Refinement of Disorder with SHELXL

A tutorial by Peter Müller

Director, X-Ray Diffraction Facility
MIT Department of Chemistry
pmueller@mit.edu

Introduction

A crystal is a potentially endless, three-dimensional, periodic discontinuum built up by atoms, ions or molecules. Because of the periodicity, every object is regularly repeated in three-dimensional space; *i.e.* every unit cell has exactly the same orientation with all molecules in the same conformation as in the cells to its left, right, top, bottom, front, and back. However, an *ideal crystal* does not exist; in most *real crystals* there are several lattice defects and/or impurities. Frequently, parts of molecules (or in some extreme cases whole molecules) are found in more than one crystallographically independent orientation[1]. One can distinguish three cases:

1.   more than one molecule per asymmetric unit

2.   twinning

3.   disorder

In the case of disorder, the orientations of some atoms differ randomly in the different unit cells. Picture a thousand soldiers lined up very neatly in a square, which are supposed to all look to the right, but some 20 percent of them misunderstood the order and turn their heads to the left. This arrangement would be much like a two dimensional crystal with an 80:20 disorder.

The structure determined from the diffraction pattern is the *spatial average* over the whole crystal. In by far most cases, disorder only affects small parts of molecules like organic side chains or $SiMe_3$-groups, or the heads of the soldiers from the above example. Another typical case is the *tert*-butyl group, which is normally almost free to rotate. Disorders of free solvent molecules located in holes in the crystal lattice are also very common. In principle, the presence of disorder is in contradiction with the definition of the crystalline state given above. Yet normally the order predominates, especially when only two different conformations are present in the crystal. Therefore, the conditions for

---

[1] Owing to space group symmetry, the molecules forming a crystal always possess more than one orientation (except for the space group $P1$). This, of course, is not a disorder.

X-ray diffraction are fulfilled, and the diffraction pattern looks unsuspicious. Normally, the solution and initial refinement of a partially disordered structure are not problematic. However, the ellipsoids derived from the anisotropic displacement parameters (ADPs) may be of pathological shape because the program tries to describe two or more atom sites with only one ellipsoid, and the presence of relatively high residual electron density peaks or holes close to the disordered atoms is not unusual.

Types of disorder

Substitutional disorder

Substitutional disorder describes a situation in which the same site in two unit cells is occupied by different types of atoms. This type of disorder is known from minerals and salt-like crystals, in some zeolites the Al- and Si-atoms share the same sites. In biological structures sometimes water molecules share a site with sodium, chloride or other ions. The refinement of substitutional disorder is relatively easy. Nevertheless one should know about it and should be able to recognize substitutional disorder. The clearest warning sign (in most cases the only one) is a too small or too large anisotropic (or isotropic) displacement parameter.

Partial occupancy of atom sites is a relatively common special case of substitutional disorder, and non-coordinating solvent molecules are frequently found to occupy only about half of the holes in the crystal lattice. The presence of "half waters" in protein structures is a typical example. Unusually high displacement parameters are a sign for partially occupied solvent molecules; however, one should take into account that, due to their mobility, even fully occupied non-coordinating solvent molecules tend to show relatively high displacement parameters. Therefore, the ADPs should be drastically larger to justify a reduction of the occupancy factors. The residual electron density map, which shows negative electron density at or around the nuclear positions if the true occupancy is lower than one is a better criterion. In some cases, mixed crystals can be treated as positional disorders if two similar molecules crystallize together at a single site in the same unit cell.

Positional disorder

Positional disorder is the "normal case" of disorder: one atom occupies more than a single site. This can happen in a single unit cell (dynamic disorder, a real motion in the solid state) or distributed among different unit cells (static disorder, a look-alike motion). Both dynamic and static disorders are treated in the same way during refinement.

In a case of discrete disorder, the molecule can possess two (seldom more) well defined energetically similar conformations. The example with the soldiers, some looking to the left and the others to the right, is such a case. In the spatial average, *i.e.* in the structure to be refined, one sees a superposition of both cases. The two positions appear as split atomic sites. Once recognized, such a disorder is refined relatively easily, as we will see below.

Continuous disorder is much more annoying (the soldiers from the example would all be shaking their heads, unable to decide whether to look to the left or to the right). If all rotational angles of *e.g.* a *tert*-Butyl group are energetically similar and there are no steric hindrances, this group of atoms might rotate virtually freely in the crystal (at least at room temperature), and in the spatial average one sees this group as a rotational toroid. It is hard to describe this situation to the refinement program. Normally, one reduces the problem to a refinement of only two or three sites per atom and accepts elongated ADPs. Fortunately, in many cases continuous disorder can be avoided or at least reduced by collecting low temperature data.

Refinement of Disorder

In most real life cases it is sufficient to describe a disorder by formulating two different positions per disordered atom. The principle of disorder refinement is simple. The program needs to know the two sets of coordinates (*i.e.* positions) for each atom together with the relative occupancies (*i.e.* the ratio). The relative occupancies can either be given or refined.

Finding the coordinates for both components is frequently more complicated than formulating the disorder. In this context it is always a good idea to refine disorders at first *isotropically*, as anisotropic displacement parameters tend to compensate for the disorder, which makes it difficult to find additional positions.

Refinement of disorder with SHELXL

SHELXL refines disorder by dividing the disordered atoms into groups. The occupancies of disordered groups are allowed to be refined freely. Together with the **PART** instruction, the introduction of so-called *free variables* makes the refinement of disorders both easy and universal. For a better understanding, the refinement of positional disorder with only two components will be described. The refinement of disorders over several positions is done similarly.

The PART instruction

To begin with, the **PART** instruction in the .ins file divides the disordered atoms into two (or more) groups. Thus, each group represents one component of the disorder, *i.e.* both groups contain the same atoms but on different sites. Practically, in front of the first disordered atom one writes **PART 1**, directly followed by all atoms of the first component. Directly before the atoms of the second component one writes **PART 2**. After all disordered atoms one writes **PART 0** to end the area of split sites. It is always a good idea to make sure that in both parts the atoms are in the same order. This enhances clarity and allows the use of **SAME** (see below).

The second free variable

In the next step, the relation or ratio of the two components has to be taken into account. If the disorder does not involve any special positions, the occupancies of both

components are allowed to possess any ratio. It is important, however, that the site occupancy factors (*sof*) sum up to exactly one.

The occupancy is refined with the help of a free variable, given in the .ins file. The line which directly precedes the first atom starts with **FVAR** and contains the overall scale factor (*osf*), also known as first free variable. For the refinement of a disorder, the *osf* should be followed by a second free variable whose value is between 0 and 1, describing the fraction of unit cells in the crystal showing the conformation described under **PART 1**. This means the second free variable is equivalent to the occupancy of the atoms in component one. *E.g.* a value of 0.6 for the second free variable corresponds to a ratio of 0.6:0.4, describing a 60 % to 40 % disorder. The values of the free variables are refined, but one must guess the initial value or estimate it from the peak height in the difference Fourier map. When in doubt, 0.6 is almost always a reasonable starting value.

