financial assistance (project SR/FT/CS-163/2011), R.D. thanks the Indian Institute of Science for a senior research fellowship, and G.R.D. thanks the DST for the award of a J.C. Bose fellowship.

LITERATURE CITED

1. Kitaigorodskii AI. 1973. *Molecular Crystals and Molecules*. New York: Academic
2. Desiraju GR. 1989. *Crystal Engineering: The Design of Organic Solids*. Amsterdam: Elsevier
3. Price SL. 2009. Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism. *Acc. Chem. Res.* 42:117–26
4. Day GM. 2011. Current approaches to predicting molecular organic crystal structures. *Crystallogr. Rev.* 17:3–52
5. Kendrick J, Leusen FJJ, Neumann MA, van de Streek J. 2011. Progress in crystal structure prediction. *Chem. Eur. J.* 17:10736–44
6. Bragg WH. 1921. The structure of organic crystals. *Proc. Phys. Soc. Lond.* 34:33–50
7. Robertson JM. 1953. *Organic Crystals and Molecules: Theory of X-Ray Structure Analysis with Applications to Organic Chemistry*. Ithaca, NY: Cornell Univ. Press
8. Pauling L. 1929. The principles determining the structure of complex ionic crystals. *J. Am. Chem. Soc.* 51:1010–26
9. Robertson JM. 1935. The structure of benzoquinone: a quantitative X-ray investigation. *Proc. R. Soc. Lond. A* 150:106–28
10. Robertson JM. 1936. The structure of resorcinol: a quantitative X-ray investigation. *Proc. R. Soc. Lond. A* 157:79–99
11. Robertson JM, Ubbelohde AR. 1938. A new form of resorcinol. II. Thermodynamic properties in relation to structure. *Proc. R. Soc. Lond. A* 167:136–47
12. Authier A. 2013. *Early Days of X-Ray Crystallography*. Oxford, UK: Oxford Univ. Press
13. Khotsyanova TL, Kitaigorodskii AI, Struchkov YT. 1953. Crystal structure of iodoform. *Z. Fiz. Khim.* 27:647–56
14. Lommerse JPM, Motherwell WDS, Ammon HL, Dunitz JD, Gavezzotti A, et al. 2000. A test of crystal structure prediction of small organic molecules. *Acta Crystallogr. B* 56:697–714
15. Motherwell WDS, Ammon HL, Dunitz JD, Dzyabchenko A, Erk P, et al. 2002. Crystal structure prediction of small organic molecules: a second blind test. *Acta Crystallogr. B* 58:647–61
16. Day GM, Motherwell WDS, Ammon HL, Boerrigter SXM, Della Valle RG, et al. 2005. A third blind test of crystal structure prediction. *Acta Crystallogr. B* 61:511–27
17. Day GM, Cooper TG, Cruz-Cabeza AJ, Hejczyk KE, Ammon HL, et al. 2009. Significant progress in predicting the crystal structures of small organic molecules: a report on the fourth blind test. *Acta Crystallogr. B* 65:107–25
18. Bardwell DA, Adjiman CS, Arnautova YA, Bartashevich E, Boerrigter SXM, et al. 2011. Towards crystal structure prediction of complex organic compounds: a report on the fifth blind test. *Acta Crystallogr. B* 67:535–51
19. Kitaigorodskii AI. 1961. The interaction curve of non-bonded carbon and hydrogen atoms and its application. *Tetrahedron* 14:230–36
20. Kitaigorodskii AI. 1965. The principle of close packing and the condition of thermodynamic stability of organic crystals. *Acta Crystallogr.* 18:585–90
21. Williams DE. 1966. Nonbonded potential parameters derived from crystalline aromatic hydrocarbons. *J. Chem. Phys.* 45:3770–78
22. Nyburg SC, Wong-Ng W. 1979. Anisotropic atom-atom forces and the space group of solid chlorine. *Proc. R. Soc. Lond. A* 367:29–45
