Disorder in crystal structures:

new approaches in finding the best model

Inauguraldissertation

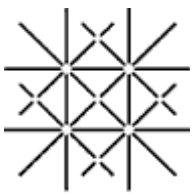
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Markus Neuburger

aus Basel

Basel, 2012



Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Edwin. C. Constable

Dr. David Watkin

Basel, den 21.6.2011

Prof. Dr. Martin Spiess
Dekan

Thanks

This work could not have been completed successfully without many people's help and support.

First of all I would like to thank Dr. David Watkin for his assistance and guidance in all the different phases of the realization of this project. Providing help and advice at crucial points on the way of this development work was very valuable, and different visits in the crystallography lab in Oxford University were very useful in order to communicate the state, the advances and the problems to solve. Also the exchange with his coworkers Dr. Amber Thompson and Dr. Richard Cooper was always very fruitful, and their hints and opinions opened the view on new possible routes to go.

I also would like to give my very big thanks to Prof. Dr. Edwin Constable and Prof. Dr. Catherine Housecroft. To work in and for their research group has always been and continues to be a great experience. To solve the structures for the Chemistry Department including those of the Constable/Housecroft group was not only providing lots of sample data to test and improve the code on, some of the encountered problems and their solutions also gave the ideas for new features and enhancements, ideas that can only be furnished by true every day data. Such a great resource of true cases cannot be simulated by a series of test structures. This very fruitful environment together with the freedom to find solutions using own ways contributed essentially to the success of this project.

Many thanks go to all colleagues in the group and all students from the chemistry department that bring their crystals to the lab in order to know their three dimensional structure. Some of them had occasionally to wait a moment longer for their results if their structure showed a problem as I had to decide whether to just finish their structure in the old way or to try and invest the time in a more durable solution. So sometimes a piece of code had to be finished in order to go another step towards a final solution. This always included the risk that it could take initially longer to finish the work than if I just would have solved and refined the structures in question in the traditional and well known but quite time consuming way.

Thanks go also to Dr. Jennifer Zampese for proof reading the manuscript. She helped to make the text more concise and clear as a look from outside always has a different focus on the subject.

Last but not least I would like to thank Prof. Dr. Ulrich Druwe, vice-director of the University of Basel at the time this project started. He decided to accept my application to start the work on this PhD thesis project at the University of Basel, and with his decision he opened the way for the project to be carried out in the way it presents itself today.

Table of contents

Summary

A new approach has been implemented in the refinement program CRYSTALS [1] that enables the crystallographer to treat structures with disordered parts in an intuitive way, improving the refined models and saving a considerable amount of time. Roughly half of all crystal structure analyses suffer from some kind of problems. Many of them are problems involving disorder, which keep the crystallographer busy, often for many hours and days. Many disordered structures can be described rationally as the molecule taking advantage of one or more of its degrees of freedom. However, an equivalent easy approach is not provided by the traditional way to build crystallographic models for refinement.

New scripts in CRYSTALS now help the scientist to deal with disorder with more ease. The scripting environment in CRYSTALS provides a mechanism for formalising procedures which have to be repeated frequently and partially automating them. The new scripts help the crystallographer to first regularize the geometry of structural fragments that have become distorted during initial refinement of the starting model. After this step, disorder models with two components are set up by applying a non-crystallographic symmetry operator to the prototype regularised component in order to create a duplicate in the alternative position. Finally, the user is assisted in the different steps of refining this two-component disorder model before integrating it into the final refinement of the whole structure.

On the way through the refinement of the disorder the new scripts keep the model consistent, even in the case where extensions of the disorder assembly or modifications to it are necessary. This feature too results in saving time. Searching errors in long hand edited instruction lists used to be tedious and time consuming.

While first tests show promising results the current development is aimed to extend the number of cases this approach can be adapted to. Disorder models with more than two components, assisted model building for disordered solvent molecules, inclusion of non-atomic electron density descriptors in the scripts or the development of validation criteria and tools for the way disorder has been refined are possible fields for future work on the base of this concept.

1 Description and origin of disorder in crystal structures

1.1 Introduction

Structure determination by X-ray diffraction has become a standard analysis in chemical research. While the quality and efficiency of the equipment has improved considerably over the past years and decades and while the power of programs and computers has made remarkable progress in that same period there has remained one domain where crystallographers still spend lots of their time, and this is when the structures under investigation are *disordered*. Disorder is a physical phenomenon leading to uncertainty in the chemical composition, or in the spatial arrangements of atoms.

Unfortunately the CCDC [2] is not a very good source of information when it comes to disorder as authors could fear problems publishing disordered structures. This is why the estimate about the overall percentage of disordered structures that could be obtained from a database search will tend to be too low. In order to get a better estimate of the importance of this widespread problem about 400 structures solved and refined over the last years in the laboratory for Chemical Crystallography of the University of Basel have been examined regarding this issue. Taking into account all cases where the phenomenon can be observed, from the case where the adp's (anisotropic displacement parameters, sometimes also called temperature parameters) are larger than usual without the need of intervention to the extreme case of different kinds of disorder in the same structure, the estimate of about half of the structures being affected by some kind of problems is confirmed.

The laboratory at the University of Basel carries out a few hundred data collections per year, from which 200 to 300 lead to successful structure determinations. The samples come mainly from the field of organic and metal-organic chemistry, while minerals and inorganic samples are rather rare. Samples are returned if the scattering power is not sufficient or if the sample contains no suitable single crystals. The large number of datasets collected contains lots of examples from different kinds of problematic structure solutions and refinements. If the problem is disorder we try to find the best model to describe it. Even though there is no time limit imposed we try to keep the list of unfinished structures short as the overall efficiency tends to suffer from long "to do" lists.

Finding successful solutions for disordered structures is not only useful in terms of time economy, it is also rewarding as it reveals details of the structure under investigation that would otherwise remain hidden. Very often the resolved disorder model opens a new view on the structure that has its own order that was not visible at first sight.

1.2 The standard structure determination: an overview

A structure determination can be subdivided in four steps. The first step in determining a crystal structure is to get crystals of good quality, one of the crucial requirements for a successful structure determination. Even though this point is beyond the scope of this

text and is therefore just mentioned for the sake of completeness the importance of good quality crystals for a successful structure determination cannot be overestimated.

In the second step the crystallographer chooses a sample and mounts it on the diffractometer for the measurement. Materials and techniques have improved to make the process straight forward for most samples. Mounting of the crystals in oil drops that are frozen in the cold gas stream of a cryostat is a widespread method to prevent decay. Moreover this method makes mounting easy, keeps the sample stable during the data collection, and it helps to prevent possible solvent loss that would destroy the crystal. All this helps to create the best possible conditions for a successful experiment. Modern area detector systems are able to collect large amounts of data in a very short time. While traditional serial diffractometers could collect a maximum of about 2000 reflections per day the typical area detector system of today collects easily 10 to 100 times this amount of data in the same time. Advances in computer technology make out of the process of integration of the raw diffraction images a matter of minutes. Like this the data collection time has decreased in an astonishing way from days to hours, and in favorable cases the first results can be achieved in less than one hour. Even if the data collection time is longer the crystallographer is only involved in setting up the experiment and is free to complete other tasks while waiting for the completion, so even if data collection times may vary this is in most cases not critical to the overall efficiency of a crystallographic laboratory.

Also in the field of structure solution, the third step on the way to a successful structure determination, there are big advances, some of them due to new strategies [3], some of them due to improvements to the existing ways of solving [4, 5, 6]. Most times these programs work in a quite robust way being more tolerant than earlier versions to wrong prerequisites like the composition of the compound. Getting a starting model good enough to permit completion and refinement of a structural model is a necessary step on the way of a successful structure determination. Nevertheless it has to be mentioned that the influence of the crystallographer on how well these programs work lies mainly in providing good quality data from a good quality crystal. The computations themselves are very fast, and in most cases one program out of the available choice will work well enough to be able to continue.

In the last step in the structure determination the structure has to be refined in order to get the best fit between a structural model with the structure factors calculated from it and the structure factors observed during the experiment. With ordered structures this task may be finished in a short time with standard tools. If the structure contains disordered parts then the time needed for the completion of the refinement becomes a quantity that is very hard to estimate. It is a fact that crystallographers have spent and will spend days and weeks on the refinement of difficult structures. Figure 1.1 illustrates how this fact happens to make the refinement part the most unpredictable step in terms of the time invested on the way to a finished structure determination.

Some numbers about the time needed to complete a structure determination are

summarized in the following graph. The given amounts of time for each step are estimates based on many years of experience. The important message this graph illustrates is that refinement of a few problematic disordered structures is keeping the crystallographer busy for most of her or his time, and the question is now how this time can be shortened. This is not only an economical issue. Making the solution of disorder problems easier and faster the crystallographer is again free to apply her or his knowledge to the solution of new problems. Disorder is of course not the only reason for difficulties. In those cases where it is indicated that disorder could be the reason of the problem, some support to deal with them faster will clearly be appreciated.

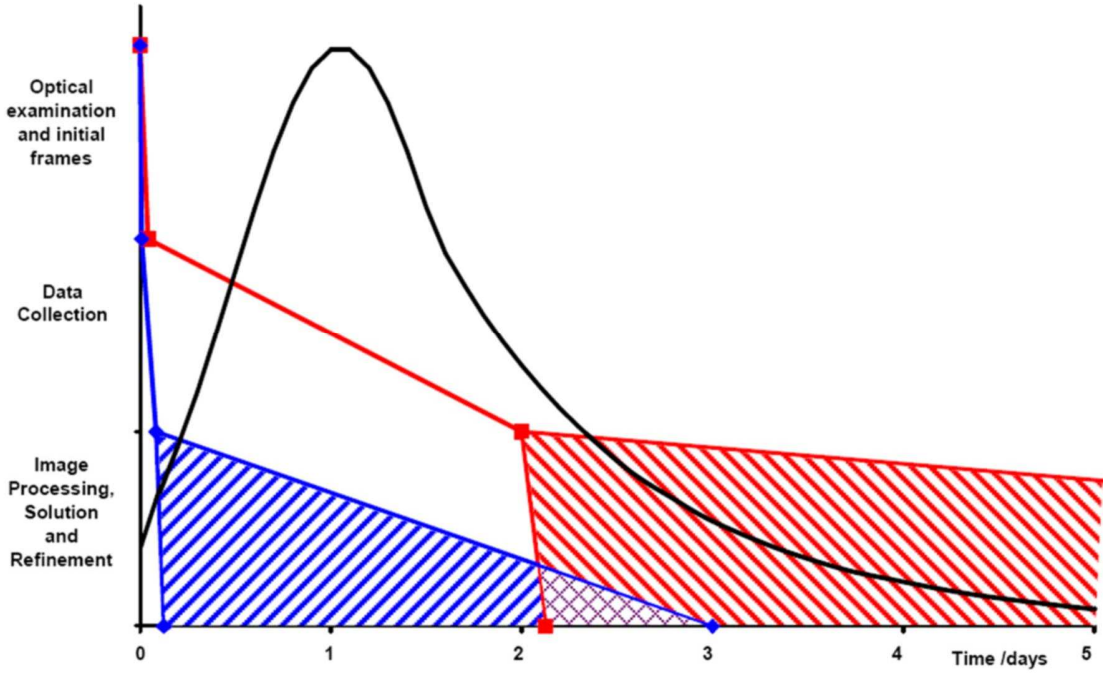


Fig. 1.1: Estimate of minimal and maximal times needed for a structure determination as determined by the crystallography lab in Oxford University

1.3 The “perfect” crystal vs. the real crystal

During the data collection for a single crystal X-ray diffraction experiment using a modern area detector system the crystal, after having been mounted and centered on the diffractometer, is rotated while being exposed to the radiation. While most of the primary beam is flying straight through the crystal and is absorbed by the beam stop mounted behind the crystal the diffracted radiation is detected by the instrument and recorded as images that are called frames. During this process the Bragg condition is usually fulfilled various times for each frame, and each time this results in a spot being detected by the system. In the standard case to each of these spots a set of Miller indices can be assigned, and intensity information can be extracted from the recorded frames.

Using the ideal and perfectly ordered crystal for our experiment, all atoms would be perfectly in line with all other equivalent atoms of the structure. Diffraction intensities would only be observed when the Bragg condition is fulfilled and would be zero anywhere else. The width of the Bragg peak would be very narrow, ideally point like.

Real crystals grown in the laboratory or created by nature are never perfect. First of all there is the phenomenon called mosaicity that describes the fact that a crystal always consists of its mosaic blocks that are misaligned in relation to the idealized lattice by typically 0.3 degrees. Together with the divergence of the primary beam that is typical for most laboratory X-ray sources it is mainly the mosaicity that is responsible for the width of the reflections, thus their size in the diffraction image recorded by the diffractometer. Moderate mosaicity is not a bad feature of the crystal as zero mosaicity would lead to problems like multiple scattering of the X-ray beam. The crystal with the right mosaicity is therefore also sometimes called the “perfectly imperfect crystal”.

Other imperfections of the crystal may consist in impurities that have been incorporated during the growth of the crystal, or two initially independent crystallization nuclei can grow together having thus a zone where the lattice is not periodic. Other heterogenities and imperfections may lead to a considerable part of the diffracted intensity that is not following Bragg’s law and that is found in between the Bragg peaks as diffuse scattering. The kind of imperfections or disorders that cause diffuse scattering is beyond the scope of this work and will not be dealt with in more detail due to that reason.

1.4 Classification of disorders

There are three major classes of disorder that will be described in the following pages: dynamic, static and substitutional disorder. After that disordered structures are classified following the criterion where the disorder occurs.

1.4.1 Dynamic disorder

Atoms and molecules are never fixed. They are vibrating, and these movements are increasing with temperature. The so called Brownian motion has first been observed in 1827 by the Scottish botanist Robert Brown observing the unexplainable motion of small particles from plants held in an aqueous suspension. Later it could be shown that the description of this motion is a good model to describe random physical processes. In the case of molecules these movements are very fast, usually in the order of magnitude of 10^{12} and 10^{14} Hz. As exposure times with standard equipment in laboratories are between a few seconds and some minutes it is clear that these experiments cannot resolve these motions of the molecules. The refined structure therefore represents the average of all the possible positions of all the atoms being involved.

In the case of a standard structure determination the motion of the atoms is described as a harmonic oscillation of the atom around the coordinates that have been attributed to that atom. This atomic model is called the “independent atom model” or IAM model.

The electrons are assumed to be distributed spherically around the nucleus in this model, and the only deformations it describes are those caused by thermal motion. Its big advantage is that it is simple and uses only nine parameters to describe the electron distribution around the atom: three coordinates and six displacement parameters. Depending on the environment of a given atom and on the temperature the thermal motions can become very large. As the IAM model does not account for effects imposed by the chemical bonding it comes quickly to its limits. While the effective motion of the atoms is usually on some circular path it tries to describe the electron density of the vibrating atom using a linear harmonic motion resulting in residual electron density higher than desired. This situation is called **dynamic disorder**. It is usually sensitive to temperature and its effects can be decreased by lowering the ambient temperature of the data collection. Figure 1.2 shows an example of a data collection carried out at a temperature too high to achieve good separation of the individual atomic sites.

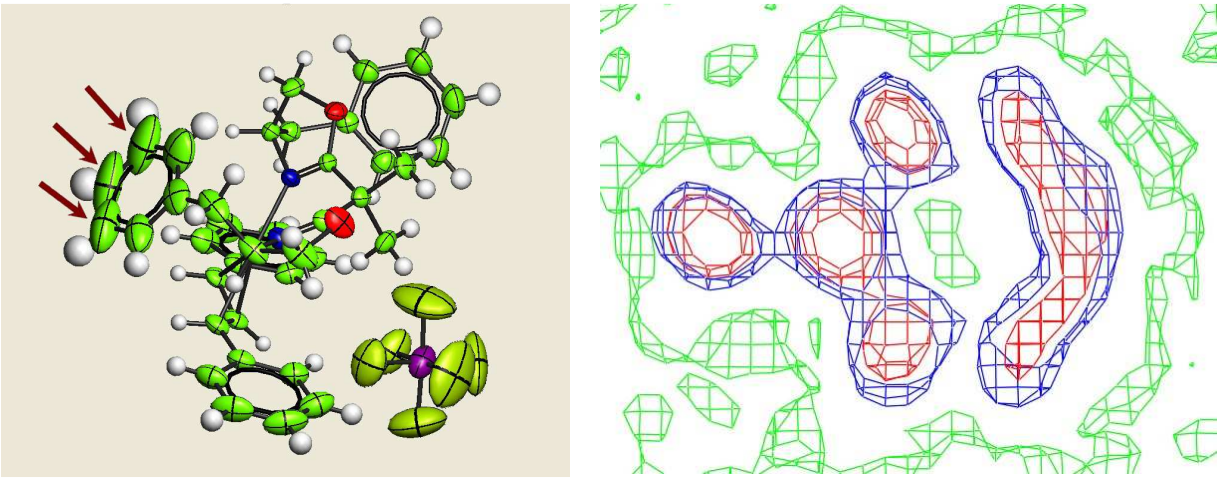


Figure 1.2: The structure of a Palladium complex measured at 250K. In particular the phenyl ring shown left is affected by dynamic disorder. The slant Fourier map through the plane defined by the ring shows how the electron density is spread over the three peripheral positions of the ring. The electron density for the three atoms indicated by the arrows on the left side shows no separation between the atoms. (Structure 1, published in 1995, [7].)

1.4.2 Static disorder

During the process of crystallization every molecule that is joining the crystal has to find the best position with lowest energy itself. It is possible that there is more than one way with comparably low energy to join the crystal. A peripheral group with rotational freedom may have similar space requirements in different orientations. There may be more than one possibility to interact with the local environment for a co-crystallizing solvent molecule, or the absence of interactions may give a molecule or molecular fragment the freedom to orient randomly.

As the structural fragments that vary are not explainable as the oscillation around a common position caused by thermal motion their contributions to the final diffraction pattern are different, even if only the resulting sum can be determined. This is called **static disorder** as the disordered parts will remain in their positions even if ambient conditions are changed. A dynamic disorder will be influenced if temperature is lowered. Decreasing the temperature in the case of a static disorder will of course also decrease the displacement parameters of the disordered part, but the disorder itself will not be affected.

This was the case for the complex shown in figure 1.3. The allyl group can bind to the Palladium atom in two orientations distinguished by the position of the central carbon atom pointing up or down respectively. In solution these two conformations are in equilibrium. By NMR studies in solution at different temperatures it could be observed that lowering the temperature changed the equilibrium in favor of one of the two conformations, and at about 200K the signals of the less frequent conformation disappeared completely. Crystals grown at room temperature were used for the structure determination. Repeating the experiment at different temperature showed no change in the ratio of the site occupancy factors of about 3:1 which illustrates nicely the static nature of the disorder in the crystal structure.

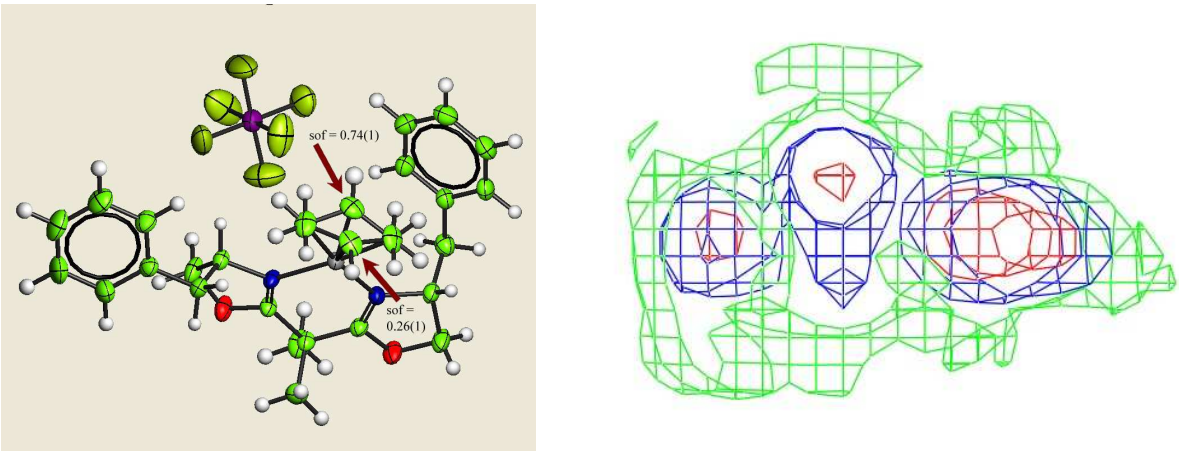


Figure 1.3: The structure of a Palladium complex measured at 250K. The allyl group is statically disordered, the site occupancy factors of the two positions of the central Carbon atom refined to 0.74(1) and 0.26(1), the two positions are marked by arrows. The Carbon atom with higher site occupancy can be seen well in the slant Fourier map and the position of the less occupied central carbon atom is separated from it. (Structure 2, published in 1995, [7].)

In the absence of long range order there is no way of separating the molecules with different conformation and the result is that we see the variations of the structure as a superimposition of all conformations present. It is not entirely clear why in some cases there is long range order in the local variations and in others not. The speed of crystal growth, the pureness of the compound to crystallize and the ambient conditions during

crystallization may be reasons that can be in favor or in disfavor of building long range order.

1.4.3 Substitutional disorder

There is another case of disorder that can be observed mostly with inorganic materials like minerals. Different elements with similar binding properties can occupy equivalent places in a structure. This type of disorder is therefore called **substitutional disorder**.

In order to cope well with this type of disorder accurate knowledge of the composition is extremely helpful. Substitutional disorder is static by nature. Amongst the difficulties to refine such disorders is the fact that the bond length to the environment may vary resulting in bad displacement parameters. Figure 1.4 shows the structure of the mineral Jentschite where Antimony can take the place of Arsenic.

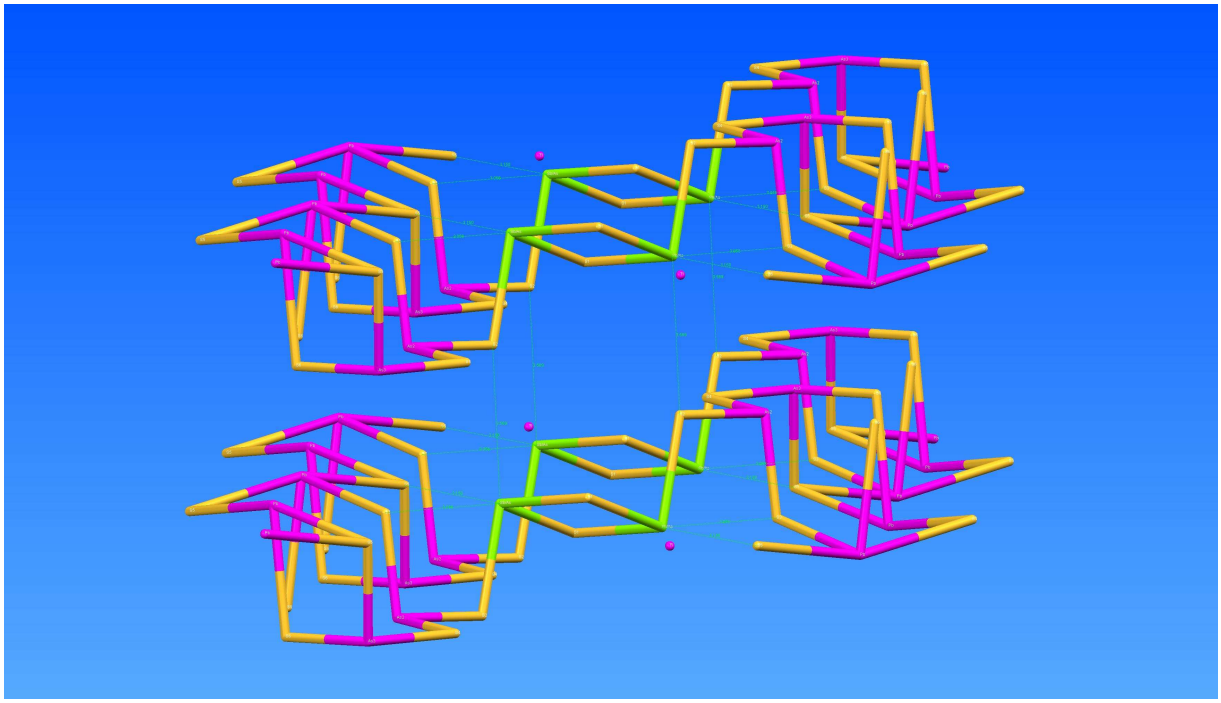


Figure 1.4: Jentschite, the substitution of Arsenic and Antimony can be observed at the atomic site colored in green. (Structure 3, published in 1996 [8], intensity data measured in the X-ray laboratory of Basel University.)

There are examples in organic structures that could be described as substitutional disorder. The frequent case of a thiophene ring with the sulfur position being disordered between the two adjacent positions to the connecting bond could be described as a substitution of a Carbon atom by a Sulfur atom. Figure 1.5 shows a typical example of this disorder. But it is probably better to describe this phenomenon as the result of effects taking place during crystallization. As the space requirements for both

orientations are very similar and as there are no significant interactions guiding the crystallization process some of the thiophene rings will be oriented in one way while the rest will take the opposite orientation. Substitutional disorder suggests the substitution in the moment of the synthesis or crystal growth. Even though a Sulfur atom takes the place of a carbon atom and vice versa the different orientations of the thiophene ring are a result of the crystallization process and would therefore be rather classified as a static than a substitutional disorder.

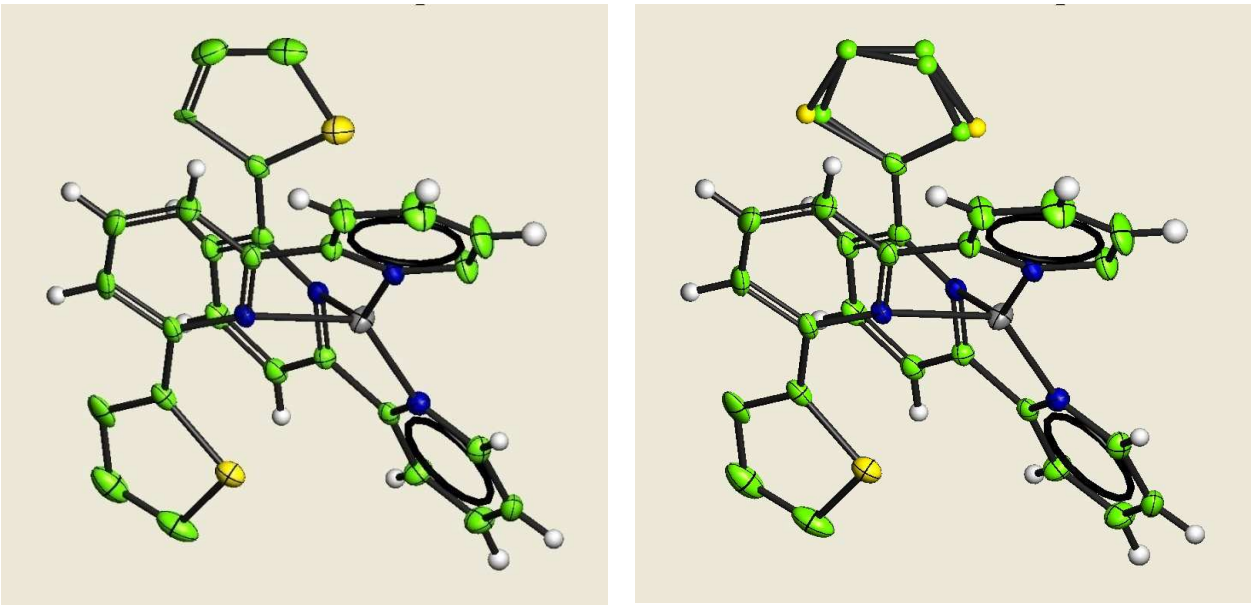


Figure 1.5: Disordered thiophene ring as seen after initial refinement and a disorder model with the original positions rotated around the bonding axis by 180 degrees. The second thiophene ring visible in the bottom of the image shows similar signs of the disorder described, and in fact it is possible to refine also this ring with a similar disorder model. The result is a second group of atoms with a site occupancy factor of about 5%, and not much can be said about the correctness of the model. 5% of a carbon atom is in the region of electron density where we only expect the residual electron density, therefore it is difficult to conclude if the assumption is correct or not. (Structure 4, unpublished work.)