Note that the refined value of any free variable has a calculated standard uncertainty, which is listed in the .lst file. The value of this standard uncertainty is supposed to be much smaller than the value for the free variable, or the disorder represented by the free variable would not be very meaningful.[2]

The site occupancy factor (*sof*)

Finally, the site occupancy factors of the disordered atoms must be manipulated to refer to the second free variable. This is done by changing the value of the *sof* instruction from **11.0000** to **21.0000** for the atoms in **PART 1** and to **-21.0000** for the **PART 2** atoms. The *sof* is given for each atom in the sixth column of the .ins file. "**21.0000**" means that the *sof* is set to "*1.0000 times the value of the second free variable*", while "**-21.0000**" sets the *sof* to "*one minus the value of the second free fariable*", completing the disorder. Thus, the *sof* of both components add up to exactly

---

[2] If the free variable coupled to a disorder should refine to 0.95 ±0.1 – this corresponds to an occupancy of the minor compound of 5(±10) % – it is very reasonable to assume that there is no disorder represented by the coordinates coupled to the free variable in question. In such a case, the atoms from the second component should be deleted, the *sof* instruction of the atoms of the first component should be changed back to **11.0000** and the **PART** instructions should be removed.

1.0000, while the ratio can be refined freely. The following example shows excerpts from an .ins file, describing the disorder of two carbon atoms. It was assumed that the component represented by **PART 1** would possess an occupancy of about 60 %. The use of the **PART** instruction, the second free variable and the change of the *sof* instruction are highlighted (**boldface**).

```
FVAR          0.11272  0.6
(...)
PART 1
C1A    1    0.255905     0.173582    -0.001344     21.00000     0.05
C2A    1    0.125329     0.174477     0.044941     21.00000     0.05
PART 2
C1B    1    0.299373     0.178166    -0.015708    -21.00000     0.05
C2B    1    0.429867     0.176177    -0.062050    -21.00000     0.05
PART 0
```

Sometimes, an atom lies on a special position, and therefore its occupancy for the ordered case has to be reduced to 0.5 (*e.g.* an atom on a twofold axis or on an inversion center) or 0.25 (atom on a fourfold axis), which corresponds to a *sof* instructions of **10.5000** or **10.2500**, respectively.[3] If such an atom is involved in a disorder, the *sof* instruction has to be changed to *e.g.* **20.5000** or **20.2500** instead of **21.0000**.

How to find the second site

Disorder may be obvious if a second set of peaks appears in the difference Fourier map, or subtle if the ellipsoids stretch. If the ADPs of an atom behave strongly anisotropically, SHELXL writes a suggestion for the two possible sites of this atom into the .lst file. This message is to be found in the list of "Principal mean square atomic displacements U" (located in the .lst file after the R value calculation following the last least squares cycle and right before the K-factor analysis and the list of most disagreeable reflections). However, not all 'may be split' atoms should be split; sometimes the

---

[3] Other special positions or the combination of several special positions can lead to even lower occupancies. SHELXL recognizes atoms on or very close to special positions and automatically generates the constraints for all special positions in all space groups, which includes the reduction of the *sof* .

anisotropic motion of an atom on a single position is a better description. In cases in which the positions of the disordered atoms are too far from each other to allow one ellipsoid to cover both sites, the anisotropy of the ADPs may not be strong enough for SHELXL to generate the message. In such cases, one can frequently use the coordinates of residual electron density peaks for the second site, or sometimes for both sites. These peaks are named Q by SHELXL and can be found at the very bottom of the .res file.

Sometimes, one does not see residual electron density near an atom and the ADPs are suspiciously elongated but not anisotropic enough for SHELXL to generate the 'may be split' message. In such a case one can use the same initial coordinates for both sites of a split atom; SHELXL separates them during the refinement. However, it is helpful to slightly "move" one of the two sites by hand to prevent a mathematical singularity.

Disorder about special positions

If a molecule lies on a special position of higher symmetry than the molecule can possess, there are only two possibilities to eliminate this geometrical problem: either one changes the space group to one of lower symmetry without this particular special position, or – in most cases far better – one assumes a disorder of the molecule about this particular special position. A typical example is a toluene molecule on an inversion center: none of the atoms lie on the special position, nevertheless in the *spatial average* the whole molecule is disordered in a ratio of 0.5 to 0.5 about the inversion center.

The refinement of such disorders is relatively easy: the second site of each atom can be calculated directly from the positions of the atoms of the first component *via* the symmetry operator of the special position. Therefore, it is not necessary to have two parts in the .ins file. Instead of **PART 1**, **PART 2**, and **PART 0**, the disordered atoms are flanked with **PART -1** and **PART 0**. The negative part number suppresses the generation of special position constraints, and bonds to symmetry-related atoms are excluded from the connectivity table. Moreover, the use of the second free variable is not indicated in such a case, as the ratio between the components is determined be the multiplicity of the special position.

The site occupancy factors must take into account the multiplicity of the special position. For example, in the case of a mirror plane, a twofold axis and an inversion center, the *sof* instruction has to possess the value **10.5000**. A threefold axis causes a *sof* instruction of **10.3333** and a fourfold axis one of **10.2500**, etc. SHELXL generates these site occupancy factors automatically only for atoms on or very close to special positions, but not necessarily for all atoms involved in a disorder about a special position.

Molecules that are located very close to special positions, so that the symmetry would lead to chemically unreasonable arrangements, are treated the same way. In such a case the **SPEC** instruction, which generates all appropriate special position constraints for the specified atoms, may be helpful too.

Disorders with more than two components

In some rare cases it can be appropriate to refine three components of a disorder. The atoms of the three components are grouped in **PART 1**, **PART 2** and **PART 3**, and each component is associated with its own free variable, *e.g.* the free variables number 2, 3 and 4 (the first free variable is always the overall scale factor). Accordingly, the *sof* instructions need to be changed to **21.0000**, **31.0000**, and **41.0000**, and the sum of the three free variables must be one. With the help of the **SUMP** instruction, SHELXL combines free variables in the following way: the weighted sum of the specified free variables is restrained to possess a certain target value within a given standard deviation. Both the target value and the standard deviation can be chosen freely. In the case of a three component disorder associated to the free variables one, two and three, the correct **SUMP** instruction is the following:

SUMP 1.0 0.0001 1.0 2 1.0 3 1.0 4

Right after the **SUMP** command, the target value is given (1.0, as the three components must add up to precisely one), followed by the standard deviation (0.0001).

Thereafter, one finds pairs of weighting factors (here the weighting factors are all 1.0) with the numbers of the free variables (2, 3, and 4).

More than one disorder in one structure

If there are more than one independent disorders in a structure, one also has to use more then one additional free variable. Accordingly, the *sof* instructions are than to be changed to **21.0000** and **-21.0000**, **31.0000** and **-31.0000**, **41.0000** and **-41.0000**, etc. For each disorder one uses **PART 1**, **PART 2** and **PART 0**. Higher part numbers are only used to formulate disorders with more than one component. The format of the .ins file limits the number of free variables to 999, which should be enough to describe even very complicated structures.