23. Price SL, Stone AJ. 1982. The anisotropy of the $C1_2$–$C1_2$ pair potential as shown by the crystal structure: evidence for intermolecular bonding or lone pair effects? *Mol. Phys.* 47:1457–70

24. Williams DE, Hsu L-Y. 1985. Transferability of nonbonded Cl···Cl potential energy function to crystalline chlorine. *Acta Crystallogr. A* 41:296–301
25. Sakurai T. 1965. The crystal structure of the triclinic modification of quinhydrone. *Acta Crystallogr.* 19:320–30
26. Sakurai T. 1968. On the refinement of the crystal structures of phenoquinone and monoclinic quinhydrone. *Acta Crystallogr. B* 24:403–12
27. Wen S, Beran GJO. 2012. Accidental degeneracy in crystalline aspirin: new insights from high-level ab initio calculations. *Cryst. Growth Des.* 12:2169–72
28. Wen S, Beran GJO. 2012. Crystal polymorphism in oxalyl dihydrazide: Is empirical DFT-D accurate enough? *J. Chem. Theory Comput.* 8:2698–705
29. Neumann MA. 2008. Tailor-made force fields for crystal-structure prediction. *J. Phys. Chem. B* 12:9810–29
30. Beran GJO, Nanda K. 2010. Predicting organic crystal lattice energies with chemical accuracy. *J. Phys. Chem. Lett.* 1:3480–87
31. Hirata S, Gilliard K, He X, Li J, Sode O. 2014. Ab initio molecular crystal structures, spectra, and phase diagrams. *Acc. Chem. Res.* 47:2721–30
32. Yang J, Hu W, Usvyat D, Matthews D, Schütz M, Chan GK-L. 2014. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science* 345:640–43
33. Stoll H, Paulus B, Fulde P. 2005. On the accuracy of correlation-energy expansions in terms of local increments. *J. Chem. Phys.* 123:144108
34. Manby FR, Alfe D, Gillan MJ. 2006. Extension of molecular electronic structure methods to the solid state: computation of the cohesive energy of lithium hydride. *Phys. Chem. Chem. Phys.* 8:5178–80
35. Li J, Sode O, Voth GA, Hirata S. 2013. A solid-solid phase transition in carbon dioxide at high pressures and intermediate temperatures. *Nat. Commun.* 4:2647
36. Kitaigorodskii AI. 1968. Calculation of molecular conformation and packing in crystals by means of atom-atom potentials. *Polym. Sci. U.S.S.R.* 10:3097–100
37. Williams DE. 1968. Computer calculation of molecular crystal structures. *Science* 159:645–46
38. Dauber P, Hagler AT. 1980. Crystal packing, hydrogen bonding, and the effect of crystal forces on molecular conformation. *Acc. Chem. Res.* 13:105–12
39. Leiserowitz L, Hagler AT. 1983. The generation of possible crystal structures of primary amides. *Proc. R. Soc. Lond. A* 388:133–75
40. Maddox J. 1988. Crystals from first principles. *Nature* 335:201
41. Perlstein J. 1992. Molecular self-assemblies: Monte Carlo predictions for the structure of the one-dimensional translation aggregate. *J. Am. Chem. Soc.* 114:1955–63
42. Karfunkel HR, Gdanitz RJ. 1992. Ab initio prediction of possible crystal structures for general organic molecules. *J. Comput. Chem.* 13:1171–83
43. Hofmann DMF, Lengauer T. 1998. Crystal structure prediction based on statistical potentials. *J. Mol. Model.* 4:132–44
44. Holden JR, Du Z, Ammon HL. 1993. Prediction of possible crystal structures for C-, H-, N-, O-, and F-containing organic compounds. *J. Comput. Chem.* 14:422–37
45. Gavezzotti A, Filippini G. 1996. Computer prediction of organic crystal structures using partial X-ray diffraction data. *J. Am. Chem. Soc.* 118:7153–57
46. Hofmann DWM, Lengauer T. 1997. A discrete algorithm for crystal structure prediction of organic molecules. *Acta Crystallogr. A* 53:225–35