1.4.4 Solvent disorder

Disordered structures can also be classified on the basis of what part of the structure is disordered. As a rule of thumb it may be stated that the smaller the molecule that we want to look at the bigger the probability that it moves with temperature. Thus disorder in solvent molecules tends to be more frequent than disorder in bigger molecules. This problem is of particular interest and importance in macromolecular crystallography and in protein crystallography where it is quite frequent that up to 40% of a structure is simply filled with water molecules. The crystal is behaving like a sponge. Of course hydrogen bonds will build up, but they will, as long as the crystal is not cooled down,

change their orientation and with this the whole network structure will be different.

Even though it is rare to find such high percentages of included solvent molecules in small molecule crystallography these structures exist as shown in figure 1.6. Solvent disorder sometimes presents tough problems to solve, and the difficulty is mostly due to the fact that disorder together with high thermal motion leads to flat electron density distributions where it is difficult to get started by assigning atom types to residual electron density maxima.

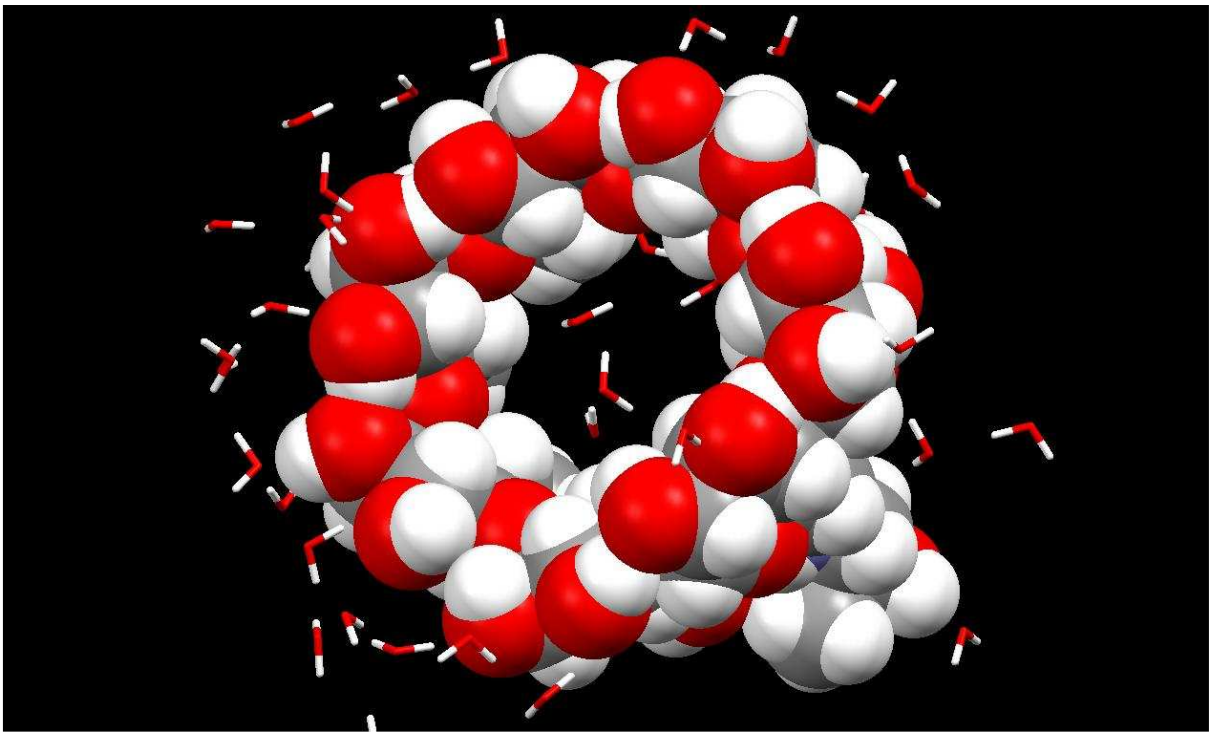


Figure 1.6: This structure contains 16 water molecules for one of b-cyclodextrine moiety. The molecules shown are only those that could be localized in the density map. (Structure 5, published in 2004 [9].)

As the refinement of disordered solvent molecules, in particular in macromolecular structures, may remain without good results special tools have been developed to find solutions that do not require a true atomic model for these solvent accessed areas. These will be described later in this work.

1.4.5 Partial fragment disorder

If a larger molecule is affected by disorder then it is usually only one or more parts that are disordered while the bulk structure is ordered. Looking at the magnitudes of the displacement parameters it can be shown that usually U-values tend to be smaller in the center of the molecules where the degrees of freedom are restricted while peripheral

groups make more frequent use of their degrees of freedom. But this observation is too general as there are examples of disorder being found in the very centre of molecules. A very nice example is shown in figure 1.7.

The fact that there are large parts of the structure that are ordered is usually a good indication that it is a true case of disorder that is under investigation. When large areas become suspect of being disordered it is increasingly difficult not only to handle the problem as a disorder problem, but it becomes also more and more likely that another reason causes that problem, for instance a wrong choice of the space group.

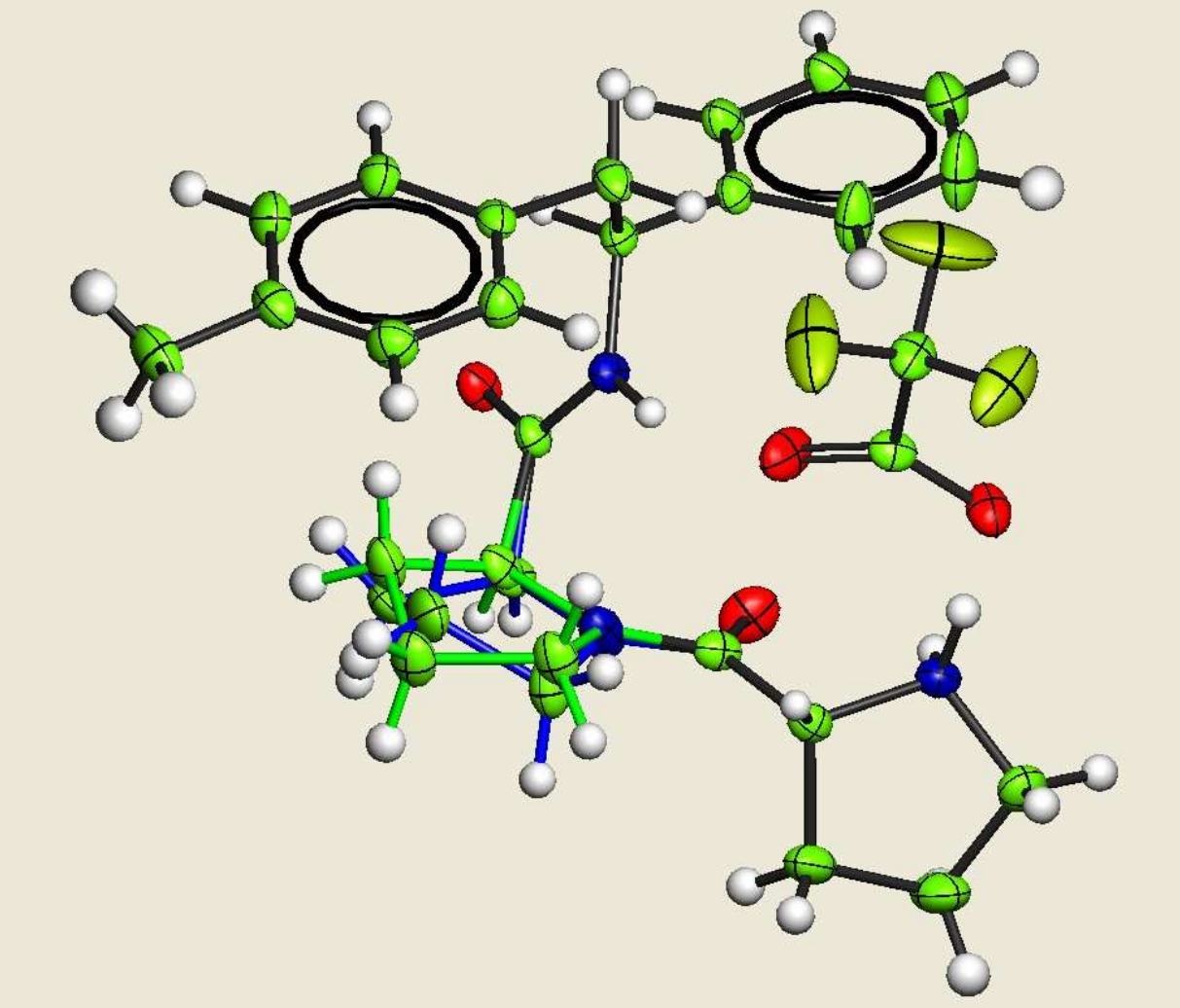


Figure 1.7: In this structure the disorder is located in the center of the molecule. The site occupancy factors of the two disordered parts refined to 0.509(5) (green) and 0.491(5) (blue). (Structure 6, unpublished work.)

1.4.6 Whole molecule disorder

It is in principle also possible that a whole structure is affected by disorder problems. However there is, as already mentioned, the strong suspicion that something else than disorder might be the true reason for the trouble and that limitations in the data collection equipment or in the crystal quality makes it impossible to see the true reason. In the hypothetic case of a modulated structure from a poorly diffracting crystal with satellites making up 1% of the total diffracting power of the crystal it is very easy that the investigator simply misses the extra observations, but without the additional data the modulation functions cannot be determined and what we see is apparently a disordered structure.

The structure with the highest percentage of disordered atoms observed in the crystallography lab in Basel was the structure of a metal organic framework built of metal ions and a bridging ligand. It is shown in figure 1.8. While the metal ions and the counter-ions were ordered the halogenated ligand showed two orientations throughout the whole structure.

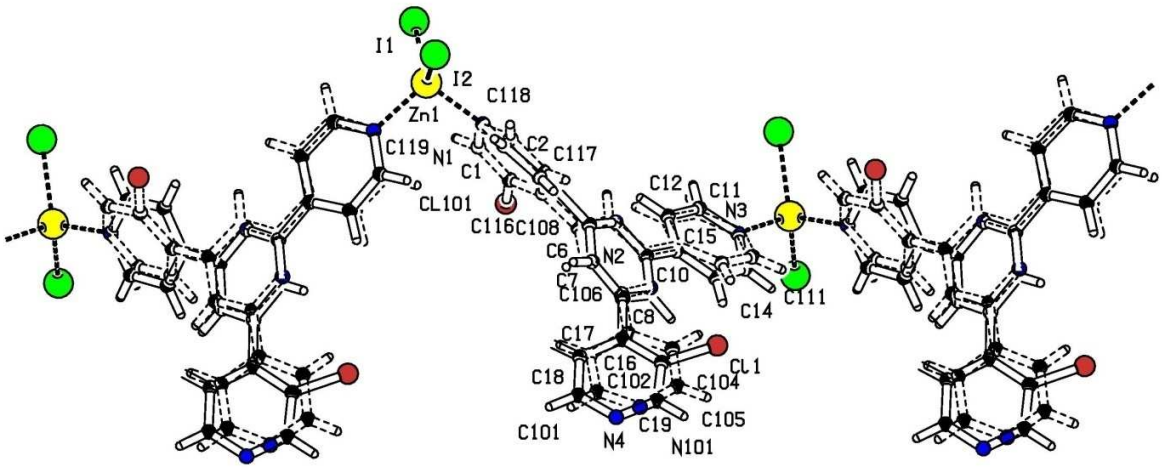


Figure 1.8: One entire ligand molecule is disordered in this structure. The less frequent arrangement of the ligand is shown with dashed bonds. (Structure 7, unpublished work.)

1.5 Disordered structures and structures without obvious long range order

A structure may look as if it would be disordered, but in fact the current description of the structure is incomplete and does not take into account other possible elements of ordering that may not be obvious to spot. In the following a few of the most frequent ones will be described.

1.5.1 Structures with Z'>1

The static disorder has been described as the result of competing environmental conditions when the different possibilities to join the crystal are chosen randomly. It is possible that every possible arrangement is equally frequent, and their occurrence is alternating. If the structural features of the different contributors are similar and if the molecules are arranged in a pseudo-translational way, reflections along the direction of the alternation will be weakened in a systematic way. If the number of variants is n every reflection with the Miller index x*n along the involved axis will be observed while the ones in between will be weakened. The correct way of treating this case is to re-determine the unit cell constants taking into account weak reflections and to extract the intensity information again with the correct unit cell. The solution of such a structure will show more than one molecule in the asymmetric unit that cannot be related by crystallographic symmetry. The number of formula units in the asymmetric unit is also referred to as Z', and in this case we have a structure with a value of **Z' > 1**.

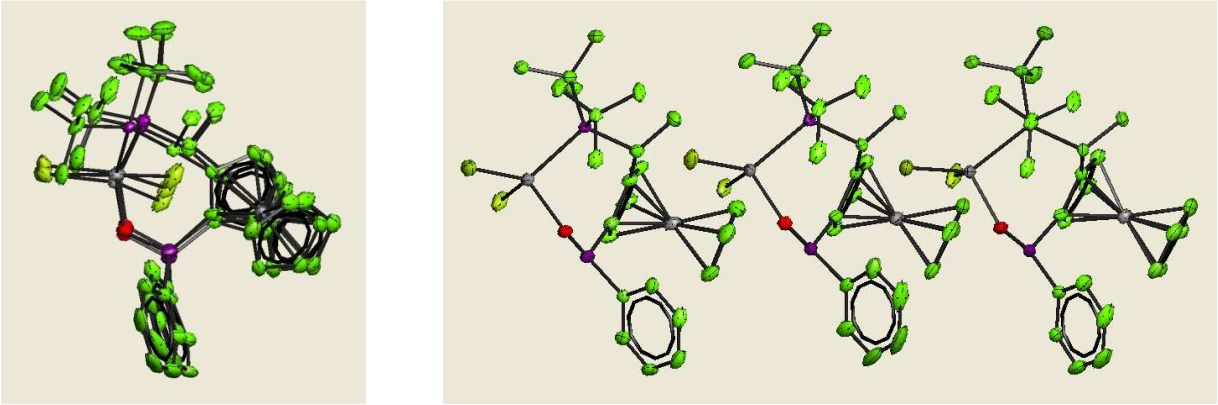


Figure 1.9: The unit cell of this structure was not determined correctly in a first run resulting in a structure that looked as if it would be disordered. Here the correct solution is shown. The apparently disordered structure appeared as a superimposition of all three slightly different molecules, the picture on the left hand side gives a good impression of this situation. (Structure 8, unpublished work.)

Even though the percentage of structures with Z' > 1 has probably always been equal it seems that modern equipment improves the chances of detecting the additional observations needed to determine these structures correctly. This apparently results in an increase of the number of structures with Z' >1. It is always recommended to test carefully if the pseudo symmetry is not matching missed higher space group symmetry.

1.5.2 Z'>1 and commensurately modulated structures

Alternatively the same structure can be described as a modulated structure. The difference is that our model is made up of an average position for each atom and a modulation of the atoms around their corresponding average position. In the case of the

structure shown in the figure above we would have the case of a commensurately modulated structure as the modulation has a fixed sequence of three states and returns to the origin. The original small unit cell would have to be taken for the integration of the frames and the weakened peaks in between the strong peaks would have to be integrated as so called satellites. In order to describe a modulation a new dimension has to be added to 3D space. For each direction in 3D space along which a modulation is detected a new dimension is created. The number of indices used to index reflections and satellites depends on the number of dimensions needed to describe both three dimensional space and the modulations and is restricted to three plus three dimensions as a maximum. The concept of modulated structures has been first described by P. M. de Wolff in 1974 [10], and in this paper the term superspace is used to give the construction a name. Figure 1.10 illustrates how the modulation works and why, if not treated, the structure may look as if it would be disordered.

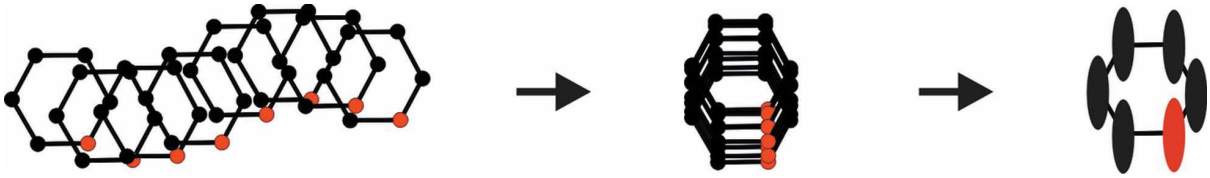


Figure 1.10: A simple case of a modulation and its effect on the structure if the modulation is not detected and not taken care of properly. (Illustration by T. Wagner and A. Schönleber, published in 2009 [11]).

In this case there is the choice of how to describe the structure. The description as a structure with Z' > 1 has the advantage that the usual techniques of structure solution can be applied. To describe the structure as commensurately modulated the integration of the frames needs advanced options of the integration program to assign the additional indices to the satellite peaks. Structure solution and refinement programs must be chosen accordingly. There are special space groups in superspace in order to take into account the additional dimensions, and the refinement of a modulated structure needs specialized skills and training. In the case of the example with Z' = 3 shown in figure 1.9 where the refinement in three dimensional space is straight forward there is no need to choose the more complicated way of refining as a modulated structure. If Z' is growing the situation may change quickly.

If the molecules in a structure with Z' > 1 line up in almost perfect translational symmetry a part of the reflections along the axis in which the near translational symmetry is observed will be weakened. This can be the reason why these reflections are missed when indexing the crystal. These weak reflections can in extreme cases become too weak to be measured accurately. If the number of parameters to refine is growing while the number of observed intensities is small it may become difficult to refine such a structure in three dimensions. Choosing to refine the structure as commensurately modulate has a great advantage in this case. There is one molecule with its physical parameters to be refined, in addition there are the parameters of the atomic modulation functions to be refined for those atoms where a modulation is

detected. This is in general a net advantage compared to refining three sets of possibly highly correlated parameters.

It is worth noting at this point that there are also examples where the description as commensurately modulated structures can be useful to get a better understanding of a structure or of a series of structures even if the single structures can be determined in three dimensional space. The advantage of such a description is clear as it makes it much easier to spot common features and differences in the structures when describing the differences as modulations of a common base structure.

1.5.3 Incommensurately modulated structures

If the modulation function does not have a periodicity that falls together with a grid point of the average structure after a finite number of unit cells the structure is said to be incommensurately modulated. In this case there is no choice. The satellite peaks need to be indexed and the modulation needs to be determined for a proper description of the structure. Usually the modulations are not very big, and this means that the satellite peaks are weak in relation to the main peaks. The diffraction power of all satellites may be in the range of a few percent in relation to the main peaks, which can make it very difficult to observe them. The average structure may look weird in such a case and it will in most cases not be possible to find a good explanation. New detectors that are more sensitive will probably make the number of recognized modulated structures increase.

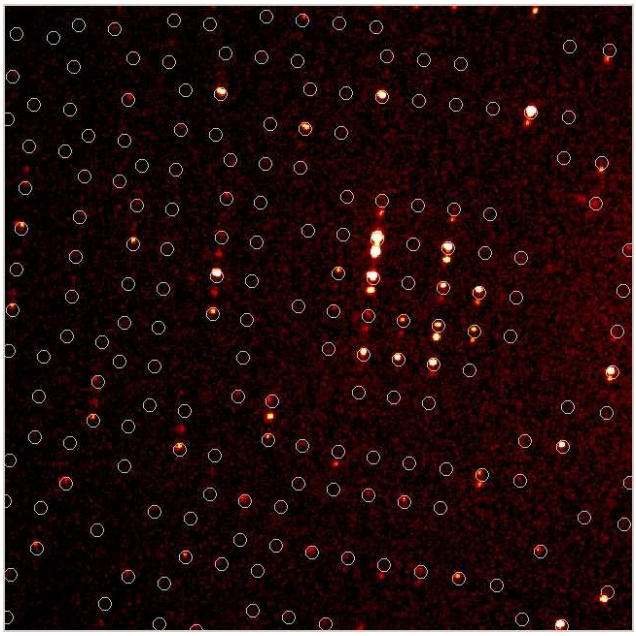


Figure 1.11: One frame out of the data collection of an incommensurately modulated structure. The white circles show where the main spots are expected, the satellites are located in equal distances from the main reflections. (Unpublished work)

It is also possible to find incommensurately modulated structures where the satellite reflections are very well visible and cannot be overlooked. Figure 1.11 is a frame of a data collection that has been carried out in the X-ray laboratory of Basel University. The so called average structure could be solved while the refinement of the modulated structure is still to be finished.

For this thesis, it is important to state that not everything that looks like disorder is disorder in reality.

1.5.4 Twinned structures

Another problem related with symmetry may be the reason why a structure looks as if it would be disordered. But in this case, in contrast to the case with wrong space groups where too much symmetry is imposed on the structure, there is not enough symmetry present to describe the structure fully. This is the case with twinned structures, which are crystals in which there are differently oriented unit cells related by symmetry called the twin laws. Again we see a superimposition of two or more possible orientations of the same structure. The elegant thing with the twin law is that, once determined, it solves the problem by refining only one additional parameter for the twin element scale factors.

Sometimes the shape of the crystals reveals that the sample is composed of more than one lattice, some examples are shown in figure 1.12. This can also be a hint when doing structure determinations, but of course the absence of external features is no proof that the crystal is a single crystal.

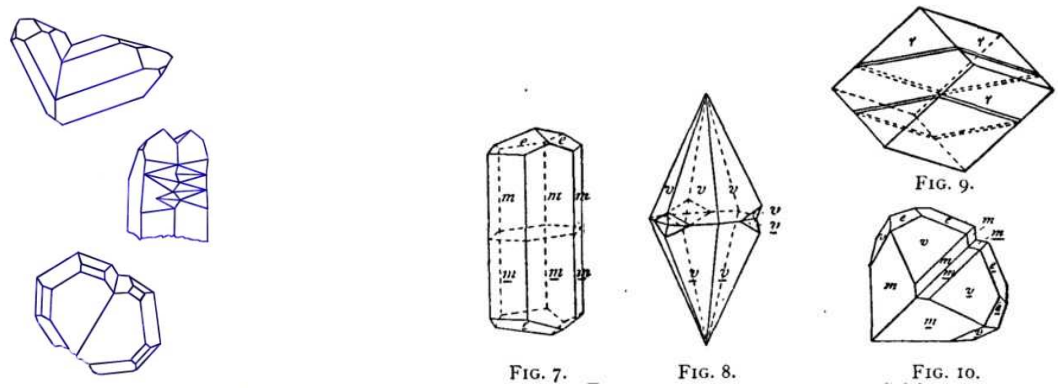


Figure 1.12: Drawings of twinned crystals of quartz and calcite found in Encyclopedia Britannica and on a website called “Virtual Geology Museum” respectively.

Twinned crystals can be subdivided in merohedral and non-merohedral twins. While for merohedral or pseudo merohedral twins there is for each reflection in one component a reflection in the other component that overlaps with it, even though the reflections have their own miller indices in their respective unit cells, and the intensity they contribute to the final observation can be very different. Because of this such twins have an unexceptional diffraction pattern that shows no traces of the twinning. In spite of this problem it is possible to detect and resolve the twinning in later stages of the structure determination (Rotax [12]).

For non-merohedral twins, that is, twins where the contributions of the different components do not overlap, twinning may become obvious when the diffraction data fails to index. With persistence it may be possible to individually index the interpenetrating lattices. All components can then be indexed and integrated separately which gives a good chance to solve and refine the corresponding structure successfully.

1.5.5 Composite structures

At the end of this overview one type of structures that has also a different concept of ordering which goes beyond our intrinsic understanding of order in three dimensions should be mentioned. In the case of this type of structures there are host and guest molecules, and they have both their own periodic structure, but these two structures behave mostly independent from each other. Host and guest structure have each their own unit cell with independent Miller indices, both have their own space group and for each lattice the data must be extracted separately. The refinement program has to cope with the presence of the two datasets, and for each physical parameter it must be stated to which of the two substructures it is contributing.

1.5.6 Conclusions

Composite structures are not that frequently found, but the concept of composite structures is possibly a starting point for the solution of the problem that we find structures that are perfectly ordered in a high symmetry, but have solvent accessible areas that show almost no symmetry. As this type of disorder is not the focus of this work the idea is not developed further at this point. The concept of the two structures in one single place will be taken up again later as a possible way of dealing with areas filled with solvent molecules.

Apart from that the detailed description of structures with Z' > 1, modulated structures, twins or composite structures is beyond the scope of this work as these structures can be treated as ordered as soon as the ordering principle can be identified.

1.6 Pitfalls and borderline cases

When dealing with disordered structures it is sometimes difficult to make the right distinction for structures where the order has not been detected yet. In the following some cases from everyday work are describes as examples.

1.6.1 Crystal quality

In the crystallography laboratory at Basel University we happened to have crystals of tolylterpyridine several times. The first crystals showed disorder with two orientations of the phenyl ring suggesting edge to face contacts between these rings in the crystal packing. New crystals showed unit cell parameters with one axis doubled in size in respect to the original unit cell parameters and the structure appeared to be ordered. Now also the adp's in the part formerly declared as disordered were looking much more reasonable. As the different experiments were carried out distributed over a period of several years it is probably impossible to trace back if the first crystals would have permitted the measuring of the additional observations needed to find the correct unit cell if more time was invested in data collection, or if the quality of the crystals was just better in the last attempt.

This example shows that even with modern equipment wrong unit cells are causing problems from time to time. Some years ago when serial diffractometers with point detectors were widely used it was common to miss weak reflections in the initial stages of an experiment when determining the unit cell, especially as the peak search routine was searching strong reflections. A series of reflections weakened by a near translational symmetry could easily be missed in this way.