Disorder and Restraints

Introduction of disorder into a model can increase the number of refined parameters quite considerably. Therefore, the refinement of disorders should always include restraints. Restraints are treated like experimental observations in the refinement and provide target values for particular parameters or link certain parameters. Thus, they allow the crystallographer to introduce chemical and physical information derived from sources other than the diffraction experiment into the refinement process. In addition, restraints help to reduce or break the correlation between displacement parameters and coordinates of disordered atoms. The following paragraphs give an overview of the restraints commonly used in connection with the refinement of disorder.

Similarity Restraints

Equivalent bond lengths and angles in the two (or more) components of a disorder are assumed to be equal. If the atoms are in the same order in both components of a disorder, one may use the **SAME** instruction. The command **SAME**, followed by a list of atom

names, must be located at the correct position within the .ins file. A **SAME** instruction makes the first atom in the list of atom names equivalent to the first atom immediately following the **SAME** command, the second atom equivalent to the second following, *etc*. "Equivalent" means here that the 1,2- and 1,3-distances of corresponding atoms are restrained to be equal within certain standard deviations (default values are 0.02 for 1,2- and 0.04 for 1,3-distances). The program automatically sets up the $n \cdot (n-1)/2$ restraint equations that are required when n atoms should be equal. For a disordered tetrahydrofuran (*thf*) molecule (see Figure 1 for the atomic numbering scheme) the .ins file would look as follows (the letters **A** and **B** in the atom names refer to the components of the disorder: **A** for atoms in **PART 1**, **B** for atoms in **PART 2**).

```
FVAR  ....   0.6
(...)
PART 1
SAME O1B C1B C2B C3B C4B
SAME O1A C4A C3A C2A C1A
O1A   4  ....  ....  ....    21.000
C1A   1  ....  ....  ....    21.000
C2A   1  ....  ....  ....    21.000
C3A   1  ....  ....  ....    21.000
C4A   1  ....  ....  ....    21.000
PART 2
O1B   4  ....  ....  ....   -21.000
C1B   1  ....  ....  ....   -21.000
C2B   1  ....  ....  ....   -21.000
C3B   1  ....  ....  ....   -21.000
C4B   1  ....  ....  ....   -21.000
PART 0
```
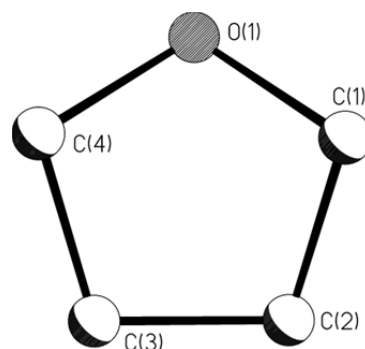

Figure 1: Tetrahydrofuran molecule with atomic numbering scheme.

In this example the first **SAME** instruction precedes the atoms of the first component, listing the atoms of the second component. The oxygen atom O(1B), which is listed first after the word **SAME**, is made equivalent to O(1A), the first atom following the **SAME** command. Next, the carbon atom C(1B), the second atom listed in the **SAME** instruction is made equivalent to C(1A), the second atom following in the .ins file. Similarly, C(2B) is made equivalent to C(2A), C(3B) to C(3A) and C(4B) to C(4A), thus making the two components of the disorder equivalent. *I.e.* all equivalent 1,2- and 1,3-distances between the two components are restrained to be the same. Thereby, in both components the atoms must be in the same order. The second **SAME** instruction also

precedes the atoms of the first component, while the atoms that are listed with it are from the same (the first) component, but in backwards order. This assumes the oxygen atom O(1A) to be equivalent with itself, the carbon atom C(4A) equivalent to C(1A) of the same component, C(3A) equivalent with C(2A) and so forth, thus reflecting the symmetry within the *thf* molecule (see Figure 1). The combination of the two **SAME** instructions restrains equivalent 1,2- and 1,3-distances within each of the components *and* between the components to be the same. The second **SAME** instruction is not disorder specific but can also be used for *thf* molecules which are not disordered.

The list of atom names given in the **SAME** instruction may also contain the '<' or '>' symbols, meaning all intervening non-hydrogen atoms in a forward or backward direction, respectively. Thus, the two **SAME** commands in the example above could also have been formulated as follows:

```
SAME O1B > C4B
SAME O1A C4A < C1A
```

The **SAME** command is very powerful and by no means limited to the refinement of disorders. Whenever there is more than one molecule or group of atoms of the same kind in one structure (*e.g.* several *thf* molecules or SiMe$_3$ groups, or more than one molecule per asymmetric unit) – disordered or not – the **SAME** instruction efficiently restrains the bond lengths and angles to be similar. However, the **SAME** instruction is at the same time a sitting duck for mistakes. If the atoms in the two components (or independent molecules or groups of atoms) are not precisely in the same order, the restraints generated by the **SAME** command may do more harm than good. Typing errors in the list of atom names that follow the **SAME** command are also often fatal.

Alternatively to the **SAME** command, the distances between arbitrary atom pairs can be restrained to possess the same value using the **SADI** instruction. **SADI** is given together with a list of atom pairs. The distances between all pairs mentioned in a single **SADI** instruction are restrained to be equal within a certain standard deviation (default value is 0.02 Å). Restraining distances to a certain target value can be done using **DFIX** or **DANG**. Formulating exactly the same restraints for all equivalent distances with **SADI**

as generated by the two **SAME** commands in the *thf* example above would require the following six lines:

```
SADI O1A C1A O1A C4A O1B C1B O1B C4B
SADI C1A C2A C3A C4A C1B C2B C3B C4B
SADI C2A C3A C2B C3B
SADI 0.04 O1A C2A O1A C3A O1B C1B O1B C3B
SADI 0.04 C1A C3A C2A C4A C1B C3B C2B C4B
SADI 0.04 C1A C4A C1B C4B
```

   The 0.04 for the last three commands changes the standard deviation from the default value 0.02, which is suitable for 1,2-distances, to 0.04, a value more reasonable for 1,3-distances.

SIMU / DELU

Disordered atoms tend to show problems when the first attempts are made to refine them anisotropically. Figure 2 shows ellipsoids representing problematic anisotropic displacement parameters and the effect of applying restraints to the ADPs. The similar-ADP restraint **SIMU** and the rigid-bond restraint **DELU** should be used in disorders to make the ADP values of the disordered atoms more reasonable. **SIMU** restrains the anisotropic displacement parameters of adjacent atoms to be similar, and **DELU** enforces that the main directions of movement of covalently bonded atoms are the same. The default values for the standard deviations are 0.04 for **SIMU** (0.08 for terminal atoms, which tend to move more strongly) and 0.01 for **DELU**. Note that **DELU** is only meaningful for anisotropically refined atoms and is ignored by SHELXL[4] if the specified atoms are still anisotropic. **SIMU**, in contrast, can be applied to isotropically refined atoms as well.