47. Gavezzotti A. 1991. Generation of possible crystal structures from the molecular structure for low-polarity organic compounds. *J. Am. Chem. Soc.* 113:4622–29
48. Chin DN, Palmore GTR, Whitesides GM. 1999. Predicting crystalline packing arrangements of molecules that form hydrogen-bonded tapes. *J. Am. Chem. Soc.* 121:2115–22
49. Wawak RJ, Gibson KD, Liwo A, Scheraga HA. 1996. Theoretical prediction of a crystal structure. *Proc. Natl. Acad. Sci. USA* 93:1743–46
50. Willock DJ, Price SL, Leslie M, Catlow CRA. 1995. The relaxation of molecular crystal structures using a distributed multipole electrostatic model. *J. Comput. Chem.* 16:628–47

51. van Eijck BP, Mooij WTM, Kroon J. 1995. Attempted prediction of the crystal structures of six monosaccharides. *Acta Crystallogr. B* 51:99–103
52. Chaka AM, Zaniewski R, Youngs W, Tessierj C, Klopman G. 1996. Predicting the crystal structure of organic molecular materials. *Acta Crystallogr. B* 52:165–83
53. Rarey M, Kramer B, Lengauer T, Klebe G. 1996. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261:470–89
54. Rarey M, Wefing S, Lengauer T. 1996. Placement of medium-sized molecular fragments into active sites of proteins. *J. Comput.-Aided Mol. Des.* 10:41–54
55. Williams DE. 1996. Ab initio molecular packing analysis. *Acta Crystallogr. A* 52:326–28
56. Asmadi A, Neumann MA, Kendrick J, Girard P, Perrin MA, Leusen FJJ. 2009. Revisiting the blind tests in crystal structure prediction: accurate energy ranking of molecular crystals. *J. Phys. Chem. B* 113:16303–13
57. Dunitz JD, Gavezzotti A. 2005. Towards a quantitative description of crystal packing in terms of molecular pairs: application to the hexamorphic crystal system, 5-methyl-2[(2-nitrophenyl)amino]-3-thiophenecarbonitrile. *Cryst. Growth Des.* 5:2180–89
58. Desiraju GR, Steiner T. 1999. *The Weak Hydrogen Bond in Structural Chemistry and Biology*. Oxford, UK: Oxford Univ. Press
59. Aparna V, Rambabu G, Panigrahi SK, Sarma JARP, Desiraju GR. 2005. Virtual screening of 4-anilinoquinazoline analogues as EGFR kinase inhibitors: importance of hydrogen bonds in the evaluation of poses and scoring functions. *J. Chem. Inf. Model.* 45:725–38
60. Jetti RKR, Boese R, Sarma JARP, Reddy LS, Vishweshwar P, Desiraju GR. 2004. Searching for a polymorph: second crystal form of 6-amino-2-phenylsulfonylimino-1,2-dihydropyridine. *Angew. Chem. Int. Ed. Engl.* 42:1963–67
61. Roy S, Matzger AJ. 2009. Unmasking a third polymorph of benchmark crystal-structure-prediction compound. *Angew. Chem. Int. Ed. Engl.* 48:8505–8
62. Chan HCS, Kendrick J, Leusen FJJ. 2011. Molecule VI, a benchmark crystal-structure-prediction sulfonimide: Are its polymorphs predictable? *Angew. Chem. Int. Ed. Engl.* 50:2979–81
63. Grimme S, Antony J, Ehrlich S, Krieg H. 2010. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* 132:154104
64. Marom N, DiStasio RA Jr., Atalla V, Levchenko S, Reilly AM, et al. 2013. Many-body dispersion interactions in molecular crystal polymorphism. *Angew. Chem. Int. Ed. Engl.* 52:6629–32
65. Reilly AM, Tkatchenko A. 2014. Role of dispersion interactions in the polymorphism and entropic stabilization of the aspirin crystal. *Phys. Rev. Lett.* 113:055701
66. Braun DE, Bhardwaj RM, Florence AJ, Tocher DA, Price SL. 2013. Complex polymorphic system of gallic acid: five monohydrates, three anhydrates, and over 20 solvates. *Cryst. Growth Des.* 13:19–23