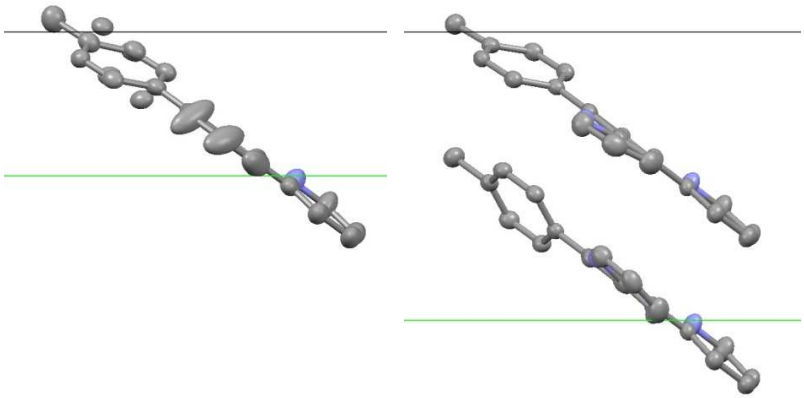


Figure 1.13: Disordered and ordered structure of tolylterpyridine: the ordered structure reveals more details and gives the explanation for the strange adp's of the disordered structure where only part of the disorder has been resolved (Structure 9, published in 2007 [13])

1.6.2 Shock freezing

Another point arises from the very popular method of shock freezing crystals on the diffractometer when mounting the crystal for data collection. Michel Dusek, co-author of the crystallographic computing system Jana2006 [14], spoke at the "5th Workshop on Structural Analysis of Aperiodic Crystals" that took place in March 2007 in Bayreuth about a structure of a calixarene that showed disorder when shock freezing the crystal. By chance, as there was a technical problem with the cooling device, the crystal which had remained mounted on the diffractometer was slowly cooled down for a second data collection, and this time, with the same crystal, the structure was modulated, thus ordered. (Michal Dusek, fzu, Prague, unpublished work.)

During mounting of a sample in the crystallographic laboratory in Basel a crystal was literally destroyed by the shock freezing. The change could be observed with the optical system of the diffractometer, and the shock frozen crystals did not show any diffraction. Trying to mount and measure the crystal at a higher temperature gave a good dataset, and the structure could be identified as a new polymorph of a structure determined earlier. By cooling down gently the crystal survived the phase transition and the structure showed to be the known polymorph. The phase transition could be repeated two times in both directions while the crystal did not survive the shock freezing.

Unfortunately most of the time the parallel experiment will not be carried out in order to know to which extent the cooling is affecting the order in the sample. It is still good to keep in mind that shock freezing may affect the crystal in unexpected ways.

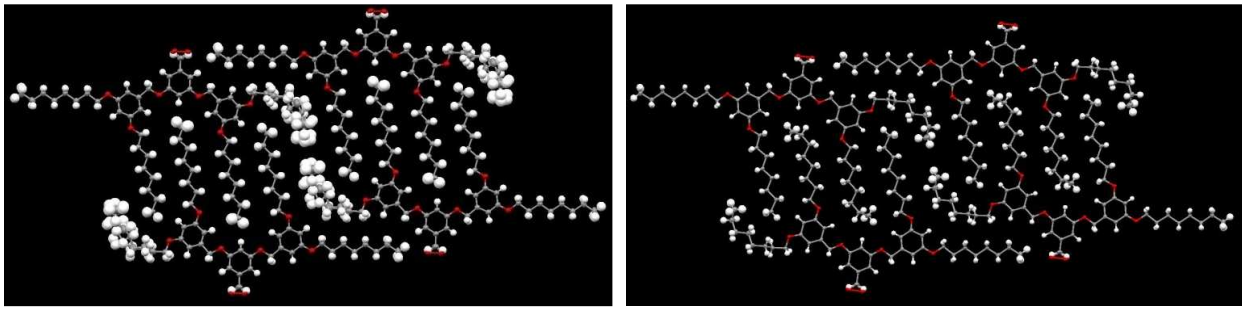


Figure 1.14: Both phases of a structure where a reversible phase transition from a disordered high temperature structure measured at 223K (left hand side) to an ordered low temperature structure measured at 123K (right hand side) could be observed. (Structure 10, unpublished work)

1.6.3 Wrong space group

It can happen that space group determination is ambiguous. Two chiral molecules may look at a first glance like they are related by an inversion center. It is clear that if the compound is in fact chiral it is not possible that the molecules are related by inversion as this would imply the presence of both enantiomers. But the bulk structure may behave as if there would be an inversion. To add an inversion centre would imply a space group change, and we know from the chemical information we have that the centrosymmetric space group must be wrong. Such a case of an ambiguity between the space groups P1 and P-1 is shown in figure 1.15.

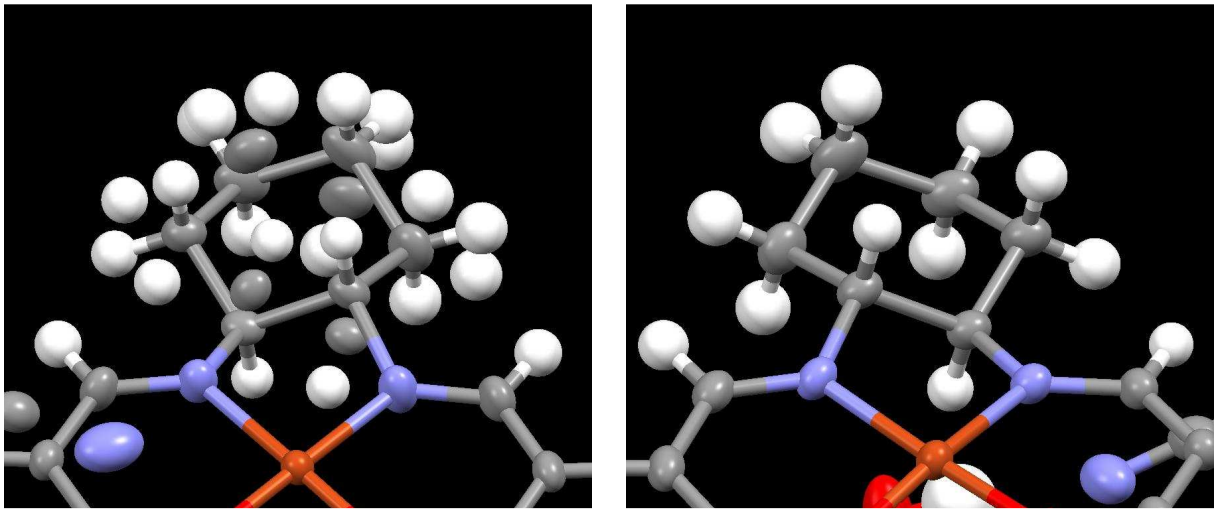


Figure 1.15: Refinement in the space group P-1 led to apparent disorder, while refinement in the correct space group P1 gave a perfectly ordered structure. The apparently disordered atoms are shown on the left image as isolated atoms. (Structure 11, publication in preparation)

Nevertheless it may be possible to solve the structure in the centrosymmetric space group. The refinement possibly shows at some advanced stages that something is wrong. Supposedly sp3 hybridized carbon atoms may appear to have the geometric features of sp2 hybridized atoms, and their displacement parameters may look like those of disordered atoms. Thus it is always worth checking if a space group change could be the solution to the problem before investing time in a laborious disorder refinement.

It needs to be pointed out that the ambiguity between space groups can also occur with two space groups that are both chiral, and the choice may not be as clear as desirable.

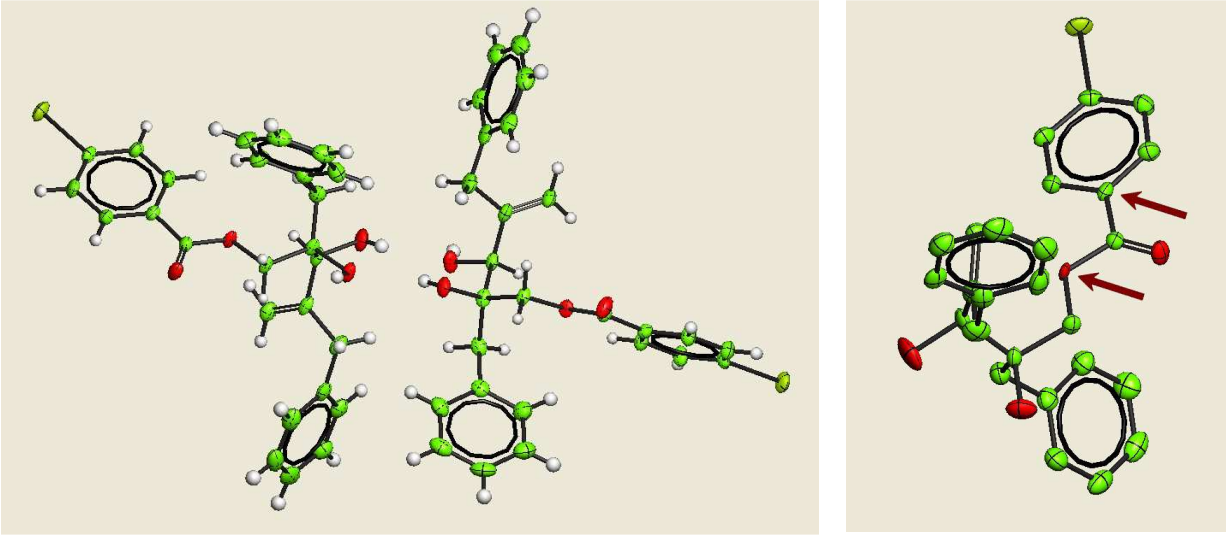


Figure 1.16: The structure on the left hand side is refined in P1. The twofold axis between the two molecules added by the space group C2 would cause disorder. Moreover the refinement in the space group C2 that is shown on the right hand side has strange anisotropic displacement parameters. (Structure 12, unpublished work)

The structure shown in figure 1.16 was solved and refined in P1 with Z' = 2. The validation suggested a transformation to C2, as the two molecules seemed to be images of each other produced by a two-fold axis. This change could be made at the cost that the ordered structure in P1 became disordered in C2. One Oxygen position would not refine in a satisfactory way, and the hydrogen bonding became disordered. Merging the data in the monoclinic system gave a merging R factor of 11.8%, while in the triclinic system it was 3.6%. With +0.9 and -1.9 e/Angstrom^3 residual electron density maxima and minima were higher in C2 than in the refinement obtained in P1 where the same values were +0.7 and -1.0. R-values were comparable for both space groups with 7.1% in C2 and 6.2% in P1. Also the Flack parameter was in the same region with 0.09(3) for C2 and 0.034(18) for P1. All reported values are improved going from C2 to P1, but it is the merging R value together with the fact that the structure is ordered that are most in favor of the lower symmetry space group.

1.7 How to detect disorder

It has already been mentioned in different places that it is very often it is the displacement parameters that give the first hint that the structure could be disordered. But the adp's do not give the answer to the question of the nature of the disorder. Sometimes static disorder can be spotted directly by residual electron density peaks that cannot be explained in a chemically sensible manner within the molecular fragment in question. But when two contributors come very near it is almost impossible to really see any difference from the electron density pattern produced by a dynamic disorder. There are a few possibilities that can allow a clearer distinction. They are all demanding in terms of experimental time and skills.

1.7.1 Experiments at different temperatures

One way of getting hints on the nature of disorder is to try to collect data at different temperatures. While a dynamic disorder problem should become less pronounced with temperature decrease a static disorder will mostly remain untouched by temperature changes with the exception of the temperature dependent dynamics of the individual disorder groups. Nevertheless it is time consuming to carry out different data collections from the same compound at different temperatures, and the efficiency requirements of most labs will not allow this kind of investigations on a regular basis. Most of the time low temperature will be used always or never, depending on the availability, and if low temperature data collection is used, then it will be at a standard temperature that the data will be collected. So it will be rare, apart from special opportunities, that we will get a series of the same structure from the same crystal at various temperatures.

In Basel University such a series has been recorded of a structure going through a reversible phase transformation from a statically disordered high temperature structure to an ordered low temperature structure. This experiment has been briefly discussed in the section about shock freezing, see also figure 1.14. It is very rare that a static disorder disappears at low temperature, and probably it was only possible because there was the phase transformation in between. For this reason it was justified to spend one week of diffractometer time on this series of measurements while usually the time schedule of an X-ray diffraction laboratory providing a service does not allow this.

1.7.2 Solid state NMR

Another way of having a different look at crystals than irradiating them with X-rays is to perform solid state NMR experiments. The different orientations of a structural fragment in relation to the bulk structure have the potential to represent changes in the chemical environment that can be detected by NMR techniques. In the case of static disorder these different orientations are conserved when preparing the sample for the NMR experiment. Solid state NMR is therefore a possibility to get hints about the nature of the disorder under investigation. However, there are a few difficulties to overcome before getting results.

Magnetic dipole-dipole interactions, a phenomenon that averages to zero in solution, is preserved in the solid state case as the relative positions of the molecules are not anymore random. In order to get interpretable spectra the sample is rotated at high speed (1 to 70 kHz) at an angle of 54.74 degrees, a process called magic angle spinning. This averages the dipolar couplings found in solids and simulates, so to say, the situation that we find in solution. This makes it possible to record spectra with a usual line width. In the case of static disorder two or more sets of signals with slight changes in their chemical shifts should be detectable as soon as the chemical environments of the disordered fragments exhibit enough differences, and their relative occurrence should be interpretable by the integrals of the signals.

From the point of view of the experiment and the instrumentation solid state NMR is demanding because of the high speed of the spinning mechanism and of the cooling. As it needs special equipment and skills it is not as readily available as the standard NMR at many sites. It is maybe for this reason that in small molecule crystallography the cases where disorders are looked at with solid state NMR techniques are rather difficult to find [15].

Recent advances in all fields of data collection and structure solution of single crystal X-ray structures often encourage crystallographers of today to go ahead without waiting for additional information about their problem. In a similar way as described for data collections at different temperature the time and effort to get experimental evidence for the observed phenomenon is often abbreviated. The problem that arises is evident. After increasing efforts spent on a disorder refinement with possibly wrong assumptions it gets more and more difficult to stay objective and to reject those models if they do not coincide with the collected data, no matter how much time has been invested to make them fit.

2 A bit of theory about disorder

In this chapter the basic elements that allow refinement of disordered structures will be presented and discussed. Usually the first hint about disorder comes from the anisotropic displacement parameters (adp's). These are the result of the fact that we see all variations that can occur at equivalent places in all unit cells superimposed on top of each other. The following paragraphs try to illuminate how the superimposition can be explained on the basis of diffraction theory and modeled for the use in refinement programs.

2.1 Twinned structures: sum of intensities

In order to make the distinction we first have to look at a twinned crystal where we can also observe the superimposition of multiple structures caused by the fact that the sample examined is not a single crystal. The sample on the diffractometer is composed of two or more components, each of them with their own unit cell and orientation matrix, and each reflection having its own set of Miller indices relating it to the unit cell it has been diffracted from. The intergrowth of the twin elements results in observing the sum of more than one experiment in one single place. Because there is no coherence between the various components of the twin the emergent beams are not able to interfere with each other. The intensities of the individual twin components are added up to one single observation as soon as we have overlapping reflections, and because of this it is not straight forward to calculate the contribution of each twin component to the resulting intensity. It is only the finished structural model that allows the precise determination of the contributions of each twin component and then, knowing the twin law and the relative volume fractions, the resulting intensity for each resulting reflection can be calculated by adding up the different contributions from the twin components.

Formula 1 shows the expression for the structure factor using the atomic model as base to calculate the electron density. Formula 2 illustrates the twin case with different twin fractions contributing to the total intensity measured.

$$F_{hkl} = \sum_j f_j e^{2\pi i\left(hx + ky + lz\right)}$$

Formula 1: The structure factor expression

$$F_{twin}^2 = k_a F_a^2 + k_b F_b^2$$

Formula 2: This expression shows the situation for a twinned crystal with two components, k_a and k_b are the corresponding twin fractions that sum up to 1

2.2 Disorder: phased structure factors

In the case of disorder it is a single crystal that is mounted on the diffractometer. Each unit cell of the crystal is part of the same lattice. The disordered structural parts have their own, basically independent physical parameters, and the contributions are summed up from all ordered and disordered parts and scaled by their corresponding site occupancy factors. In this sense the disordered case, from the point of view of the structure factor calculations, is not too much different from the ordered case. Partial site occupancy factors need to be allowed in refinement in order to get the intensities right.

Assessing the disorder it may be useful to have an impartial look at what the experiment tells us. Formula 3 represents what is happening in the experiment. The continuous electron density in the crystal is described without interpretation in terms of atoms.

$$F_{hkl} = \int_V \rho_{xyz} e^{2\pi i(hx+ky+lz)} \partial v$$

Formula 3: The electron density in the crystal as observed during the experiment

The situation becomes different when refining a disordered structure. In the twinned case it is only the twin element scale factor that needs to be refined in addition to the physical parameters of the structural model. In the disorder case there are additional physical parameters for each atom that is part of an additional disorder group. It is the sum of all phased structure factors taking into account all physical parameters and their relative occurrence that gives the final result. This situation is summarized in formula 4.

$$F_{hkl} = \sum_j f_j e^{2\pi i(hx+ky+lz)} + occ_1 f_1 e^{2\pi i(hx_1+ky_1+lz_1)} + occ_2 f_2 e^{2\pi i(hx_2+ky_2+lz_2)}$$

Formula 4: Site occupancy factors are needed to accommodate the different contributors in the case of disorder if the disordered area is modeled using multiple contributors. The sum of the site occupancy factors corresponding to one atomic site, in this case the sum of occ_1 and occ_2, is usually one.

The refinement of disordered parts will be discussed in more detail later in this work. One detail may be clarified at this point. Having disordered groups occupying the equivalent space in different unit cells it seems to be reasonable to apply one constraint. In most cases the number of atoms that can be found in one single place should not exceed unity. In the case of disordered solvents the sum of atoms present may be lower, but values higher than one give, also in accordance to IUCr publication guidelines, not very much physical sense.

2.3 Techniques that allow treatment of disorder

When disorder is found in a structure the preparation of the model and the refinement needs to be changed in order to get to a successful end. These basic elements allowing the treatment of disordered structures will be summarized in the following pages.

2.3.1 Partial site occupancy factors

It has already been mentioned that amongst the basic features needed to describe disordered structures there is the partial occupancy of atomic sites that needs to be allowed for in refinement. This is fairly obvious as in the disordered case we still have one molecule in one single place, but parts of it may be oriented in a different way causing the physical parameters of these parts to be different. From this observation follows that the sum of the site occupancy factors of all atoms contributing to a single atomic site in the disordered area should be one.

2.3.2 Restraints

As the different parts of the disordered assembly are interpenetrating some of the atomic positions of atoms belonging to different parts may happen to be very near to each other, too near to be visible as separate electron density maxima in the Fourier map. Restraints establish relations between physical parameters and allow in this way the use of chemical knowledge like bond distances or angles in the refinement. They are of great help to keep geometrical parameters in sensible ranges when the electron density map is ambiguous.

Restraints can be applied for bond distances or angles, for anisotropic displacement parameters, or a group of atoms can be restrained to be planar. Restraints may be created by the user to set into relation parameters of her or his choice. While in structure determinations from powder X-ray diffraction data the use of restraints is widespread as data quality usually does not allow refinement without them, scientists working on X-ray structures from single crystal data use them less frequently. In the ordered case it will usually not be necessary to use them anyway, but in the disordered case they may make the difference between a better and a worse model.

Sometimes the work with restraints is referred to as “playing God” giving the idea that it is possible to arrange structures to look in a determined way using this technique. Restraints that are in disagreement with observations from the X-ray experiment will have a large residual and will not have the power to change a structure in a substantial way.

The weight of a restraint can be increased by reducing the requested standard uncertainty. A restraint with a large weight can perturb a structure. If the restraint is in conflict with the X-ray data from the experiment, the restraint residual, which is the square of the difference between the target value of the restraint and the calculated

value from the model, will be much larger than the standard uncertainty used for the computation of the restraint weight. If the residual remains large (e.g. >3esd) then the restraint may be invalid or the data may be of poor quality.

Another type of restraint that can be very helpful in the refinement of disorder is the so called shift limiting restraint. This restraint can be discussed in a controversial way, but it has clear advantages in the early steps of the refinement of disorder as the attempt to try to find the next best minimum is slowed down, and this can help to avoid false minima or the complete deterioration of a nice starting model.

2.3.3 Part numbers

Another basic feature to treat disordered structural fragments is the technique to assign part numbers that allow addressing groups of atoms with one common identifier. If atoms are added or removed the commands operating on the different parts can remain unchanged without losing the consistency of the instructions. The use of part numbers has proved to be very useful and is an important base feature used extensively during the stage of script writing in this work. The robustness of the refinement directives using part numbers is of great value in all stages of the refinement of disordered structures.

In the refinement program CRYSTALS the part numbers are used in a particular way as they are composed of two numbers. The so called assembly contains groups and the combination of assembly and group numbers is the part number. This is valuable as it can be determined automatically which parts belong together.

There is a second concept to group atoms which is the residue number. It is used to mark different moieties in a structure and helps to keep track of them during the refinement or when editing the structure. The residue number is not of particular use in the refinement of disorder.

Figure 2.1 illustrates the use of the terms residue, part, assembly and group using the example structure we already encountered in figure 1.7. As also mentioned in the figure caption the use of assembly and group numbers is of particular importance for the setup and preparation of the refinement, and these identification numbers facilitate considerably the programming work when creating refinement directives and restraints automatically as shown later.

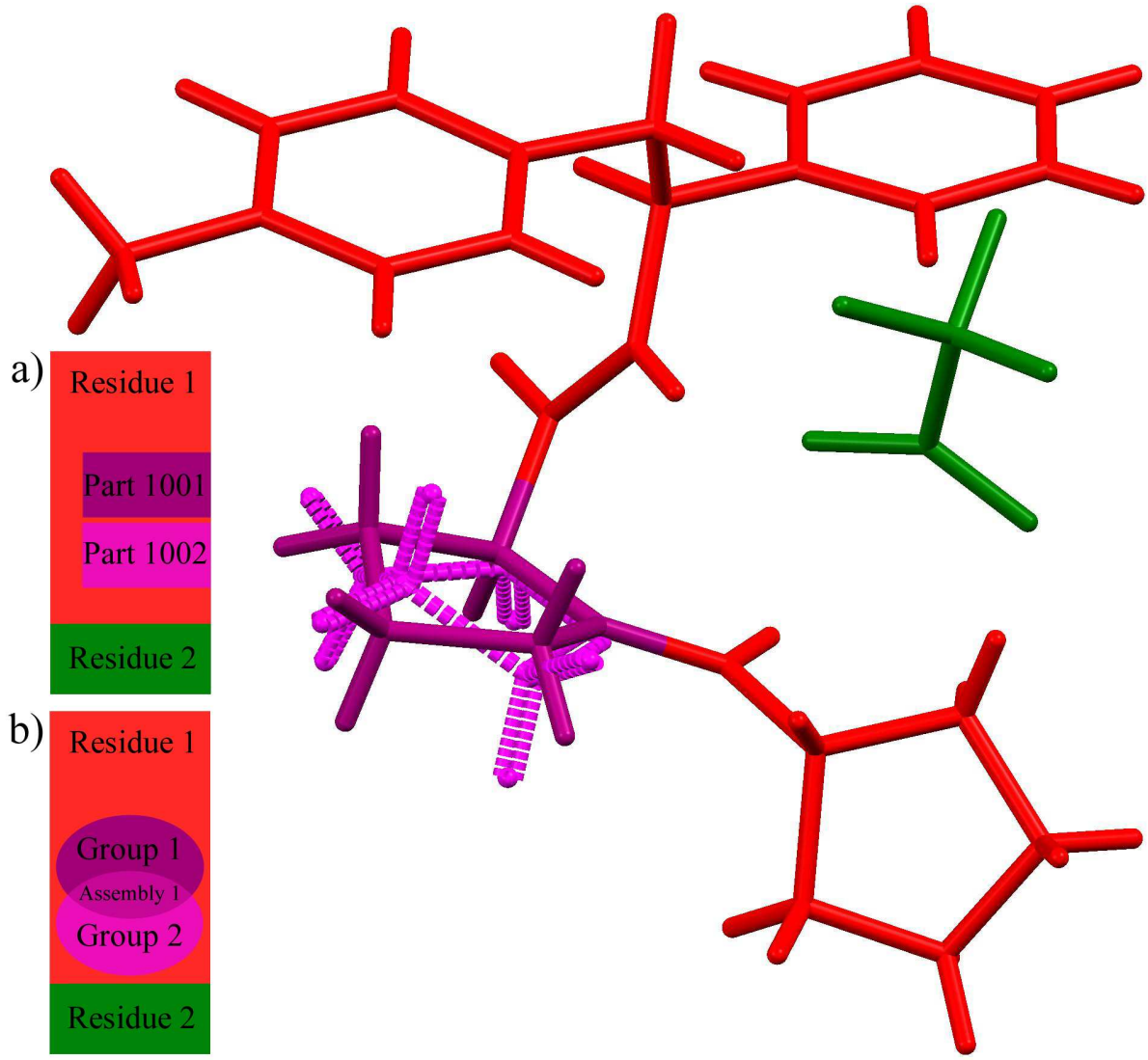


Figure 2.1: Structure 6 displays disorder in the main moiety shown in red to which the residue number 1 has been attributed. The second moiety has been attributed the residue number 2 and is shown in green. The disordered region in residue 1 has been subdivided into two parts with the corresponding part numbers 1001 and 1002 which are basically independent from each other, see a). The two part numbers have in common that integer dividing them by 1000 we get 1 for both of them, and this information is telling us that these two parts belong together. They form an assembly to which is attributed the number 1. Subtracting the assembly number multiplied by 1000 from the part numbers we get the group numbers. In this example the group numbers are 1 and 2 respectively, see b). In this way assembly 1 is made up by the groups 1 and 2. The atoms are members of the groups, and a set of groups forms an assembly. This helps the user to keep track of the special features of the molecular model. CRYSTALS, when refining, knows only residues and parts.

2.3.4 Non-atomic electron density

Some groups in chemical compounds exhibit disorder problems very often. ClO_4^-, PF_6^- and BF_4^- anions are amongst them, together with peripheral groups like CF_3 moieties. If they are disordered the electron density observed may become a continued distribution along the possible rotation axis of the group in question. The electron density of the Fluorine atoms of the CF_3 group will be found as an annulus, and the ball-like shape of the anions mentioned may lead to electron density distributed on a shell as the molecules, due to their shape, are not easily locked in the crystal lattice. In other cases an atom may move forward and backward on a line. The refinement program CRYSTALS gives the possibility to refine disordered atoms as electron density distributed on one of the special shapes line, torus or sphere (Ludger Schröder et al, 2004) [16]. It is important to point out that these special shapes are input in a similar way to an atom as they have three coordinates that are located at the center of the shape describing the electron density. They may also be created from atomic positions, but their shape directly simulates the electron density observed and is no longer a model for an atom. In the case of the CF_3 group the annulus stands for a possible set of three positions the involved Fluorine atoms can take respectively. These non-atomic descriptors of the electron density can be helpful in refining structures with dynamic disorder that do not refine well with assemblies and multiple parts.

3 Refining disordered structures

It should be obvious, but nevertheless it cannot be repeated enough times that disordered structures need to be treated with great care. Particularly in the early stages of refinement this is of great importance, as the electron density by itself usually does not guide the least squares refinement in the right direction. Therefore it is always worth investing some time in careful preparation of the model and of the strategy to refine it. A gentle stepping forward usually gives much better results than if the model needs to be corrected in later stages as the refinement has rushed into false minima.

