1

Crystallography

Susan M. Reutzel-Edens[1] *and Peter Müller*[2]

[1] *Small Molecule Design & Development, Eli Lilly & Company, Lilly Corporate Center, Indianapolis, IN, USA*
[2] *X-Ray Diffraction Facility, MIT Department of Chemistry, Cambridge, MA, USA*

1.1 Introduction

Functional organic solids, ranging from large-tonnage commodity materials to high-value specialty chemicals, are commercialized for their unique physical and chemical properties. However, unlike many substances of scientific, technological, and commercial importance, drug molecules are almost always chosen for development into drug products based solely on their biological properties. The ability of a drug molecule to crystallize in solid forms with optimal material properties is rarely a consideration. Still, with an estimated 90% of small-molecule drugs delivered to patients in a crystalline state [1], the importance of crystals and crystal structure to pharmaceutical development cannot be overstated. In fact, the first step in transforming a molecule to a medicine (Figure 1.1) is invariably identifying a stable crystalline form, one that:

- Through its ability to exclude impurities during crystallization, can be used to purify the drug substance coming out of the final step of the chemical synthesis.
- May impart stability to an otherwise chemically labile molecule.
- Is suitable for downstream processing and long-term storage.
- Not only meets the design requirements but also will ensure consistency in the safety and efficacy profile of the drug product throughout its shelf life.

The mechanical, thermodynamic, and biopharmaceutical properties of a drug substance will strongly depend on how a molecule packs in its

Pharmaceutical Crystals: Science and Engineering, First Edition.
Edited by Tonglei Li and Alessandra Mattei.

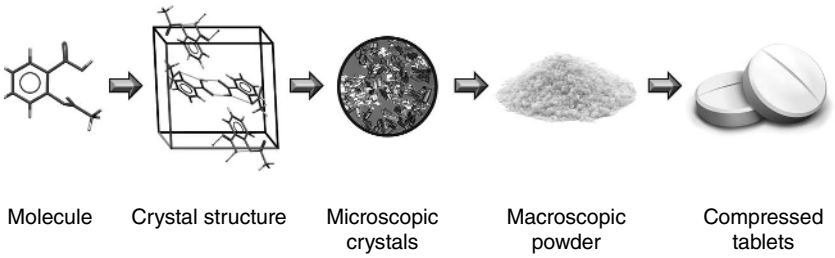


Figure 1.1 Materials science perspective of the steps involved in transforming a molecule to a medicine.

three-dimensional (3D) crystal structure, yet it is not given that a drug candidate entering into pharmaceutical development will crystallize, let alone in a form that is amenable to processing, stable enough for long-term storage, or useful for drug delivery. Because it is rarely possible to manipulate the chemical structure of the drug itself to improve material properties,[1] pharmaceutical scientists will typically explore multicomponent crystal forms, including salts, hydrates, and more recently cocrystals, if needed, in the search for commercially viable forms. A salt is an ionic solid formed between either a basic drug and a sufficiently acidic guest molecule or an acidic drug and basic guest. Cocrystals are crystalline molecular complexes formed between the drug (or its salt) and a neutral guest molecule. Hydrates, a subset of a larger class of crystalline solids, termed solvates, are characterized by the inclusion of water in the crystal structure of the compound. When multiple crystalline options are identified in solid form screening, as is often the case for ever more complex new chemical entities in current drug development pipelines, it is the connection between internal crystal structure, particle properties, processing, and product performance, the components of the materials science tetrahedron, [3] that ultimately determines which form is progressed in developing the drug product. Not surprisingly, crystallography, the science of shapes, structures, and properties of crystals, is a key component of all studies relating the solid-state chemistry of drugs to their ultimate use in medicinal products.

Crystallization is the process by which molecules (or ion pairs) self-assemble in ordered, close-packed arrangements (crystal structures). It usually involves two steps: crystal nucleation, the formation of stable molecular aggregates or clusters (nuclei) capable of growing into macroscopic crystals; and crystal growth, the subsequent development of the nuclei into visible dimensions. Crystals that successfully nucleate and grow will, in many cases, form

1 There is good interest in using small-molecule crystallography to address the solubility limitations of lead compounds by disrupting crystal packing through chemical modification, with some success reported in the literature. See Ref. [2].

distinctive, if not spectacular, shapes (habits) characterized by well-defined faces or facets. Commonly observed habits, which are often described as needles, rods, plates, tablets, or prisms, emerge because crystal growth does not proceed at the same rate in all directions. The slowest-growing faces are those that are morphologically dominant; however, as the external shape of the crystal depends both on its internal crystal structure and the growth conditions, crystals of the same internal structure (same crystal form) may have different external habits. The low molecular symmetry common to many drug molecules and anisotropic (directional) interactions within the crystal structure often lead to acicular (needle shaped) or platy crystals with notoriously poor filtration and flow properties [4]. Since crystal size and shape can have a strong impact on release characteristics (dissolution rate), material handling (filtration, flow), and mechanical properties (plasticity, elasticity, density) relevant to tablet formulation, crystallization processes targeting a specific crystal form are also designed with exquisite control of crystal shape and size in mind.

Some compounds (their salts, hydrates, and cocrystals included) crystallize in a single solid form, while others crystallize in possibly many different forms. *Polymorphism* [Greek: *poly* = many, *morph* = form] is the ability of a molecule to crystallize in multiple crystal forms (of identical composition) that differ in molecular packing and, in some cases, conformation [5]. A compelling example of a highly polymorphic molecule is 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile, also known as ROY, an intermediate in the synthesis of the schizophrenia drug olanzapine. Polymorphs of ROY, mostly named for their red-orange-yellow spectrum of colors and unique and distinguishable crystal shapes, are shown in Figure 1.2 [6]. Multiple crystal forms of ROY were first suggested by the varying brilliant colors and morphologies of individual crystals in a single batch of the compound. Confirmation of polymorphism later came with the determination of many of their crystal structures by X-ray diffraction (Table 1.1) [7]. In this example, the color differences were traced to different molecular conformations, characterized by θ, the torsion angle relating the rigid *o*-nitroaniline and thiophene rings in the crystal structures of the different ROY polymorphs [8].

The current understanding of structure in crystals would not be where it is today without the discovery that crystals diffract X-rays and that this phenomenon can be used to extract detailed structural information. Indeed, it is primarily through their diffraction that crystals have been used to study molecular structure and stereochemistry at an atomic level. Of course, detailed evaluation of molecular conformation and intermolecular interactions in a crystal can suggest important interactions that may drive binding to receptor sites, and so crystallography is a vital component early in the drug discovery process when molecules are optimized for their biological properties. Crystallography plays an equally important role in pharmaceutical development, where material properties defined by 3D crystal packing lie at the heart of transforming

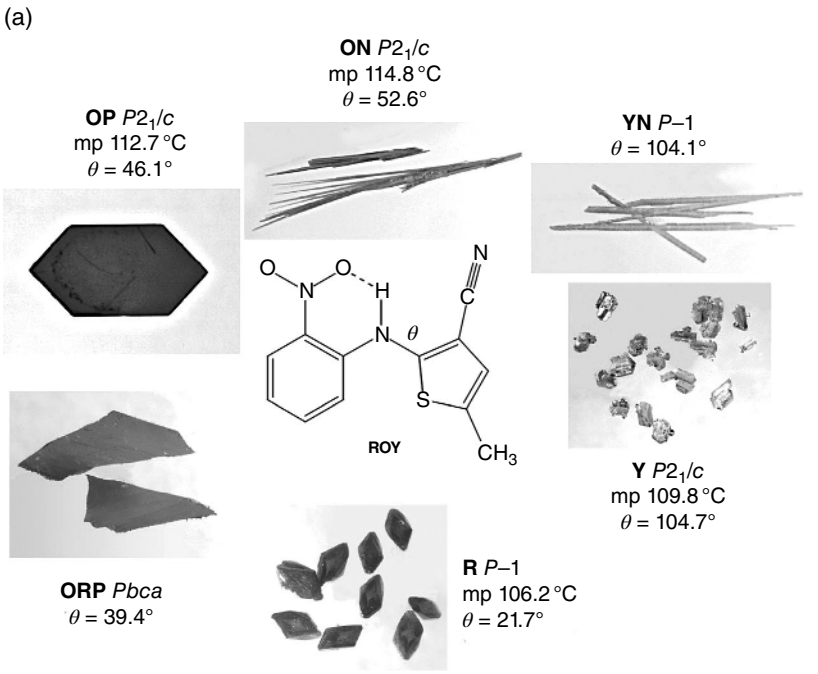
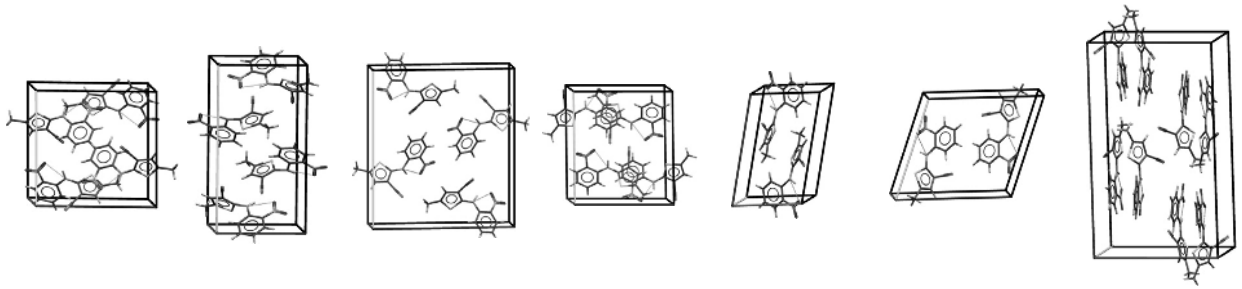


Figure 1.2 (a) Crystal polymorphs of ROY highlighting the diverse colors and shapes of crystals grown from different solutions and (b) photomicrographs showing the concurrent cross nucleation of the R polymorph on Y04 produced by melt crystallization and (c) single crystals of YT04 grown by seeding a supersaturated solution. *Source:* Adapted with permission from Yu et al. [6], copyright 2000, and from Chen et al. [7], copyright 2005, American Chemical Society. (*See insert for color representation of the figure.*)

a molecule to a medicine. Thus, this chapter considers small-molecule crystallography for the study of molecular and crystal structure. Following a brief history of crystallography, the basic elements of crystal structure, the principles of X-ray diffraction, and the process of determining a crystal structure from

Table 1.1 Crystallographic data from X-ray structure determinations of seven ROY polymorphs.



Form	YT04	Y	ON	OP	R	YN	ORP
CSD refcode	QAXMEH12	QAXMEH01	QAXMEH	QAXMEH03	QAXMEH02	QAXMEH04	QAXMEH05
Crystal system	Monoclinic	Monoclinic	Monoclinic	Monoclinic	Triclinic	Triclinic	Orthorhombic
Space group	$P2_1/n$	$P2_1/n$	$P2_1/c$	$P2_1/n$	P-1	P-1	$Pbca$
Color	Yellow	Yellow	Orange	Orange	Red	Yellow	Orange-red
Habit	Prism	Prism	Needle	Plate	Prism	Needle	Plate
a, Å	8.2324(4)	8.5001(1)	3.9453(1)	7.9760(1)	7.4918(1)	4.5918(1)	13.177(3)
b, Å	11.8173(5)	16.413(2)	18.685(1)	13.319(2)	7.7902(1)	11.249(2)	8.0209(10)
c, Å	12.3121(6)	8.5371(1)	16.3948(1)	11.676(1)	11.9110(1)	12.315(2)	22.801(5)
α, deg	90	90	90	90	75.494(1)	71.194(1)	90
β, deg	102.505(1)	91.767(1)	93.830(1)	104.683(1)	77.806(1)	89.852(1)	90
γ, deg	90	90	90	90	63.617(1)	88.174(1)	90
Volume, Å^3	1169.36(9)	1190.5	1205.9	1199.9	598.88	601.85	2409.8
Z	4	4	4	4	2	2	8
D_{calc}, g cm^{-3}	1.473	1.447	1.428	1.435	1.438	1.431	1.429
T, K	296	293	293	295	293	296	296

diffraction data are described. Complementary approaches to single-crystal diffraction, namely, structure determination from powder diffraction, solid-state nuclear magnetic resonance (NMR) spectroscopy (NMR crystallography), and emerging crystal structure prediction (CSP) methodology, are also highlighted. Finally, no small-molecule crystallography chapter would be complete without mention of the Cambridge Structural Database (CSD), *the* repository of all publicly disclosed small-molecule organic and organometallic crystal structures, and the solid form informatics tools that have been developed by the Cambridge Crystallographic Data Centre (CCDC) for the worldwide crystallography community to efficiently and effectively mine the vast structural information warehoused in the CSD.

1.2 History

Admiration for and fascination by crystals is as old as humanity itself. Crystals have been assigned mystic properties (for example, crystal balls for future telling), healing powers (amethyst, for example, is said to have a positive effect on digestion and hormones), and found uses as embellishments and jewelry already thousands of years ago. Crystallography as a science is also comparatively old. In 1611, the German mathematician and astronomer Johannes Kepler published the arguably first ever scientific crystallographic manuscript. In his essay *Strena seu de nive sexangula* (a new year's gift of the six-cornered snowflake), starting from the hexagonal shape of snowflakes, Kepler derived, among other things, the cubic and hexagonal closest packings (now known as the Kepler conjecture) and suggested a theory of crystal growth [9].