---

[4] That means there is no harm in using **SIMU** on atoms that are still refined isotropically. **SIMU** will become effective as soon as the atoms corresponding to the restraint are refined anisotropically.
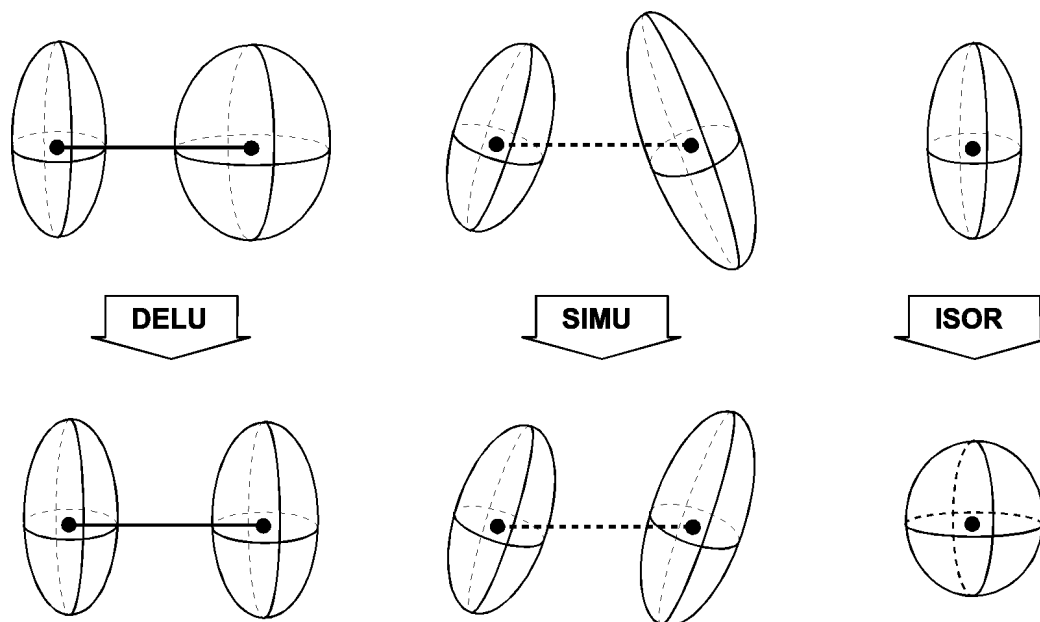
Figure 2: Effect of the restraints DELU, SIMU and ISOR.

ISOR

In particularly tough cases of disorders, especially for disordered solvent molecules, it can be useful to restrain the anisotropic U^{ij}-values of the atoms to behave more isotropically with **ISOR**. As with **DELU** SHELXL ignores **ISOR** commands for atoms that are refined isotropically. **ISOR** is helpful for certain special cases (*e.g.* a disordered atom close to a special position, or anisotropic refinement of a protein against 1.5 Å data) and should almost always be applied to the water molecules of a protein model, but is otherwise less appropriate than **SIMU** or **DELU**. Figure 2 illustrates the effect the restraints **DELU**, **SIMU** and **ISOR** have on anisotropic displacement parameters.

FLAT

If four or more atoms are supposed to lie on a common plane (*e.g.* atoms of an aromatic system) one can use **FLAT** to restrain them to do so within a given standard deviation (default value 0.1 $Å^3$).

Disorder and constraints

In some cases, even *constraints* can be used to refine disorders. As restraints, constraints improve the data to parameter ratio, however not by contributing observations but by decreasing the number of parameters to be refined. Constraints are exact mathematical relationships between certain parameters and have no standard uncertainty.

      `EXYZ` (for equal XYZ) followed by a list of atom names forces the named atoms to possess the same coordinates as the first atom of the list. This can be useful for some types of substitutional disorder, *e.g.* a phosphate and a sulfate ion sharing the same site in a structure.

      `EADP` (for equal ADP) followed by a list of atom names makes the anisotropic displacement parameters of all named atoms equal to those of the first atom in the list.

      If one encounters "geometrical problems", say a phenyl ring does not want to be hexagonal, the `AFIX` constraints can help: `AFIX 66` forces the six following non-hydrogen atoms to form a regular hexagon, while `AFIX 56` defines a regular pentagon. This should be done with care and preferably only in early stages of a disorder refinement. Whenever a similarly satisfying effect can be reached by the use of restraints, the restraints should be given the preference.

General remarks

To make sure that the two sites of an atom are clearly separated and not fitted by an ellipsoid, it is necessary to make all disordered atoms isotropic (if they are not already) prior to the formulation of the disorder. Once the two positions seem to be stable, one can proceed to refine them anisotropically, preferably with the help of restraints.

      In any case, a disorder must be chemically reasonable. Not every significant residual electron density peak is caused by disorder. High residual electron density can also be caused by inadequately corrected absorption, Fourier series truncation errors (*e.g.* when strong reflections are missing) or radiation damage. Such artifacts often lead to the accumulation of spurious electron density at special positions.

Examples

In the following sections three examples deomnstrate how to parameterize disorder for refinement with SHELXL. All files you may need in order to perform yourself the refinements are given on a CD-ROM and are also pre-installed on the computer in front of you. The first example is an easy and straightforward case of static positional disorder that should acquaint you with the **PART** command, the free variables and overall scale factors, as well as the use of restraints and the addition of hydrogen atoms to disordered molecules. The second case is a difficult static positional disorder that affects most of the molecule and involves a special position. You will learn from this example the use of restraints involving symmetry equivalents of atoms (using the **EQIV** command) and you will develop a thorough understanding of the phenomenon of disorder by refining this structure step by step. The third example describes disordered of solvent molecules and you will learn the use of the **PART -1** instruction for two molecules on twofold axes not fulfilling the symmetry of the special position.

Example 1: Gallium-Iminosilicate – Disorder of Two Ethyl Groups

The gallium-iminosilicate $[RSi(NH)_3GaEtGaEt_2]_2$, where R is 2,5-tPr$_2$C$_6$H$_3$NSiMe$_2$$i$Pr, crystallizes in the monoclinic space group $C2/c$ with half a molecule per asymmetric unit. The other half of the molecule is generated from the first one by the crystallographic inversion center. The core of the molecule can be described as a bi-truncated square bi-pyramid. It is formed by four six-membered SiGa$_2$N$_3$-rings in the boat conformation, sharing the three stern-atoms and the three prow-atoms with each neighbor ring. Thus, two planar four-membered SiGaN$_2$-rings are formed. In all rings, metal and nitrogen atoms occupy alternating sites. This kind of cage is also known from tetraasteran (C$_{12}$H$_{16}$, see H. M. Hutmacher, H. G. Fritz, H. Mussow, *Angew. Chem.* (1975) **78**, 174). A more detailed description of the molecule and the chemistry behind it can be found in C. Rennekamp, P. Müller, J. Prust, H. Wessel, H. W. Roesky, I. Usón, *Europ. J. Inorg. Chem.* (2000) **5**, 1861.