3.1 Split atoms

When the adp's of an atom become very big they tend to lose their physical meaningfulness. If this happens and if it is assumed that the reason is disorder the distinction should be made if the nature of the disorder is dynamic or static.

Assuming dynamic disorder the model is possibly not optimal as the harmonic oscillations that are the theoretical base of the adp's are reaching their limits. Atoms bound to each other will never move in a straight line in reality. The movement will rather be on a circular pathway around the atom the dynamically disordered atom is bonded to. As long as the movement is small enough the approximation with linear harmonic oscillations is good enough to describe the electron density found in the experiment. If the movement is getting larger, for instance due to temperature, the calculated electron density that has the shape of an ellipsoid subtracted from the experimental electron density that has a banana-like shape, will create residual electron density that cannot be explained by the model anymore. Moreover trying to fit the ellipsoidal calculated electron density on the experimental electron density leads to shortening of the bond distances calculated from the model.

Probably the best solution would be to use anharmonic displacement parameters [17]. Most programs do not offer an implementation for these more complex displacement parameters as there are more parameters needed to refine them, and it is rare that enough data of sufficiently high quality can be collected in order to get reliable information from such a refinement. Moreover it is difficult to distinguish between the effects of anharmonic motion and the deformations in the electron distributions caused by the bonding (P. R. Mallinson et al, 1988) [18]. All these difficulties make the use of simpler solutions very popular. Splitting the dynamically disordered atom position and refining two overlaying models instead may describe the experimental electron density more accurately even if in the case of dynamic disorder the real maximum is expected to be found at the center rather than in two positions somewhere left and right from the center.

In the case of static disorder, where two or more parts of molecules are overlaying that are by themselves ordered, there is sometimes the chance to find the true electron density maxima as separate peaks, but most of the time partially overlapping atomic

positions will be seen as one broad maximum. The new positions have to be calculated in order to find starting models for refinement. One way to do this is by splitting the positions with adp's that show high anisotropy. This technique has been available for many years and is a first step to build models for refinement where the resolution of the X-ray experiment does not give the possibility to separate positions. When the problem is getting complicated it can become very tedious to keep track of the different parts of the disorder, and in particular changes may cause trouble and long searching for errors in the setup. Nevertheless atom splitting gives the possibility to build molecular models that go beyond what can be found straight away in the density maps of disordered structures.

3.2 Postulation of models

Unfortunately there is rarely evidence at early stages of refinement if the disorder is dynamic or static. If no other hint is available the careful analysis of the direction and size of the ellipsoids can be useful to get more information. If the adp's are such that some kind of back and forward motion could produce them then the chances are that the disorder under investigation is dynamic. In this case so called TLS calculations [19] can be carried out that simulate the motion of the particular molecular fragment as a rigid body and calculate the adp's that this rigid body would produce. T stands for translation, L for libration and S for screw. The calculated adp's can now be compared to the refined ones, and an R-value can be defined that tells the user how well the refined adp's can be explained by a rigid body motion. This can give some hint if the assumption about the nature of the disorder being dynamic is correct.

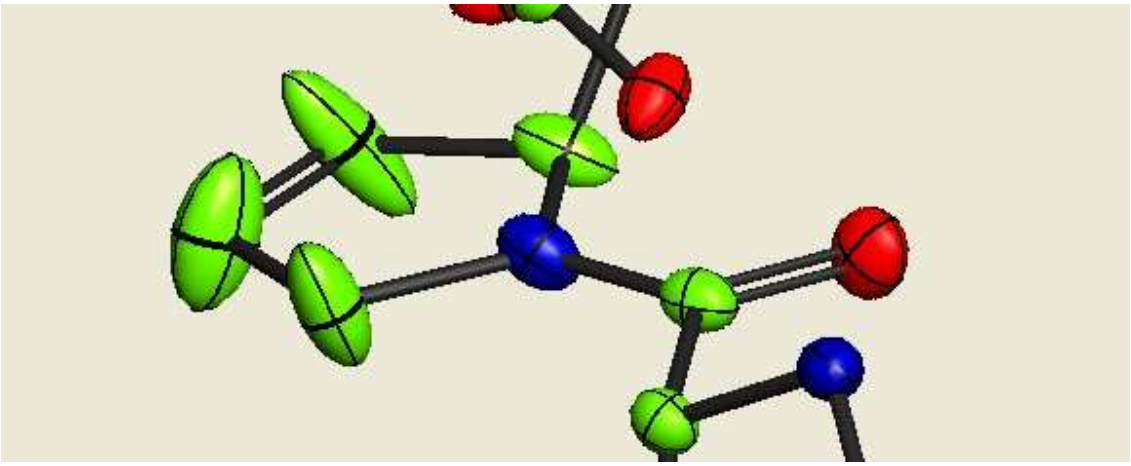


Figure 3.1: A detail view from structure 6 before disorder has been resolved. This is a typical case where the static nature of the disorder can be deduced from the shape and direction of the ellipsoids. See figure 1.7 for the refined structure with the disorder resolved.

If the adp's indicate that a transformation between possible alternating positions of the atoms need a substantial rearrangement and that some swinging movement from one

to the other position is unlikely, then it is likely that the nature of the disorder is static. The structure shown in figure 1.7 in chapter 1 containing a disordered five membered ring is a very nice example where the shape of the ellipsoids of the refinement without resolving the disorder, as visible in figure 3.1, are not explainable by a continuous vibration of the molecule.

As observed frequently in crystallography a good guess is often a good starting point for a successful refinement. Following the assumptions concerning the disorder under investigation the starting model is created and refined in order to be validated against the experimental data.

3.3 Creation of models

There have been implemented many aids to facilitate the creation of models. Splitting of atoms and part numbers to keep track of groups are the most important ones to mention besides the possibility of having restraints to define geometrical features of the model. They have all been around for many years.

The traditional way to create starting models used to be tedious, mainly due to the following reason. Interpenetrating parts that appear as a superimposition in the electron density map are difficult to separate. For instance will one Carbon atom of the first part overlap with a Nitrogen atom of the second part? To keep track of all details like keeping site occupancy factors and displacement parameters consistent is, even though it is pure housekeeping, a quite tough job. Making errors is to be expected and creates long idle times to repair the input data. A good knowledge of how to edit the structural parameters as well as a good overview of all the atoms with their names that make up that structure is needed to successfully build a model by these means. This makes the barrier to treat disorder quite high, in particular for beginners.

Apart from these practical considerations a weak point of this approach was that it treated the alternative positions of a disordered structural fragment as basically independent. More physical parameters without new experimental observations means that the overdetermination of the system is lowered and the stability of the refinement can suffer. The general rule to try to collect the best dataset possible is of course also valid in the disordered case. The more data is available the better are the chances to successfully refine a model that needs more complexity in order to describe and reconstruct the experiment adequately.

3.4 Refinement

Once a model has been constructed, it needs to be refined to make the whole structure fit in the best possible way to the experimental data. Here the problem is that in a zone with flat electron density distribution and partially occupied positions near to each other does not help the crystallographer to find convergence quickly. The changes have to be made in a gentle way in order to prevent the refinement from deteriorating. Usually the

refinement will start with tight restraints, and attempts will be made to release the restraints bit by bit once the refinement approaches convergence. In difficult cases the restraints will not be sufficient in the early stages of the refinement and constraints have to be used to reduce the degrees of freedom. A common constraint at this stage is to refine the coordinates of the disordered parts as groups. Once the parts are sufficiently near to their final position the constraints can be removed and the refinement should converge if the case is favorable.

3.5 Impact of restraints on final model

There is still a lot of discussion about the use of restraints. Some people claim that a structure that is in its minimum should not need any restraint anymore. There is the fear that the restraints would introduce some human bias into the structure and that this would decrease the reliability of the whole structure determination. Looking at the facts as a whole changes the emphasis on the single elements. It is usually in the maximum a few hundred restraints that are competing with at least a few thousand reflections. This is the main reason why we can say that restraints that are in disagreement with the reflection data will simply not work in the expected way. As a rule of thumb it can be stated that a restraint that has its effect without a big cost in terms of R-value is safe to use.

Nevertheless it is true that the tightness of the restraints should be tuned with care. It is to be observed that releasing a restraint whose standard uncertainty has been estimated too low is much easier than to correct than trying to improve a degraded model by lowering the standard uncertainty values.

Special care has been invested to provide easy ways to change the standard uncertainty values for the restraints of a disorder assembly. As it is recommended to start with tight restraints it is good if the values can be released easily as keeping the restraints tight will produce a nicely refined model, but the final standard uncertainties written to the cif file will be very small and may cause referees to object to the structure claiming that the given values would not make sense physically.

3.6 Use of constraints instead of restraints

During the development work it was tested if at least some of the restraints could be replaced by constraints, as one part of the disordered assembly could be described as the image of the other part created by some mathematical operator. In fact the first proof of concept that was realized in the scope of this thesis project was an external program that simply wrote CRYSTALS commands to a file after analyzing some CRYSTALS output. Later these commands could be executed as a batch job, and this program writing the batch files used exclusively constraints in the first stages of refinement.

In figure 3.2, assuming that S(2) and S(12) are in the same assembly but in different

groups, the commands were issued in the following way:

```
EQUIVALENCE S(2, X) S(12, X)
WEIGHT -1 S(12,X)
EQUIVALENCE S(2, Y) S(12, Y)
WEIGHT -1 S(12,Y)
EQUIVALENCE S(2, Z) S(12, Z)
WEIGHT -1 S(12,Z)
```

The EQUIVALENCE instruction creates one common parameter for X, Y and Z of S(2) and S(12) respectively. The WEIGHT instruction applies the calculated shift to X, Y and Z of S(12) by first multiplying it by -1. This keeps the two atoms at equal distance from their centroid.

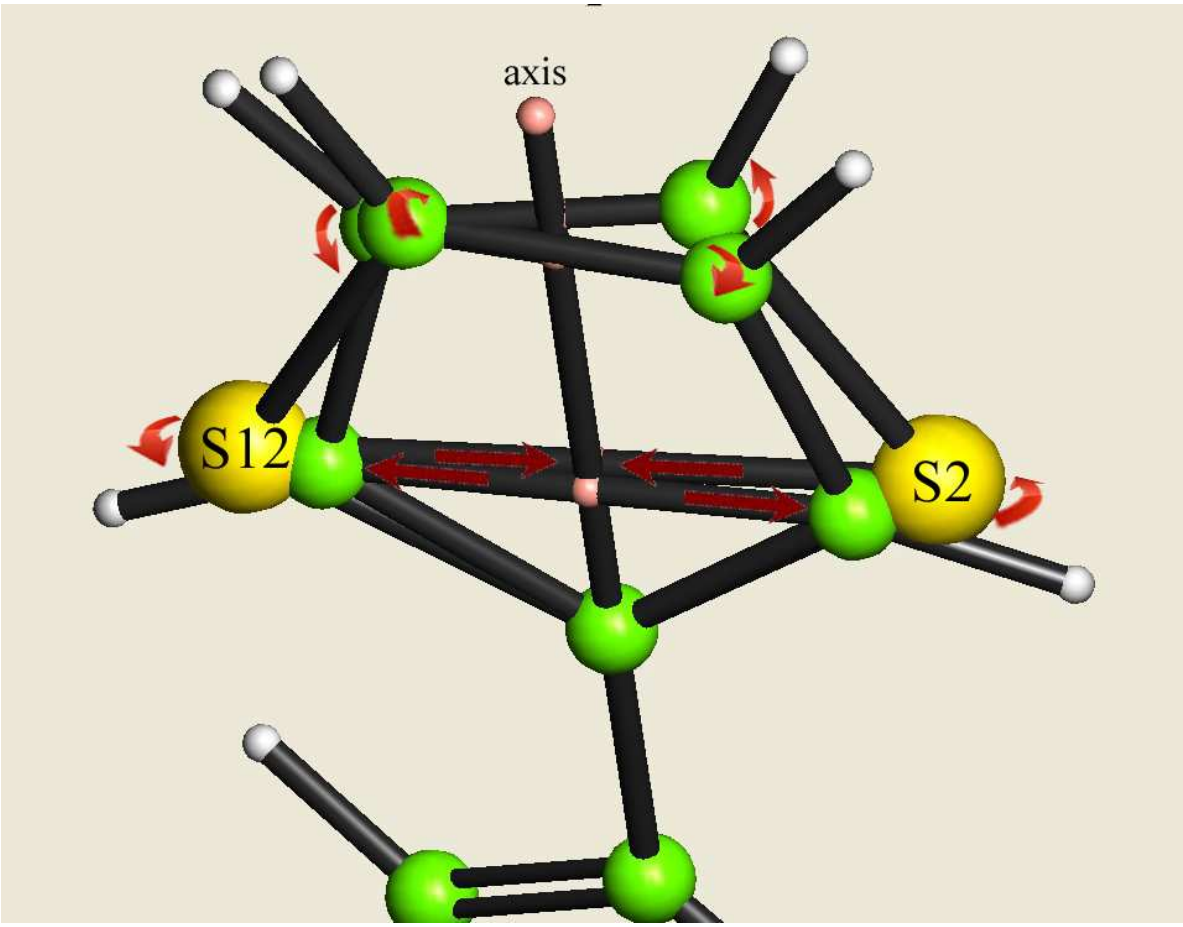


Figure 3.2: A view of the disordered thiophene ring of structure 4. The axis around which the disorder model has been created is displayed together with links between the constrained atoms. The pairs of arrows show, as an example, how S2 and S12 refine together keeping their centroid on the axis.

In order to allow the refinement to correct for an incorrectly estimated central position between the two atoms the WEIGHT instructions were, in an alternating way, left out and used again so that the refinement could relax and find the true minimum.

The big advantage of this technique was that the number of refined parameters could be reduced. Unfortunately equally simple instructions could not be used to constrain the anisotropic displacement parameters in a useful way. The LINK instruction provides an easy way to create one common least squares parameter for corresponding parameters of the atom list linked together by this instruction. For the example shown earlier this could look as follows:

LINK S(2, U'S) AND S(12, U'S)

This works well as soon as the displacement parameters can be assumed to be numerically identical or very similar. As soon as this cannot be assumed anymore to be true the use of these constraints becomes problematic as it leads to physically unreasonable models.

As the number of additional parameters coming from the adp's could not be lowered in this way this approach was not further developed even though there would be some potential to improve the refinement of structures with a limited amount of observed reflection data available.

Another difficulty that could be seen in the testing phase is that this operator describing the non-crystallographic symmetry, even though clearly visible in a qualitative way, was usually quite fuzzy. Figure 3.2 shows the disordered side chain of the high temperature structure already shown in chapter 1, figure 1.14. From the point of view of an attentive observer of patterns the way the atoms arrange is not casual, as it follows some systematic pattern. The successful refinement using constraints in the initial stages depends crucially from how good the first approximations could be made. Moreover it is not rare that the successful refinement shows in the end that a rotation axis is not really lying on a straight line, but is bent, or a mirror plane is not flat, but distorted as shown in figure 3.3. Therefore a system using only constraints would in high probability only work for a few favorable cases.

Figure 3.3: A detail view of the disordered side chain of structure 10. The atoms of the two disordered groups arrange pair wise in a mirror-like way, but with the “mirror” being far away from planarity.

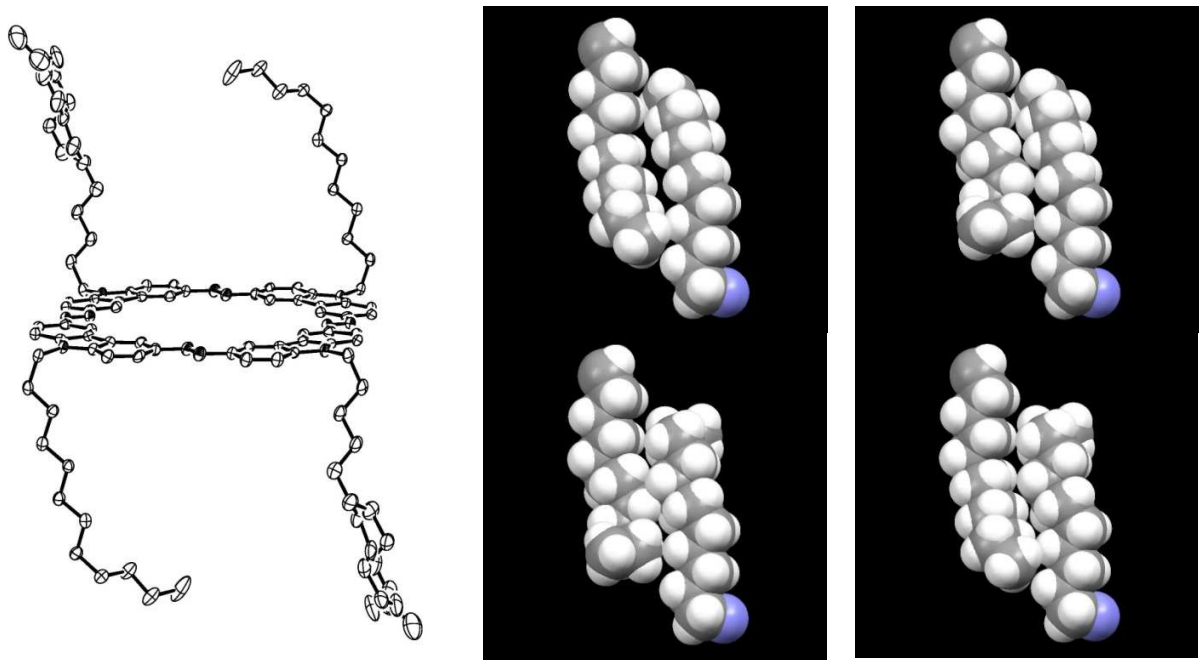
Nevertheless it seems indispensable to get this symmetry set up in the correct way as otherwise the restraints derived from and generated using these operators would suffer from averaging values that are indeed not to be averaged. In the outlook [section 7] there will be an over-view of plans to reintroduce the constraints in the final stage instead of the initial stage of the refinement. The already optimized model then could profit from the greater stability the constrained refinement gives, and the analysis of the centroids of atomic pairs in the disorder assembly could be used as a way to test the correctness of the initial assumptions about the operators used to build the model. Rearranging the already refined model following the analysis of the local symmetry could help to improve the fit of the restraints in case the restrained refinement is the final stage or the constraints are only used for the positions while restraints still stabilize the refinement of the anisotropic displacement parameters.

3.7 Validation of the results

Validation in the disordered case has its limits, as usually there is no exact experimental evidence that proves the correctness of the proposed model. The resolution of the experimental data simply does not allow seeing the details. It is clear that there is the hope that the more complex model will explain the electron density better than the simpler model that does not take into account the disorder. This should be reflected also by a lower R-value and by lower maxima in the residual electron density. Of course these indicators have limited value or usefulness in absolute terms.

One of the key points remains the claim of a chemically sensible model. If we can, in the case of a static disorder with two parts, show that two chemically sensible models can describe the electron density better than one with huge adp's then this should be a valid reason to prefer the model with disorder. If all indicators are improving including the sensibility of the model, then the refinement is on a good track.

A possible way of evaluating if the model is chemically sensible is shown using an example. One of the side chains of the molecule shown in figure 3.4 is clearly disordered. The atomic positions were split and part numbers 1001 and 1002 were assigned to the atomic sites. Restraints were set up to reflect the expected geometry for sp3 hybridized Carbon atoms. The refined positions showed site occupancy factors near 0.5. As there is one of the inversion centers of the space group P-1 near the disordered chain the symmetry generated positions were added to the coordinate file while lowering the symmetry to the space group P1. After this it became clear that only the combination of the chain with part number 1001 with the symmetry generated chain with part number 1002 were without contacts too near to be possibly existing. So the explanation of this disorder was that in this part of the structure the symmetry of P-1 cannot be fulfilled which explains why it is disordered in this area. The combinations that create clashes are displayed in the figures 3.5 and 3.6 (central column), and the ones that are without near contacts are shown in figures 3.7 and 3.8 (right column). The packing diagram shown in figure 3.9 illustrates how the side chains at the corners come very near to each other at the inversion centers located in the corners of the unit cell, while at the other inversion centers in the middle of the edges there seems to be much more space.



Figures 3.4 until 3.8: A disordered side chain and the explanation why it is disordered.

The column in the center shows how the two chains behave using the inversion: There are holes (top) or clashes (bottom). The column on the right hand side shows the solution combining one disorder group with the inversion of the other, and the two chains stay side by side without creating van der Waals contacts. (Structure 13, unpublished work)

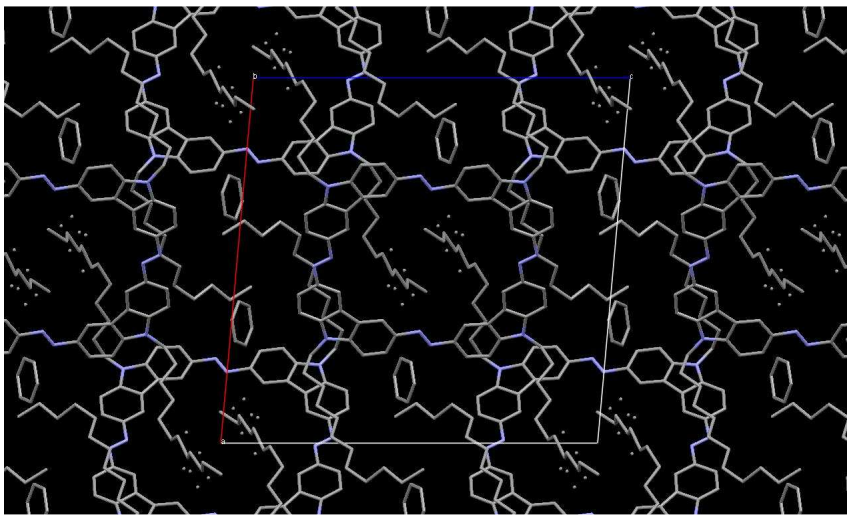


Figure 3.9: On the packing diagram that shows the structure looking down the b axis it can be seen that at the inversion centers near the corners the chains come nearer than at the inversion centers in the middle of the edges. The second group is displayed as dots only. (Structure 13, unpublished work)

Sometimes it will remain the decision of the crystallographer if she or he prefers to explain why the residual electron density is too high and why bond distances and angles are strange, or why she or he uses a model exhibiting disorder. A disorder model always depends to some extent on the decisions the crystallographer takes during the refinement work. The scripts presented should help to treat some of them in a more standardized way which takes away at least some of the human bias.

3.8 SQUEEZE and SWAT

In order to close this chapter some words should be said about techniques widely used in macromolecular crystallography or in small molecule structures with extensive solvent disorders. They provide a non-specific solution that is no longer aimed to create sensible atomic models, but these solutions try to find acceptable solutions where other methods have failed.

The refinement of disordered solvent molecules can be one of the most unsatisfactory jobs in crystallography. In particular in macromolecular crystallography the phenomenon of diffusely distributed water molecules in the cavities formed by the molecules building the lattice is widespread. On the base of Babinet's principle algorithms have been developed to fill the solvent accessible areas with water molecules without trying to locate them one by one in the difference Fourier map (Moews & Kretsinger, 1975, [20], SWAT instruction, G. Sheldrik, Shelx manual) . Babinet's principle says that the diffraction pattern of an object and of its complement are, apart from the areas where

the direct beam can be observed, equal in amplitude, but opposite in phase. Considering the ordered part of a structure as the object and the disordered solvent around as its complement the phase information from the ordered part can be used to obtain phase information for the disordered part by inverting the phases. While such approaches are comprehensible in macromolecular crystallography with large unit cells and moderate possibilities to reach sufficient primary resolution these techniques are less popular in small molecule crystallography.

Disordered solvent molecules also create big problems in structures of small molecules. If the electron distribution is too irregular the search for electron density maxima in the difference Fourier map will become almost featureless even if there are still areas where solvent molecules could be found. The least squares refinement will adjust the existing atoms to try to flatten the difference electron density map which leads to phase angles that are not optimal as they are not calculated on the complete structural model. Wrong phase angles as a consequence of diffuse contributions in areas accessed by disordered solvent molecules were the point where Ton Spek's work began. Thinking of ways of how to improve phase angles the idea came up to calculate the contributions of the disordered areas and to look at the rest of the structure without the disturbing disordered part of the structure. This is also the idea of the name SQUEEZE [21] that this program has been given as it makes us think that we are just pressing the whole solvents out of the structure. This was achieved by redefining the structure factor as the sum of the contributions from all ordered atoms and the discrete Fourier transform of the unresolved electron density in the disordered areas of the structure. Formula 5 summarizes this. The aim was to get better phases for the ordered part and to get better residual electron density peaks from the complete dataset as soon as the phases would have been optimized without the solvent contributions disturbing the refinement.

$$A_{hkl} = \sum_j f_j \cos 2\pi(hx + ky + lz) + \int_v \rho_{xyz} \cos 2\pi(hx + ky + lz)\partial v$$

Formula 5: This formula illustrates the way SQUEEZE works by splitting up ordered and disordered parts of the structure and treating them separately. Where discrete atoms are available formula 1 is used, the remaining parts not yet modeled with an atomic model are included using formula 3.

The reality of today's standard use is that people often stop investigations after having applied SQUEEZE and discuss the number of electrons SQUEEZE would have taken away by calculating possible chemical formulae that would fit that number of electrons SQUEEZEd away. It must be mentioned here that in some implementations of SQUEEZE the experimental observations are changed, a point that may make its use problematic. In CRYSTALS the SQUEEZEd region contributes to A_{calc} and B_{calc} in just the same way as normal atoms. The use of the program with the appropriate discussion is certainly justified in cases where everything else fails, but the use of the program without mentioning it and leaving the structure with holes and modified Fo's is scientifically not acceptable.

Moreover the use of SQUEEZE is restricted to disordered parts of the structure outside the Van der Waals radii of the atoms of the ordered bulk structure. There is no possibility to SQUEEZE away the contribution of interpenetrating fragments.

Both presented possibilities to cope with heavily disordered solvent molecules are only useful in the context they were intended to be used. They both have in common that they try to hide the problem instead of finding some explicit model. They are fairly easy to use, and they are effective in showing if the solvent disorder is strongly influencing the final result.

4 Introduction to CRYSTALS

All programming work and all refinements presented in this work have been executed using the crystallographic refinement program CRYSTALS. It is without any doubt one of the most complete crystallographic packages that is available today.

4.1 The roots of CRYSTALS

The story of CRYSTALS goes back many decades, right to the origins of modern structure determination. Issue 1 of the CRYSTALS system was written in 1975, and the original aim was to create a program that would be able to refine twinned structures. Paul Betteridge, David Kinna, Lisa Pearce, Allan Larson and Eric Gabe were involved in giving this project its shape, but it also contains lots of contributions from students and visitors of the Crystallography laboratory. It was running on ICL 1900 series of computers. ICL stands for International Computers Limited, originally a British brand of computers that was very successful at that time and that was taken over in 2002 by Fujitsu.

Issue 2 of CRYSTALS was a rewrite of issue 1, created in the years 1977 to 1978 by J.R. Carruthers and J.S. Rollet. This version became the base on which all the later versions until today are based on. Over the last decades David Watkin has maintained and enhanced the code and has made out of CRYSTALS one of the most complete crystallographic program systems of today.