Later in history, when mineralogy became more relevant, Nicolaus Steno in 1669 published the law of constant interfacial angles,[2] and in 1793 René Just Haüy, often called the "father of modern crystallography," discovered the periodicity of crystals and described that the relative orientations of crystal faces can be expressed in terms of integer numbers.[3] Those numbers describing the orientation of crystal faces and, generally, of any plane drawn through crystal lattice points are now known as Miller indices[4] (introduced in 1839 by William Hallowes Miller). Miller indices are one of the most important concepts in modern crystallography as we will see later in this chapter. In 1891, the Russian mineralogist and mathematician Evgraf Stepanovich Fedorov and the German mathematician Arthur Moritz Schoenflies published independently a list of all 3D space groups. Both their publications contained errors, which were

2 Published in his book *De solido intra solidum naturaliter content* (1669).
3 Published in the two essays *De la structure considérée comme caractère distinctif des minéraux* and *Exposition abrégé de la théorie de la structure des cristaux* (both 1793).
4 Perhaps because Miller is easier to pronounce than Haüy.

discovered by the respective other author, and the correct list of the 230 3D space groups was developed in collaboration by Fedorov and Schoenflies in 1892.[5]

With the law of constant interfacial angles, the concept of Miller indices and the complete list of space groups, the crystallographic world was ready for the discovery of X-rays by Wilhelm Conrad Röntgen [11].[6] Encouraged by Paul Ewald and in spite of discouragement from Arnold Sommerfeld, the first successful diffraction experiment was undertaken in 1912 by Max Theodor Felix von Laue, assisted by Paul Knipping and Walter Friedrich [12].[7] Inspired by von Laue's results, William Lawrence Bragg, at the age of just 22, developed what is now known as Bragg's law [13], a simple relation between X-ray wavelength, incident angle, and distance between lattice planes. Together with his father, William Henry Bragg, he determined the structure of several alkali halides, zinc blende, and fluorite.[8] In the following few years, many simple structures were determined based on X-ray diffraction, and as the method improved, the structures became more and more complex. The first organic structure determined by X-ray diffraction was that of hexamethylenetetramine [15] and with the structures of penicillin[9] [16] and vitamin B12[10] [17], the relevance of crystal structure determination for medical research became apparent. The first crystal structure of a protein followed just a few years later[11] [18], and since then, crystal structure determination has become one of the most important methods in chemistry, biology, and medicine.

1.3 Symmetry

1.3.1 Symmetry in Two Dimensions

Symmetry is at the heart of all crystallography. There is symmetry in the crystal (also called real space) and symmetry in the diffraction pattern (also called reciprocal space), and sometimes, there is symmetry in individual molecules, which may or may not be reflected by the symmetry group of the crystal structure. An excellent definition of the term symmetry was given by Lipson and Cochran [19]: "A body is said to be symmetrical when it can be divided into

5 This is a wonderful example for constructive collaboration between scientific colleagues. There is a long communication between Fedorov and Schoenflies, which eventually yielded the correct and complete list of all space groups. For a history of the discovery of the 230 space groups. See Ref. [10].
6 In 1901 Röntgen received the Nobel Prize in Physics for this discovery.
7 Nobel Prize in Physics for von Laue in 1914.
8 Nobel Prize in Physics for father and son Bragg in 1915 [14].
9 Dorothy Hodgkin's maiden name was Crowfoot.
10 Nobel Prize in Chemistry for Dorothy Hodgkin in 1964.
11 Nobel Prize in Chemistry for Max Perutz and John Kendrew in 1962.

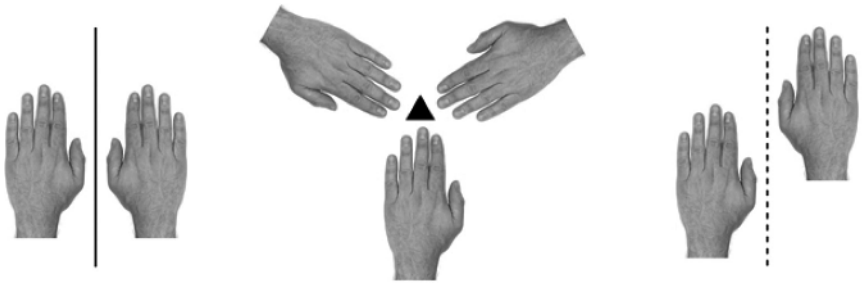


Figure 1.3 Symmetry operations of mirror, threefold rotation, and glide are depicted on a photograph of a hand. The symbol for a mirror is a solid line, for a threefold rotation a triangle (▲), and for a glide a dashed line.

parts that are related to each other in certain ways. The operation of transferring one part to the position of a symmetrically related part is termed a symmetry operation, the result of which is to leave the final state of the body indistinguishable from its original state. In general, successive application of the symmetry operation must ultimately bring the body actually into its original state again." In two dimensions, these are (besides identity) the following symmetry operations: mirror, rotation, and glide (Figure 1.3). Typically, the mirror is the easiest operation to visualize, as most people are familiar with the effect of a mirror. Rotation can be two-, three-, four-, or sixfold in crystallography.[12] The glide operation is somewhat more difficult to grasp. It consists of the combination of two symmetry operations, mirror and translation. In crystallography, glide operations shift one half of a unit cell length (except for the d-glide plane which shifts 1/4 unit cell).

The above describes local symmetry of objects. When adding translation, the following quotation from Lawrence Bragg [20] describes the situation perfectly: "In a two-dimensional design, such as that of a wall-paper, a unit of pattern is repeated at regular intervals. Let us choose some representative point in the unit of pattern, and mark the position of similar points in all the other units. If these points be considered alone, the pattern being for the moment disregarded, it will be seen that they form a regular network. By drawing lines through them, the area can be divided into a series of cells each of which contains a unit of the pattern. It is immaterial which point of the design is chosen as representative, for a similar network of points will always be obtained." To illustrate this, assume the two-dimensional (2D) pattern shown in Figure 1.4. Following the instructions given by Bragg, we can select one point, say, the eye of the light/white bird, and mark it in all light/white birds. The light/white bird's eyes are then the corner points of a 2D regular network, called a lattice. The design

12 That is, in conventional crystallography. Quasicrystals are a different story.

Figure 1.4 Wallpaper design by M. C. Escher. Lattice points are indicated by circles; the lattice is drawn as lines. It does not matter which reference point is chosen; the same lattice is always obtained. There is no symmetry besides translation. The lattice type is oblique and the plane group is p1. Each unit cell contains two birds, one black and one white. *Source:* M.C. Escher's "Symmetry Drawing E47" © 2018 The M.C. Escher Company-The Netherlands. All rights reserved. www.mcescher.com.

can now be shifted freely behind the lattice, and the lattice points will always mark equivalent points in all birds, for example, into the eye of the black bird or, for that matter, anywhere in the design. Those "cells" introduced by Bragg are commonly called *unit cells* in crystallography. The entire design or crystal can be generated by the unit cell and its content simply through translation. One can understand the crystal as built up from unit cells like a wall may be built by bricks. All bricks look the same, and all unit cells forming the crystal are the same.

The unit cell is the smallest motif from which the entire design can be built *by translation alone*; however there often is an even smaller motif that suffices to describe the entire design. This smallest motif is called the *asymmetric unit*, and the symmetry operators of the plane group generate the unit cell from the asymmetric unit. In the design with the black and white birds, there is no symmetry in the unit cell (plane group $p1$), and the asymmetric unit is identical with the unit cell. More commonly, however, one can find symmetry elements in the cell, and the asymmetric unit corresponds to only a fraction of the unit cell (for example, ½, ⅓, or, as in the example below, ⅛).

The design shown in Figure 1.5 contains several symmetry operators, which are drawn in white. Most notably there is a fourfold axis, marked with the symbol ■, but also several mirror planes (solid lines). In addition there are twofold

Figure 1.5 Wallpaper design by M. C. Escher. Assume the grey and white spiders are equivalent and a symmetry operation transforming a grey spider into a white one or vice versa is considered valid. Lattice points are indicated by black circles; the lattice is drawn as black lines. Symmetry elements are drawn in white (fourfold axes, twofold axes, mirrors, and glides). The lattice type is *square* and the plane group is *p4gm*. Each unit cell contains 4 spiders, the asymmetric unit ½ spider. *Source:* M.C. Escher's "Symmetry Drawing E86"

axes (symbol ⧫) and glides (dashed line). The crystal lattice is drawn in black; the lattice type is square, the plane group *p4gm*. Each unit cell contains four bugs, the asymmetric unit ½ bug. Careful examination of Figure 1.5 shows that there are two different kinds of fourfold axes, those on the lattice corners and those in the center of the unit cells. Although those two kinds of fourfold axes are crystallographically equivalent, they are, indeed, different, as one has the bugs

grouped around it in a clockwise arrangement, while the other one shows a counterclockwise arrangement of the bugs.

1.3.2 Symmetry and Translation

Not all symmetry works in crystals or wallpapers. The 2- or 3D periodic object must allow filling the 2- or 3D space without leaving voids. Just as one cannot tile a bathroom with tiles that are shaped like a pentagon or octagon, one cannot form a crystal with unit cells of pentagonal symmetry (Figure 1.6). This means there are no fivefold or eightfold axes in crystallography.[13] Compatible with translation are mirror, glide, twofold, threefold, fourfold, and sixfold rotation.

Combination of all allowed symmetry operations with translation gives rise to 17 possible plane groups in 2D space and 230 possible space groups in 3D space. Each symmetry group falls in one of the seven distinct lattice types (five for 2D space): triclinic (oblique in 2D), monoclinic (rectangular or centered rectangular in 2D), orthorhombic (rectangular or centered rectangular in 2D), tetragonal (square in 2D), trigonal (rhombic in 2D), hexagonal (rhombic in 2D), and cubic (square in 2D).

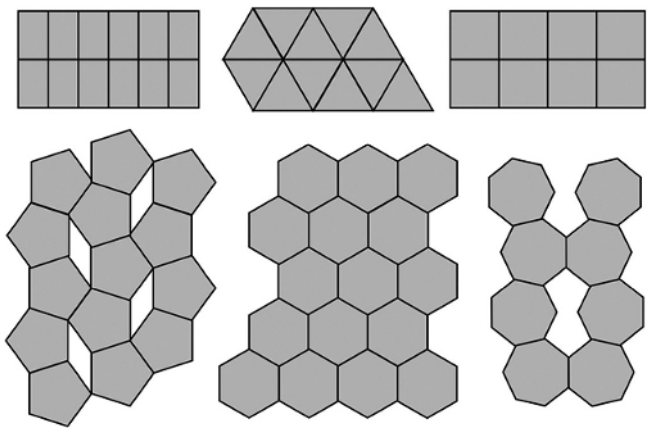


Figure 1.6 In classical crystals (ignoring quasicrystals), only twofold, threefold, fourfold, and sixfold rotation are compatible with translation. Attempts to tile a floor with, for example, pentagons or heptagons will leave gaps.

13 Fivefold and other translational incompatible symmetry can occur within unit cells; however this would always be local symmetry, and a fivefold symmetric object would be understood and treated as asymmetric. Such a symmetry operation is called "pseudo symmetry" or "noncrystallographic symmetry".

1.3.3 Symmetry in Three Dimensions

In 3D space, there are additional symmetry operations to consider, namely, screw axes and the inversion center. Screw axes are like spiral staircases. An object (for example, a molecule) is rotated about an axis and then translated in the direction of the axis. Screw axes are named with two numbers, n_m. The object rotates counterclockwise by an angle of $360^\circ/n$ and shifts up (positive direction) by m/n of a unit cell. For example, a 6_1 screw axis rotates $360^\circ/6 = 60^\circ$ counterclockwise and shifts up 1/6 of a unit cell, a 6_2 screw axis also rotates 60° but shifts up 1/3 of a unit cell. Similarly, a 6_5 screw axis rotates 60° counterclockwise, yet it shifts up 5/6 of a unit cell. In a crystal, there always is another unit cell above and below the current cell, and from any set of coordinates, one can always subtract 1 (or add 1) to any or all of the three coordinates without changing anything. Therefore, shifting up 5/6 of a unit cell is equivalent to shifting down by 1/6. This means that the 6_1 and 6_5 screw axes are mirror images of one another; they form an enantiomeric pair or, in other words, one is right handed, the other one left handed. The same is true for the 6_2 and 6_4 axes, which also form an enantiomeric pair. Figure 1.7 shows 3D models of the five different sixfold screw axes.

Inversion centers can (and should) be understood as a combination of mirror and twofold rotation. Whenever a twofold axis intersects a mirror plane, the point of intersection is an inversion center. Intersection of twofold screw axes with glide planes also creates inversion centers; however the inversion center is not located at the point of intersection. Like all symmetry operations involving a mirror operation, inversion centers change the hand of a chiral molecule.

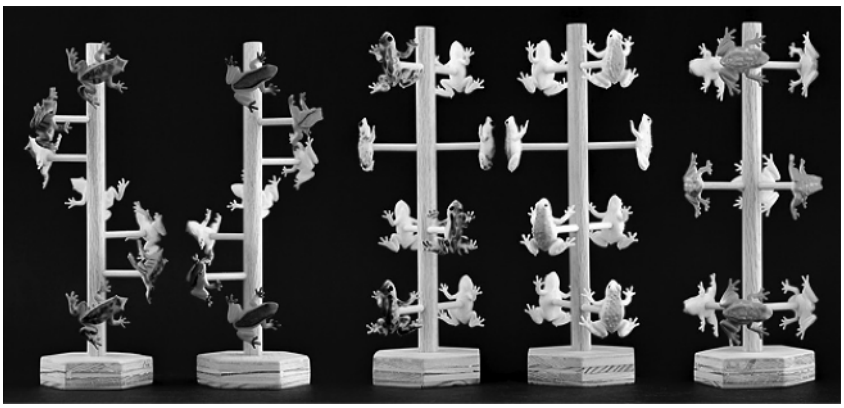


Figure 1.7 Models of all five sixfold screw axes (built by Ellen and Peter Müller in 2010). From left to right: 6_1, 6_5, 6_2, 6_4, 6_3. It can be seen that $6_1/6_5$ and $6_2/6_4$ are enantiomeric pairs, i.e. mirror images of one another or, in other words, the right- and left-handed versions of the same screw.

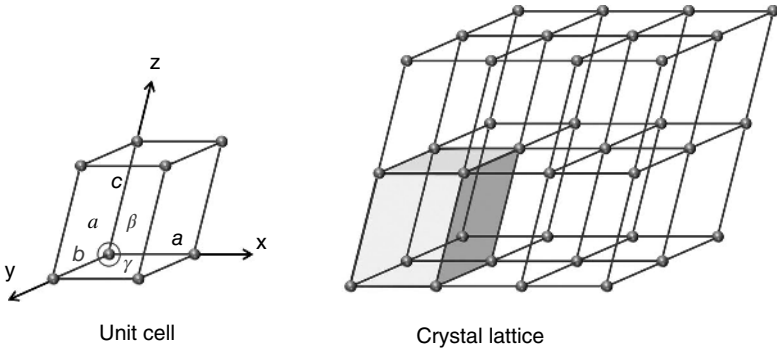


Figure 1.8 Unit cell, defined by lattice vectors (*a*, *b*, *c*) and angles (α, β, γ), the basic building block used to construct the three-dimensional crystal lattice.

In addition, mirror and glide, which are mere lines in two dimensions, become mirror planes and glide planes in 3D space. Glide planes are similar to the glide operation in two dimensions. The only difference is that the glide can be in one of several directions. Assume the mirror operation to take place on the *a*-*c*-plane. The mirror image can now shift in the *a*- or the *c*-direction or even along the diagonal in the *a*-*c*-plane. The first case is called an *a*-glide plane, the second one a *c*-glide plane, and the third case is called an *n*-glide plane.