Ethyl groups tend to be disordered. The above described gallium cage is a good example, as two of the three crystallographically independent ethyl groups – one per Ga-atom – show rotational disorder about the Ga-C-axis. The solution and first refinement steps of this structure are very straightforward, and we join the refinement at a point where the first signs of disorder appear. This is the file ga-01.res in the disorder subdirectory (also on the CD-ROM), which contains the complete isotropic model without hydrogen atoms. When you look at the atoms and difference density peaks in this file with a graphical interface such as XP or ORTEP the following becomes visible: The highest residual electron density peaks are near the two Ga atoms: Q(1) (2.28 electrons per Å³) and Q(2) (2.23 electrons) near Ga(1) and Q(7) (1.16 electrons) and Q(8) (1.10 electrons) close to Ga(2). This is a normal effect for isotropically refined heavy metals. In addition, relatively high residual density can be found near two of the three independent ethyl groups: Q(3), Q(4) and Q(5) with 2.20, 1.79 and 1.31 electrons, respectively. The latter three residual electron density maxima could indicate disorder of the two ethyl groups. However the high residual density close to the metal atoms reduces the significance of the other maxima. It should be a good idea to first refine all metal atoms

(*i.e.* gallium and silicon) in the structure anisotropically, and then examine the remaining residual electron density. For that purpose we add the following instruction directly above the first atom of the atom list.

ANIS $GA $SI

This has been done in the file ga-02.ins (the file ga-01.res was renamed to ga-02.ins, after the changes were done). After ten cycles of refinement with SHELXL, the results are in the file ga-02.res. When you examine the model carefully, you can see that the atomic displacement parameters of the atoms C(13) and C(15) are somewhat too large. In addition, next to these atoms appear the three highest residual electron density maxima: Q(1), Q(2) and Q(3), with 2.17, 1.71 and 1.33 electrons per Å³. This is typical for a disorder. The logical interpretation is: Q(1) is the second site of C(15), and Q(2) and Q(3) are two new positions for C(13), *replacing* the current C(13). The next highest density maximum, Q(4) (1.23 electrons) lies on an aromatic bond and is not part of a disorder. Q(5) and the other maxima are too weak to be relevant at this state of the refinement. To formulate the disorder, use the **PART** instruction, introduce two new free variables, change the *sof* instructions and make the following changes:

C(15)   →   C(15A)
Q(1)   →   C(15B)
Q(2)   →   C(13A)
Q(3)   →   C(13B)
Delete the old C(13)

And, as there is no disorder refinement without restraints, you should use **SAME** (or the respective **SADI** instructions) to make the 1,2- and 1,3-distances equivalent. To catch all possible 1,3-distances, start two atoms earlier. That means the **SAME** instructions should not be given immediately before the disordered atoms (C(13) and C(15)) but rather two atoms before them (right before Ga(1) and Ga(2)). Also make sure that the atoms are all in the right order! The similarity restraints **SIMU** and **DELU** (**DELU** is

ignored by SHELXL if the named atoms are refined isotropically) make the atomic displacement parameters more reasonable.

The critical portion of the new file, ga-03.ins, looks like this:

```
SIMU c12 c13a c13b c14 c15a c15b
DELU c12 c13a c13b c14 c15a c15b
WGHT     0.100000
FVAR       0.11272   0.6   0.6
same ga1 c12 c13b
GA1    5    0.447952    1.122706    0.039108    11.00000    0.01492    0.02158 =
            0.01663   -0.00262    0.00225    0.00363
C12    1    0.400303    1.237823    0.073859    11.00000    0.02906
PART 1
C13A   1    0.4379      1.3631      0.0949      21.00000    0.05
PART 2
C13B   1    0.4371      1.2955      0.1347     -21.00000    0.05
PART 0
same ga2 c14 c15b
GA2    5    0.445620    0.809823    0.031364    11.00000    0.01782    0.02043 =
            0.01730    0.00047    0.00362   -0.00096
C14    1    0.423155    0.663631   -0.020808    11.00000    0.03426
PART 1
C15A   1    0.375224    0.581908   -0.005250    31.00000    0.11712
PART 2
C15B   1    0.4151      0.5406      0.0044     -31.00000    0.05
PART 0
```

The next .lst file will contain all the information you need to check whether the restraints are used properly. If **MORE 3** is given in the .ins file the .lst file contains a list of all distances treated as equivalent be SHELXL. After ten cycles of SHELXL, the files ga-03.res and ga-03.lst contain the results of the disorder refinement.

The highest residual electron density maximum (Q(1) with 1.21 electrons per Å^3) lies on an aromatic bond – where Q(4) was located before. Taking into account that the model is still mainly isotropic, the other electron density maxima are unsuspicious. In the next step we can refine all atoms anisotropically by including **ANIS** in the .ins file, directly before the first atom. To make sure that you can find all possible hydrogen sites in the difference density, give **PLAN 60** instead of **PLAN 20**. This makes SHELXL find 60 residual electron density peaks instead of 20. The file ga-04.ins contains all these changes.

Except for the H-atoms involved in the disorder, all hydrogen positions can be seen in the difference Fourier synthesis (see the Q-peaks in the file ga-04.res). To validate the restraints, examine the file ga-04.lst (especially the lines 137 to 175). The density peaks Q(12), Q(23) and Q(24) (0.65, 0.62 and 0.61 electrons) correspond to the hydrogen atoms bonded to N(2), N(3) and N(4), respectively. The following **HFIX** instructions cause SHELXL to geometrically calculate the hydrogen positions:

HFIX 43 for all Ar-H

HFIX 13 for the CH-groups

HFIX 23 for the CH_2-groups (but *not* for C(12) and C(14); see below)

HFIX 33 for the disordered CH_3-groups

HFIX 137 for the other CH_3-groups

The H-atoms bonded to N(1), N(2) and N(3) are taken directly from the difference Fourier maxima Q(12), Q(23) and Q(24):

Q(12) → H(2N)

Q(23) → H(3N)

Q(24) → H(4N)

The N-H-distance is set to a value of 0.88 Å for N(2), which makes only two bonds to metal atoms, and to 0.91 Å for N(3) and N(4), which make three bonds to metal atoms, using the distance restraint **DFIX**:[5]

DFIX 0.88 N2 H2N
DFIX 0.91 N3 H3N N4 H4N

---

[5] These distances are sensible at this temperature (-140 °C). A list of X-H distances at the temperature defined by the **TEMP** instruction in the .ins file, can be found in the .lst file.

It is important to note that the positions of the hydrogen atoms bonded to C(12) and C(14) are disordered in the same way as the corresponding Me-groups, although C(12) and C(14) are not directly involved in the disorders themselves. To manage this problem, give after both C(12) and C(14) a **PART 1** and a **PART 2** instruction. In each part, write an **AFIX 23** and an **AFIX 0** instruction, flanking two H-atoms with the coordinates **0 0 0**. These coordinates are ignored by SHELXL, which calculates the correct positions following the geometry defined by the **AFIX** command.

```
C12   1    0.400238    1.237374    0.073651    11.00000    0.02185    0.03417 =
          0.03500   -0.01223    0.00818    0.00417
PART 1
AFIX 23
H12A  2  0  0  0   21.00   -1.200
H12B  2  0  0  0   21.00   -1.200
AFIX 0
PART 2
AFIX 23
H12C  2  0  0  0  -21.00   -1.200
H12D  2  0  0  0  -21.00   -1.200
AFIX 0
PART 1
C13A  1    0.440412    1.342682    0.107947    21.00000    0.04071    0.04595 =
          0.08781   -0.03921    0.00854    0.00640
PART 2
C13B  1    0.440475    1.309546    0.125832   -21.00000    0.03099    0.02988 =
          0.04538   -0.01776    0.02015   -0.00523
PART 0
```

Finally, we can set **PLAN** back to 20. Everything described has been changed in the file ga-05.ins, and after ten more cycles of refinement, we will have the files ga-05.res and ga-05.lst.