4.2 Interactivity

Continuously the possibilities of the program were enhanced. Already in relatively early stages the possibility to interactively enter commands was added. Now users did not have to edit the command files anymore before running a job, as it was possible to react immediately while typing the commands when the output suggested it. For the refinement cycles the batch mode could be used in order to balance more effectively the CPU time consumption while the manipulations on the model that needed more immediate control could be executed interactively.

For the graphical representation of the structure the foreign command links were used, in a similar way as when calling a direct methods program for the structure solution. Already in early times the graphics program distributed together with CRYSTALS, CAMERON, was able to store changes applied to the model in a way that permitted to import them again into CRYSTALS which gave the possibility to make edits in a graphically controlled environment.

4.3 Data storage and command language

CRYSTALS stores all data in one central binary file. It is a so-called direct access file that allows substituting single records at any place of the file, a possibility that sequential files do not offer. As this file always had to be placed on a hard disk it was

and is still called the diskfile. The diskfile is the database of CRYSTALS, and its direct access structure allows resuming work on the structure at the point it has been left when closing down CRYSTALS the last time.

The units that are stored are the so-called lists. Each object of the crystal structure gets its place in storage in a list, and the total of all lists makes up the structure as a whole. These lists may be very big, as list 6 that holds the reflection data, or they may be small, as the ones that store for instance the unit cell or the space group, lists 1 and 2. The lists can be rewritten one by one, and this assures that only the item that changes needs to be saved and written to disk. Another advantage of the lists is that following an internal flag they can be overwritten when a new version is saved, or all the versions can be kept for later backup. The latter is very useful for the storage of the coordinates, list 5, as it saves the user from data loss when something goes wrong. Stepping backwards, the last useful set of coordinates can be found and work can restart from there instead of from scratch.

Within list 5 there is storage for all physical parameters of the structure. Overall parameters like the scale factor or the Flack enantiopole parameter are stored aside from the coordinates in the same list 5. Every atom can be a member of a residue and of a part, and this feature permits to define moieties in the structure that can be addressed as a whole.

The three most significant digits of the part number are interpreted as the assembly number, the three least significant digits as the group number. This construction enables the user to see easily which groups belong to the same disordered area. The concept of the assembly containing groups which enhances the concept of part numbers has been explained in section 2.3.3, and figure 2.1 illustrates common features and differences of part numbers in relation to assembly- and group numbers.

The command structure of CRYSTALS has different levels. Each command can be made more specific by giving directives, and each directive can be fine tuned with its own set of keywords and values. Each command line can be extended, new keywords and values can be added by starting the line with CONTINUE. The input of the command is terminated by entering END.

4.4 The script processor

As the next major enhancement there came the capability to automate the way CRYSTALS was running by adding the script processor. First this was a simple construction that allowed concatenating commands, but more than that it also allowed having flow control and decision taking by adding logical structuring to it. The scripts gave the possibility to present relatively complicated tasks as one operation to the user. The possibility to have flow control gave the possibility to adapt to different situations, or to give the user the choice between several possible ways to continue. Having access to the crystallographic data at all times made it also possible to base decisions on the

data and act in this way as an expert system. Moreover the possibility to run scripts opened a way for all users of the program to contribute enhancing the functionality of the system without touching the binary core of the program.

The scripting language offers the possibility of executing quite extensive calculations. The instructions available are tailored for the purpose they have been designed for. It is therefore not as complete as a programming language, and newcomers will miss their favorite tools and features. It usually does not take long to think of a possible route using the features and functions available.

4.5 The graphical user interface (GUI)

The request for graphical user interfaces in Crystallography was growing towards the end of the last decade of the 20th century. The first release with graphical user interface (GUI) was issued in June 1999. It had been written by Richard Cooper in the scope of his PhD project (Richard Cooper, Dphil Thesis, University of Oxford, 1999). After proofs of concept of different possible ways to go it was the enhancement of the script processor that showed to be most promising. The laboratory in Basel was involved in the development of the concept of this GUI as in that time Ludwig Macko was developing a CRYSTALS version for Macintosh computers running MacOS 9 (Ludwig Macko, Dphil Thesis, University of Basel, 1997) and the two laboratories exchanged ideas and experiences.

The script code now contains not only the commands to control CRYSTALS, it also controls the appearance and the functionality of the GUI with its dialogues. Built in switches allow to redirect the script code in a way that it can be used to build dialogues, or it can be used in the traditional ways to assemble and pass commands to CRYSTALS. The dialogues pass the data the user provides back to the scripts that then operate in the known way controlling CRYSTALS.

These extensions of the script processor are linked via an abstraction layer to the operating system functions that provide the GUI functionality. All the GUI functions are running in a separate thread so that the program is not blocked when calculations are going on. This architecture ensures a minimum of operating system specific code and has therefore the potential to be more easily adaptable to other operating systems. A Linux version exists, and a Macintosh version for OSX is under development. As the abstraction layer calls system specific routines to create windows, dialogs, progress bars or other elements that are building blocks of a modern GUI, the look and feel is on the same level as that of any other application written in a native way for the operating system in question. At the same time there is no need to compile code when creating new dialogs when adding functionality to the program, and like that the way from the idea to a prototype is much shorter than using other environments to generate code.

5 Finding the best model for disorder.

In the following the emphasis will be mainly on static disorders that have been chosen as the field for development of new strategies in the treatment of disorder. This class of disorder has been chosen as it is most fruitful to find good solutions there.

Many times users are insisting too long on models that make no sense. Attempts are made to separate atom positions only where it is absolutely inevitable to do, and this leads many times to bad models. If in a phenyl ring three atoms are marked by the refinement program as possibly to be split, then it is necessary to split the whole ring as otherwise the model will be senseless. In this case this may be more obvious than in others, but the distortions on the model can also be found on the less obvious cases. So the point of departure was that it should be easier and less laborious to build good disorder assemblies.

In order to achieve this goal a series of scripts have been written that should facilitate in a substantial way the generation and refinement of disordered assemblies in structures exhibiting static disorder.

5.1 The problem

Lots of work has already been invested in providing basic and advanced functions to deal with disorder. Observations in research and teaching show that a lot of time is used when it comes to set up models for disordered structures. Manipulating the models is still tedious, and errors are frequent as the available tools mostly operate on one atom, but not on a group of atoms. Therefore it is frequent that finding the error in the setup takes more time than the effective refinement.

From this follows that the tools that can help to improve this situation are those that act on logical units of the structure and keep these units consistent. Such a logical unit can be a residue, assembly or part, or it can be simply a selection of atoms marked with mouse clicks. For disorder it will mostly be the assemblies and parts that can be built out of a selection of atoms and that are to be manipulated as a whole by such a script to be written.

5.2 Observations on the way to new solutions

Chemical knowledge and imagination in three dimensional space gives humans outstanding instruments that enable them to make good assumptions about the nature of a disorder. A series of peaks in the difference map may be very fuzzy, but still may be sufficient to recall the geometry of some molecule that could be hidden there in the electron density.

Trying to assemble a molecular model out of electron density maxima comes to a limit when we have disorder. The consequence is that other ways of building the model have to be found that, as mentioned, act on groups of atoms in a consistent way. If we

suspect a solvent molecule to be present in a second orientation, then it is best to act on the model as described by duplicating the positions and reorienting the newly created atoms according to how we suspect and define the second orientation to be.

After the successful completion of this action the model consists of two groups making up the disordered assembly. This process is fast and produces chemically reasonable models as soon as the starting model is reasonable too.

In order to get the prototype models in good shape different ways of regularization are available that allow optimizing the geometry of the model. In particular there is some interfacing code for some of the REGULARISE options in CRYSTALS. Moreover there is the possibility to put atoms on a line or make them planar, and by choosing the LINE option of the script a_calcMolax.scp, which will be described in detail in section 8, atoms can be shifted along the line in order to get better fit with the required geometry. This option has been used successfully when refining disordered aliphatic chains built using SPLIT from weird adp's.

The process of duplicating and reorienting a molecular fragment can be understood as if there would be some additional symmetry element valid only for the disordered part. In order not to create confusion the term operator will be used to denominate the local symmetry. Possible operators are inversion, rotation axis and mirror. As most disorders, including fuzzy solvent disorders, are localized in a confined space there was until now no hint about translational additional symmetry in disorder. This is also valid for disorder assemblies covalently bound to an ordered bulk structure.

5.3 The first step: disorder using a two-fold axis

In the first step the emphasis has been put on disorder where the twofold axis can be used to build a second group for the assembly. The CRYSTALS directive ROTATE in the EDIT command is the base command used for that purpose. After having selected the atoms to be rotated, the user has to choose a rotation axis. This can be a bond (in the CRYSTALS interface the line showing the bond between two atoms), but it can also be two non-bonded atoms. If only one atom is selected the rotation axis is automatically completed by adding the centroid of all atoms selected for rotation.

Now the rotation angle has to be determined. Again, in order to make the problem easier to cope with in the first stages, two frequent cases have been picked out, and it could be seen that these two cases cover a lot of the cases that turned up in the laboratory in Basel. The first one is to leave the original part unchanged and to rotate the created part by 180 degrees. The second one is to rotate the first by +15 and the second by -15 degrees. There is a script that allows the free rotation of atoms around some axis in the structure, and until now this option has only been used once when a disordered alcohol group had the best fit with a residual electron density peak using a rotation angle of 140 degrees for the starting model. Alternatively this disorder model could also have been obtained by building the mirror plane using the bisecting angle

and the atom the group is bonded to in the bulk structure. As only two atoms were involved, this would not have affected the result in this case.

5.4 Refinement strategies

When it comes to the refinement of a model generated in this way it is particularly important to refine step by step. Site occupancy factors may be far from 0.5 for the two parts, so it is important to refine them first alone. The two parts then can be refined as rigid groups to approach the final position. It may be necessary to use shift limiting restraints in the first steps of refinement in order to prevent deterioration of the model. Then a common isotropic displacement parameter can be refined for the whole assembly, and if all this is successful the next step can be to refine the position of the groups together with the common isotropic displacement parameter and the site occupancy factors.

In order to be able to refine the positions individually restraints have to be generated that describe the similarity of the two parts. In the first steps of refinement mainly restraints for the connectivity are needed. Later also the displacement parameters need to be restrained. The corresponding script gives the choice to generate restraints for bond distances and angles, for vibration and similarity restraints, and it is possible to use the SAME, DELU and SIMU restraints available in CRYSTALS. If only the standard uncertainties of the restraints are changed there is the possibility to do that too.

Now the individual positions can be refined, and this will lead to a further relaxation of the structural model. In the beginning the common displacement parameter will be kept, later individual displacement parameters may be refined, and in the end also anisotropic displacement parameters can be refined. If this step is successful then the disordered assembly can be incorporated into the refinement of the whole structure.

5.5 Validation of the operators used to build the model

Analysis of local symmetry can help to give a better base to the postulated and successfully refined model. A possible approach would be to start calculating centroids for all pairs of atoms making up the assembly. If the operator that makes the image of one part on top of the other is an inversion, all these centroids should be concentrated around a point. If the operator is a two-fold axis then we should find a straight line, and in the case of a mirror all centroids should fit on a plane.

If we take the pairs of atoms as we have defined them when building the model for the disordered assembly the results of such calculations of course depend crucially on the assumptions used at the time of the creation of the model. Evaluation needs to be simplified, and tools to reorganize a disorder assembly are needed, for instance if the pseudo-symmetry operator reveals to be a mirror while the model was created using a rotation. Some possibilities how this could be done will be described in the outlook of this work.

6 Examples of successfully refined disorder

In the following pages a few examples will be presented where these strategies have successfully been used to find good solutions for the refinement of disorder. The datasets have all been collected in the crystallography laboratory of the University of Basel. For some of them the data sets have been collected recently, others date back a few years and have been refined one more time using the new scripts.

6.1 cs_Cu-S-SMe (structure 14)

This structure was one of those structures done in the laboratory in Basel where the first experiences with disorder were collected. In 1997 the refinement of this structure took many hours, distributed over different days, to be completed successfully. The structure looked bad refining one set of coordinates. The displacement parameters, as shown in figure 6.1.1, did not make any sense assuming that they should be similar between bonded atoms. Taking into account the possibility of disorder and assuming that one carbon atom of the disordered group of atoms would be coinciding with the position of the Sulfur atom of the second orientation of this same group gave the hint to solve the problem.

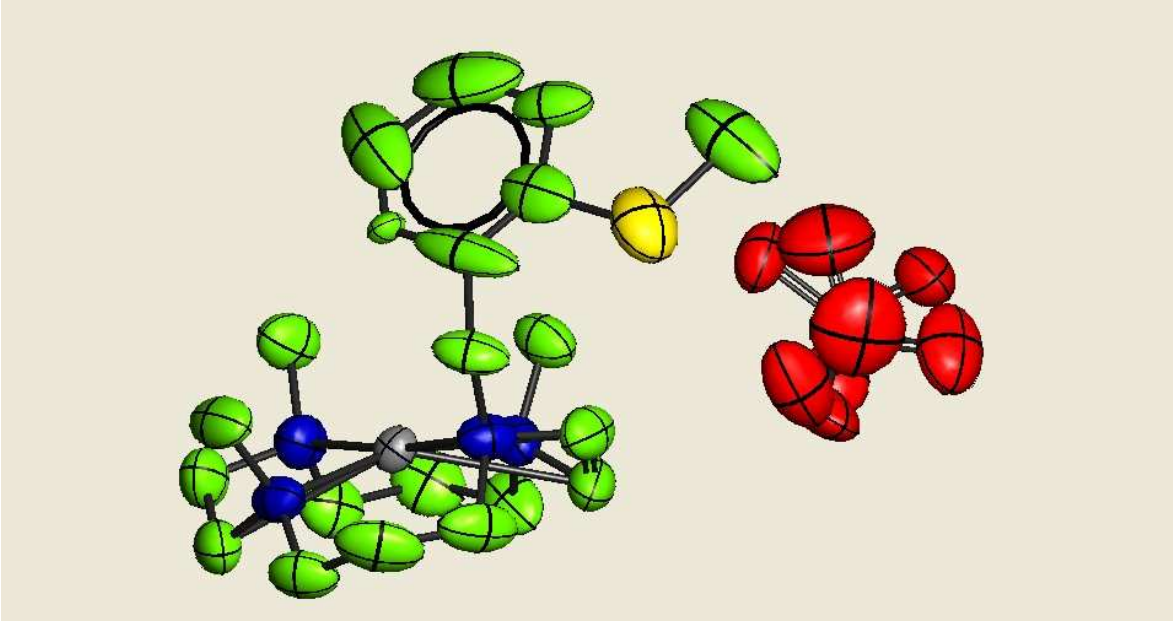
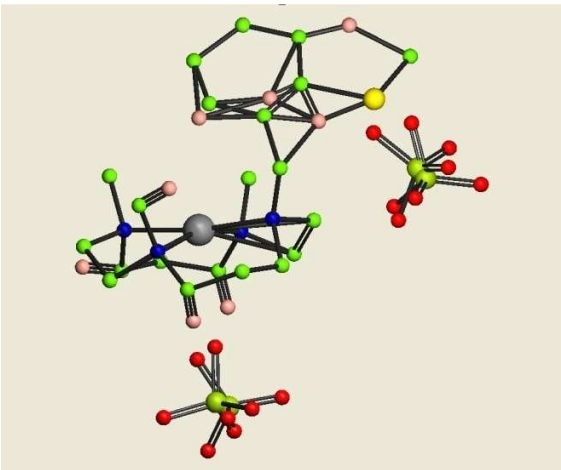
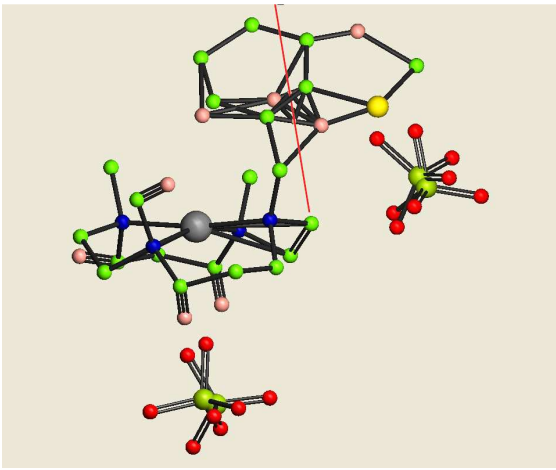


Figure 6.1.1: The weird displacement parameters after initial stages of refinement

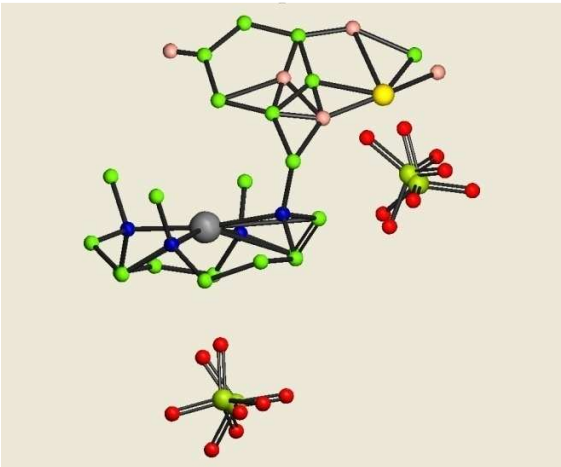
In 1997 all the new positions had to be created one by one using different methods. Some atoms could be split in order to get the new positions. For some of them electron density maxima from the difference Fourier map could be used and included in the refinement. For others the position of an atom had to be duplicated and renamed.

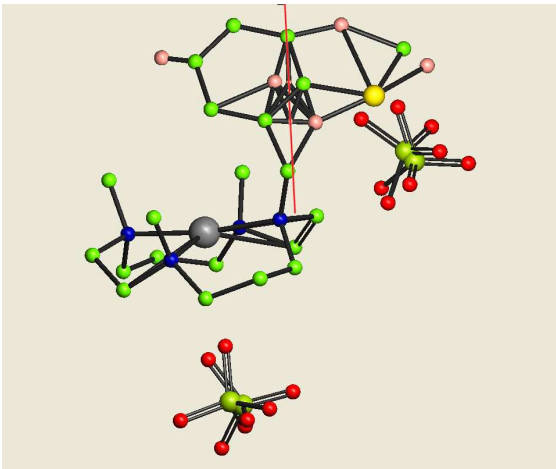




Figures 6.1.2 and 6.1.3: Difference Fourier peaks and rotation vector shown in red after first attempt.

All had to be checked for consistent site occupancy factors and sensible U-values to start the refinement. All the restraints needed to make the two groups look similar had to be written by hand, and their consistency had to be checked equally. These processes usually took a long time, mostly as searching for typing errors could easily result in a lengthy process.





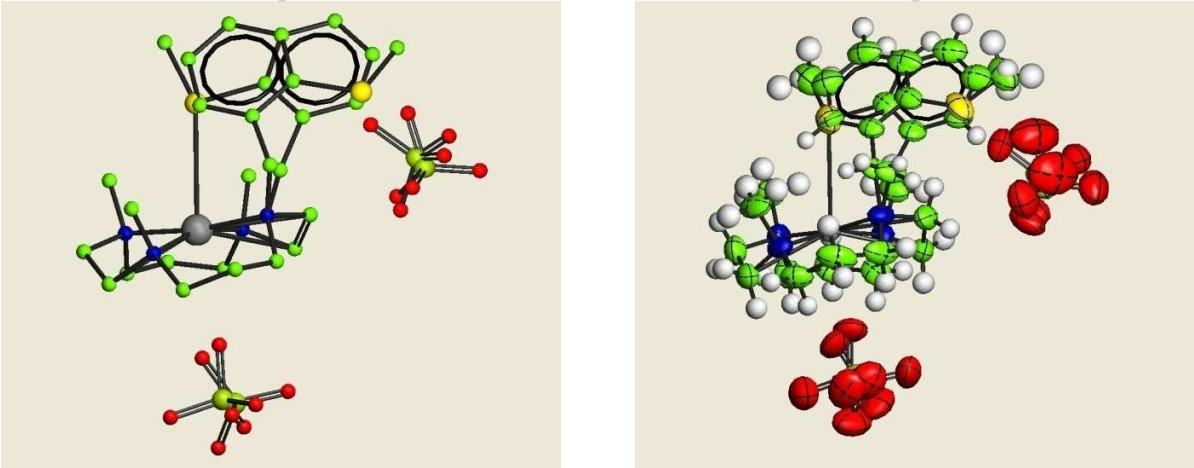
Figures 6.1.4 and 6.1.5: Difference Fourier peaks and rotation vector of the successful solution.

When redoing this structure, the disorder assembly was created using an axis defined by two dummy atoms between one position of the model under enquiry and one maximum of the residual electron density. This operation was done twice as the first tentative model created with the positions coming directly from the refinement was producing a second orientation of the structural fragment that was clearly too far off

from where it was expected to be in the end. Figures 6.1.2 and 6.1.3 show the peaks from the difference Fourier map and the rotation vector that could be derived from these positions.

Instead of correcting the disorder assembly the operation to create it from scratch was attempted a second time with an improved model that gave a better oriented axis for that purpose. Figure 6.1.4 and 6.1.5 show the regularized model with the difference Fourier peaks and the rotation vector that could be obtained after this improvement.

Creation of the restraints was without problems, and the refinement was successful when using shift limiting restraints as well as the geometrical restraints that were both released during the process of refinement. This led to a model very similar to the one obtained in 1997, with the advantage of being able to finish the refinement in about one hour.



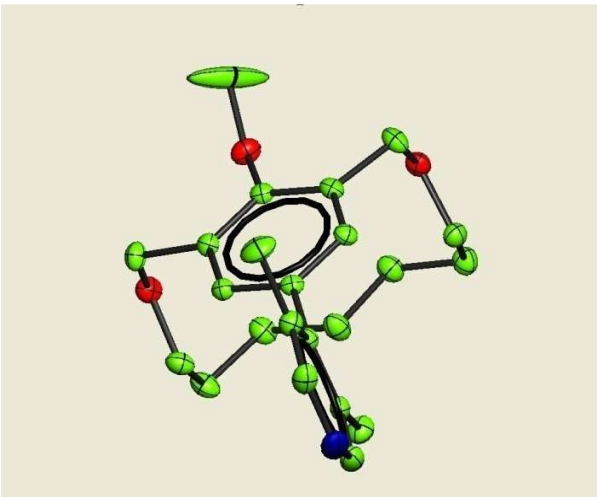
Figures 6.1.6 and 6.1.7: The model after the creation and initial refinement of all disorder assemblies and the final result after the addition of the Hydrogen atoms.

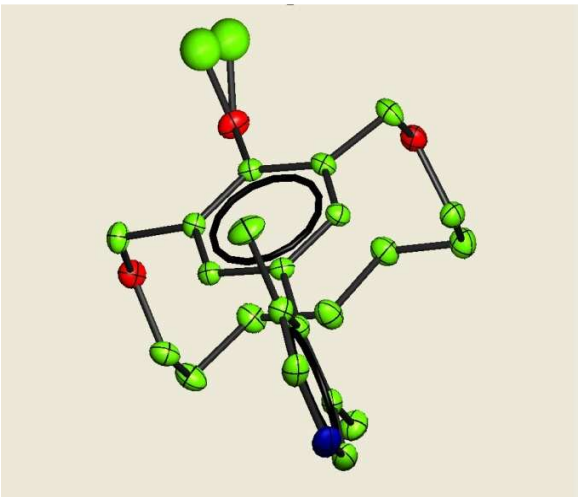
The two perchlorate ions were refined using the same technique after building the disorder assembly using the "split and join" script followed by some renaming by hand. The model as it was used before the last steps of the refinement can be seen in figure 6.1.6.

The refinements from 1997 and the one done recently to test the new scripts are very comparable, with the important difference that the final model as displayed in figure 6.1.7 could be obtained much faster. The advantage of using these scripts that help substantially in the initial steps with model building and with the creation of restraints is visible.

6.2 ts_a242-3 (structure 15)

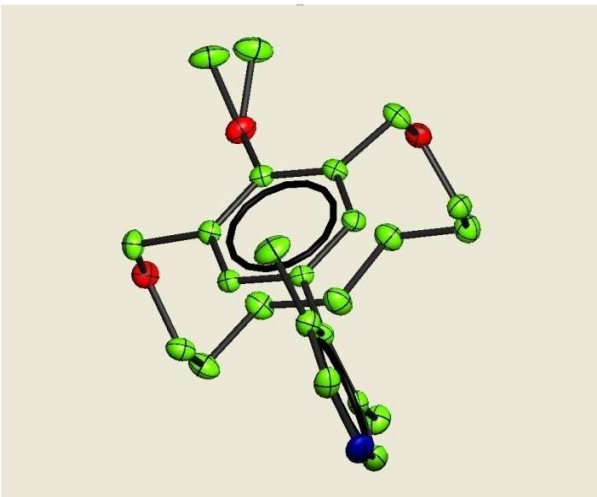
This structure shows a very simple disorder. The ester group of this molecule is not well behaved, as can be seen in figure 6.2.1. In a first approach it seemed appropriate to just split the Carbon atom. Figure 6.2.2 shows the starting model used in this refinement.

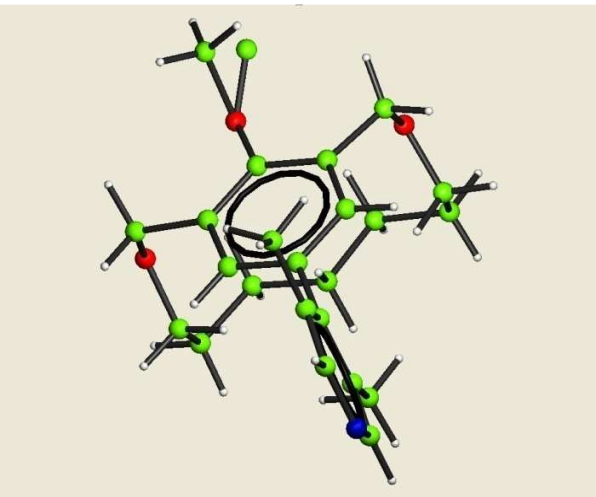




Figures 6.2.1 and 6.2.2: The disordered Carbon atom before and after splitting

This works fine in terms of improved R-value, and also the difference Fourier seemed to be improved, so the visual inspection of the model as displayed in figure 6.2.3 did not seem unreasonable. But the geometrical parameters of the ester group ended up being inacceptable. One Oxygen-Carbon bond was about 1.45 Angstrom, thus too long, and the other one was far too short with 1.33 Angstrom.

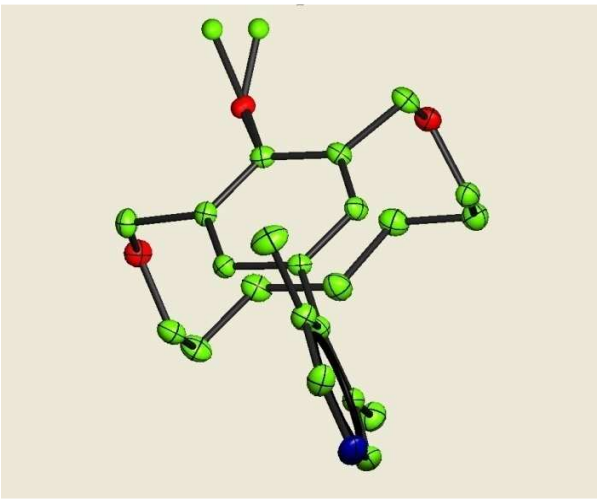




Figures 6.2.3 and 6.2.4: Initial stages of refinement and failure to place Hydrogen atoms due to bond distances being too far from the expected values.