One possible definition of a crystal is this: A crystal is a 3D periodic[14] discontinuum formed by atoms, ions, or molecules. It consists of identical "bricks" called unit cells, which form a 3D lattice (Figure 1.8). The unit cell is defined by axes *a*, *b*, *c*, and angles α, β, γ, which form a right-handed system. As described above, the unit cell is the smallest motif that can generate the entire crystal structure only by means of translation in three dimensions. Except for space group *P*1, the unit cell can be broken down into several symmetry-related copies of the asymmetric unit. The symmetry relating the individual asymmetric units is described in the space group. Typically, the asymmetric unit contains one molecule; however it is possible (and occurs regularly) for the asymmetric unit to contain two or more *crystallographically independent* molecules or just a fraction of a molecule.

1.3.4 Metric Symmetry of the Crystal Lattice

The metric symmetry is the symmetry of the crystal lattice without taking into account the arrangement of the atoms in the unit cell. Each of the 230 space groups is a member of one of the 7 *crystal systems*, which are defined by the

14 Again, this holds only for classical crystals. In quasicrystals strict periodicity is not observed.

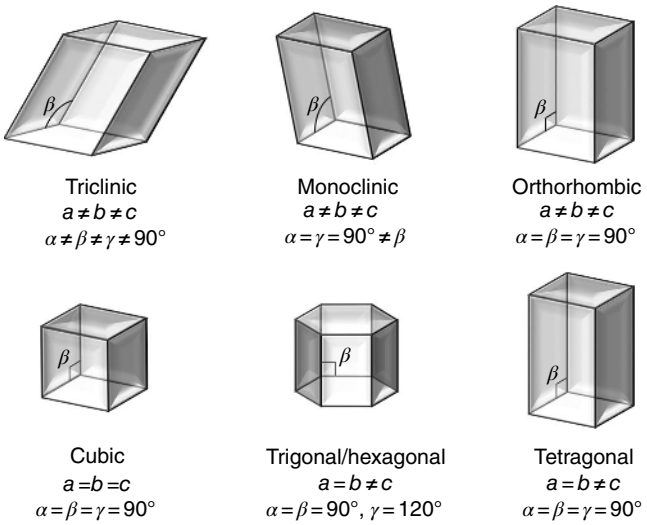


Figure 1.9 Seven crystal systems, defined by the shape of the unit cell. (Trigonal and hexagonal have the same metric symmetry, but are separate crystal systems.)

shape of the unit cell (Figure 1.9). We distinguish the triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, and cubic crystal systems.[15]

As will be shown below, the shape and size of the unit cell, its *metric symmetry*, in real space determines the location of the reflections in the diffraction pattern in reciprocal space. Considering the metric symmetry of the unit cell alone, ignoring the unit cell contents (that is, the atomic positions), is equivalent to looking at the positions of the reflections alone without taking into account their relative intensities. That means it is the relative intensities of the diffraction spots that hold the information about the atomic coordinates and, hence, the actual crystal structure. More about that later.

1.3.5 Conventions and Symbols

As mentioned above, the unit cell forms a right-handed system $a, b, c, \alpha, \beta, \gamma$. In the triclinic system, the axes are chosen so that $a \le b \le c$. In the monoclinic system the one non-90° angle is β and the unit cell setting is chosen so that $\beta \ge 90°$. If there are two possible settings with $\beta \ge 90°$, that setting is preferred where β is closer to 90°. In the monoclinic system b is the *unique axis*, while in the

15 Some crystallographers count rhombohedral as a separate crystal system; however it usually is understood as a special case of the trigonal system (*R*-centering). It should also be noted that trigonal and hexagonal are considered different crystal systems even though they have the same metric symmetry.

tetragonal, trigonal, and hexagonal systems, *c* is unique. If a structure is centrosymmetric, the origin of the unit cell is chosen so that it coincides with an inversion center. In noncentrosymmetric space groups, the origin conforms with other symmetry elements (for details see Volume A of the International Tables for Crystallography) [21].

1.3.6 Fractional Coordinates

In crystallography, atomic coordinates are given as fractions of the unit cell axes. All atoms inside the unit cell have coordinates $0 \le x < 1$, $0 \le y < 1$, and $0 \le z < 1$. That means that, except for the cubic crystal system, the coordinate system in which atomic positions are specified is not Cartesian. An atom in the origin of the unit cell has coordinates 0, 0, 0, an atom located exactly in the center of the unit cell has coordinates 0.5, 0.5, 0.5, and an atom in the center of the *a*-*b*-plane has coordinates 0.5, 0.5, 0, etc. When calculating interatomic distances, one must multiply the differences of atomic coordinates individually with the lengths of the corresponding unit cell axes. Thus, the distance between two atoms x_1, y_1, z_1 and x_2, y_2, z_2 is

$$d = \sqrt{((x_2 - x_1)a)^2 + ((y_2 - y_1)b)^2 + ((z_2 - z_1)c)^2} = \sqrt{(\Delta xa)^2 + (\Delta yb)^2 + (\Delta zc)^2}$$

Note that this equation is valid only in orthogonal crystal systems (all three angles 90°), that is, orthorhombic, tetragonal, and cubic. For the triclinic case the formula is

$$d = \sqrt{(\Delta xa)^2 + (\Delta yb)^2 + (\Delta zc)^2 - 2\Delta x\Delta yab\cos\gamma - 2\Delta x\Delta zac\cos\beta - 2\Delta y\Delta zbc\cos\alpha}$$

The *x*, *y*, *z* notation is also used to describe symmetry operations. If there is an atom at the site *x*, *y*, *z*, then $x + 1$, *y*, *z* is the equivalent atom in the next unit cell in *x*-direction (*a*-cell axis), and coordinates $-x$, $-y$, $-z$ are generated from *x*, *y*, *z*, by an inversion center at the origin (that is, at coordinates 0, 0, 0). In the same fashion, a twofold rotation axis coinciding with the unit cell's *b*-axis (as, for example, in space group *P*2) generates an atom $-x$, *y*, $-z$ from every atom *x*, *y*, *z*, and a twofold screw axis coinciding with *b* (say, in space group $P2_1$) generates $-x$, $y + ½$, $-z$ from *x*, *y*, *z*.

1.3.7 Symmetry in Reciprocal Space

The symmetry of the diffraction pattern (reciprocal space) is dictated by the symmetry in the crystal (real space). The reciprocal symmetry groups are called *Laue groups*. If there is, for example, a fourfold axis in real space, the diffraction space will have fourfold symmetry as well. Lattice centering and other translational components of symmetry operators have no impact on the Laue group, which means that symmetry in reciprocal space does not distinguish between,

Table 1.2 Laue and point groups of all crystal systems.

Crystal system	Laue group	Point group
Triclinic	$\bar{1}$	1, $\bar{1}$
Monoclinic	$2/m$	2, m, $2/m$
Orthorhombic	mmm	222, $mm2$, mmm
Tetragonal	$4/m$	4, $\bar{4}$, $4/m$
	$4/mmm$	422, $4mm$, $\bar{4}2m$, $4/mmm$
Trigonal/rhombohedral	$\bar{3}$	3, $\bar{3}$
	$\bar{3}/m$	32, $3m$, $\bar{3}m$
Hexagonal	$6/m$	6, $\bar{6}$, $6/m$
	$6/mmm$	622, $6mm$, $\bar{6}m2$, $6/mmm$
Cubic	$m3$	23, $m3$
	$m3m$	432, $\bar{4}3m$, $m3m$

for example, a sixfold rotation and a 6_1-, 6_2-, or any other sixfold screw axis. In addition, reciprocal space is, at least in good approximation, centrosymmetric, which means that all Laue groups are centrosymmetric even if the corresponding space group is chiral.

The Laue group can be determined from the space group via the *point group*.[16] The point group corresponds to the space group minus all translational aspects (that is, glide planes become mirror planes, screw axes become regular rotational axes, and the lattice symbol is lost). The Laue group is the point group plus an inversion center, as reciprocal space is centrosymmetric. If the point group is already centrosymmetric, then Laue group and point group are the same. Take, for example, the three monoclinic space groups $P2_1$ (chiral), Pc (noncentrosymmetric), and $C2/c$ (centrosymmetric). While those three space groups have different point groups, they all belong to the same (only) monoclinic Laue group (Table 1.2).

Space group	Point group	Laue group
$P2_1$	2	$2/m$
Pc	m	$2/m$
$C2/c$	$2/m$	$2/m$

It is important to note that the symmetry of the Laue group can be lower than the metric symmetry of the crystal system but never higher. That means that, for

16 The point group is also called the crystal class (not to be confused with crystal *system*).

example, a monoclinic crystal could, by mere chance, have a β angle of exactly 90° and, thus, display orthorhombic metric symmetry. When considering the unit cell contents, however, and when examining the symmetry of the diffraction pattern, the symmetry in both real and reciprocal space would still be monoclinic, and, hence, the metric symmetry would be higher than the Laue symmetry.[17]

1.4 Principles of X-ray Diffraction

In a diffraction experiment, the X-ray beam interacts with the crystal, giving rise to the diffraction pattern. Diffraction can easily be demonstrated by shining a beam of light through a fine mesh. For example, one can look through a layer of sheer curtain fabric into the light of a streetlamp (Figure 1.10). The phenomenon is always observed when waves of any kind meet with an obstacle, for example, a mesh or a crystal; however the effect is particularly strong when the wavelength is comparable with the size of the obstacle (the mesh size or the size of the unit cell in a crystal).

1.4.1 Bragg's Law

One way of understanding diffraction is through a geometric construction that describes the reflection of a beam of light on a set of parallel and equidistant planes (Figure 1.11). The planes can be understood as the lattice planes in a

Figure 1.10 View of streetlamps from a hotel room in Chicago in 2010. The image on the right side is the exact same view as the one on the left; only it was taken through the curtain fabric. All strong and point-like light sources show significant diffraction.

17 This occurs occasionally and is prerequisite for merohedral and pseudo-merohedral twinning.

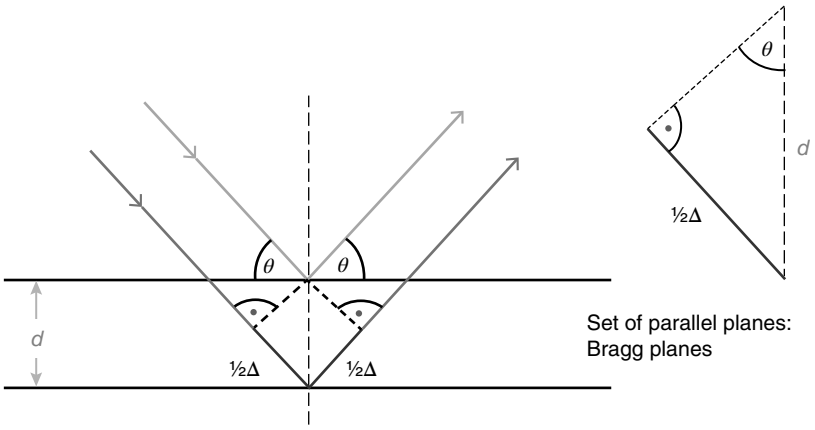


Figure 1.11 Bragg's law derived from partial reflection of two parallel planes.

crystal, the light as the X-ray beam. The beam travels into the crystal, is partially reflected on the first plane, continues to travel until being partially reflected on the second plane, and so forth. Only two planes are necessary to understand the principle. Simple trigonometry leads to an equation that relates the wavelength λ to the distance d between the lattice planes and the angle θ of diffraction:

$$\sin\theta = \frac{1/2\Delta}{d} = \frac{\Delta}{2d}$$

It is apparent that constructive interference is only observed if the path difference is the same as the wavelength of the diffracted light (or an integer multiple thereof). That means $\Delta = n\lambda$, and hence

$$n\lambda = 2d\sin\theta$$

This equation is also known as Bragg's law, and the parallel planes of the crystal lattice are called Bragg planes.

When Bragg's law is resolved for d, one can easily calculate the maximum resolution to which diffraction can be observed as a function of the wavelength used:

$$d = \frac{\lambda}{2\sin\theta}$$

The maximum resolution corresponds to the smallest value for d, which is achieved for the largest possible value of sin θ.[18]

18 The highest value the sin can ever have is 1. This corresponds to an angle of $\theta = 90°$.

$$d_{\min} = \frac{\lambda}{2\sin\theta_{\max}} = \frac{\lambda}{2}$$

Therefore the maximum theoretically observable resolution is half the wavelength of the radiation used. Practically, this resolution can never be observed, as it would require the detector to coincide with the X-ray source; however modern diffractometers get as close as ca. $d_{\min} = 0.52\,\lambda$. The two most commonly used X-ray wavelengths are Cu K_α, (λ = 1.54178 Å) and Mo K_α, (λ = 0.71073 Å). The respective practically achievable maximum resolutions are 0.80 Å for Cu and 0.37 Å for Mo radiation. As will be seen below, most crystals do not diffract to such high resolution as one could observe with Mo radiation, and some crystals will not even diffract to the 0.84 Å resolution recommended as a minimum by the International Union of Crystallography (IUCr).

1.4.2 Diffraction Geometry

Bragg planes can be drawn into the crystal lattice through the lattice points. The planes are characterized by their angle relative to the unit cell and by their spacing d, and each set of equidistant planes can be uniquely identified by a set of three numbers describing at which point they intersect the three basis vectors of the crystal lattice (i.e. the unit cell axes) closest to the origin (Figure 1.12). Those numbers are called the Miller indices h, k, and l and correspond to the reciprocal values of the intersection with the unit cell. Each set of Bragg planes gives rise to one pair of reflections in reciprocal space, which are uniquely identifiable by the corresponding Miller indices h, k, l and $-h$, $-k$, $-l$. Higher values for h, k, l correspond to smaller distances between corresponding Bragg planes, larger distances between lattice points on the planes, and higher resolution of the corresponding reflection. For each interplanar distance vector d_{hkl}, there is a scattering vector s_{hkl} with $s = 1/d$.