Look into the file ga-05.lst to verify that **HFIX 137** resulted in a defined torsion angle for all ethyl groups (the relevant part consists of lines 255 to 316). The file ga-05.res contains the final version of the refinement. Examine the final model in XP and take a look at the disordered CH_2 group at C(12) and C(14).

Finally, the weighting scheme has to be refined to convergence. This has been done in the file ga-06.res, which represents the publishable final model.

Example 2: Disorder of a Ti(III) compound

The Ti(III) complex $(\eta^5 C_5 Me_5)_2 Ti_2 (\mu\text{-F})_8 Al_4 Me_8$ crystallizes in the monoclinic space group $C2/c$ with half a molecule per asymmetric unit. The other half is generated from the first one by the crystallographic twofold axis through the atoms Al(1) and Al(3). The green crystals grow from toluene and are extremely sensitive to air: immediately after retaining them from the flask they start to decompose and lose the green color. Only cooling the crystals under the microscope and low-temperature data collection made a structure determination possible. A more detailed description of the molecule and the chemistry behind it can be found in P. Yu, P. Müller, M. A. Said, H. W. Roesky, I. Usón, G. Bai, M. Noltemeyer, *Organometallics*, **1999**, Vol. 18, No. 9, 1669 - 1674.

The problems with this structure start right at the beginning. The only information we have for model building is the following:

Cp*TiF$_3$ + 2 AlMe$_3$ → green crystals[6]
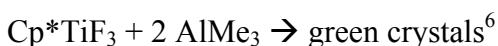
The solution from SHELXS contains a titanium atom and a list of 39 unscaled electron density peaks. The peaks Q(24) to Q(39) are significantly weaker than Q(1) to Q(23) and can therefore be deleted. Examine the arrangement of the remaining 24 atoms in XP. Clearly recognizable is the Cp* ring and the four F-atoms as well as the titanium atom in the asymmetric unit. After generating the symmetry equivalents of the atoms in order to see the whole molecule (using the GROW command in XP), the rest of the Qs form a strange cage, which, albeit pretty, does not seem to make any chemical sense. These peaks are therefore deleted. The titanium and the fluorine atoms as well as the carbon atoms forming the Cp* ligand are retained in ti-01.ins and fed into SHELXL.

The five highest residual electron density maxima in the file ti-01.res (resulting from the first refinement job) are of about the same height (12.85 to 12.47 electrons per

---

[6] Cp* is pentamethylcyclopentadienyl

Å³) and much higher then the other residual maxima (Q(6) has only 5.88 electrons). A peak height of about 13 electrons corresponds very well with aluminum, which is expected to bond to the fluorine atoms. However, the aluminum positions do not seem to be chemically reasonable. Taking into account the symmetry equivalent atoms (labeled with an additional A after their original name), one sees the peaks Q(1), Q(2), Q(4) and Q(4A) to form one component and the peaks Q(3), Q(3A), Q(5) and Q(5A) form the other one. Q(1) and Q(2) both lie on the crystallographic twofold axis.

To formulate the disorder, make the following changes: use the **PART** instruction, introduce a new free variable, change the *sof* instructions and rename the residual electron density peaks Q(1) to Q(5) into aluminum atoms as follows:

Q(1)   →   Al(1)   in **PART 1**

Q(2)   →   Al(2)   in **PART 1**

Q(3)   →   Al(4)   in **PART 2**

Q(4)   →   Al(3)   in **PART 1**

Q(5)   →   Al(5)   in **PART 2**

Note that the correct site occupancy factor instructions of Al(1) and Al(2) (both in **PART 1**) are **20.5000** and not **21.0000**, as they lie on the crystallographic twofold axis.

To restrain the *U* values of all Al atoms give **SIMU 0.04 0.08 2.5 $Al**; the $ sign means "all". The first two numbers after the **SIMU** command are the standard deviations for non-terminal and terminal atoms respectively (both the default values). The 2.5 is the radius of influence for the restraint. It is increased from its default value (1.7) as the Al-F-distances could be slightly larger than 1.7 Å.

The refinement gives rise to the files ti-02.res and ti-02.lst and shows the following results: The second free variable refines to 0.51, which is a reasonable value. Q(1) to Q(4) (5.35 to 2.70 electrons) are significantly higher than the other residual electron density maxima, and seem to represent carbon atoms. They bond to the

disordered aluminum atoms, but are not disordered themselves (only their *connectivities* are disordered not their *sites*). Next step:

Q(1)   →   C(100)

Q(2)   →   C(200)

Q(3)   →   C(300)

Q(4)   →   C(400)

It is also a good idea to use some restraints on the Cp* ligand, which is assumed to posses D_{5h} symmetry. A single **SAME** command can restrain all 1,2- and 1,3-distances in the ligand to be identical (all together 25 restraints!), provided you have the atoms in the right order.

```
same C2 > C5 C1 C20 > C50 C10
C1    1     0.193465     0.169566     0.113055     11.00000     0.02258
C2    1     0.166825     0.249754     0.066287     11.00000     0.02231
C3    1     0.191162     0.324737     0.120094     11.00000     0.01964
C4    1     0.235954     0.291536     0.202630     11.00000     0.02627
C5    1     0.235499     0.197300     0.196759     11.00000     0.02431
C10   1     0.176273     0.077515     0.082855     11.00000     0.03727
C20   1     0.121213     0.255132    -0.024279     11.00000     0.03015
C30   1     0.181023     0.419057     0.098734     11.00000     0.03716
C40   1     0.276334     0.344412     0.275982     11.00000     0.04180
C50   1     0.269335     0.134130     0.264819     11.00000     0.04437
```

The first eight residual electron density maxima as found in ti-03.res are very close to the fluorine positions. Together with the relatively high U values of the F-atoms, this result indicates that the fluorine atoms are also disordered. Therefore we delete all current F-atoms and replace them with the new sites taken from the Q-positions. To make sure that all new F-atoms belong to the right component, one should check the Al-F distances (or Al-Q distances, respectively), which are supposed to be about 1.7 Å. This is much easier after generating the symmetry equivalent atoms.

It becomes clear that Q(1) to Q(4) belong to **PART 1**, while Q(5) to Q(8) belong to the other component (don't forget to change the *sof* instructions to **21.0000** or **-21.0000** respectively, and the atom type number to 3).