Figure 6.2.4 shows the model after the attempt to add hydrogen atoms which failed as the distances were so far off the expected values for an sp3 hybridized Carbon atom. This difference could be made smaller with very tight restraints, but it was also clearly visible that this could only be achieved at some cost in terms of the R-value increasing. So this clearly could not be the best way to refine this structure.

The solution was to split the Oxygen atom too. Figure 6.2.5 shows the model after using "split and join" one more time, this time applied to the Oxygen atom. Successive reduction of the standard uncertainties lead to convergence of the refinement, and now the Oxygen-Carbon bonds refined to 1.418(4) and 1.423(4) respectively, a value that is in very good agreement with the one found for an ordered structure of the same type where the corresponding bond was found to be 1.407(2). Also the angles found about the Oxygen atom were very similar: 115.7(3) and 116.4(4) respectively in the disordered structure, and 115.37(12) in the ordered structure. Figure 6.2.6 shows the final result after the addition of the Hydrogen atoms.



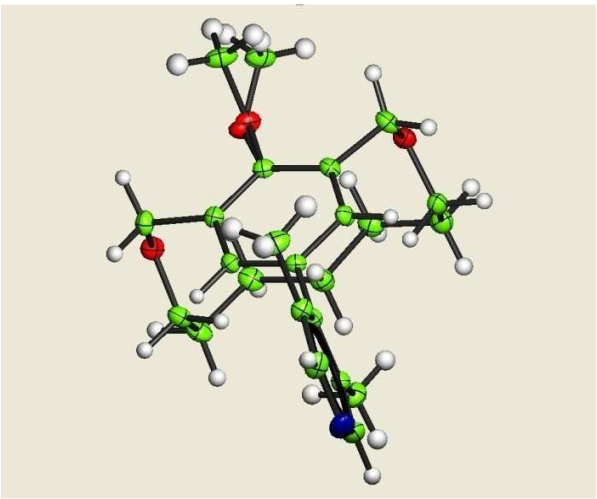


Figure 6.2.5 and 6.2.6: The model after splitting one more atom and at the end of the successful refinement.

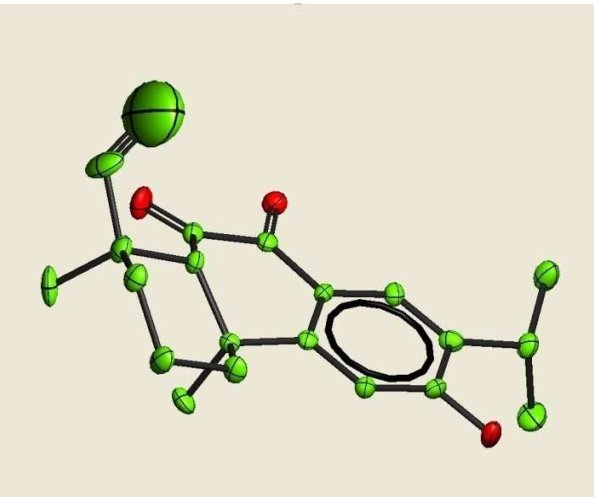
This example shows very nicely that the correct choice of what is refined as the disordered assembly is of big importance for the consistency of the final model. We do not speak about R-values or similar indicators, the question is how sensible our model is in terms of chemistry. This example is nice as it is, in all its apparent simplicity, sensitive to all sorts of possible errors made by the user.

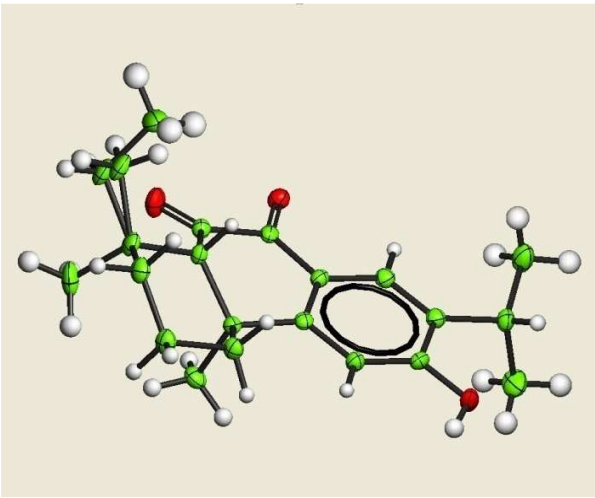
6.3 ckj_2-145b (structure 16)

The kind of disorder found in this structure is less frequently found. As visible on figure 6.3.1 one atom of the ethyl group on the left hand side of the structure seems to be ordered while the other one shows clearly that something is odd. Displacement parameters that are too big can indicate that a wrong atom type has been chosen. If the atom is assigned an atom type with more electrons than are actually present the displacement parameters are getting too small, in the opposite case they get too large. There was no lighter atom type that could have been chosen here that would have made sense in the current environment. So there remained the possibility to reduce the site occupancy factor. In order to make this work a disorder between a methyl and an ethyl group was assumed. This assumption was backed by the chemist's remark that the starting material had not been pure, and that it had not been possible to separate the two compounds having a methyl and an ethyl group respectively by chromatography.

The scripts to generate the restraints are always trying to produce sensible output, even if an unexpected situation appears. The two groups of the assembly are arranged in a way that the one containing more atoms is taken care of first. If for one of the atoms involved there is no corresponding atom in the second group, then the restraint is written for the first group and omitted for the second group. The restraints should be useful also in cases like this one preventing the user from writing or adapting restraints by hand as much as possible. Nevertheless more testing has to be done as this is the only example where this feature has been tested until now.

The refinement done on this model gave good results and site occupancy factors in the ratio of about 2:1. The resulting model is displayed in figure 6.3.2.

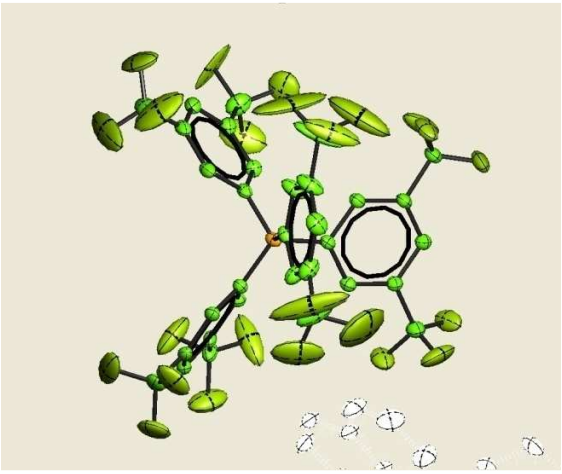


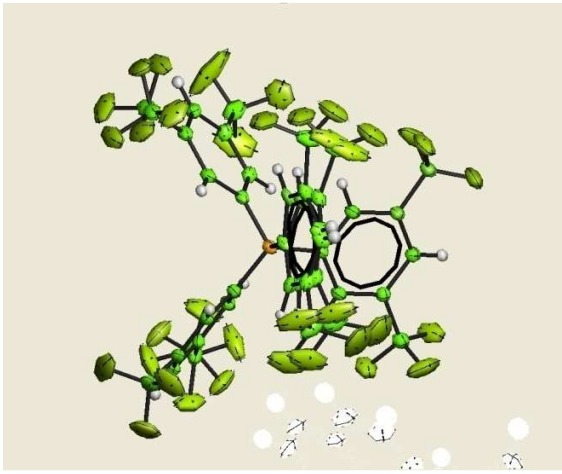


Figures 6.3.1 and 6.3.2: The problem visible after initial steps of refinement and its solution after the successful creation and refinement of a disorder model.

6.4 ek006_123k (structure 17)

This structure shows disorder in different places. Some of the CF_3 groups of the BARF counter-ion are disordered, this can be seen easily by examining the displacement parameters for the Fluorine atoms as shown in figure 6.4.1. For two of them the standard approach of building a starting model by duplicating the positions and rotating them by +15 degrees and -15 degrees respectively was quite successful. The three Fluorine atoms around C65 refined nicely and the ratio of the site occupancy factors was found to be about 2:1. The refinement of the second of these CF_3 groups around C56 was successful too and the refinement converged. Nevertheless the ratio of the site occupancy factors of about 5:1 suggests that with some probability the use of the non-atomic model of the torus to represent the electron density would probably have given better results.



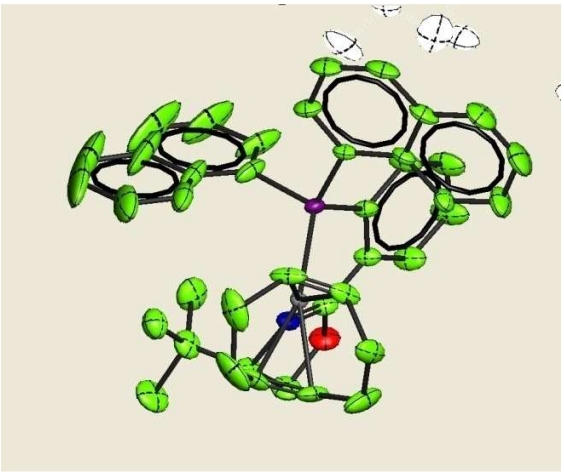


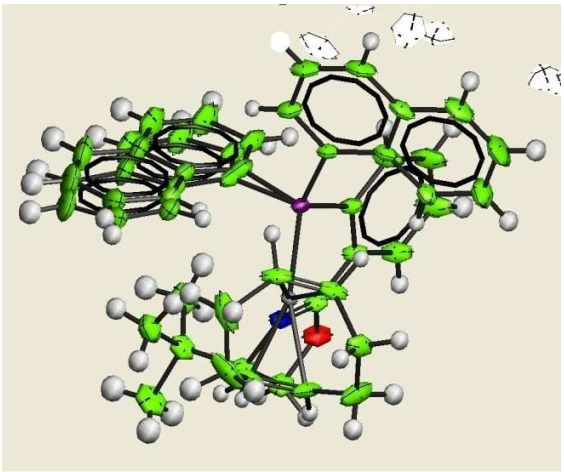
Figures 6.4.1 and 6.4.2: The disorder problem and its solution with the BARF counter-ion in this structure.

The refinement of the two disordered CF_3 groups around C72 and C73 started in the same way as described earlier. When the group refinement became stable and the site occupancy factors of both CF_3 groups could be refined the similar ratios of the site occupancy factors were suggesting that the disordered assembly was bigger than expected. The aromatic ring was added to the assembly using the "split and join" script, and later the two assemblies were joined together using the function "change assembly number". The site occupancy factors had to be adjusted manually (it is planned to write a proper script for joining assemblies) and the refinement could be started. Planar restraints for the rings including the Carbon atom of the CF_3 groups were added to the restraints list manually (for this it is also planned to write another script to create planar restraints within a disorder assembly). It was very important to find a good starting value for the shift limiting restraints in order to refine this assembly successfully. Putting in too tight values did stop the refinement almost completely, using too loose starting values made the model degrade. In the end the refinement converged with a ratio of the site

occupancy factors of roughly 2:1, and the resulting model is displayed in figure 6.4.2.

The disordered naphthalene group bonded to the Phosphorus atom of the ligand created similar problems when refining it. The displacement parameters as shown on the left upper corner of figure 6.4.3 were suggesting the group was disordered. The starting model was constructed using "split and join", and careful tuning of the esd's of all restraints made it possible to refine this disorder successfully. This included in particular generating different sets of restraints. The first set of restraints assumed all Carbon-Carbon bonds to be equal. This is for sure not correct, but as a first approximation it was successful. Later the bonds were restrained pair wise to their common mean. It is not entirely clear in this case whether this is a dynamic disorder that was just too pronounced that the harmonic oscillations could describe it well enough, or if there is some static disorder as well. The movement is not only along the plane of the naphthalene, but also perpendicular to it. The current refinement seems to predict well the electron density determined in the experiment and the model can be thought of as being chemically reasonable. The final R-factor of about 2.5% is very agreeable and the structure, as shown in figure 6.4.4, should be publishable in the current state of refinement without problems.





Figures 6.4.3 and 6.4.4: The disorder problem and its solution with the naphthalene group in this structure.

6.5 sps080b (structure 18)

This structure has been chosen as it is a typical example of a disorder that can be described roughly by a mirror plane defined by the least squares plane through all involved atoms. The refinement of this disorder, after setting up the model using the rotation around 180 degrees, fine tuning the model using the rotation menu and renaming the non matching atom labels, was straight forward and successful. As only four labels had to be changed the renaming was not difficult to do by hand this time.

Setting up the restraints using the script was straight forward, and the result from the refinement using the usual sequence from group refinement down to the individual atoms with anisotropic displacement parameters is shown in figure 6.5.1. In figure 6.5.2 the same molecule is displayed without the hydrogen atoms and in a simpler drawing style, and the coloring of the two parts shows nicely the mirror symmetry that links the two parts together.

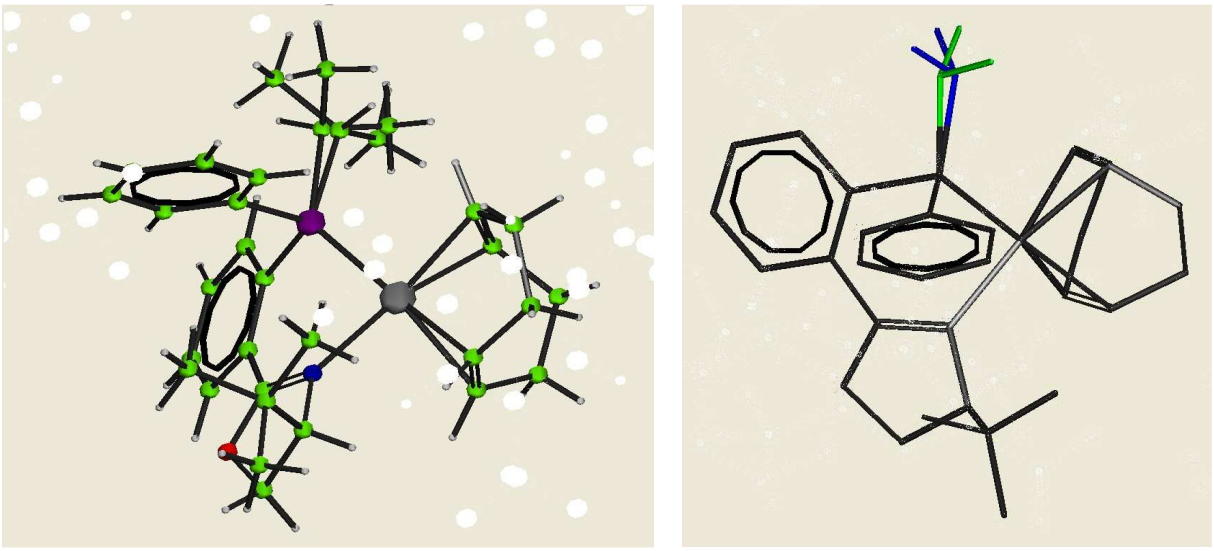
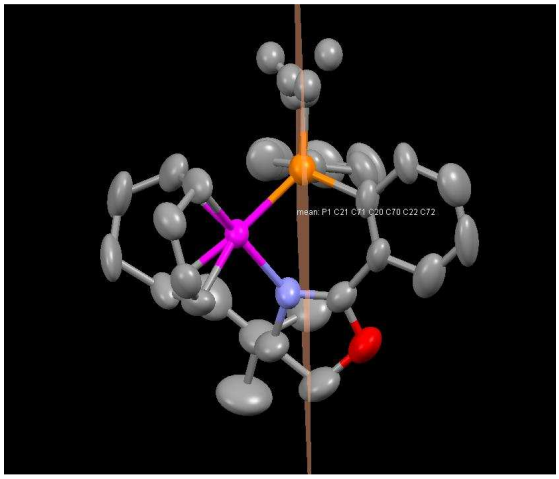
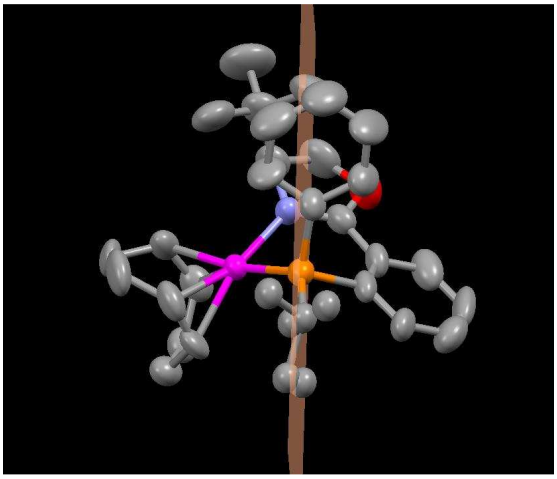


Figure 6.5.1 and 6.5.2: A view of the refined disorder assembly showing mirror symmetry. On the right hand side the disordered moiety is shown in colors, the site occupancy factors were refined to 0.526(12) (green) and 0.474(12) (blue).

The structure was done about ten years ago and the data is not of optimal quality. More work would be required to treat all the CF_3 groups of the BARF counter-ion (not displayed here) for disorder phenomena. Here the structure is retained mostly as it gives a good example for the mirror symmetry applicable to the disorder assembly, and at the same time for the typical imperfections of this kind of local symmetry in the disorder of molecular structures. Examining the two figures 6.5.3 and 6.5.4 it can be seen that the two groups are not only mirrored, but also shifted a slight bit along the mirror. All centroids between two opposite positions would lie approximately on the mirror, but the connection between them is not perpendicular to the supposed mirror.





Figures 6.5.3 and 6.5.4: Two views of the structure and the best plane created by all disordered atoms.

6.6 ep117b (structure 19)

This structure is another good example to show that disorder can be described using mirror symmetry.

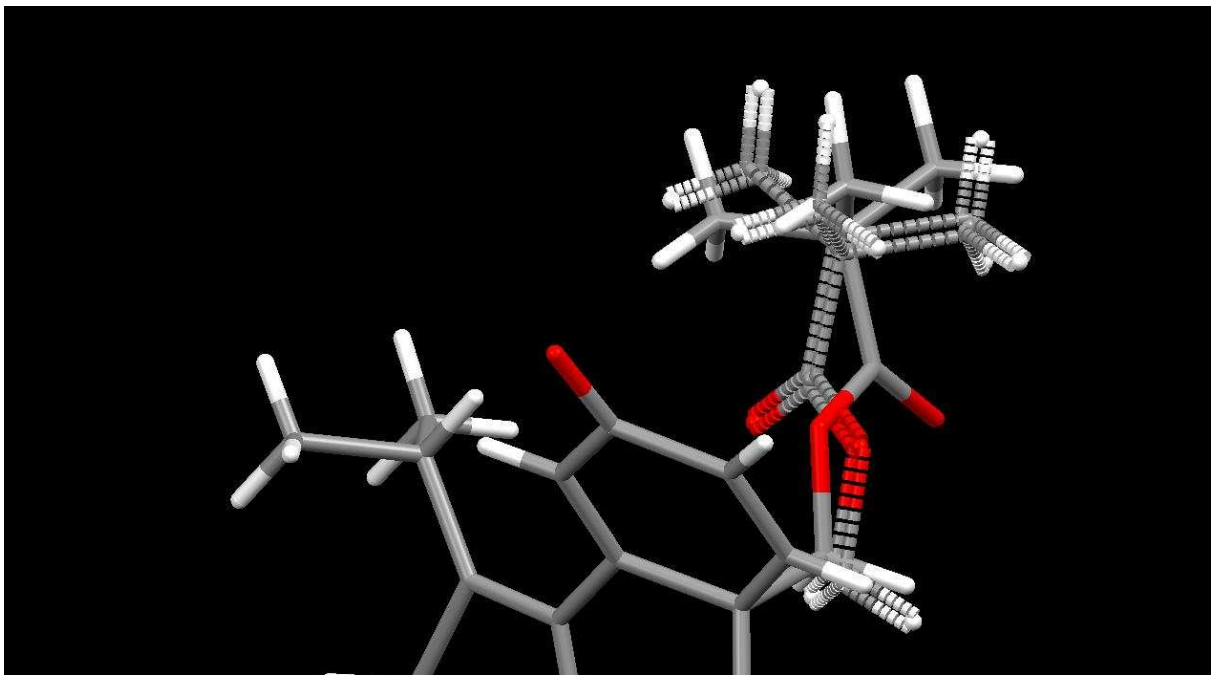
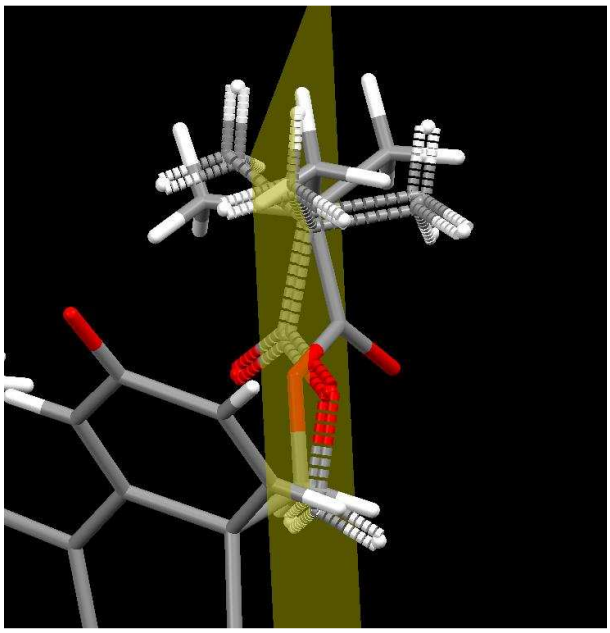
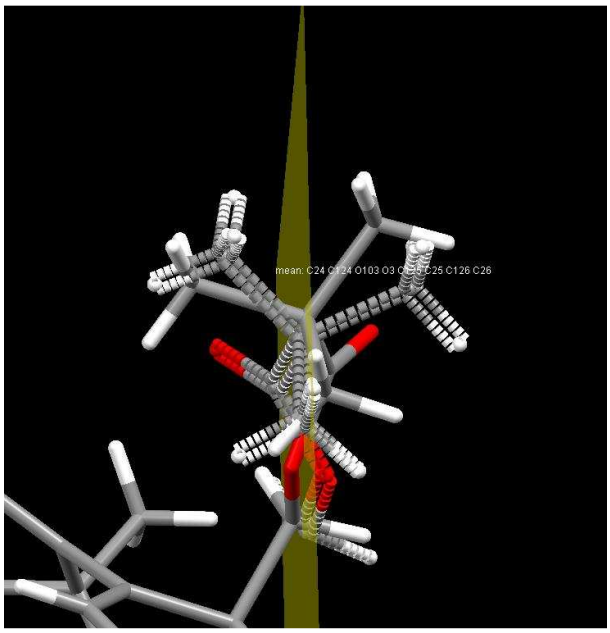


Figure 6.6.1: The final result of the refinement after the addition of planar restraints for the carboxylic ester group.

The starting model for the refinement of this disorder model was created using the rotation about one atom and the centroid of the group by +15 and -15 degrees. The refinement was straight forward as soon as the restraints were in accordance with the model. It was important to observe the planarity of the carboxylic ester group for the refinement to converge correctly to the model shown in figure 6.6.1. This is mentioned as in the model that came from direct methods both of the carbonyl Oxygen atoms could be seen, and without the proposed structure from the chemist this could have been taken easily for a sp3 hybridized Carbon atom.





Figures 6.6.2 and 6.6.3: Two views of the disordered groups being mirrored and shifted along the plane displayed in green.

As also mentioned in another occasion the local symmetry that can describe the relation between the two groups of the assembly is never perfect and is mostly of qualitative nature. Local conditions for each of the groups result in slight differences, and in this case a shift along the mirror plane is visible on both figures 6.6.2 and 6.6.3. It will be necessary to try and formalize the relations between disordered groups in assemblies in order to understand this kind of local symmetry in a more complete way.

6.7 jh120 (structure 20)

This structure is an example that the strategy proposed to treat static disorder more efficiently, as soon as the mathematical operation is recognized that describes this disorder, can also be extended to solvent molecule disorder. This structure would probably have passed reviewing without treating the disorder. The last maxima in the residual electron density could be explained by just mentioning that the solvent

molecule is disordered, even though the last maximum was too high with 1.65 electrons per cubic Angstrom. The R-value was slightly higher than 5%, but in view of the possible disorder this would probably have passed the reviewing process.

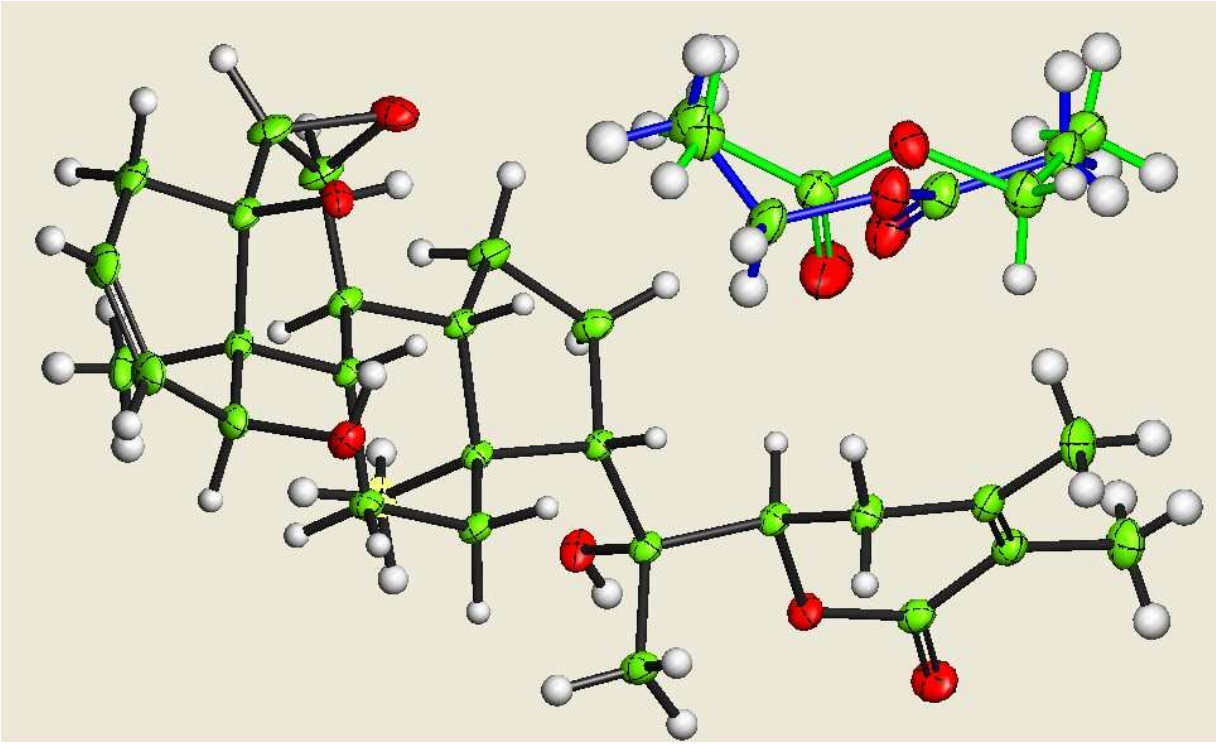


Figure 6.7.1: The final result of the refinement of a disordered solvent molecule. The two parts of the solvent molecule are shown with green and blue bonds respectively. The site occupancy factors refined to 0.7911(8) (green) and 0.2089(8) (blue) respectively.