67. Desiraju GR, Sarma JARP. 2002. The supramolecular synthon approach to crystal structure prediction. *Cryst. Growth Des.* 2:93–100
68. Dey D, Kirchner MT, Vangala VR, Desiraju GR, Mondal R, Howard JAK. 2005. Crystal structure prediction of aminols: advantages of a supramolecular synthon approach with experimental structures. *J. Am. Chem. Soc.* 127:10545–59
69. Gavezzotti A. 2002. Structure and intermolecular potentials in molecular crystals. *Model. Simul. Mater. Sci. Eng.* 10:R1–29
70. Desiraju GR. 1995. Supramolecular synthons in crystal engineering: a new organic synthesis. *Angew. Chem. Int. Ed. Engl.* 34:2311–27
71. Desiraju GR. 2005. Chemistry: the middle kingdom. *Curr. Sci.* 88:374–80
72. Groth P. 1906–1919. *Chemische Kristallographie*, Vol. 3–5. Leipzig: Verlag von Wilhelm Engelmann
73. Deffet L. 1942. *Répertoire des Composés organiques polymorphes*. Liége: Ed. Desoer
74. Staab E, Addadi L, Leiserowitz L, Lahav M. 1990. Control of polymorphism by 'tailor-made' polymeric crystallization auxiliaries: preferential precipitation of a metastable polar form for second harmonic generation. *Adv. Mater.* 2:40–43
75. Desiraju GR. 2002. Cryptic crystallography. *Nat. Mater.* 1:77–79
76. McCrone WC. 1965. Polymorphism. In *Physics and Chemistry of the Organic Solid State*, Vol. 2, ed. D Fox, MM Labes, A Weissberger, pp. 725–67. New York: Wiley Intersci.

77. Dunitz JD, Bernstein J. 1995. Disappearing polymorphs. *Acc. Chem. Res.* 28:193–200
78. Sarma JARP, Desiraju GR. 1999. Polymorphism and pseudopolymorphism in organic crystals. In *Crystal Engineering: The Design and Application of Fundamental Solids*, ed. KR Seddon, M Zaworotko, pp. 325–56. Dordrecht: Kluwer Acad.
79. Mukherjee A, Desiraju GR. 2011. Synthon polymorphism and pseudopolymorphism in co-crystals: the 4,4′-bipyridine–4-hydroxybenzoic acid structural landscape. *Chem. Commun.* 47:4090–92
80. Sreekanth BR, Vishweshwar P, Vyas K. 2007. Supramolecular synthon polymorphism in 2:1 co-crystal of 4-hydroxybenzoic acid and 2,3,5,6-tetramethylpyrazine. *Chem. Commun.* 23:2375–77
81. Amadei E, Carcelli M, Ianelli S, Cozzini P, Pelagatti P, Pelizzi C. 1998. Ligand behaviour and reactivity of phenyl 2-pyridyl ketone azine: structures of two polymorphic forms of the azine and a copper complex of the 3-phenyltriazolo[1,5-*a*]pyridine. *J. Chem. Soc. Dalton Trans.* 1998:1025–30
82. Desiraju GR. 1997. Crystal gazing: structure prediction and polymorphism. *Science* 278:404–5
83. López-Mejías V, Kampf JW, Matzger AJ. 2012. Nonamorphism in flufenamic acid and a new record for a polymorphic compound with solved structures. *J. Am. Chem. Soc.* 134:9872–75
84. Boonstra EG, Herbstein FH. 1963. Composite hexabromobenzene crystals. *Acta Crystallogr.* 16:252–55
85. Bond AD, Boese R, Desiraju GR. 2007. On the polymorphism of aspirin. *Angew. Chem. Int. Ed. Engl.* 46:615–17
86. Bond AD, Boese R, Desiraju GR. 2007. On the polymorphism of aspirin: crystalline aspirin as intergrowths of two "polymorphic" domains. *Angew. Chem. Int. Ed. Engl.* 46:618–22
87. Bhatt PM, Desiraju GR. 2007. Tautomeric polymorphism in omeprazole. *Chem. Commun.* 2007:2057–59
88. Bernstein J, Dunitz JD, Gavezzotti A. 2008. Polymorphic perversity: crystal structures with many symmetry-independent molecules in the unit cell. *Cryst. Growth Des.* 8:2011–18