Q(1)   →   F(1A)

Q(2)   →   F(2A)

Q(3)   →   F(3A)

Q(4)   →   F(4A)

Q(5)   →   F(1B)

Q(6)   →   F(2B)

Q(7)   →   F(3B)

Q(8)   →   F(4B)

Give also some similarity restraints for the disorder. **SIMU** and **DELU** should be given for the whole structure. In addition, equivalent 1,2- and 1,3 distances should be restrained to the same value using **SADI**. Be very careful when dealing with symmetry equivalents (use the **EQIV** instruction). All these changes have been made in the file ti-04.ins. The **EQIV** / **SADI** instructions look like this:

```
EQIV $1 -x, y, -z+1/2
SADI 0.04 Ti1 Al1 Ti1 Al2 Ti1 Al3 Ti1 Al4 Ti1 Al5
SADI F1A Al1 F2A Al2 F3A Al3_$1 F4A Al3 F1B Al5_$1 F2B Al4 F3B Al4_$1 F4B Al5
SADI C100 Al3_$1 C100 Al4 C200 Al5 C200 Al3_$1 C300 Al2 C300 Al4_$1 =
     C400 Al5_$1 C400 Al1
SADI F1A Ti1 F2A Ti1 F3A Ti1 F4A Ti1 F1A Ti1 F1B Ti1 F2B Ti1 F3B Ti1 F4B Ti1
```

After eight cycles of refinement, the *R*-values have already much improved and the highest residual electron density peak is at 0.99 electrons. In the next step, we can refine all atoms anisotropically by writing **ANIS** right before the first atom in ti-05.ins.

The hydrogen positions for the Cp* ligand can be seen in the difference Fourier synthesis (see the Q-peaks in the file ti-05.res). Now we can add **HFIX 137** for the Cp* Me-groups. The hydrogen atoms of the Al-CH$_3$ groups are to be treated like the disordered CH$_2$-groups in the Ga-Iminosilicate (first example). To avoid problems during the generation of the disordered hydrogen positions, which can occur in this rare case, it is useful to have all atoms of the model in the same asymmetric unit. Therefore, for the

file ti-06.ins the coordinates of some atoms are transformed, and the distance restraints are changed accordingly. It is a little tedious to actually generate the symmetry equivalent atoms and change all the restraints correctly. It is, however, very educational and explicitly recommended.[7] Only if you don't manage to do this yourself move on to the provided file ti-06.ins.

The file ti-06.res contains the complete anisotropic model with all hydrogen atoms. Finally, the weighting scheme has to be refined to convergence. This has been done in the file ti-07.res.

Comparing the first solution coming from SHELXS to the final model, we can see that the solution (ti-00.res) already contained all the non-hydrogen atoms. It was, however, not exactly easy to interpret that solution correctly.

---

[7] If you are using XP, the commands **ENVI** and **SGEN** will prove very helpful.

Example 3: Benzoic Acid on a Twofold Axis

The file benz-01.res on the accompanying CD-ROM and in the disorder subdirectory contains a complete anisotropic model of a diabetes drug, which crystallizes in the monoclinic space group *C*2 with two molecules in the asymmetric unit. The model contains hydrogen atoms but not yet the solvent. The two independent molecules are related by a pseudo inversion center, only violated by the chiral carbon atom. This one atom per molecule makes the difference between space group *C*2/m with one molecule per asymmetric unit and space group *C*2 with two independent molecules; but this point is not part of the disorder and will be the subject of another tutorial.

Let's examine the situation with XP: The 30 highest residual electron density peaks in benz-01.res, Q(1) to Q(30), with intensities of 2.24 to 2.17 electrons per Å^3, are significantly stronger than the others and seem to be solvent. From the crystallization conditions we suspect the presence of benzoic acid or benzoate, and, indeed, the shapes formed by the peaks support this hypothesis. Two molecules of benzoic acid (or benzoate) are easily identifiable and – when assuming the atomic numbering scheme as shown in Figure 3 – we can make the following assignments:
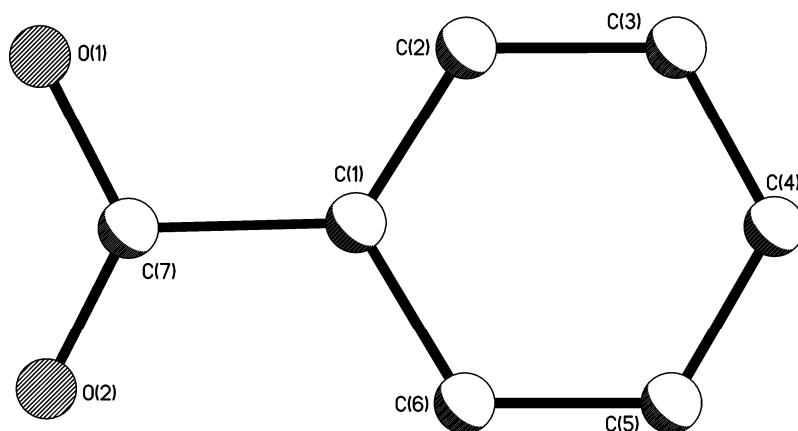
Figure 3: Suggestion for an atomic numbering scheme for benzoate.

Q(3)	→	C(11)	Q(4)	→	O(21)
Q(5)	→	O(12)	Q(7)	→	C(27)
Q(10)	→	C(17)	Q(8)	→	C(21)
Q(11)	→	O(11)	Q(9)	→	O(22)
Q(15)	→	C(14)	Q(13)	→	C(26)
Q(16)	→	C(12)	Q(14)	→	C(25)
Q(18)	→	C(16)	Q(17)	→	C(24)
Q(20)	→	C(13)	Q(21)	→	C(22)
Q(24)	→	C(15)	Q(22)	→	C(23)

Two more molecules are visible, however only partially. Taking into account the symmetry equivalents of the residual density peaks (*e.g.* using the GROW command in XP), one sees that these two molecules are located very close to crystallographic twofold axes. But even though benzoate possesses a twofold axis, in this case the molecules are not oriented along the crystallographic twofold axis but slightly tilted: the molecules are disordered. As only the carboxyl group and three atoms of the aromatic ring are visible, we need to find a way to generate the missing atoms. The easiest is to use geometrical constraints as described above. **AFIX 66** generates a perfect hexagon. The following residual density maxima can easily be assigned.