1.4.3 Ewald Construction

Paul Ewald described Bragg's law geometrically, and it is his construction (Figure 1.13) that most crystallographers see in front of their inner eye when they think about a diffraction experiment. The core of the construction is a sphere with radius $1/\lambda$, and the X-ray beam of wavelength λ intersects the sphere along its diameter. The crystal and hence the origin of real space are located in the center of the sphere (point C), while the origin of the reciprocal lattice (point O) is located at the exit point of the X-ray beam. The scattering vector s is drawn as footing in point O. For each set of Bragg planes with spacing d, there is one s-vector with length $1/d$ and direction perpendicular to the planes. If the crystal were represented by the s-vectors, it would be reminiscent of a sea urchin with spines of different lengths, each spine corresponding to one s-vector. Rotation

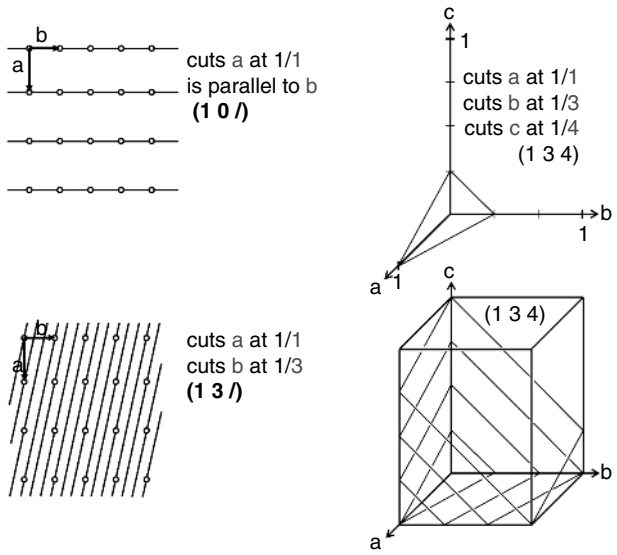


Figure 1.12 Between the points of a crystal lattice in real space, there are Bragg planes. Each set of Bragg planes corresponds to one set of Miller indices. The Miller indices *h*, *k*, *l* correspond to the reciprocal values of the points at which the planes cut the unit cell axes closest to the origin. Each set of Bragg planes corresponds to one reflection. Each reflection is identified by the corresponding Miller indices *h*, *k*, *l*. The positions of the reflections form another lattice, the reciprocal lattice. There is a vector *d* perpendicular to each set of Bragg planes; its length is equivalent to the distance between the corresponding Bragg planes. Each reflection *h*, *k*, *l* marks the endpoint of the scattering vector $s = 1/d$. The length of *s* is inversely related to the distance between the Bragg planes.

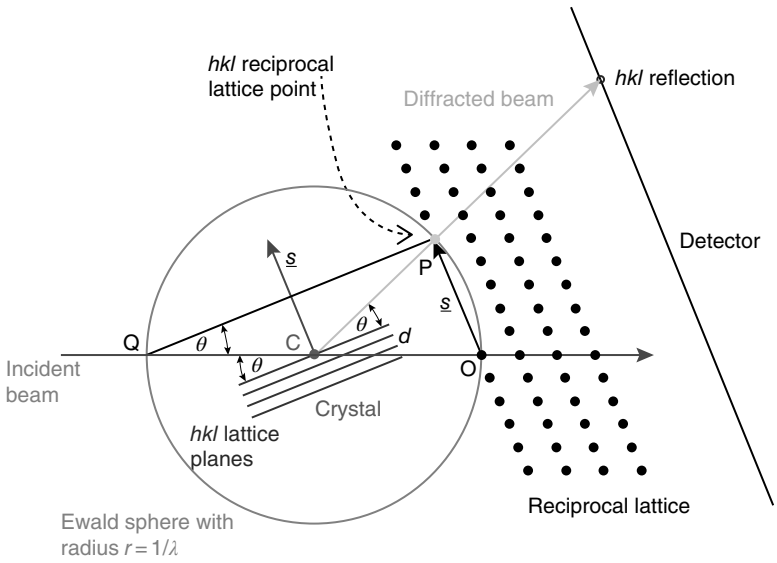


Figure 1.13 Ewald construction. The Ewald sphere has the radius $1/\lambda$. Points C, O, P, and Q mark the position of the crystal, the origin of the reciprocal lattice, the point where the diffracted beam exits the Ewald sphere (corresponding to the endpoint of *s* on the surface of the sphere), and the point where the primary beam enters the Ewald sphere, respectively. Through rotation of the crystal, all *s*-vectors that are shorter than $2/\lambda$ can be brought into a position in which they end on the surface of the Ewald sphere.

of the crystal corresponds to the rotation of the sea urchin located in point O. Depending on crystal orientation, the various *s*-vectors will, at one time or other, be ending on the surface of the Ewald sphere. It can be demonstrated that Bragg's law is fulfilled exactly for those *s*-vectors that end on the Ewald sphere.[19] That means, for each crystal orientation, those and only those reflections can be observed as projections onto a detector whose *s*-vectors end on the surface of the Ewald sphere.

1.4.4 Structure Factors

With the help of Bragg's law and the Ewald construction, we can calculate the place of a reflection on the detector, provided we know the unit cell dimensions. Indeed, the position of a spot is determined alone by the metric symmetry of the unit cell (and the orientation of the crystal on the diffractometer). The relative intensity[20] of a reflection, however, depends on the contents of the unit cell, i.e. on the population of the corresponding set of Bragg planes with electron density. If there are many atoms on a plane, the corresponding reflection is strong; if the plane is empty, the reflection is weak or absent.[21] Whether or not there are many atoms on a specific set of Bragg planes in a given unit cell depends on the shape, location, and orientation of the molecule(s) inside the unit cell. Every single atom in the unit cell is positioned in some specific way relative to every set of Bragg planes. The closer an atom is to one of the planes of a specific set and the more electrons this atom has, the more it contributes constructively to the corresponding reflection. Therefore, every single atom in a structure has a contribution to the intensity of every reflection depending on its chemical nature and on its position in the unit cell.

Two other factors influencing the intensity of observed reflections are the thermal motion of the atoms (temperature factor) and the atomic radius (form factor). Only if atoms were mathematical points could they fully reside on a

19 Since the triangle OPQ is a right triangle and since $\sin\alpha = \frac{\text{adjacent side}}{\text{hypotenuse}}$ and the diameter of the Ewald sphere is $2/\lambda$, it follows that $\sin\theta = \frac{s}{2\lambda}$. Since $s = 1/d$, it follows that $2d \sin\theta = \lambda$, which is Bragg's law.

20 The absolute intensity also depends on many other factors such as exposure time, crystal size, beam intensity, detector sensitivity, etc.

21 It is slightly more complicated than that, as "destructive interference" *alone* leads to observable intensity as well (interference is only destructive if there is something to be destroyed…). That means if the Bragg planes for a specific reflections are empty but many atoms can be found exactly halfway between the Bragg planes, the reflection will be just as strong as if the atoms were all on the planes instead of halfway in between. This can be understood when one realizes that the exact position (not orientation or spacing!) of the Bragg planes depends on the origin of the unit cell, which is established merely by conventions. If, in this example, the unit cell origin were to be shifted so that the Bragg planes moved in such a fashion to coincide with the atoms, thus vacating the space between the planes, all electron density would reside on the planes and not in between, yet the structure would remain unchanged.

Bragg plane. Yet because they have an appreciable size and, in addition, vibration, an atom residing perfectly on a Bragg plane will have electron density also above and below the plane. This density above and below will contribute somewhat destructively to the corresponding reflection, depending on the motion and size of the atoms and on the resolution of the reflection in question. As explained above, the distance d between Bragg planes is smaller for higher resolution reflections. That means that at higher resolution, the electron density above and below the Bragg planes will extend closer to the center between the planes and, hence, weaken the corresponding reflection more strongly than it would for a lower resolution reflection with a larger d. When d becomes small enough that atomic motion will lead to so much electron density between the planes and that perfect destructive interference is achieved, no reflections beyond this resolution limit will be observed. This is a crystal-specific resolution limit, and crystals in which the atoms move more than average will diffract to lower resolution than crystals with atoms that move less. This circumstance also explains why low-temperature data collection leads to higher resolution datasets, as at lower temperatures atomic motion is significantly reduced.

Strictly speaking, "reflections" should be called "structure factor amplitudes." Every set of Bragg planes gives rise to a structure factor $\vec{F}$, and the observed reflection is the structure factor amplitude $|F|^2$.[22] The structure factor equation describes the contribution of every atom in a structure to the intensity of every reflection:

$$F_{hkl} = \sum_i f_i [\cos 2\pi(hx_i + ky_i + lz_i) + i \sin 2\pi(hx_i + ky_i + lz_i)]$$

The structure factor F for the set of Bragg planes specified by Miller indices h, k, l is the sum over the contributions of all atoms i with their respective atomic scattering factors f_i and their coordinates x_i, y_i, z_i inside the unit cell. Note that the i in $i \sin 2\pi$ is $\sqrt{-1}$ and not the same i as the one in f_i or x_i, y_i, z_i. Temperature factor and form factor are, together with electron count, contained in the values of f_i for each atom.[23]

1.4.5 Statistical Intensity Distribution

In a diffraction experiment, we measure intensities. As described above, the intensities correspond to the structure factor amplitudes (after application of corrections, such as Lorenz and polarization correction and scaling and a few other minor correction terms). It turns out that the variance of the intensity

22 Structure factors are vectors in a complex plane. They have intensity and a phase angle.

23 That means that the value of f_i is a function of scattering angle θ and, hence, the resolution of the reflection h,k,l.

distribution across the entire dataset is indicative of the presence or absence of an inversion center in real space (remember: in good approximation reciprocal space is always centrosymmetric). This variance is called the $|E^2 - 1|$-statistic, which is based on normalized structure factors E. To calculate this statistic, all structure factors are normalized in individual thin resolution shells. In this context, normalized means every squared structure factor F^2 of a certain resolution shell is divided by the average value of all structure factors in this shell: $E^2 = F^2/<F^2>$ with E^2, squared normalized structure factor; F^2, squared structure factor; and $<F^2>$, mean value of squared structure factors for reflections at same resolution.

The average value of all squared normalized structure factors is one, $<E^2> = 1$; however $< | E^2 - 1 | > = 0.736$ for noncentrosymmetric structures and 0.968 for centrosymmetric structures.

Heavy atoms on special positions and twinning tend to lower this value, and pseudotranslational symmetry tends to increase it. Nevertheless, the value of this statistic can help to distinguish between centrosymmetric and noncentrosymmetric space groups.

1.4.6 Data Collection

An excellent introduction to data collection strategy is given by Dauter [22]. In general, there are at least five qualifiers describing the quality of a dataset: (i) maximum resolution; (ii) completeness; (iii) multiplicity of observations (MoO[24], sometimes called *redundancy*); (iv) I/σ, i.e. the average intensity divided by the noise; and (v) a variety of merging residual values, such as R_{int} or R_{sigma}. A good dataset extends to high resolution the International Union of Crystallography (IUCr) suggests at least 0.84 Å, but with modern equipment 0.70 Å or even better can usually be achieved without much effort)[25] and is complete (at least 97% is recommended by the IUCr, yet in most cases 99% or even 100% completeness can and should be obtained). The MoO should be as high as possible (a value of 5–7 should be considered a minimum), and "good data" have I/σ values of at least 8–10 for all data. As usual with residual values, the merging R-values should be as low as possible, and most small-molecule datasets have R_{int} (also called R_{merge}) and/or R_{sigma} values below 0.1 (corresponding to 10%)

24 "This term was defined at the SHELX Workshop in Göttingen in September 2003 to distinguish the MoO from redundancy or multiplicity, with which the MoO has been frequently confused in the past. In contrast to redundancy, which is repeated recording of the same reflection obtained from the same crystal orientation (performing scans that rotate the crystal by more than 360°), MoO, sometimes also referred to as "true redundancy," describes multiple measurements of the same (or a symmetry equivalent) reflection obtained from different crystal orientations (i.e. measured at different Ψ-angles)" [23].

25 Note that resolution describes the smallest distance that can be resolved. Therefore, smaller numbers mean higher resolution, and 0.70 Å is a much higher resolution than 0.84 Å.

for the whole resolution range. In general, diffraction data should be collected at low temperature (100 K is an established standard). Atomic movement is significantly reduced at low temperatures, which increases resolution and I/σ of the diffraction data and increases order in the crystal.

1.5 Structure Determination

The final goal of the diffraction experiment is usually the determination of the crystal structure, which means the establishment of a crystallographic model. This model consists of x, y, and z coordinates and thermal parameters for every atom in the asymmetric unit as well as a few other global parameters. After data collection and data reduction, the steps described in the following paragraphs lead to this model, which is commonly referred to as *the crystal structure*. In this context it is worth pointing out that a crystal structure is not only the temporal average, averaged over the entire data collection time, but also always the spatial average over the whole crystal. That means the crystal structure shows what the molecules making up the crystal look like on average. Crystal structure determination is, therefore, not an ideal tool for looking at molecular dynamics or single molecules. Real crystals are neither static nor perfect, and atoms can be misplaced (packing defects or disorders) in some unit cells. On the other hand, it is easy to derive information about interactions between the individual molecules in a crystal. Through application of space group symmetry and lattice translation, packing diagrams reveal the positioning of all atoms within a portion of the crystal larger than the asymmetric unit or unit cell, and interactions of neighboring molecules or ions become readily apparent.

1.5.1 Space Group Determination

The first step in crystal structure elucidation is typically the determination of the space group. The metric symmetry is a good starting point; however, considering that the true crystal symmetry could be lower than the metric symmetry, it is important to determine the Laue group based on the actual symmetry of the diffraction pattern, i.e. in reciprocal space. Having determined the Laue symmetry, the number of possible space groups is significantly reduced. The value of the $|E^2-1|$-statistic allows reducing the number of space group further by establishing at least a trend toward centrosymmetric or noncentrosymmetric symmetry.