89. Desiraju GR, Calabrese JC, Harlow RL. 1991. Pseudoinversion centers in space group $P\bar{1}$ and a redetermination of the crystal structure of 3,4-dimethoxycinnamic acid: a study of non-crystallographic symmetry. *Acta Crystallogr. B* 47:77–86
90. Nangia A. 2008. Conformational polymorphisms in organic crystals. *Acc. Chem. Res.* 41:595–604
91. Desiraju GR. 2007. On the presence of multiple molecules in the crystal asymmetric unit ($Z' > 1$). *CrystEngComm* 9:91–92
92. Anderson KM, Steed JW. 2007. Comment on "On the presence of multiple molecules in the crystal asymmetric unit ($Z' > 1$)" by Gautam R. Desiraju, *CrystEngComm*, 2007, 9, 91. *CrystEngComm* 9:328–30
93. Thakur TS, Sathishkumar R, Dikundwar AG, Guru Row TN, Desiraju GR. 2010. Third polymorph of phenylacetylene. *Cryst. Growth Des.* 10:4246–49
94. Dikundwar AG, Sathishkumar R, Guru Row TN, Desiraju GR. 2011. Structural variability in the monofluoroethynylbenzenes mediated through interactions involving "organic" fluorine. *Cryst. Growth Des.* 11:3954–63
95. Ridout J, Price LS, Howard JAK, Probert MR. 2014. Polymorphism arising from differing rates of compression of liquids. *Cryst. Growth Des.* 14:3384–91
96. Gavezzotti A. 2002. Ten years of experience in polymorph prediction: what next? *CrystEngComm* 4:343–47
97. Laio A, Parrinello M. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* 99:12562–66
98. Raiteri P, Martoňák R, Parrinello M. 2005. Exploring polymorphism: the case of benzene. *Angew. Chem. Int. Ed. Engl.* 44:3769–73
99. Karamertzanis PG, Raiteri P, Parrinello M, Leslie M, Price SL. 2008. The thermal stability of lattice-energy minima of 5-fluorouracil: metadynamics as an aid to polymorph prediction. *J. Phys. Chem. B* 112:4298–308
100. Mukherjee A, Grobelny P, Thakur TS, Desiraju GR. 2011. Polymorphs, pseudopolymorphs, and co-crystals of orcinol: exploring the structural landscape with high throughput crystallography. *Cryst. Growth Des.* 11:2637–53
101. Tothadi S, Desiraju GR. 2012. Unusual co-crystal of isonicotinamide: the structural landscape in crystal engineering. *Philos. Trans. R. Soc. A* 370:2900–15
102. Desiraju GR. 2007. Crystal engineering: a holistic view. *Angew. Chem. Int. Ed. Engl.* 46:8342–56
103. Friščič T, MacGillivray LR. 2009. Engineering cocrystal and polymorph architecture via pseudoseeding. *Chem. Commun.* 773–75

104. Blagden N, Davey RJ. 2003. Polymorph selection: challenges for the future? *Cryst. Growth Des.* 3:873–85
105. Dubey R, Pavan MS, Guru Row TN, Desiraju GR. 2014. Crystal landscape in the orcinol: 4,4′-bipyridine system: synthon modularity, polymorphism and transferability of multipole charge density parameters. *IUCrJ* 1:8–18
106. Dubey R, Pavan MS, Desiraju GR. 2012. Structural landscape of benzoic acid: using experimental crystal structures of fluorobenzoic acids as a probe. *Chem. Commun.* 48:9020–22
107. Dubey R, Desiraju GR. 2014. Structural landscape of the 1:1 benzoic acid:isonicotinamide cocrystal. *Chem. Commun.* 50:1181–84
108. Hammond GS. 1955. A correlation of reaction rates. *J. Am. Chem. Soc.* 77:334–38
109. Curtin DY. 1954. Stereochemical control of organic reactions: differences in behavior of diastereoisomers. I. Ethane derivatives: the *cis* effect. *Rec. Chem. Prog.* 15:111–28