Q(1)	→	O(32)	Q(2)	→	O(42)
Q(6)	→	C(31)	Q(12)	→	C(41)
Q(25)	→	C(32)	Q(19)	→	C(42)
Q(26)	→	O(33)	Q(23)	→	C(43)
Q(27)	→	C(37)	Q(28)	→	O(41)
Q(30)	→	O(31)	Q(29)	→	C(47)

The remaining three atoms of each ring are generated geometrically in the following way. Write **AFIX 66** in front of the first atoms of the phenyl rings (C(31) and C(41) respectively), complete the number of atoms (three atoms are missing in each of the two incomplete molecules) with atoms with the coordinates **0   0   0**, and write

AFIX 0 after the last atom of the two phenyl ring. Use **SAME** to restrain equivalent distances within and between the four solvent molecules. The site occupancy factor instructions for the two disordered molecules must be changed to **10.5000** and all disordered atoms must be in **PART -1**. As the second site of each disordered atom can be calculated directly from the positions of the atoms of the first component *via* the symmetry operator of the respective twofold axis, it is not necessary to have two parts in the .ins file. The negative part number suppresses the generation of special position constraints, and bonds to symmetry-related atoms are excluded from the connectivity table. In addition, the use of the second free variable is not indicated, as the ratio between the components is determined by the multiplicity of the special position, which is expressed by the *sof* instruction **10.5000**. **SIMU** and **DELU** have already been given for the entire structure earlier in the refinement, so we do not need to do it here for the disordered atoms. It is also important to add an **AFIX 0** after the last hydrogen atom. This hydrogen atom was the last line before the **HKLF 4** card, therefore an **AFIX 0** would have been meaningless and SHELXL would automatically remove it. If, however, other atoms follow, which is now the case, the **HFIX 0** becomes important. Taking all

this into account, the solvent part of the next .ins file (benz-02.ins) should look as follows (take some time to find out the meaning of the **SAME** commands):

```
AFIX 0
O11   4    0.36900    1.09370    0.50540   11.00000    0.05000
O12   4    0.41820    1.22620    0.47870   11.00000    0.05000
SAME C17 C11 C16 < C12
C17   1    0.38020    1.18030    0.49350   11.00000    0.05000
C11   1    0.33280    1.26340    0.49820   11.00000    0.05000
C12   1    0.28610    1.22030    0.51360   11.00000    0.05000
C13   1    0.24140    1.30520    0.51520   11.00000    0.05000
C14   1    0.24620    1.40130    0.50460   11.00000    0.05000
C15   1    0.29080    1.44490    0.48790   11.00000    0.05000
C16   1    0.33400    1.37520    0.48590   11.00000    0.05000
SAME O11 > C16
O21   4    0.07960    0.71580    0.01790   11.00000    0.05000
O22   4    0.13020    0.84580   -0.00650   11.00000    0.05000
C27   1    0.11870    0.75770    0.00450   11.00000    0.05000
C21   1    0.16800    0.67320    0.00210   11.00000    0.05000
C22   1    0.16490    0.56930    0.01220   11.00000    0.05000
C23   1    0.20910    0.49700    0.01270   11.00000    0.05000
C24   1    0.25450    0.53820   -0.00170   11.00000    0.05000
C25   1    0.25730    0.63580   -0.01390   11.00000    0.05000
```

```
C26    1    0.21400     0.71580    -0.01220    11.00000    0.05000
SAME O11 > C16
PART -1
O31    4    0.58950     0.88470     0.48860    10.50000    0.05000
O32    4    0.50000     0.95790     0.50000    10.50000    0.05000
C37    1    0.53770     0.87270     0.49580    10.50000    0.05000
AFIX 66
C31    1    0.51920     0.75480     0.49610    10.50000    0.05000
C32    1    0.55470     0.66620     0.49100    10.50000    0.05000
C33    1    0.51570     0.55780     0.49670    10.50000    0.05000
C34    1    0           0           0           10.5        0.05
C35    1    0           0           0           10.5        0.05
C36    1    0           0           0           10.5        0.05
AFIX 0
SAME O11 > C16
O41    4   -0.08800     0.05130     0.00970    10.50000    0.05000
O42    4    0.00000    -0.01740     0.00000    10.50000    0.05000
C47    1   -0.03700     0.07310     0.00530    10.50000    0.05000
AFIX 66
C41    1   -0.01990     0.17860     0.00210    10.50000    0.05000
C42    1   -0.05710     0.27550     0.00820    10.50000    0.05000
C43    1   -0.01870     0.38040     0.00340    10.50000    0.05000
C44    1    0           0           0           10.5        0.05
C45    1    0           0           0           10.5        0.05
C46    1    0           0           0           10.5        0.05
AFIX 0
PART 0
```
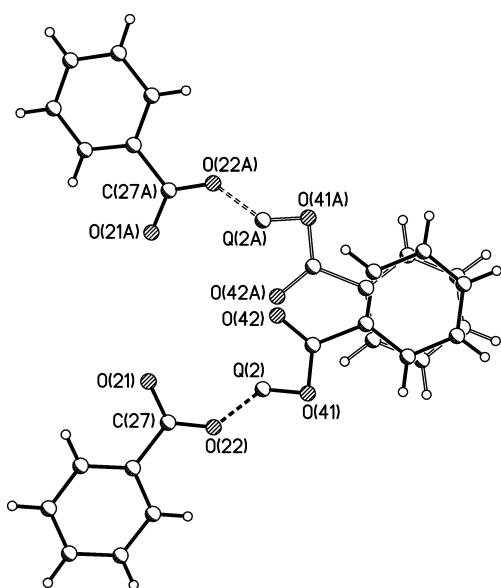
The file benz-02.res contains the complete disorder. Now the **AFIX 66** and **AFIX 0** lines can be removed,[8] and the atoms can be refined anisotropically (add **ANIS**), as the disorder is stable. This has been done in the file benz-03.ins.

In the peak list of residual density maxima found in the file benz-03.res, the hydrogen atoms on the non-disordered phenyl rings appear quite clearly. **HFIX 43** applied to these positions in the next step (benz-04.ins) generates the hydrogen atoms geometrically. In the file benz-04.res the refinement of the disorder is complete. Before publishing the structure, however, there are a couple of questions to be addressed. First: how many benzoic acid (or benzoate) molecules are there per molecule of the drug? The answer is one-and-a-half.[9] There are clearly two complete independent molecules of the drug and two fully occupied and not disordered solvent molecules per asymmetric unit.

---

[8] You will find that the last **AFIX 0** has automatically been removed by SHELXL, as it was in the last line before the **HKLF 4** card and hence meaningless. So there are only three lines left to be removed.

[9] And not two, as many – even experienced – crystallographers might answer. Disordered molecules on special positions are a famous and infamous trap and it is sometimes hard to picture such a scenario correctly.

Also, there are the two disordered half molecules in the asymmetric unit. In the complete unit cell, there are four drug molecules and six solvent molecules, two of the latter located on special positions in a disordered way. The second question is: are the solvent molecules benzoic acids or benzoates? To answer this, the total charge of the rest of the asymmetric unit needs to be taken into account. The drug molecules each bear one positive charge on the nitrogen atom N(1) (the three hydrogen atoms were clearly visible in the difference Fourier for both independent molecules), which makes it necessary that two of the three solvent molecules in the asymmetric unit be benzoate ions and the third one benzoic acid. Looking in the residual electron density for the one missing hydrogen atom (it might actually be disordered over the eight possible positions) and thinking about possible hydrogen-bonding patterns can be a nice weekend-pastime for the inclined reader. Taking *e.g.* the two strongest residual density maxima, Q(1) and Q(2) each as half a hydrogen atom leads to a scenario in which both the disordered benzoic acid molecules connect two benzoate ions in the unit cell, as shown in Figure 4 for one of the two independent benzoate ions.

Figure 4: Possible hydrogen bonding pattern for one of the two independent benzoate molecules. Atoms of symmetry equivalent atoms are labeled with an A after the original atom name.

More practical refinement exercises and SHELXL examples in the same style as this workshop can be found in this book.