Having a careful look at the maxima from the difference Fourier map, the positions of the two Oxygen atoms from the second orientation of the disordered solvent molecule could be identified. Now it was possible to build the vector for the rotation of the second group to be generated by the script that creates disorder assemblies by applying rotations to the second group of atoms. The rotation vector was constructed by creating two dummy atoms half-way between the two possible positions of the two Oxygen atoms.

By using the shift limiting restraints in the beginning to make the refinement process proceed slowly the disordered assembly could be refined successfully. At one stage one position of the newly created group had to be exchanged with a Fourier peak as the true position of this Carbon atom was too far away from the position that could be achieved by refining the newly generated positions. Once this process was finished the difference Fourier map was basically featureless with exception of the proton positions. By adding these protons to the model the structure refined down to 2.9%, and the residual electron

density had gone down to +0.28 and -0.14 electrons per cubic Angstrom. The final result is shown in figure 6.7.1.

6.8 gq214 (structure 21)

Similarly to the structure gq213 almost one entire ligand of this metallo-organic framework is disordered. The two orientations of two of the pyridine rings could be detected by the peaks found in the difference fourier maps. Soon it could be detected that at least with one of the two rings the disorder was not a rotation along a bonding axis, but the rotation axis was off the central atomic positions. The whole rings had to be split in order to improve the model. The “split and join” script was used to do that.

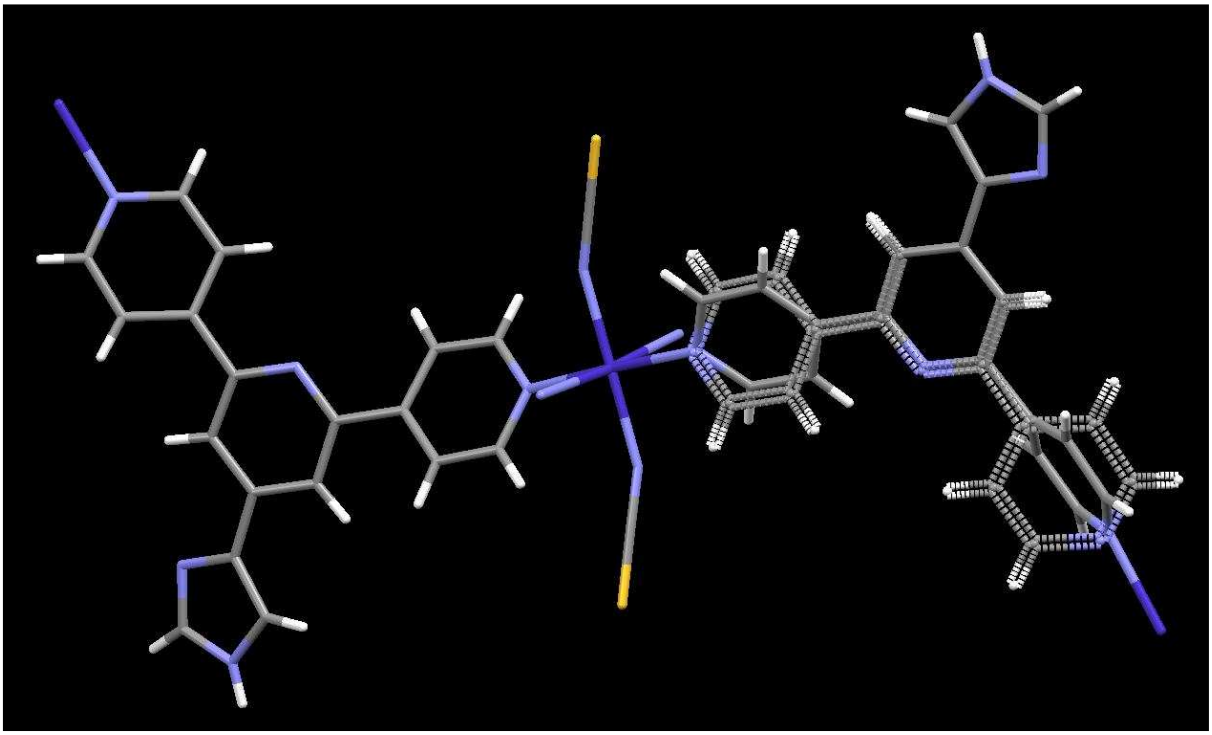


Figure 6.8.1: The ordered and the disordered ligand in the structure of this metallo-organic framework.

Assuming the bond to the neighboring rings was in plane with the aromatic ring soon showed another problem. The model leaving the central pyridine ring ordered while refining the two peripheral rings as disordered gave a quite deficient model. The central ring had quite high U-values perpendicular to its best plane, while the bonding axis to the neighboring rings made strange angles. Moreover the site occupancy factors of the two disordered rings suggested that these two disorders were correlated. Five of the six positions were split and the two assemblies were joined using the script to change assembly numbers in addition to some hand editing in order to get the site occupancy

factors consistent. To complete the starting model the “best plane” script was used to make the central pyridine ring perfectly planar before continuing the refinement. In addition to the restraints that could be generated automatically planar restraints were added to list 16 manually to improve the chances of a successful refinement.

The usual procedure was applied that tries to improve the fit continuously by adding to the refinement directives more parameters to refine while releasing the shift limiting restraints and the standard uncertainties of the geometrical restraints step by step. The site occupancy factors refined to a ratio of about 3:1 and the refinement converged at about 5.5% using 3σ data. The result is displayed in figure 6.8.1.

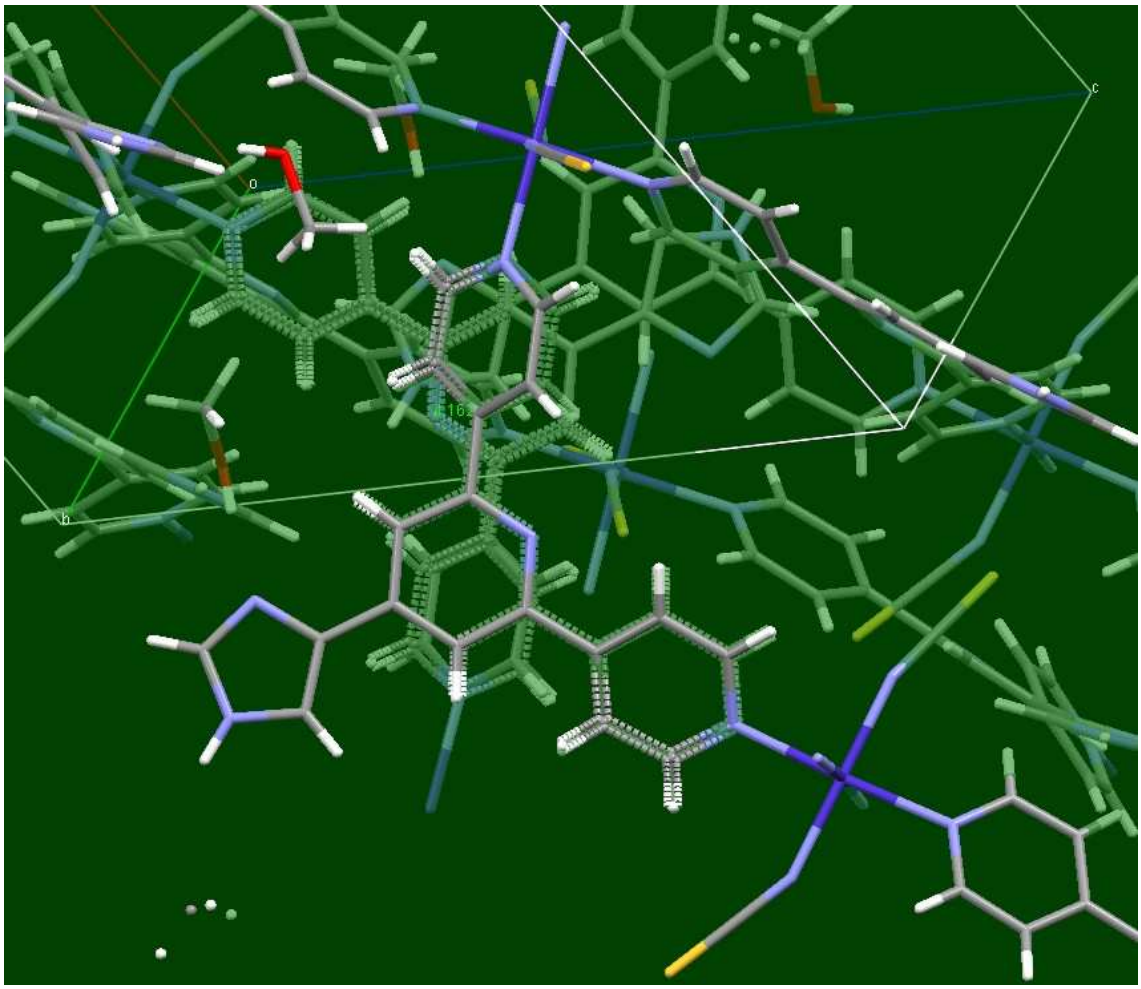


Figure 6.8.2: Two disordered ligands in proximity of each other.

A solvent molecule is disordered as well, and its site occupancy factors seem to suggest that this disorder is correlated with the disorder of the ligand too. As this is much less obvious than the ligand disorder the site occupancy factors have been refined independently until the final stages and have converged at a ratio of about 4:1.

The analysis of the structure showed interesting details concerning the understanding of the disorder situation. Firstly it could be observed that there are strands of disordered ligands interchanging with ordered strands of ligands as visible in figure 6.8.3. The environment of the disordered ligand was examined very carefully, and the reason for the disorder could be identified in one of the neighboring imidazole rings. One proton of the ring is pointing towards the pyridine ring and there is no space for it to be placed in plane with its neighboring ring. It has to fold up or down in order to find its place without causing contacts too near to be able to exist. The two orientations of the pyridine ring can be seen in figure 6.8.4 directly behind the Cobalt atom depicted in deep blue. On the left hand side of the disordered pyridine ring the imidazole ring can be seen, and one of its protons points right in between the two orientations of the rings.

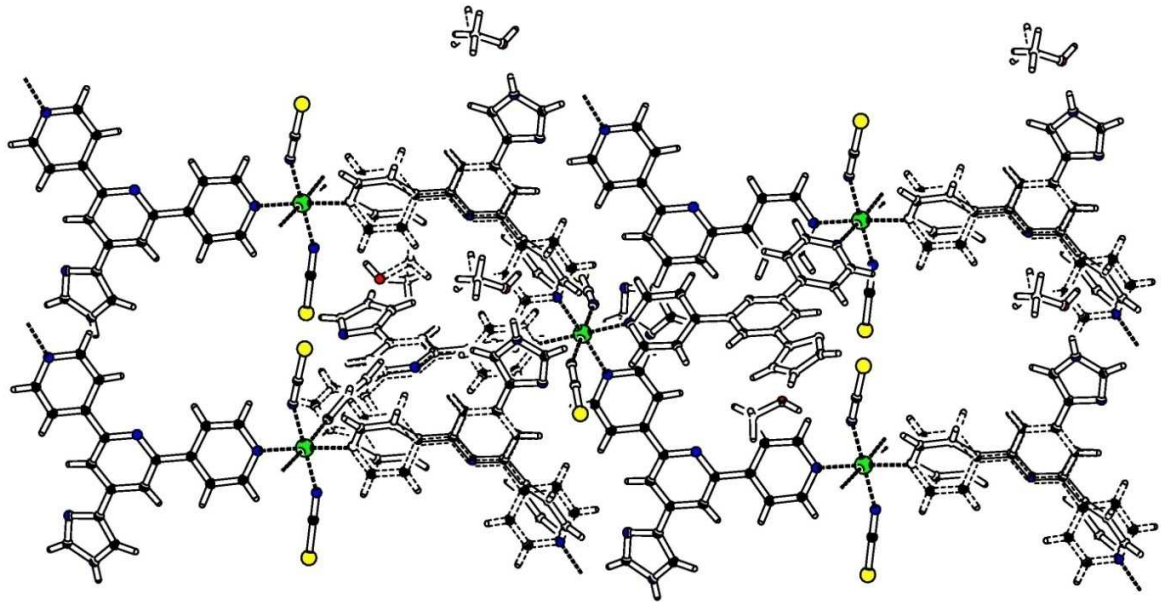


Figure 6.8.3: Strands of ordered and disordered ligands.

The plane of the ligand molecule directly above the molecule under investigation has been colored in green. The distance to the central Nitrogen atom of the ligand under investigation was measured to be 4.16 Angstrom. This is clearly more than we expect for π-π-stacking. As the ligand cannot arrange in a planar way this higher distance is well understandable. Figure 6.8.2 shows that two of the three pyridine rings are placed almost on top of each other.

The disorder can be understood in this way. It is local interactions that are the cause for this disorder, through which the structure can find its energetic minimum even though space is tight. The refinement has been carried out using a disorder model as the diffraction images gave no rise to the assumption that the structure would be modulated.

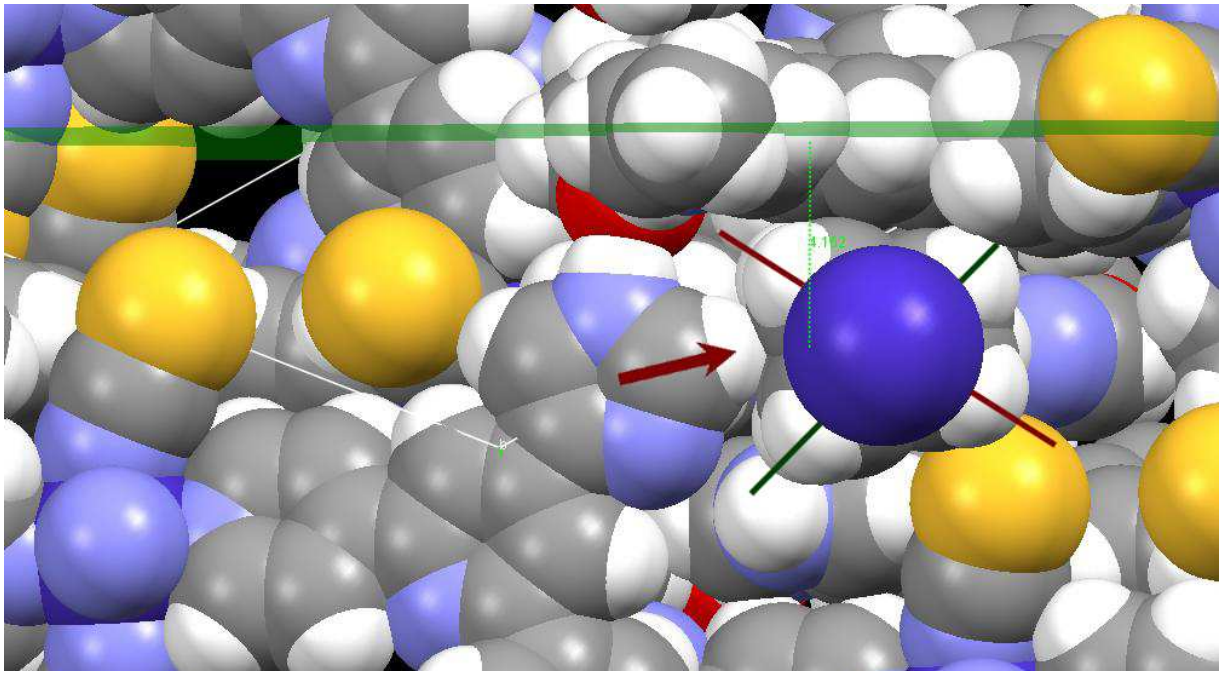
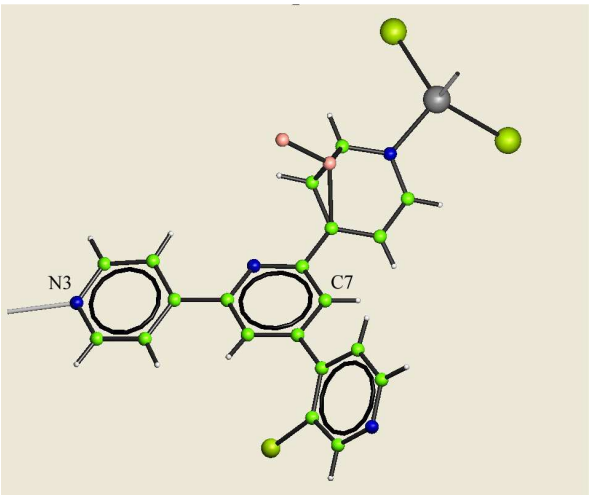


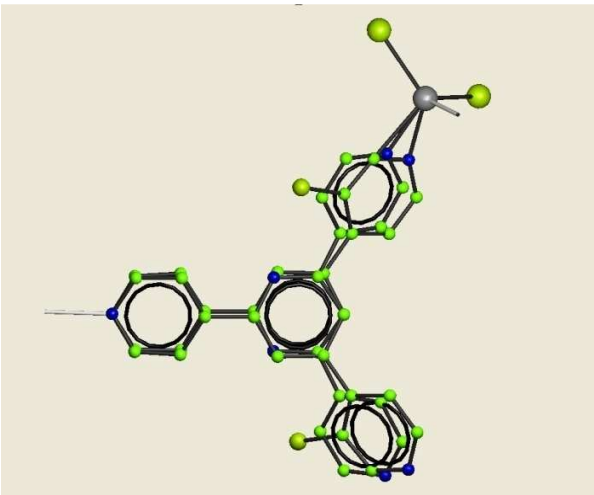
Figure 6.8.4: The imidazole ring in proximity of the disordered pyridine ring. The two orientations of the disordered pyridine rings are indicated by a green and a red line, the red arrow shows where the imidazole ring is coming near to the pyridine ring making the planar arrangement impossible.

6.9 gq213 (structure 7)

The highest peak in the fourier map of this structure was about 3 e/A^3 high, and it was in a position that suggested a second orientation of the ligand could be created by a rotation of 180 degrees around the bonding axis roughly defined by the atoms C(7), C(10), C(13) and N(3), which are oriented horizontally on figure 6.9.1. Its distance to the next Fourier peak, equally visible on figure 6.9.1, was 1.69 Angstrom, a distance that suggested a Chlorine-Carbon bond. The axis C(7) to N(3) was taken to produce a rough starting model by rotating the whole ligand aroung this axis by 180 degrees that is shown in figure 6.9.2.

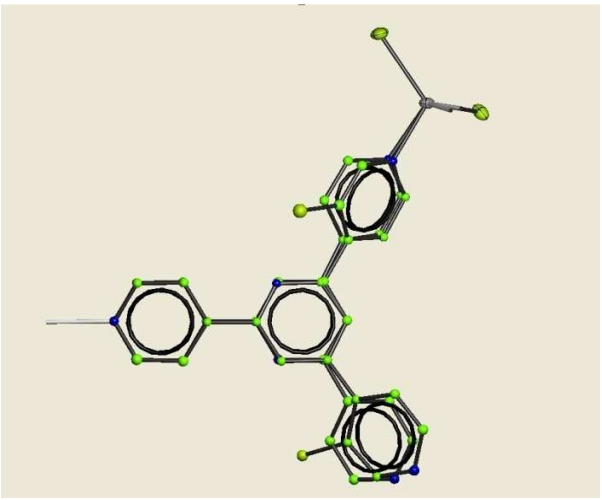
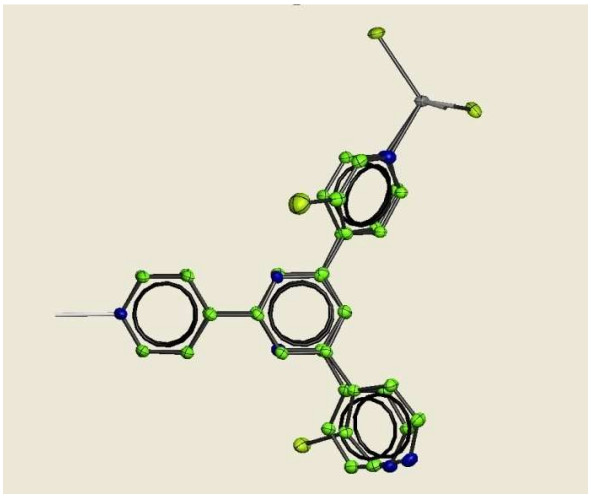
After generating the restraints the refinement could be started by increasing the complexity of the model in slow steps. It seemed that the geometry of the two variants of the ligand differed substantially. Planar restraints were added for the aromatic rings including the neighboring atoms and for both ligands the six membered rings were refined as two times four groups of atoms initially.





Figures 6.9.1 and 6.9.2: Residual electron density peak and the disorder model that could be derived from this peak. N3 and C7 were used to create the rotation axis that is shown on the right hand side.

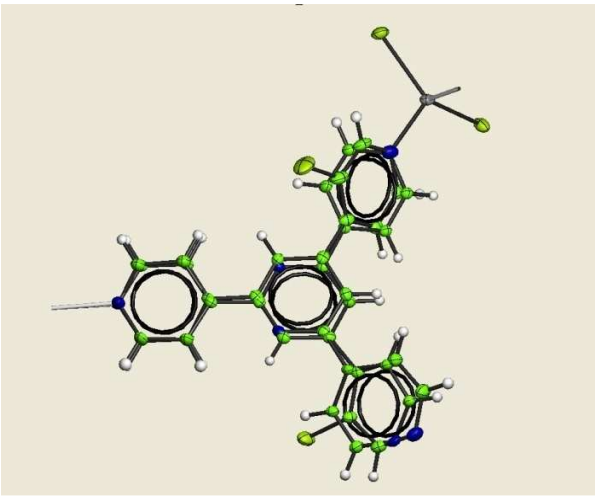
The improvement came slowly, the atoms could be refined individually and the standard uncertainties of the corresponding restraints could be released step by step. In particular the fit of the angle restraints improved on release of the standard uncertainties, and in the end the isotropic model had arrived at an R-value of about 6%. U-values could be refined, first the isotropic ones and later the anisotropic ones. After addition of the hydrogen atoms the structure could be refined to about 3.6 % of R-value. The isotropic model can be seen in figure 6.9.3. After anisotropic refinement the structure looked as shown in figure 6.9.4, and the final result is displayed in figure 6.9.5.

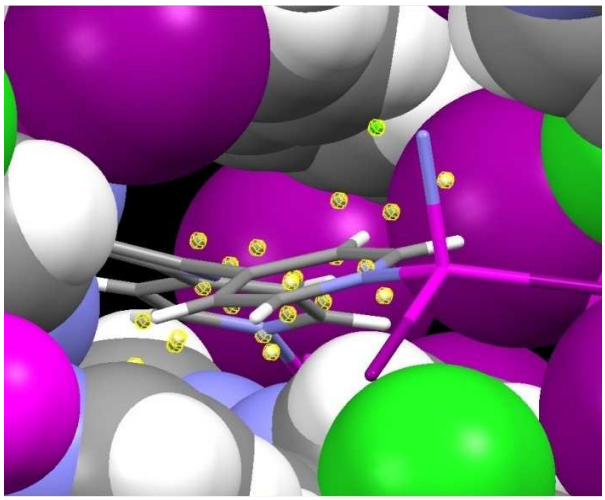



Figures 6.9.3 and 6.9.4: Isotropic and anisotropic refinement of this complex and experimental disorder model.

This is not that much of an improvement considering the initial R-value of about 3.8 % without treating the disorder. Also the residual electron density is still in the same order of magnitude. The advantage is that the highest residual electron density peak is now located where we expect it to be, near the Iodine atoms.

It can be discussed if a disordered part of a molecule can be refined if the site occupancy factor is indicating that the second orientation is only present in a ratio of about 1:20. Here the test was to see if it was possible, and this could be confirmed together with the fact that this ligand could also stay in a different orientation in that framework. Figure 6.9.6 shows the environment of one end of this ligand, and it can be seen why the less frequent arrangement of the ligand needs to be twisted in order to fit into the lattice.





Figures 6.9.5 and 6.9.6: The finished refinement. The disorder can be understood as a way to use space efficiently for two possible orientations of this ligand.

Even if the ordered structure would be submitted to the database the disordered refinement that could be finished in a relatively short time could be added to the supplementary material explaining this peak in a zone where no electron density should be found.

7 Outlook

Test runs with the new scripts show mostly positive results. In the successful case the disordered fragment is ready to be incorporated into the final refinement steps within minutes, while it used to be frequent to spend hours and days on building the model that finally would do the job. Also in cases where success is not gained on the first try it is much easier to attempt another approach if retrying does not mean to bury hours of hard work. The scripts will also prevent the user from typing errors that used to be very frequent setting up models and in particular restraints lists manually. In this way the new scripts will help to finish part of the refinement of disordered structures earlier, and in this way they will have the potential to improve the overall performance of a crystallographic laboratory and the quality of the refined models.

Of course no tool is suitable for all problems. Future work will be focused on extending the number of cases where this collection of scripts can provide a new possibility to get to a successful end of the refinement.

7.1 Adapt for the use with solvent disorder

One class of structures the functionality could be extended to, as far as it is not already possible, would be structures with disordered solvent molecules. Tools to match a solvent molecule template on a series of electron density maxima otherwise difficult to interpret are on the list for future developments. Refining solvent molecules using group refinement, thus with only one set of coordinates for the whole molecule and three parameters that define the orientation would be much more robust in refinement than single atom positions that are held together by some restraints.

In order to successfully manipulate such solvent accessed areas the routines to rotate a series of atoms needs to be enhanced. Rotations about a point in space to allow the free orientation of the solvent molecule would be most welcome. For instance water molecules could be placed and positioned to allow for hydrogen bonds.

7.2 New functions to create and edit assemblies

Scripts to allow for more possibilities and consistent ways to change defined assemblies and groups will need to be created. More ways to add atoms to assemblies or remove them again later while keeping the associated restraints and refinement directives consistent would be welcome. This should facilitate the work when the first attempt is not successful and the user tries again with changed settings. Also tools to give the possibility to join or separate assemblies could prove to be useful.

Also the way the assemblies are created can be optimized. Here is just one example. The starting value for the site occupancy factor is set in most automatic procedures to 0.5 for assemblies with two groups. If there is evidence that the actual values lie far away from 0.5 then it is useful to change the defaults in order to give the refinement a smoother start.

Moreover it is worth thinking of what can be done in cases where the total site occupancy does not add up to one. Thinking of ways of how to get good estimates for the displacement parameters and site occupancy factors are important as these two values are correlated and their contributions cannot be looked at separately. For this reason a mechanism that enables the user to refine the site occupancy factor relating its value to external values like the overall U[iso] value would be helpful in order to get a chemically reasonable structure.