Finally, there are systematic absences that point out specific symmetry elements present in the crystal. While, as described above, lattice type and other translational components of the space group have no influence on the corresponding Laue group, those symmetry operations do leave their traces in

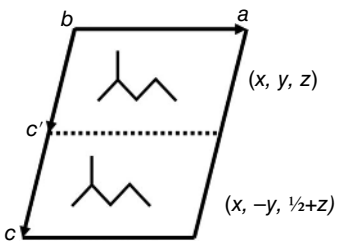


Figure 1.14 Projection of a unit cell along the crystallographic *b*-axis (i.e. in [*h*, 0, *l*] projection) in presence of a *c*-glide plane coinciding with the *a*-*c*-plane. In this projection the unit cell seems to be cut in half which, in turn, doubles the volume of the corresponding reciprocal unit cell. Reflections corresponding to this projection will be according to the larger reciprocal cell, which means that reflections of the class *h* 0 *l* with $l \neq 2n$ are not observed, i.e. systematically absent.

reciprocal space in the form of systematic absences. Assume, for example, a *c*-glide plane in the space group *Pc*. Figure 1.14 shows the unit cell in projection along the *b*-axis, i.e. onto the *a*-*c*-plane. For every atom *x*, *y*, *z*, the *c*-glide plane at $y = 0$ generates a symmetry-related atom x, $-y$, $z + ½$. In this specific 2D projection, the molecule is repeated at $c/2$, and the unit cell seems to be half the size ($c' = c/2$) because one cannot distinguish the height of the atoms above or below the *a*-*c*-plane when looking straight at that plane. This doubles the apparent reciprocal cell in this specific projection $h\ 0\ l : c^{*\prime} = 2c^*$. Therefore, the reflections corresponding to this projection will be according to the larger reciprocal cell, which means that reflections of the class *h* 0 *l* with $l \neq 2n$ (that is, reflections with odd values for *l*) are not observed or, in other words, systematically absent. Similar considerations can be made for all screw axes and glide planes as well as for lattice centering.

Combination of all these considerations can narrow the choice of space groups down to just a few possibilities to be considered and sometimes even to just one possible space group. Knowing the space group means knowing all symmetry in real space. This knowledge can help to solve the phase problem.

1.5.2 Phase Problem and Structure Solution

Crystals are periodic objects, which means that each unit cell has the same content in the same orientation as every other unit cell. Molecules inside the unit cell consist of atoms, and atoms, simply put, consist of nuclei and electrons. X-rays interact with the electrons of the atoms, not the nuclei, and – at least from the perspective of an X-ray photon – an atom can be described as a more or less localized cloud of electron density. Therefore, to the X-rays, the unit cell looks like a 3D space of variable electron density, higher electron density at the atom sites, and low electron density between atoms. Jean-Baptiste Joseph Fourier stated that any periodic function can be approximated through superposition of sufficiently many sine waves of appropriate wavelength, amplitude, and phase. The example in Figure 1.15 is taken with permission from Kevin

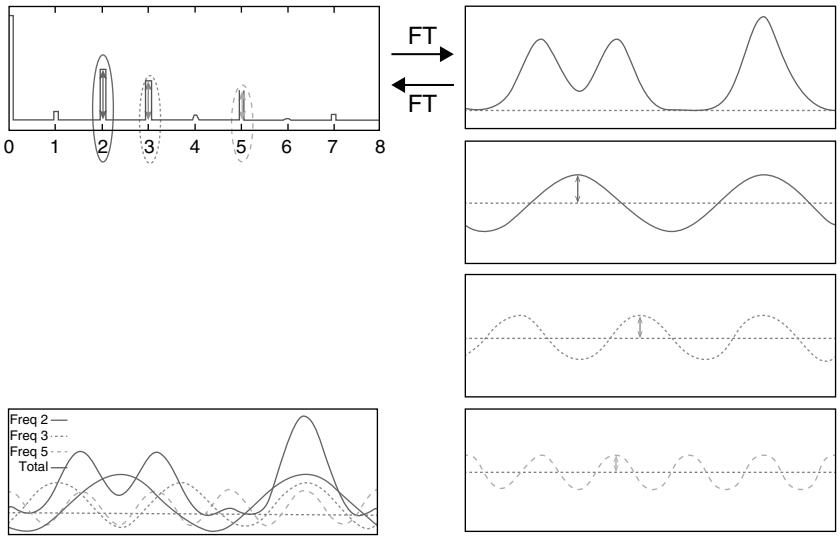


Figure 1.15 Electron density of a hypothetical one-dimensional crystal with a three-atomic molecule in the unit cell (top right). This density function can be represented fairly well in terms of just three sine waves: The first sine wave has a frequency of 2 (i.e. there are two repeats of the wave across the unit cell); its phase is chosen that one maximum is aligned with the two lighter atoms on the left of the unit cell and the other one is with the heavier atom on the right. The second one has a frequency of 3; it has a different amplitude and also a different phase (one maximum is aligned with the heavier atom on the right of the unit cell). The third sine wave with a frequency of 5 also has a different amplitude, and its phase is chosen so that two of this wave's peaks are lined up with the two lighter atoms to the left of the unit cell. Adding up the three sine waves results in the thick curve at the bottom left of the figure. These sine waves are the "electron density waves" mentioned in the text above, and the frequencies of 2, 3, or 5 correspond to the "electron density wavelengths." The top left of the figure shows the Fourier transformation of the unit cell, corresponding to the diffraction pattern, together with the one-dimensional Miller indices. The three sine waves can be identified as the three strongest reflections. The intensities of the reflections correspond to the amplitudes of the sine waves in the right-hand side of the figure, and the frequencies of the sine waves correspond to the respective Miller indices (2, 3, and 5). Unfortunately, the phases are not encoded in the diffraction pattern. *Source:* Reproduced with permission of Kevin Cowtan's Book of Fourier. http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

Cowtan's online *Book of Fourier*[26] and illustrates how a one-dimensional (1D) electron density function can be represented reasonably well by three sine waves, assuming the amplitudes and phases are chosen correctly. The wavelengths of those sine waves used are all in integer fractions of the unit cell length in accordance with the Miller indices of the corresponding reflections. These

26 http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html

wavelengths are referred to as "electron density wavelengths"[27] and have nothing to do with the wavelength of the x-radiation used in the diffraction experiment.

All reflections together form the diffraction pattern, which can be understood as the Fourier transform of the 3D electron density function in the crystal. That means that a Fourier transformation allows going from one space to the other and back. Every independent reflection in the diffraction pattern is one Fourier coefficient. As we saw above, structure factors F are vectors in the complex plane[28] with amplitude and phase angle. The amplitude of the reflection corresponds to the amplitude of the electron density wave, its Miller indices correspond to the electron density wavelength, and the reflection's phase corresponds to the phase in the Fourier summation. In order to perform a Fourier summation as in Figure 1.15, all three properties of the structure factors are needed: amplitude, phase, and frequency (that is, the reciprocal of the electron density wavelength). The structure factor amplitudes are measured as intensities in the diffraction experiment, the reflections' Miller indices directly lead to the frequencies, yet, unfortunately, the phase angles cannot be determined in a standard diffraction experiment. This unlucky circumstance is typically referred to as the "crystallographic phase problem," and it has to be solved individually for every crystal structure. Assigning a tentative and sometimes only approximate phase to the structure factors is called "solving the structure" or "phasing the structure," as together with amplitude and frequency, knowledge of the phases (even if only approximate) affords an electron density map in which atoms may be located. There are several methods for solving structures; two of which will be described here, the Patterson function and direct methods.

The Patterson function goes back to Arthur Lindo Patterson who discovered that a convolution of reciprocal space (that is, a Fourier transformation of the measured intensities in the diffraction pattern without phases) gives rise to a 3D map, the Patterson map [24]. This map is not unlike an electron density map; however the maxima in the Patterson map do not correspond directly to atoms but rather they represent interatomic distances. The distance of a Patterson peak from the origin of the Patterson map corresponds to the distance between two atoms in the crystal structure. Therefore, for every peak with coordinates u, v, w in the Patterson map, there is a pair of atoms in the unit cell that reside on coordinates x_1, y_1, z_1 and x_2, y_2, z_2, such that $u = x_2\text{-}x_1$, $v = y_2\text{-}y_1$, and $w = z_2\text{-}z_1$. The height of a Patterson peak corresponds to the number of electrons involved in this interatomic distance. That means that distances between heavier atoms (which have more electrons than light ones) will result in stronger Patterson peaks than distances between lighter atoms. The Patterson map is typically fairly noisy, and Patterson peaks tend to be fuzzy and overlap with one another.

27 This term was introduced by Jenny Glusker.

28 Also called the Argand plane.

Therefore, it is difficult (though not impossible) to extract information about light atoms from the Patterson map, and in the absence of any heavier atoms, Patterson methods can fail. Knowledge of the symmetry in reciprocal space (Laue group) and real space (space group) allows deriving coordinates for the heaviest atom or atoms in a crystal structure. Based on those coordinates, phases for all reflections can be calculated as if the structure consisted only of those heavy atoms. In by far most cases, the structure has more atoms than just the heavy ones located by the Patterson method, and, therefore, the first set of phases is only approximate. Nevertheless, the phasing power of just a few heavy atoms is usually sufficient to locate more atoms in the electron density map calculated from the measured structure factor amplitudes and the approximate phases. Including the newly found atoms into the crystallographic model gives rise to better phases and, therefore, a clearer electron density map, which will show more features than the one before. At this stage of structure determination, we are no longer solving the structure but already refining it (see below).

Direct methods are based on probabilistic relationships between specific groups of structure factors and their phases. The foundations of classical direct methods are a few simple and sensible assumptions, most importantly (i) that electron density is never negative and (ii) that a structure consists of discrete atoms resolved from one another. The first assumption gives rise to a set of phase relationships, the Harker–Kasper inequalities, which allow assigning phases to some select strong reflections. The second assumption leads to the finding that the squared electron density function is similar to the electron density itself times a scaling factor (Sayre equation). Derived from the Sayre equation is the triplet phase relation, which states that the sum of the three phases of three strong structure factors is approximately zero if the three structure factors in question are related to one another in such a fashion that three values for the h, the three values for k, and the three values for l all add up to zero (h, k, and l are the Miller indices of the reflections in question). An excellent introduction to direct methods can be found in Chapter 8 of the book *Crystal Structure Analysis A Primer* [25].

Since direct methods assume that atoms are discrete and resolved from one another, comparatively high resolution of the diffraction data is required for those methods to work (ca. 1.1 Å as a practical minimum requirement). Luckily, most small-molecule crystals easily diffract to this limit.

1.5.3 Structure Refinement

"Refinement is the process of iterative alteration of the molecular model with the goal to maximize its compliance with the diffraction data" [26]. The term

structure refinement describes everything that leads from the initial structure solution to the complete, publishable crystal structure. With the phase problem solved, a Fourier synthesis using all diffraction spots as Fourier coefficients and the freshly determined phases gives rise to a first electron density map. This type of map is called the F_o-map, where F_o stands for the observed structure factors. The F_o map gives the electron density at any given point inside the unit cell, and it shows maxima where the atoms are located. The height of the individual maxima is proportional to the number of electrons of the corresponding atom. Naively put, a high-electron density peak is a heavy atom, a weaker peak is a light atom, and a very weak peak is usually no atom at all but noise. The initial map is often noisy, and sometimes only the heavier atoms can be located with confidence; however one can calculate a new set of phases from the so-determined substructure. This new phase set is usually better than the initial phases, and a new Fourier transformation gives rise to a new and better F_o map. At this point, a second type of electron density map is calculated, the so-called difference map or F_o–F_c map. F_c stands for the structure factors calculated from the existing model, and, therefore, the F_c map corresponds to the electron density distribution as described by the current model. The difference map is calculated by subtracting the F_c map from the F_o map (hence the name F_o–F_c map). This map is essentially flat at places where the molecular model is correct, as the difference between model and crystal is small. In contrast, the F_o–F_c map has electron density maxima where the model is still lacking atoms or where it contains an atom that is too light. Similarly, the F_o–F_c map shows minima (negative electron density) where the model accounts for too much electrons (if an atom in the model is heavier than it should be or if the model contains an atom where there should not be one). Based on the F_o–F_c map, the initial model can be improved, and phases calculated from the improved model lead to even better F_o and F_o–F_c maps. The improved maps allow improving the model further, and another electron density map can be calculated, which is better still. This iterative process continues until all nonhydrogen atoms are found and one has arrived at what is called the *complete isotropic nonhydrogen model*.

Crystallographers distinguish between nonhydrogen atoms and hydrogen atoms. Hydrogen atoms, which have only one electron and are more difficult to detect in the electron density map, receive special treatment and are introduced into the model toward the end of the refinement process. When all nonhydrogen atoms are included in the model, the next step is to refine the structure anisotropically. In an anisotropic model, the individual nonhydrogen atoms are allowed to move differently in different directions, and atoms are no longer described as spheres but rather as ellipsoids (Figure 1.16). Expanding the model to anisotropic atomic motion dramatically increases the number of parameters to be refined. For an isotropic description, there are four parameters

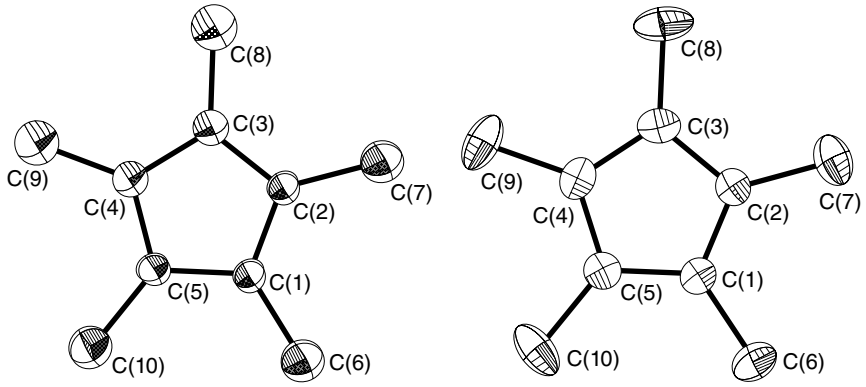


Figure 1.16 Molecular model of a Cp^* ring in a crystal structure refined with isotropic (left) and anisotropic (right) displacement parameters.

per atom (atom coordinates x, y, z and the radius of the sphere), while an ellipsoid needs six parameters (a symmetric 3 × 3 matrix), in addition to the coordinates, for a total of nine parameters per atom. Therefore, anisotropic refinement is only possible for datasets with sufficiently high resolution (the cutoff is between 1.0 and 1.5 Å of resolution).

The IUCr recommends a data-to-parameter ratio of 10 : 1. For a fully anisotropic model, this ratio is reached at a resolution of 0.84 Å. Many small-molecule datasets extend to resolutions of 0.7 Å or even beyond; however, as mentioned above, not all crystals diffract well enough to meet this IUCr standard. The use of restraints and constraints can help improve the data-to-parameter ratio. Constraints are mathematical equations rigidly relating two or more parameters or assigning fixed numerical values to certain parameters, thus reducing the number of independent parameters to be refined. For example, two atoms could be constrained to have the same thermal ellipsoid, or the coordinates of an atom located on a mirror plane could be constrained to keep the atom from leaving the plane. Or, to give a third example, the six atoms of a phenyl ring could be constrained to form a perfect hexagon. Restraints, in contrast, are treated as additional data and, just as data, have a standard uncertainty. In the absence of restraints, the model is refined solely against the measured diffraction data, and the minimization function M looks like this:

$$M = \sum w\left(F_o^2 - F_c^2\right)^2$$

In this equation w is a weighting factor applied to every structure factor expressing the confidence in the corresponding observation[29]; F_o and F_c are

29 In good approximation, $w = 1/\sigma$, where σ is the standard uncertainty of the corresponding reflection.

the observed and calculated structure factors, respectively. Restraints allow including additional information (for example, that aromatic systems are approximately flat or that the three C–F bond distances in a CF_3 group are approximately equivalent). These additional bits of information can be added to the diffraction data, and the function M in the presence of restraints changes to

$$M = \sum w\left(F_o^2 - F_c^2\right)^2 + 1/\sigma^2\left(R_t - R_o\right)^2$$

In this equation σ is the standard uncertainty (also called elasticity) assigned to a specific restraint, R_t is the target value the restraint assigns to a specific quantity, and R_o is the actual value of the restrained quantity as observed in the current model. Comparison of the two minimization functions above shows that restraints are treated exactly like data in a structure refinement. Some structures, perhaps even most, do not require any restraints at all; however when the data-to-parameter ratio is low or disorders or twinning cause strong correlations between certain parameters that should not be correlated, restraints can be essential. "In general, restraints must be applied with great care and only if justified. When appropriate however, they should be used without hesitation, and having more restraints than parameters in a refinement is nothing to be ashamed of" [27].

It is important to critically inspect a graphical representation of the anisotropic thermal parameters, as the shape, orientation, and relative size contain important information about the quality of the model. Usually, those graphical representations are called thermal ellipsoid representations or thermal ellipsoid plots,[30] and the word "thermal" implies that the ellipsoids represent the thermal motion of the individual atoms. Most commonly, the volumes or boundaries of the ellipsoids are chosen so that each ellipsoid contains 50% of the electron density of the atom in question, and a typical description of such a plot would be "thermal ellipsoid representation at the 50% probability level". In a good structure, all thermal ellipsoids should have approximately the same size,[31] and their shapes should be relatively spherical. Strongly prolate or oblate ellipsoids point to problems with the data or incorrect space group. Strongly elongated ellipsoids usually indicate disorder that needs to be resolved, and noticeable small or large ellipsoids suggest that the wrong element was assumed for the atom in question.

30 Often, people call them "ORTEP plots". ORTEP is the name of the first program that could generate those graphical representations. The program was written by Carol Johnson, and ORTEP stands for Oak Ridge Thermal Ellipsoid Plot. One should never call a thermal ellipsoid representation an ORTEP plot unless the program ORTEP was actually used to generate them.
31 One should consider, however, that terminal atoms move more than central ones.

Once the complete anisotropic nonhydrogen model is established, the hydrogen atoms can be included. As hydrogen atoms have only one electron, which is delocalized, they are difficult to place based on the difference density map. Luckily, in most cases (especially with carbon-bound hydrogen atoms), it is straightforward to calculate the positions of the hydrogen atoms and to include them into their calculated positions.[32] The hydrogen model derived in this fashion is often better than it would be if the hydrogen positions were taken from the difference map. An additional advantage of calculating hydrogen atom positions is that no additional parameters need to be refined. In contrast, potentially acidic hydrogen atoms (for example, H bound to oxygen or nitrogen), hydrogen atoms in metal hydrides, or other chemically unusual or special hydrogen atoms should be included into the model from the difference map.[33]

When all hydrogen atoms are included into the model, the refinement is essentially complete. Before the structure is published, however, a structure validation step should be performed. Freely available software such as Platon [28] or the online tool checkCIF[34] analyze the final model for typical problems (symmetry, thermal ellipsoid shape, data integrity, etc.) and create a list of alerts that should be examined critically.

1.5.3.1 Resonant Scattering and Absolute Structure

It was mentioned above that reciprocal space is, in good approximation, centrosymmetric. This centrosymmetry of reciprocal space was described independently by Georges Friedel [29] and Johannes Martin Bijvoet [30], and the equation $|F_{h,k,l}|^2 = |F_{-h,-k,-l}|^2$ is called Friedel's law or Bijvoet's law. This law only holds for strictly elastic interactions between photons and electrons, and in the presence of resonant scattering (often also called "anomalous scattering" or inelastic scattering), the centrosymmetry of reciprocal space is slightly disturbed in noncentrosymmetric space groups.[35] The strength of resonant scattering depends on atom type and X-ray wavelength: Heavier atoms and longer

32 During the subsequent refinement cycles, the hydrogen positions are updated continuously as the positions of the nonhydrogen atoms change. This treatment is called a *riding model*, as the hydrogen atoms sit on the molecule as a rider on a horse and where the horse goes, the rider follows. (The author of these lines made a different experience when attempting to ride a horse, but in structure refinement this description of a riding model usually holds.)

33 Refinement of such hydrogen atoms is usually aided by application of X–H distance restraints (X is any atom type) and by constraining the hydrogen atoms' thermal parameter to, for example, 150% of the thermal motion of the atom X. Such a treatment is called a "semi-free refinement" of the hydrogen atoms.

34 http://checkcif.iucr.org/

35 In centrosymmetric space groups where for every atom x, y, z there is another atom $-x, -y, -z$, the effects of inelastic scattering for every such pair of atoms cancel each other out.

wavelengths give rise to more inelastic scattering. To observe this effect with Mo radiation, atoms heavier than Si should be present in the structure, while for Cu radiation even oxygen gives enough resonant scattering to observe a slight violation of Friedel's law. Those weak differences, often only a few percent of the absolute intensity, allow determining the absolute structure of a crystal and, thus, the absolute configuration of chiral molecules. During structure refinement, the model is treated mathematically as if it were a mixture of both hands, and the ratio between the two hands is refined. This ratio is called the Flack x parameter [31], and its value ranges from zero to one. A Flack parameter of zero indicates that the hand of the molecule in the structure is correct, and a value of one means that the structure should be inverted. Values between zero and one indicate mixtures of both hands, and a value of 0.5 corresponds to a perfectly racemic mixture. It must be noted that the Flack x comes with a standard uncertainty, which is as important as the value itself. For an absolute structure to be considered determined correctly and confidently, the Flack x should be zero within two to three standard uncertainties, and the standard uncertainty should be smaller than 0.01. If it is known that a compound is enantiopure, racemic twinning can be ruled out, and the Flack x can only be one or zero but not in between. In this case, a higher standard uncertainty of, say, 0.1 can be accepted [32].

1.6 Powder Methods

Single-crystal X-ray diffraction is unequivocally the most definitive technique for determining crystal structures. All too often, however, the structures of small-molecule crystal forms are elusive because of the single-crystal size/quality requirements of the X-ray methods or the methods of preparation. For example, solution methods of crystallization were used to produce single crystals of seven of the ROY polymorphs (Figure 1.2). However, the four most recently discovered polymorphs, YT04, Y04, RPL [33], and R05 [34], were not initially crystallized from solution, having instead been discovered many years later through melt crystallization, vapor deposition, and solid-state phase transitions. None of these methods are conducive to generating single crystals, and only by introducing YT04 seeds obtained by melt crystallization into a supersaturated solution of ROY were single crystals of this polymorph ultimately produced for its structure determination. Fortunately, in cases where single-crystal substrates are not available, powder methods may be used to solve crystal structures. Two approaches, namely, structure solution from powder diffraction and NMR crystallography, are increasingly used for crystal structure analysis in pharmaceutical development and will be briefly described in the following sections.

1.6.1 Powder Diffraction

Powder X-ray diffraction (PXRD) patterns provide 1D fingerprints of 3D crystal packing arrangements dispersed among randomly oriented polycrystallites. Owing to the ease with which powder patterns can be collected, PXRD is extensively used in pharmaceutical development to identify crystal forms based on their unique diffraction peaks and intensities. PXRD patterns can also be used for structure determination when suitable single crystals cannot be grown to sufficient size (on the order of ~100 μm) or quality [35, 36]. Although single-crystal and powder diffraction provide the same intrinsic information, when 3D diffraction in reciprocal space is compressed into a 1D powder pattern, information is inevitably lost, particularly at shorter d-spacings (higher diffraction angles). The loss of intensity information for individual peaks in a powder pattern due to peak overlap increases both the difficulty and uncertainty of structure solution from powders. Therefore, to ensure that the structure model is as accurate and precise as possible, measures must be taken to ensure that the powder sample quality is high and that the PXRD data are properly collected. To this end, PXRD data are usually collected in transmission mode for carefully prepared, highly crystalline, and preferably phase-pure powders placed between polymer films or packed in thin-walled capillaries. To minimize preferred particle orientation effects and to give good powder averaging, the samples are spun or rotated in the incident X-ray beam during data collection. For the high accuracy needed for structure solution, the diffraction pattern is typically collected over a wide 2θ range, usually up to 70 or 80°.

Structure determination from powder diffraction data involves three steps:

1) Indexing the peaks in the experimental pattern to determine the size and shape of the unit cell, along with the space group symmetry.
2) Using the diffraction peak intensities to generate a good approximation to the atomic positions in the crystal structure.
3) Refining (usually by the Rietveld method [37]) the trial structure to fit the simulated PXRD pattern of the model to the full experimental PXRD pattern.

Indexing programs that are widely used in the first step to determine the lattice parameters (a,b,c,α,β,γ) include X-Cell [38], DICVOL [39], and singular value decomposition [40]. Sensible indexing solutions are generally identified based on the molecular volume, cell volume, and number of unindexed reflections (checked using either Le Bail [41] or Pawley [42] fitting). Once the powder pattern has been successfully indexed, the space group can be assigned by identifying systematically absent reflections. Here, it should be noted that of all of the steps in the powder structure solution process, the first indexing step tends to be the most problematic, and without a correct unit cell, structure solution is impossible.

In the second step of the structure determination process, trial crystal structures are generated in direct or real space independent of the experimental PXRD data. Search algorithms, such as simulated annealing [43–45], Monte Carlo [46, 47], or genetic algorithms [48–50], are used as implemented in commercially available programs (e.g. PowderSolve,[36] DASH,[37] and TOPAS[38]) to generate trial structures with input of the chemical structure, the unit cell parameters, and space group. The positions, orientations, and internal degrees of freedom of molecular fragments are stochastically varied within a unit cell until a match between the simulated and experimental PXRD patterns is obtained. The approximate structure solution(s) from the second step serves as a starting point for the subsequent structure refinement in step 3.

At the third and final stage, the structure model, along with peak profile and background parameters, temperature factors, zero-point error, preferred orientation, etc. are Rietveld refined to a more accurate, higher-quality description of the structure, as shown for fexofenadine hydrochloride in Figure 1.17 [54]. The correctness of a powder structure solution is assessed by comparing its calculated powder pattern with the experimental pattern, the fit being qualitatively visualized by the difference curve (black curve at the bottom of Figure 1.17) and quantified by either a weighted powder profile R-factor (R_{wp}) or full profile χ^2. It is generally recommended that the crystal structure solution be subsequently verified by dispersion-corrected density functional theory (DFT-D) energy minimization [55]. With this approach, a powder structure is judged correct when the root mean square Cartesian displacement (RMSCD) value is 0.35 Å or less.

1.6.2 NMR Crystallography

NMR spectroscopy is universally recognized for its unparalleled ability to characterize molecular structure, conformation, and bonding in solution. A key to the early and enormous success of solution NMR methods has been the ease with which high-resolution spectra are acquired, made possible in part because the orientation-dependent (anisotropic) interactions that affect NMR spectra are normally averaged to single isotropic values by rapid molecular tumbling in solutions.[39] The molecular mobility in solids is, by contrast, highly restricted, and therefore strong nuclear-spin interactions are not dynamically averaged. This means that NMR spectra of solids acquired under the same (as solution)

36 PowderSolve – a complete package for crystal structure solution from powder diffraction patterns [51].

37 DASH: a program for crystal structure determination from powder diffraction data [52].

38 TOPAS [53].

39 A single crystal would produce a comparably simple NMR spectrum, in this case not because of Brownian motions but instead because only one crystal orientation is present with respect to the direction of the external magnetic field.

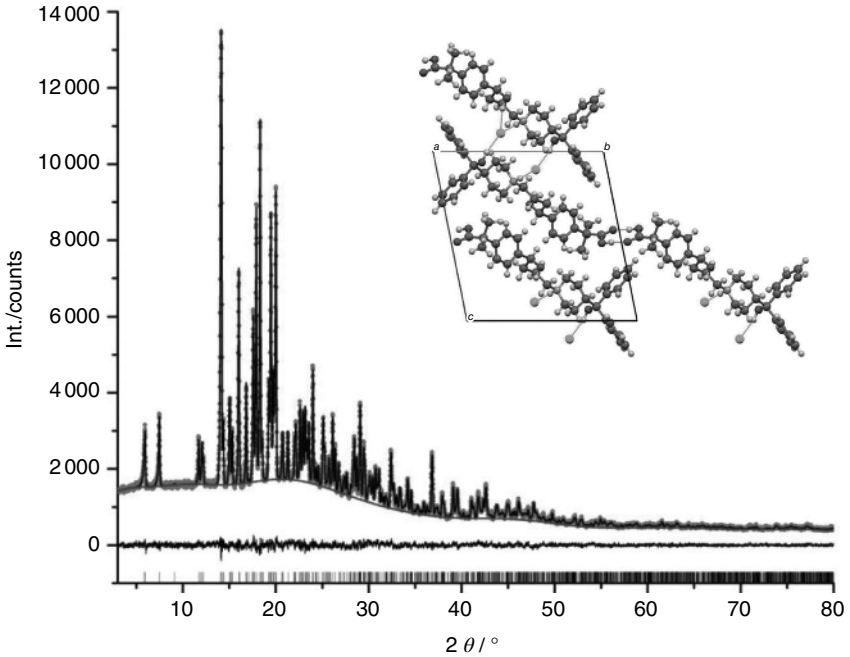


Figure 1.17 Rietveld plot of racemic fexofenadine hydrochloride showing the fit of the experimental PXRD pattern (dots) to the simulated pattern (solid line) for the powder structure model [inset]. The vertical tick marks represent the theoretical peak positions. *Source:* Adapted with permission from Brüning and Schmidt [54]. Reproduced with permission of John Wiley & Sons.

conditions are poorly resolved, owing to the simultaneous observation of nuclei in all possible orientations with respect to the external magnetic field. Not surprisingly, the widespread use of solid-state NMR spectroscopy would await the development of methods to remove (and, in some cases, reintroduce on demand) dipolar and scalar (*J*) spin–spin coupling interactions, as well as to additionally overcome the poor sensitivity associated with the detection of nuclei at low natural abundance.

Solid-state NMR methods were foreseen as a way to derive even more detailed structural information for molecules in solution, based on the premise that they would bridge solution-state NMR spectra and precisely determined molecular structures and conformations derived from X-ray diffraction. However, molecular structure in solution can be rather different, and where material properties are of interest, these attributes will be more relevant as they exist in the solid state. Either way, from the time that cross polarization (CP), magic-angle spinning (MAS), and high-power ^{1}H decoupling techniques were first combined to produce high-sensitivity, high-resolution ^{13}C spectra [56], solid-state NMR

spectroscopy has become an indispensable technique for chemical analysis, structure determination, and studying dynamic processes in the solid state. The CP/MAS experiment with its variations and extensions allows local electronic environments of different NMR-active nuclei (^{1}H, ^{13}C, ^{31}P, ^{15}N, ^{17}O, ^{19}F, etc.) that are common to pharmaceutical molecules and their formulations to be uniquely probed over a large timescale without the requirement of single crystals. Solid-state NMR spectroscopy therefore nicely complements X-ray crystallography, rendering the combination of the two techniques highly powerful for providing a complete determination of structure and dynamics at an atomistic level.

Advances in hardware and probe technology, higher magnetic field strengths, and the development of a range of specialized multinuclear and multidimensional solid-state NMR experiments, along with quantum mechanical methods for computing NMR parameters (e.g. shielding constants for chemical shift prediction), have fueled interest within the pharmaceutical community in applications of NMR crystallography, that is, the use of solid-state NMR spectroscopy for determining or refining structural models [57, 58]. For the fundamentals underpinning solid-state NMR spectroscopy, along with descriptions of the spectrometer hardware, pulse sequences, and operational aspects involved, the reader is referred to comprehensive monographs on the subject [59–61]. We focus herein on the practical application of solid-state NMR spectroscopy for the structural characterization of pharmaceutical materials, with particular attention to how this technique can be used to assist in crystal structure determination from diffraction data.

Early pharmaceutical applications of solid-state NMR spectroscopy relied on the basic CP/MAS experiment to fingerprint drug crystal forms (akin to PXRD), mainly through their unique isotropic chemical shifts. In this capacity, not only has NMR spectroscopy been invaluable for characterizing the solid-state form landscapes of drug molecules en route to selecting the crystalline delivery vehicle for a given drug product, but it also has secured its place as a research tool in drug development, ensuring that crystallization processes deliver and preserve the correct form and formulation processing and long-term storage preserve it. Solid-state NMR spectroscopy has also been used to good advantage in claiming drug crystal forms as intellectual property in patents, and in a number of cases, to later prove patent infringement of those forms in generic drug products.

NMR crystallography has evolved from the fingerprinting applications described above into what is now the derivation of precise bond lengths and angles within a molecule, and the determination of intermolecular bond lengths and angles associated with packing patterns. An impressive demonstration of solid-state NMR spectroscopy for determining 3D structure at natural isotopic abundance has been reported for simvastatin, the active ingredient in Zocor® [62]. In this work, a combination of state-of-the-art through-bond and through-space NMR correlation experiments was used to establish the

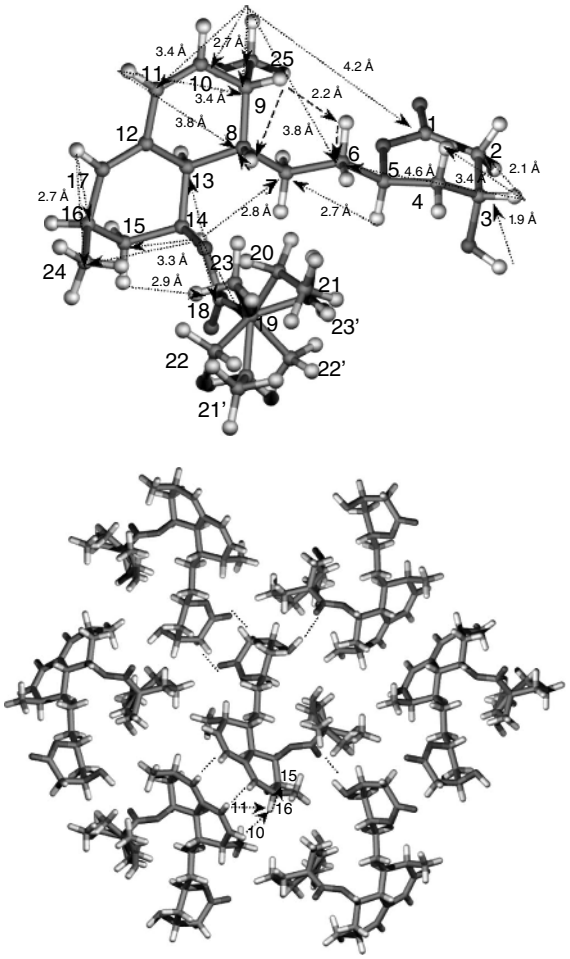


Figure 1.18 Conformation of a single molecule of simvastatin (left) and molecular packing in crystalline simvastatin (right) with interatomic $^{1}H–^{13}C$ distances and intermolecular contacts (marked by arrows) established by solid-state NMR spectroscopy. Disorder of the terminal ester was proposed by X-ray diffraction. *Source:* Reproduced from Brus and Jegorov [62] with permission of American Chemical Society. (*See insert for color representation of the figure.*)

molecular conformation of the drug molecule in its crystalline structure and also to identify close contacts between near-neighbor molecules (Figure 1.18). Such information, which nowadays is supported by first principles density functional theory (DFT) computations of chemical shielding [63, 64], may be used to validate structure solutions derived from PXRD; it may even contribute to the crystal structure determination process, either by providing restraints for structure refinement [65, 66], or in combination with CSP (vide infra), and eliminate putative but incorrect structures [67].

A recent extension to chemical-shift-based NMR crystallography has combined MD simulations and DFT calculations to quantify the distribution of

atomic positions in a crystal [68]. With this approach, NMR parameters are computed for a range of configurations taken from MD snapshots (simulating thermal motions above 0 K), effectively allowing the dynamic contributions to peak broadening in solid-state NMR spectra to be modeled and anisotropic displacement parameters (depicted in thermal ellipsoid plots, cf. Figure 1.16) to be derived to even greater accuracy than X-ray diffraction. Today, this application is neither trivial nor commonplace, but it shows the great promise that extension of the computational methods of solid-state NMR spectroscopy, in combination with experiment, has for extracting ever more detailed and accurate structural and dynamic information to reinforce or complement X-ray crystallography.

1.7 Crystal Structure Prediction

Another route to molecular and crystal structure models that has emerged in recent years is ab initio CSP, a computational methodology wherein 3D crystal packing arrangements are calculated from first principles, starting with a chemical diagram of the molecule [69, 70]. Owing to the heavy demands of the computational methods involved, CSP is generally performed in two stages. The first uses algorithms to generate trial structures that sample different crystal packing possibilities, holding the molecular conformations rigid. At this stage, anywhere from a 1000 to 1 000 000 or more plausible structures may be calculated, depending on the size and flexibility of the molecule, how many space groups and independent molecules (Z') are included in the search, the chirality of the molecule, and available computational resources and time. The low-energy local minima among the computed crystal structures identified in the first stage are then subjected in the second stage to more accurate (and computationally expensive) lattice energy minimizations, this time refining the molecular conformation within the crystal structure (obeying space group symmetry) to identify those that are lowest in energy. All successful CSP methods use electronic structure calculations, albeit in different ways. One approach involves first optimizing the geometry of the isolated molecule in a range of conformations and then selecting input structures for the global structure search among the low-energy conformational minima [71]. The computationally expensive but very powerful method of Neumann and coworkers uses a molecule-specific force field that is parameterized to reproduce DFT-D crystal structures, Monte Carlo parallel tempering to generate structures, and solid-state DFT-D calculations for the final energy minimization/ranking [72].

The output of a CSP is a crystal energy landscape, a collection of putative crystal structures, all at 0 K, which are usually ranked in order of their lattice energy and separated in the second dimension by their crystal packing efficiency (or density), as shown in Figure 1.19 [73]. In this example, one of the earliest

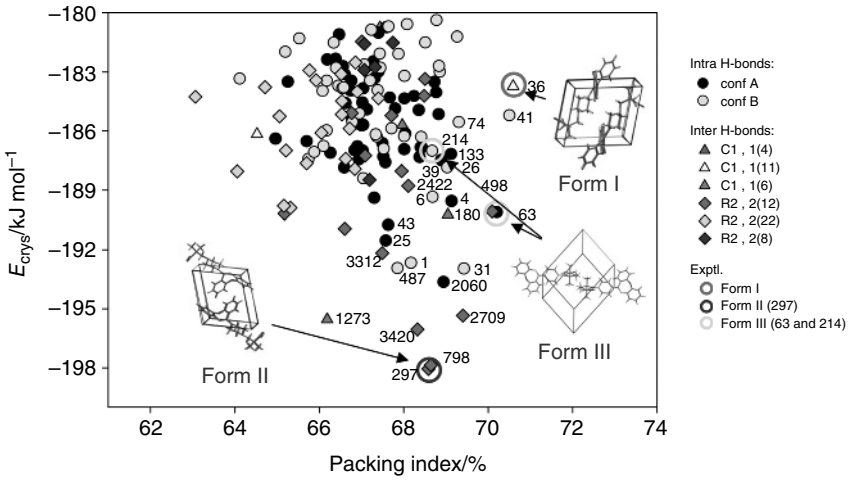


Figure 1.19 Crystal energy landscape of a model pharmaceutical. Each point represents a mechanically stable 3D structure ranked in order of lattice energy and crystal packing efficiency or packing index. Experimentally observed crystal structures found by solid form screening are encircled. *Source:* Adapted from Braun et al. [73]. https://pubs.acs.org/doi/abs/10.1021/cg500185h. Licensed under CC BY 4.0. Reproduced with permission of American Chemical Society. (*See insert for color representation of the figure*.)

anticipated uses of CSP has been realized – to provide plausible structure models for refinement in cases where crystal structure determination was not possible from available experimental data. Unlike the Form I and II crystal structures, which were solved from well-grown single crystals (but also found on the crystal energy landscape), Form III is a metastable polymorph produced exclusively by dehydration and thus was impossible to grow as a single crystal. Using CSP-generated structures, a disordered structure model, giving a promising match to both PXRD and solid-state NMR data, was proposed.

Progress in the development of CSP methods has been tracked over the last 18 years by blind test competitions hosted by the CCDC. Developers of the methods are provided a series of target molecules (or salts), for which crystal structures have not been published, and asked to predict the spatial arrangement of molecules in crystal structures given only the molecular structure diagram. Computed crystal structures are returned, usually ranked in order of their 0 K lattice energy, although most recently attempts have been made to provide a Gibbs free energy ranking to compare stability at crystallization process-relevant temperatures. With each blind test, the complexity of the challenge has increased in terms of the space groups considered, number of molecules in the asymmetric unit ($Z' > 1$), molecular size and flexibility, and inclusion of less common elements and multicomponent and ionic (salt) targets, commensurate with the development of the algorithms. The results of the most

recent sixth blind test [74], which was published in late 2016, show that much progress has been made in dealing with the challenges presented by flexible molecules, salts, and hydrates. All of the targets, apart from one, were predicted by at least one submission. However, this benchmark of CSP methodologies has shown the need for further improvement of the structure search algorithms, especially for large, flexible molecules. Furthermore, with the relative energy differences between crystal polymorphs being small (typically less than 2 kJ mol^{-1} [75]), it is clear that continued development of ab initio and DFT methods will be required if lattice energies, let alone free energies, are ever to be calculated to the accuracy (and efficiency) needed for reliable ranking of the structures.

The blind test benchmarks of CSP, along with early successes in predicting crystal structures of "small" pharmaceuticals, have generated enormous interest within the pharmaceutical community to develop CSP methods as a complement to experimental solid form screening and to increase access to crystallographic data [76]. The ability to reliably predict how a molecule will crystallize in the solid state, in particular, the range of solid-state forms (polymorphism), would not only confirm that the most stable form is known but could also help design experiments to find new polymorphs, rationalize disorder, and estimate the possible range of properties among different solid forms. These more ambitious goals of using computed crystal energy landscapes to aid solid form development are being realized to a limited extent today, with CSP not only complementing pharmaceutical solid form screening but also helping to establish molecular-level understanding of the crystallization behaviors of active pharmaceutical ingredients [77].

1.8 Crystallographic Databases

Any given crystal structure may hold the key to unlocking important details of chemical structure, conformation, stereochemistry, or intermolecular interactions that improve our understanding of how structure underpins properties. The crystallography community recognized long ago, however, that information gleaned from data *collections* would far exceed that derived from individual experiments and set out to share their data through the creation of crystal structure databases for all researchers to use. A number of such compilations exist today, including the Inorganic Crystal Structure Database,[40] Protein Data Bank,[41] and Crystallography Open Database,[42] the latter attempting to combine all classes of compounds. However, for the discovery and development

40 Inorganic Crystal Structure Database (icsd.fiz-karlsruhe.de).
41 Protein Data Bank (rcsb.org/pdb/home/home.do).
42 Crystallography Open Database (www.crystallography.net/).

of small-molecule pharmaceuticals, there is no more important database than the CSD,[43] the world's repository for small-molecule organic and metal–organic crystal structures.

Curated and maintained by the CCDC, the CSD contains as of the time of this writing over 900 000 entries from X-ray and neutron diffraction analyses, with updates released to the public every six months. Each entry in the CSD is the result of a structure determination from single-crystal or, in some cases, powder diffraction data, and each is identified by its six-letter REFCODE, appended in some cases by two numbers in reference to its publication history. The CSD has grown exponentially over many (50+) years, as has the interest in using the database for structural research. This is due in part to the CCDC's commitment to develop tools to efficiently mine and analyze the structures. The CCDC now offers a suite of CSD System software for searching the database (*ConQuest*); visualizing and analyzing 3D structures (*Mercury*); comparing bond distances, angles, and torsions against statistical distributions of those geometrical parameters within the CSD (*Mogul*); and interrogating noncovalent interactions in the context of the CSD (*Isostar*). As a service to the worldwide crystallography community, the CCDC has also made available programs for checking the syntax and format of crystallographic information files (CIF) (*enCIFer*), curating in-house (proprietary) structure databases (*PreQuest*), and others.

In recent years, the CCDC, in partnership with pharmaceutical and agrochemical companies, has developed knowledge-based tools to aid solid form development [78]. Two such structural informatics tools that are being increasingly applied in pharmaceutical development to assess the risk of polymorphism (among other applications [79]) are the logit hydrogen-bond propensity (HBP) tool [80] and full interaction maps (FIMs) [81]. The HBP tool computes the likelihood of H-bonds forming between specific donor and acceptor groups in a target molecule, while FIMs are used to assess the geometries of noncovalent interactions using various chemical probes, as shown for trimethoprim Forms I and II in Figure 1.20. Collectively, these tools can be used to identify "weaknesses" in a crystal structure, such as statistically less favorable hydrogen bond donor–acceptor pairings or unusual geometries that might warrant further investigation, possibly extending the search for alternate polymorphs.

1.9 Conclusions

Crystallography is the cornerstone of all structure-based science. While single-crystal X-ray diffraction remains the "gold standard" by which molecular and crystal structures are established, powder methods, including X-ray diffraction

43 Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, UK CB2 1EZ (www.ccdc.cam.ac.uk).

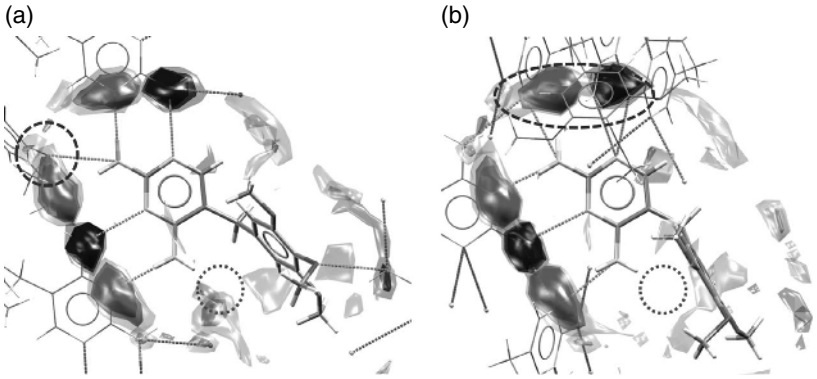


Figure 1.20 Full interaction maps for trimethoprim polymorphs, (a) Form I (AMXBPM12) and (b) Form II (AMXBPM13), showing hydrogen bond acceptor, hydrogen bond donor, and hydrophobic CH "hot spots." The solid-dashed circles highlight hydrogen bonding partners just outside the hot spots, indicating that the interaction geometries are not well represented in the CSD. The dashed circles point to where a hot spot near an NH donor is missing, presumably due to steric hindrance within the crystal conformers.

and solid-state NMR spectroscopy (NMR crystallography) and more recently crystal structure prediction, have provided unprecedented access to structural information for a broad range of materials. Today, the experimental, computational, and informatics tools for crystal structure analysis are having an enormous impact on the design of molecules with optimal biological *and* material properties. As the structure analysis toolbox continues to expand, so too will the role of crystallography in discovering, developing, and delivering safe and efficacious medicines.

References

1 Variankaval, N., Cote, A.S., and Doherty, M.F. (2008). *AIChE J.* 54: 1682–1688.
2 Scott, J.S., Birch, A.M., Brocklehurst, K.J. et al. (2012). *J. Med. Chem.* 55: 5361–5379. Wenglowsky, S., Moreno, D., Rudolph, J. et al. (2012). *Bioorg. Med. Chem. Lett.* 22: 912–915.
3 Sun, C.C. (2009). *J. Pharm. Sci.* 98: 1671–1687.
4 Hursthouse, M.B., Huth, L.S., and Threlfall, T.L. (2009). *Org Proc Res Dev.* 13: 1231–1240.
5 Cruz-Cabeza, A.J. and Bernstein, J. (2014). *Chem. Rev.* 114: 2170–2191.
6 Yu, L., Stephenson, G.A., Mitchell, C.A. et al. (2000). *J. Am. Chem. Soc.* 122: 585–591.
7 Chen, S., Guzei, I.A., and Yu, L. (2005). *J. Am. Chem. Soc.* 127: 9881–9885.
8 Yu, L. (2002). *J. Phys. Chem. A* 106: 544–550.

9 Kepler, J. (1611). Strena seu de Nive Sexangula. A translation of excerpts into English appears in *Crystal Form and Structure* (ed. C.J. Schneer). Stroudsburg, PA: Dowden, Hutchinson & Ross, 1977.

10 Burckhardt, J.J. (1967). *Arch. Hist. Exact Sci.* 4 (3): 235–246.

11 Röntgen, W. C. (1895). Über eine neue Art von Strahlen. Vortrag vor der Physikalisch-Medizinischen Gesellschaft zu Würzburg on 23 January 1896.

12 Friedrich, W., Knipping, P. and von Laue, M. (1912). Interferenz-Erscheinungen bei Röntgenstrahlen, Sitzungsber. (Kgl.) Bayerische Akad. Wiss., 303–322. A translation into English appears in *Structural Crystallography in Chemistry and Biology* (ed. J.P. Glusker), 23–39. Stroudsburg, PA: Hutchinson Ross Publishing Company, 1981.

13 Bragg, W.L. (1913). The diffraction of short electromagnetic waves by a crystal. *Proc. Camb. Philos. Soc.* 17: 43–57.

14 Bragg, W.L. and Bragg, W.H. (1913). *Structure of Some Crystals as Indicated by their Diffraction of X-rays*. *Proc. R. Soc. Lond.* A89: 248–277.

15 Dickinson, R.G. and Raymond, A.L. (1923). The crystal structure of hexamethylene-tetramine. *J. Am. Chem. Soc.* 45: 22–29.

16 Crowfoot, D., Bunn, C.W., Rogers-Low, B.W., and Turner-Jones, A. (1949). The X-ray investigation of the structure of penicillin. In: *The Chemistry of Penicillin* (ed. H.T. Clarke, J.R. Johnson and S.R. Robinson), 310–367. Princeton, NJ: Princeton University Press chapter XI.

17 Hodgkin, D., Pickworth, J., Robertson, J.H. et al. (1955). The crystal structure of the hexacarboxylic acid derived from B_{12} and the molecular structure of the vitamin. *Nature* 176: 325–328.

18 Kendrew, J.C., Bodo, G., Dintzis, H.M. et al. (1958). *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis*. *Nature* 181: 662–666.

19 Lipson, H. and Cochran, W. (1966). The determination of crystal structures. In: *The Crystalline State*, vol. III (ed. S.L. Bragg), 21. Ithaca, NY: Cornell University Press chapter II.

20 Bragg, W.L. (1965). *The Crystalline State, Volume I*, 2. Ithaca, NY: Cornell University Press chapter I.

21 Hahn, T. (ed.) (2011). *International Tables of Crystallography, Volume A*. Wiley.

22 Dauter, Z. (1999). *Acta Cryst.* D55: 1703–1717.

23 Müller, P., Sawaya, M.R., Pashkov, I. et al. (2005). *Acta Cryst.* D61: 309–315.

24 Patterson, A.L. (1934). A Fourier series method for the determination of the components of interatomic distances in crystals. *Phys. Rev.* 46: 372–376.

25 Glusker, J. and Trueblood, K. (2010). IUCr texts on crystallography. In: *Crystal Structure Analysis A Primer*, 3e, vol. 14. Oxford: Oxford University Press.

26 Müller, P. (2009). *Crystallogr. Rev.* 15: 57–83.

27 Müller, P. (ed.) (2006). IUCr texts on crystallography. In: *Crystal Structure Refinement*, vol. 8. Oxford: Oxford University Press.

28 Spek, A.L. (2009). *Acta Cryst.* D65: 148–155.

29 Friedel, G. (1913). *Comptes Rendus* 157: 1533–1536.

30 Bijvoet, J.M. (1949). *Proc. K. Ned. Akad. Wet. Ser. B* 52: 313–314.
31 Flack, H.D. (1983). *Acta Crystallogr.* A39: 876–881.
32 Flack, H.D. and Bernardinelli, G. (2000). *J. Appl. Cryst.* 33: 1143–1148.
33 Mitchell, C.A., Yu, L., and Ward, M.D. (2001). *J. Am. Chem. Soc.* 123: 10830–10839.
34 Chen, S., Xi, H., and Yu, L. (2005). *J. Am. Chem. Soc.* 127: 17439–17444.
35 Harris, K.D.M., Tremayne, M., and Kariuki, B.M. (2001). *Angew. Chem., Int. Ed.* 40: 1626–1651.
36 Tremayne, M. (2008). *Engineering of Crystalline Materials Properties* (ed. J.J. Novoa, D. Braga and L. Addadi), 477. Springer.
37 Rietveld, H.M. (1969). *J. Appl. Crystallogr.* 2: 65–71.
38 Neumann, M.A. (2003). X-cell: a novel indexing algorithm for routine tasks and difficult cases. *J. Appl. Cryst.* 36: 356–365.
39 Boultif, A. and Louër, D. (2004). *Powder pattern indexing with the dichotomy method*. *J. Appl. Cryst.* 37: 724–731.
40 Coelho, A.A. (2003). *J. Appl. Cryst.* 36: 86–95.
41 Le Bail, A., Duroy, H., and Fourquet, J.L. (1988). Ab-initio structure determination of LiSbWO6 by X-ray powder diffraction. *Mater. Res. Bull.* 23: 447–452.
42 Pawley, G.S. (1981). Unit-cell refinement from powder diffraction scans. *J. Appl. Cryst.* 14: 357–361.
43 Trask, A.V., van de Streek, J., Motherwell, W.D.S., and Jones, W. (2005). *Cryst. Growth Des.* 5 (6): 2233–2241.
44 Lapidus, S.H., Stephens, P.W., Arora, K.K. et al. (2010). *Cryst. Growth Des.* 10: 4630–4637.
45 Florence, J., Shankland, N., Shankland, K. et al. (2005). *J. Appl. Cryst.* 38: 249–259.
46 Harris, K.D.M., Tremayne, M., Lightfoot, P., and Bruce, P.G. (1994). *J. Am. Chem. Soc.* 116: 3543–3547.
47 Tremayne, M., MacLean, E.J., Tang, C.C., and Glidewell, C. (1999). *Acta Crystallogr., Sect. B* 55: 1068–1074.
48 Kariuki, B.M., Psallidas, K., Harris, K.D.M. et al. (1999). *Chem. Commun.* 16 (771): 678.
49 Hanson, J., Cheung, E.Y., Habershon, S., and Harris, K.D.M. (2005). *Acta Cryst.* A61: C162.
50 Harris, K.D.M. (2009). *Comp. Mat. Sci.* 45 (1): 16.
51 Engel, G.E., Wilke, S., König, O. et al. (1999). *J. Appl. Cryst.* 32: 1169–1179.
52 David, W.I.F., Shankland, K., van de Streek, J. et al. (2006). *J. Appl. Cryst.* 39: 910–915.
53 Coelho, A.A. (2000). *J. Appl. Crystallogr.* 33: 899–908.
54 Brüning, J.M. and Schmidt, M.U. (2015). *J. Pharm. Pharmacol.* 67 (6): 773–781.
55 van de Streek, J. and Neumann, M.A. (2014). *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater.* 70 (6): 1020–1032.

56 Schaefer, J. and Stejskal, E.O. (1976). *J. Am. Chem. Soc.* 98: 1031.
57 Harris, R.K. (2006). *Analyst* 131: 351–373.
58 Vogt, F.G. (2010). *Fut. Med. Chem.* 2 (6): 915–921.
59 Harris, R.K., Wasylishen, R.E., and Duer, M.J. (eds.) (2009). *NMR Crystallography*. Hoboken: Wiley.
60 Harris, R.K. (2004). *Solid State Sci.* 6: 1025–1037.
61 Geppi, M., Mollica, G., Borsacchi, S., and Veracini, C.A. (2008). *Appl. Spectroscopy Rev.* 43: 202–302.
62 Brus, J. and Jegorov, A. (2004). *J. Phys. Chem. A* 108 (18): 3955–3964.
63 Pickard, C.J. and Mauri, F. (2001). *Phys. Rev. B* 63 (245101): 1–13.
64 Ashbrook, S.E. and McKay, D. (2016). *Chem. Commun.* 52: 7186–7204.
65 Middleton, D.A., Peng, X., Saunders, D. et al. (2002). *Chem Commun (Camb).* 7: 1976–1977.
66 Smith, E., Hammond, R.B., Jones, M.J. et al. (2001). *J. Phys. Chem. B* 105: 5818–5826.
67 Baias, M., Dumez, J.N., Svensson, P.H. et al. (2013). *J. Am. Chem. Soc.* 135: 17501–17507.
68 Hofstetter, A. and Emsley, L. (2017). *J. Am. Chem. Soc.* 139: 2573–2576.
69 Day, G.M. (2011). *Crystallogr. Rev.* 17: 3–52.
70 Price, S.L. (2014). *Chem. Soc. Rev.* 43: 2098–2111.
71 Vasileiadis, M., Kazantsev, A.V., Karamertzanis, P.G. et al. (2012). *Acta Crystallogr., Sect. B: Struct. Sci.* 68: 677–685.
72 Neumann, M.A. and Perrin, M.A. (2005). *J. Phys. Chem. B* 109: 15531–15541.
73 Braun, D.E., McMahon, J.A., Koztecki, L.H. et al. (2014). *Cryst. Growth Des.* 14: 2056–2072.
74 Reilly, A.M., Cooper, R.I., Adjiman, C.S. et al. (2016). *Acta Crystallogr.* B72: 439–459.
75 Nyman, J. and Day, G.M. (2015). *CrystEngComm* 17: 5154–5165.
76 Price, S.L. (2004). *Adv. Drug Del. Rev.* 56: 301–319.
77 Price, S.L. and Reutzel-Edens, S.M. (2016). *Drug Discov. Today* 21 (6): 912–923.
78 Feeder, N., Pidcock, E., Reilly, A.M. et al. (2015). *J. Pharm. Pharmacol.* 67: 857–868.
79 Chisholm, J., Pidcock, E., van de Streek, J. et al. (2006). *CrystEngComm* 8: 11–28.
80 Galek, P.T.A., Allen, F.H., Fábián, L., and Feeder, N. (2009). *CrystEngComm* 11: 2634–2639.
81 Wood, P.A., Olsson, T.S.G., Cole, J.C. et al. (2013). *CrystEngComm* 15 (1): 65–72.