7.3 Make non-atomic electron density descriptors available in the scripts

Menu driven programs tend to take away some of the freedom of choice of the user. The programmer offers a choice that is often only one amongst others. Finding the possibility to give more choices when building the disorder assembly is another long term aim. In this category falls the alternative to use the atomic positions for the first group of the assembly and a descriptor for non-atomic electron density for the second group of the same assembly. This would extend the possibilities of successful use with dynamic disorders where the construction of multiple models results only in marginal improvements.

7.4 Grouping commands together

Moreover it is planned to identify places where more single commands can be grouped sensibly together to make operation easier in particular for new users. The prototype is the script that creates assemblies. This script proposes a few tested sequences of commands that can be used as a whole instead of issuing multiple commands in a sequence memorized by the user. Every disordered structure has its particularities, and the aim must always be to be specific enough to handle it properly while staying general enough to be successful with as many disordered structures as possible.

7.5 Validation tools

In order to better analyze and validate the results possibilities have to be studied how the analysis of the operators linking multiple parts of a disordered assembly together can be made more robust. At the moment it can be checked if the assumption can be confirmed or not, but in the negative case the test ends there and gives no hint on how to change the setup in order to test successfully. Moreover the operator may be correct, but if for instance the rotation angle for the two-fold axis is not exactly 180 degrees the analysis of the centroids will in all probability fail.

A first step on this way is to calculate all the centroids of all pairs of atoms, then create dummy atoms and examine the result in the model window visually. Proper answers to the questions about what operator links the parts together would enforce the trust people can have in the use of this series of new scripts.

7.6 Partial refinement in order to speed up the calculations

The last point to be mentioned here is the idea to take up some old script written many years ago during a visit to the Crystallography laboratory in Oxford. This script, which is still part of the distribution of CRYSTALS, uses the feature of CRYSTALS to store A and B parts of the structure factors for one part of the structure and start the summations with these pre-calculated values instead of calculating everything from scratch every time. While the speed of modern computers has made us forget that we could be more economical with CPU time the issue reappears now when we are arranging the disorder of a few atoms in a large structure. The convergence is usually slower when the electron density map is fuzzier, and if we need to use shift limiting restraints the number of refinement cycles is higher anyway. Here the option of partial refinement offered by CRYSTALS could again become very useful as it could reduce substantially the time used for the refinement of a disordered assembly in a big structure. The reincorporation of this old script in the new context of disorder treatment could be a good way to save time when refining large structures with disordered areas.

An evaluation needs to be made if this approach really pays off in terms of time saving. While in a small structure the time to recalculate the structure factors from scratch is not an issue it may be more interesting when working on a small disorder assembly in a large structure. But still a lot of efforts are needed to set up the partial refinement while keeping all information necessary to restore the complete model later. It may turn out that it is better to go the safe way and use more computer time instead of using partial refinement.

7.7 Personal conclusions

There are two features that I missed during the work on the scripts. One is the possibility to define array variables, especially when passing multiple parameters from one script to another, and the other is the distinction between global and local variables. Instead of array variables I pass multiple parameters using a scratch file, and I think of creating a naming convention that allows to have better control over what variables are global and what others are local. The work with scripts is rewarding as the progress and the success is immediately visible. The integration in the CRYSTALS system and the availability of all crystallographic details within the scripting environment is invaluable when working on solutions for crystallographic problems.

It can be seen easily that the development of all these ideas will take its time. Solving one problem at a time will bring the best solution.

8 Description of the current implementation

In the following pages the way the current set of scripts is working will be described in detail. After a general description with a flow diagram and some remarks on important points to observe when using the system there follows a function reference describing what the single scripts do. For all scripts it will be listed where they are called from and what scripts they call.

8.1 General remarks

The new functionality implemented can be subdivided in two groups. There are functions to help preparing the initial model for refinement, and there are other functions to guide the user through the refinement of the model.

To start with it is necessary to recall where the scripts are placed in the CRYSTALS system. Since 1999 CRYSTALS provides a graphical user interface (GUI) that is implemented using the script processor that has been extended to support the GUI functions (Richard Cooper, Dphil Thesis, University of Oxford, 1999). All the GUI elements including the main window of CRYSTALS are invoked by the scripts, and this concept provides a good separation between the core functions coded in FORTRAN, the hardware specific GUI functions written in C++ and the functionality provided by the scripts. Writing scripts is therefore safe in the sense that no binary code is altered. Nevertheless the scripts have full access to all the crystallographic information stored by CRYSTALS. Like this complex functions can be built and tested using and reassembling the functions coded in the CRYSTALS core together with the functionality provided by the script processor.

The functions for the model building make extensive use of the graphics window provided by CRYSTALS. Model building should be as easy as just describing the nature of the disorder.

If there is for instance the suspicion that the disorder is caused by a peripheral group making use of its rotational freedom, then we should be able to apply this rotation to the part of the crystallographic model involved in the disorder observed. The rotation operation needs an axis to be defined. In the GUI this axis can be defined in different ways, for instance by clicking on two atoms or by right-clicking on a bond. The axis can even be defined by one atom and the centroid of the group to be rotated. As different as all these situations may look like, they can all be dealt with by the same piece of code as soon as the input is brought to some common format. The emphasis lies on finding intuitive ways of dealing with the difficulties we may encounter.

In a similar way there is the requirement to run other pieces of code from different points of the interface. This requirement leads to the following concept. The model window accepts mouse clicks in four different contexts: atom, bond, selection, and the so called “noatom” context concerning the whole structure. No matter what context is chosen the

system variables that store the mouse actions are available at any time. For instance it is possible to select a few atoms by clicking on them, and this selection will be available to some script also if it is finally invoked using the atom context of an atom not being part of the selection. Considering this, three interfacing scripts were written that do nothing else but accept the input the user provides by the GUI, standardize it and call the scripts that will do the work. Like that the same scripts can be called from different points and contexts giving the possibility of very flexible use. Having less need for writing specific code also gives a better starting position for maintaining or extending the script code.

There was no need for such interfacing in the scope of this project for the so called noatom context as the scripts that are usually are placed there are referring to the whole structure, and the only scripts written for that context were those to sort the atom list.

The three scripts aGetParamsAtom.scp, aGetParamsBond.scp and aGetParamsSelection.scp get their input from the three corresponding files popup-atom.srt, popup-bond.srt and popup-selection.srt. There the calls to the interfacing scripts are number coded, and each of the scripts can accept one option coded in some additional integer parameter.

To access all new scripts via these three scripts makes it also easy to keep track of what files are added in the context of the development of this project, and it facilitates it to keep track of global and local variables in the newly written scripts.

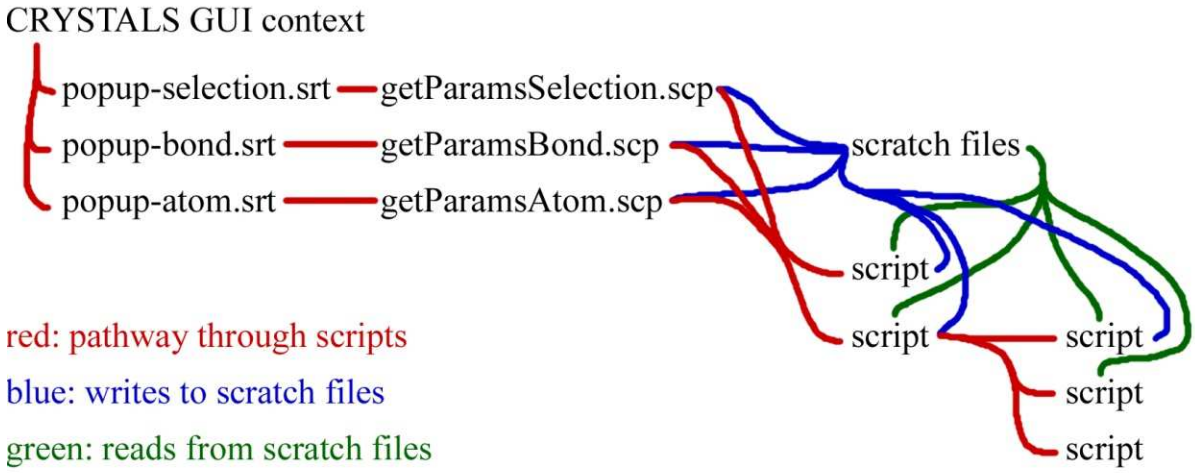


Figure 8.1: The flow scheme visualizes pathways and data flow in the scripts

In addition to these scripts that are hooked in the model window bringing the new functionality to the user via the context menus it is planned to create a new main menu in the menu bar. This menu will provide the same functionality as the scripts linked with the selection context without the use of the context menus. As all these scripts need the

_G context variable in order to get the selection of atoms, work on the realization of this menu depends on whether this context variable will be available for the menu entries in the menu bar. Alternatively a script can be created that gets the required input for these calls.

The new menu entries have been subdivided in four groups:

- Selection of atoms, preparation and regularization of the model
- Creation of the assembly using an operator
- Modification of the assembly
- Refinement of the assembly, generation of lists 12 (refinement directives) and 16 (restraints)

Later, when the different scripts will be presented one by one they will be ordered as a function of the kind of operation they execute.

8.2 The concept of assemblies and groups coded in the part number

Part numbers have already been mentioned earlier. Some details on the technical implementation will be explained in the following paragraph.

The number that identifies a group of atoms in the atom list is the part number. From the six digit integer the three leftmost digits are interpreted as the assembly number while the three rightmost digits are used as the group number. All parts whose three leftmost digits are identical are understood to be part of one assembly. It is important to know that on the level of CRYSTALS commands there is only the part number that can be used to address a group of atoms. In order to address the whole assembly the different parts setting up the assembly have to be addressed one by one.

In theory you can have 999 assemblies with 999 groups each. This mechanism gives a good means of knowing which parts are belonging together, which is essential when generating restraints or refinement directives automatically.

8.3 Some words about the strategy used to refine disordered structures

Setting up a disordered assembly consists basically of the following five steps.

First the new positions have to be defined. Usually this is a combination of splitting atoms and taking electron density peaks from the difference map.

The next step is to assign part numbers to these positions. Moreover the site occupancy factor has to be checked, and it is advisable to check if the starting values for the isotropic temperature factors are within reasonable ranges. In order to keep track of the atomic positions it is advisable to choose a good numbering scheme which helps to understand which positions correspond to each other.

Now is the moment to set up restraints for all these new positions. Having atom positions located very near to each other makes it indispensable to give some guidance to the refinement, as unrestrained refinement would in almost all cases not lead to chemically reasonable models.

Now refinement directives can be set up to optimize both models. Using the part numbers this is not excessively complicated as the single atoms do neither need to be in a specific place in the atom parameter list (LIST 5) nor do they need to have specific atom labels. Nevertheless some caution is needed to refine with care in the initial steps and to correct the directives and restraints before everything starts to diverge.

If convergence is detected, the refined assembly can be integrated in the refinement of the whole structure.

It is not extremely difficult to execute all these commands. Nevertheless experience shows that there are some problems if doing all this manually. The most severe limitation is that all these commands need to be done consistently. There are a lot of operations to be done over and over, and this work is prone to typing errors. There are pitfalls of that kind in all of the steps described, and sometimes it is difficult to find the error in the long text files that are needed to prepare that input.

This is the main reason why scripts that are helping the user to set up the disorder model in a consistent way and to produce all the commands necessary to refine it afterward are helpful in improving the efficiency of dealing with this kind of problems.

8.4 Rules and limitations

The menu controlled context often gives the user the impression that all combinations of commands are allowed. In analogy to and in agreement with many professional systems found in the most different areas of science and technology this is never true. During the programming work on the scripts and the following testing phase it became again very clear that the user can and must expect that errors are detected and handled in some way by the system. The programmer can only handle the errors he can think of or he is told about, and at the same time it is inevitable that the user observes some rules imposed by the strategy coded in the program and in the scripts. By getting feedback from users it will be possible to improve the systems in order to handle more situations, but probably never all of them. It is popular to say that a program that is free of bugs is probably obsolete.

There are features that can only be understood knowing the underlying concepts of the CRYSTALS core that are used by the interfacing scripts. During the tests one of these features showed to be of particular importance to observe, and this is linked with the way of defining atoms in the refinement directives (list 12).

8.4.1 Explicit and implicit definition of the atoms to be refined

Using the implicit way of defining parameters for refinement the hydrogen atoms are automatically excluded. This has proved to be very useful. A line like

FULL X'S U'S

will do what we usually expect the refinement to do, which is anisotropic refinement of all non-hydrogen atoms, as CRYSTALS will automatically exclude the hydrogen atoms if they should be present at that moment.

If we want to refine all atoms that are part of an assembly, it will be very useful to be able to address them by their corresponding PART numbers. In particular in early stages of refinement this will be very handy as the user may add atoms to or remove atoms from the assembly while the directives for refinement remain valid. Thus writing

PART(1001, X'S, U'S) PART(1002, X'S, U'S)

will work nicely as soon as we are aware that this is an explicit atom definition and therefore the automatic exclusion of the hydrogen atoms is not active. This means that the user has to observe the rule to add hydrogen atoms only in the last step, after having treated all disordered parts of the structure in such a way that the only extra parameter that has to be added to list 12 is the site occupancy factor.

8.4.2 Group refinement and RIDE instruction

Another feature has a deeper cause in the way the existing scripts of the CRYSTALS distribution are working. It has to be observed that if the scripts for finalizing the refinement will be used it is necessary that the group refinement is only used in the initial steps of the refinement of a disordered assembly. The possible problem that may appear later is if the user wishes to have the hydrogen atoms riding on the model refined as group. The geometric positioning of the hydrogen atoms will work as expected, but CRYSTALS will issue error messages claiming that constraints are applied twice. The command

GROUP PART(1001)

includes again the hydrogen atoms, so there would be no need to issue RIDE instructions. Having both the GROUP and the RIDE instruction puts in fact twice constraints on the same physical parameters of the hydrogen atom which leads to the problem described. In this case there would be no problems as all atoms belong to the same group that is refined together, and the RIDE instruction is simply not necessary as it would do nothing. If such internal dependencies are not recognized correctly the consequences are unpredictable. Therefore the input checking of CRYSTALS is strict and does not allow applying constraints more than once to atomic parameters.

It is possible that one is short of observed reflections the group refinement should be maintained until the final steps of refinement. If RIDE instructions are present for atoms that are part of a group, then these need to be deleted before refining, or the individual atom names of the group excluding the hydrogen atoms need to be used in the GROUP instruction instead of the PART with its part number. This could be done by a script that checks the refinement directives (list 12). This script could filter out the RIDE instruction out as soon as the atom in question is part of a group, or the GROUP instruction could be rewritten as described.

For the moment this is not done, so the structure can only be finalized using the scripts if the atoms are refined individually with the restraints.

8.4.3 Hydrogen placing at the boundary between bulk structure and assembly

The existing script to add hydrogen atoms keeps track of the individual groups of the assemblies very well when calculating the hydrogen atom positions. Caution is only needed when there is a carbon atom acting as a bridge between the ordered bulk structure and the disordered assembly. As both atoms of the two groups will form a bond to this atom the assumptions about the hybridization will not work correctly. There are different ways to overcome this problem. In the most favorable case the hydrogen atoms expected for this carbon atom can be located in the difference map. They then can be very easily included in the final structural model.

The new scripts call from time to time other scripts that have been in the distribution for a quite long time, but with the exception of one called xwrite5.scp they could all be left unchanged.

8.4.4 Generating list 12 for structures containing assemblies

The only script from the distribution that needed changes in order to have the possibility to refine the whole structure until the end is the script that generates list 12, the refinement directives. The way to implement this was to check if the structure contains assemblies. If there are assemblies with groups the user gets an additional checkbox where the refinement of the site occupancy factors can be requested, and this checkbox is checked by default. For each part number the command to refine the corresponding site occupancy factor is issued, and for the last part number the following command is added:

WEIGHT -1 PART(XXXYYY, OCC)

XXX stands here for the assembly number and YYY for the group number. If there should only be one group in the assembly then this command is suppressed. If there are more than two groups in the assembly then the WEIGHT command is issued for the last part.

If the functionality will be extended to assemblies with more than two groups then the code must be changed in order to use the SUMFIX instruction. This change will be needed for all scripts writing lists 12 for disorder assemblies.

8.5 Function reference

The scripts that permit to create or modify and refine disorder assemblies and groups will now be described one by one after a short description of the naming of the scripts. This list will be ordered in groups of scripts whose functionality is similar. A general description of how the algorithm is designed will be followed by a list of the scripts that call this script and a list of all scripts called by this script.

8.5.1 Naming of the scripts

In order to keep track of the files added to the CRYSTALS system all file names were chosen to start with a_ and this proved to be very useful as a simple ordering in alphabetical way was sufficient to identify the files added to the system. Ordering following the modification dates then gave the recent additions, and this permitted to keep track of what had been changed recently.

In a similar way all new entries in the menus or context menus were marked with my first name so that it was always clearly visible what was added and what was part of the original distribution.

8.5.2 Naming of the variables

In order to keep track of what are global and what are local variables the place where all variables are declared is marked with comments. During the test phase it could be seen that finding a better system to distinguish and keep track between the global and the local scope is advised in order to improve the reliability of the code.

8.5.3 Scripts that build the interface to the current distribution

aGetParamsSelection.scp, aGetParamsBond.scp and aGetParamsAtom.scp

These three scripts do basically the same task. This is why they are described together. The files defining the context menus which all have the file extension .srt pass the context sensitive parameters _G and _A to the scripts. _G stands for “group” and is the list of the currently selected atoms, _A stands for “atom” and contains the atom the mouse pointer is positioned on or the two atoms defining a bond, also this one pointed at by the mouse pointer.

There are more of these context sensitive parameters. One of them is _S which stands for “symmetry” and contains the second atom of a bond and can contain symmetry

information. This parameter is used a few times in the menu entries. If a script, for instance a_transSelection.scp, needs the symmetry information it extracts it from the _A context variable. _A contains two atoms defining a bond when called from the bond context menu, and_S is identical with the second atom held by the variable _A.

The three interfacing scripts write the parameters passed to them to scratch files which then are available to other scripts. In the case of a single atom or of two atoms of a bond the parameters are also stored in variables in order to have the alternative way of passing these variables to other scripts, but in fact it creates more confusion than it helps if different ways to pass are available, and because of this all atom pointer passing will be via file in the future.

Two additional parameters provided by the files defining the context menus are read in and stored in the global variables IMODE and IOPTION. While IMODE is used as a switch to call the script corresponding to the required action, IOPTION is passed on and defines different options that can be chosen from the menu entries.

Usually all scratch files are cleared after a successful run of a script (see script a_tidyup.scp). There is one exception. If a selection is stored for use in the next step the corresponding file is kept. If the script called next detects that some input is missing it takes the stored file and uses its content for the further execution of the script.

The scripts called by these three interfacing scripts are listed in the following table. The letter X means that this option is operational, the letter P stands for planned or possible.

	aGetParamsSelection	aGetParamsBond	aGetParamsAtom
a_resetSelection	X		
a_calcMolax	X	X	X
a_regularise	X		
a_refineSelection	X		
a_createAssembly	X	X	X
a_createCentroid	X	P	
a_bondSValMenu	X	X	
a_rotateAtomsMenu	X	X	X
a_exchangePos	X (if two atoms sel.)	P	
a_refineAssembly	P (taking first atom)		X
a_changeRAG	P (taking first atom)		X
a_setUisoRAG	P (taking first atom)		X
a_splitJoin	P (taking first atom)		X
a_changeGrp	P (taking first atom)		X
a_doNothing	P	P	P
a_refineAssMenu	P (taking first atom)		P (script tested)
a_selectAtoms.scp	P (taking first atom)		X
a_selectSymmAtom.scp		X	
a_transSelection.scp		X	

As visible in this table most scripts called by these three interface scripts are called from different contexts. This gives wider possibilities for the use of the scripts.

There are a few old scripts that are called directly without passing through the interfacing scripts. It is planned to move them to the common context in order to make maintenance easier.

In the atom context the only script called directly is:

a_infoAboutAtom.scp

In the so called noatom context there are three scripts called directly and one sub-menu that contains directly CRYSTALS commands. It is planned to write the corresponding interface script aGetParamsNoatom.scp. The scripts called by this interface script will be:

a_infoAboutStructure.scp
a_sortAtomsMenu.scp
a_removeDummyAtoms.scp
a_invert.scp (this script is not part of the project)

The script a_xdscman.scp could also be included in the noatom context. It gives the possibility to see the different atom lists in sequence.

There are also CRYSTALS commands issued to sort the atom list. The script a_sortAtomsMenu.scp will replace these calls on completion.

The three interfacing scripts are called by the GUI in the mode defined in the three files popup-selection.srt, popup-bond.str und popup-atom.srt. They call various scripts described in the following.

8.5.4 Scripts to prepare the model

a_selectAtoms.scp

This script gives the user an easy way to select a group of atoms. The atom name provided is searched in the atom list, and depending on the option chosen all atoms belonging to the same residual, assembly or group are selected. This selection may then be passed to other scripts in the same way as if we would have selected the atoms by clicking on them. A_selectAtoms.scp is called by different scripts that need to get hold of a whole predefined group of atoms. The script is invoked mainly from scripts in the atom context menu and takes the residue, assembly or group number of this atom to select all atoms that belong to the same group of atoms.

This script is called by the following scripts: *This script calls no other scripts.*
aGetParamsAtom.scp
a_changeGrp.scp
a_splitJoin.scp
a_xInfoAboutAtom.scp
a_xInfoAboutStructure.scp

a_selectSymmAtom.scp

With the help of this script it is possible to highlight the atom that is at the other end of a bond showing a contact to a symmetry related position. It gives the possibility to see easily which positions are involved and can eventually be moved to the adjacent position using a_transSelection.scp.

This script is called by the following script: *This script calls no other scripts.*
aGetParamsBond.scp

a_regularise.scp

This script is an interface to some options of the CRYSTALS command #REGULARISE. In its current state it is incomplete. The options that work are PHENYL, HEXAGON and CPD. There is a considerable amount of coding needed in order to get the atoms into the right order, and as this functionality is not central to the question of refining disorders the completion of this script has been postponed. Nevertheless the three options mentioned can be used, even though only when all positions are present. It is a plan to include the AUGMENT option, but the same consideration applies that has been stated earlier.

This script is called by the following script: *This script calls the following other scripts:*
aGetParamsSelection.scp
a_restorePoint.scp
a_orthoAtoms.scp
a_sortRing.scp
a_resetToRP.scp

a_createCentroid.scp

This script takes a selection of atoms and calculates the mean value for the coordinates that will then be attributed to a dummy atom of type QH. There is another script that can create centroids, but it works in a different way.

The option “centroid of group” that calls the script xcentro.scp creates a dummy atom of type QC with anisotropic displacement parameters that cover the whole selection used to create the centroid. This should make it easier to understand the way the lattice is constructed when viewing these dummy atoms with some molecular graphics program.

The centroids created by a createCentroid.scp get attributed an isotropic displacement parameter of 0.005 so that they are visible in the graphics window as small spots that can be used, for instance to perform rotations.

This script is called by the following scripts:	*This script calls the following other script:*
aGetParamsSelection.scp	nextfree.scp
a_addCentroid.scp	

a_resetSelection.scp

This script is called at the menu entry “save selection”. Saving the current selection means that now the selection is written to some scratch file, and this operation has been done already by the interfacing script aGetParamsSelection.scp. In order to give the user the possibility to make a new selection to be used as input along with the previous selection the model window has to be reset to its normal state without any selection. This script is therefore only used to do this reset after having stored a selection to file.

This script is called by the following script:	*This script calls no other scripts.*
aGetParamsSelection.scp	

8.5.5 Scripts to create assemblies

a_createAssembly.scp

This script helps the user to create the two groups of a disorder assembly by creating a copy of the original positions and by applying a rotation to the model and its “clone”. In order to have consistent atom names an offset is applied to all serial numbers of the atoms. This offset is by default 100 (and can be changed if clashes are detected). At the moment two different modes of operation are implemented. The first is to rotate the original positions by -15 degrees and the positions of the “clone” by +15 degrees in respect to the original position before the execution of this script. The second mode is to leave the original positions unchanged and to rotate the “clone” by 180 degrees. In order to work the script needs a rotation axis to be defined by the user. This can be a bond or the line through one atom and the centroid of the atoms to be duplicated and rotated.

The script assigns a new assembly and new group numbers for the two groups created by this operation, makes the atoms isotropic if they should already have been refined

anisotropically and keeps track of site occupancy factors to be changed to 0.5 for the starting model.

In fact the mirroring could also be used as an operation to create the new model, and examples have been found where this would have been a good choice. For the moment this possibility has not been followed as firstly most of the cases can be treated well with one of the two rotation options provided, and secondly it seems to be more fruitful at the moment to improve other scripts as the definition of the mirror plane to be used with dummy atoms seems to be quite complex. In all cases this option will need much more user intervention to work as expected than the rotation options, which until now have produced reliable results with a minimum of user input.

This script is called by the following scripts:	*This script calls the following other scripts:*
aGetParamsSelection.scp	a_getResiduesAndParts.scp
aGetParamsBond.scp	a_addCentroid.scp
aGetParamsAtom.scp	a_setUiso.scp
	a_refreshselection.scp
	a_xMessage.scp
	a_removeDummyAtoms.scp

a_splitJoin.scp

In this script another way is followed when creating a disorder assembly. Users may have tried to refine anisotropically a structure and may have spotted that the anisotropic displacement parameters (adp's) suggest that a part of the structure is disordered. This often happens with aliphatic C-C chains, and with these chains the approach to define an operation that produces the second disorder group out of a first set of coordinates cannot be used.

CRYSTALS offers the possibility to split atoms. The script uses this operation and adds new functions to it that should make its use simpler and easier. The original command creates two new positions with sof of 0.5 and leaves the original atom in place. While it might be interesting to keep the original position in order to compare the solution with disorder groups with the original refinement it is necessary to get rid of them for the refinement. The script therefore deletes the original positions. One of the newly created positions gets the original atom names, for the other an offset, usually 100, is applied to the serial number.

More work is done to help the user to keep the model consistent. The script analyzes the environment of the newly split atom and searches for atoms that are already part of an assembly. If there is no assembly nearby a new assembly is created. If atoms are found that are already part of some assembly, the newly split positions are assigned to the assembly and groups of the neighboring atoms, and the eventually already refined

isotropic U-values and the site occupancy factors of the existing groups are used for the newly split atoms.

If more than one assembly should be found in the neighborhood the user gets the choice to which of the assemblies the newly split positions should be joined. In the end the user gets the possibility to choose which of the group numbers should be assigned to the new positions respectively. Changing them interactively gives the possibility to make the best choice.

In this way it is very easy and quick to split a few atoms and to build a disorder assembly with consistent groups out of the split positions.

This script is called by the following script:	*This script calls the following other scripts:*
aGetParamsAtom.scp	a_calcDistPivot.scp isfree.scp a_getResiduesAndParts.scp a_selectAtoms.scp a_setFlipParts.scp

8.5.6 Scripts to modify residues, assemblies and groups

a_changeRAG.scp

With the help of this script the residue-, assembly- or group number of the chosen group of atoms can be changed. This can be practical if sorting the atoms by residue or part numbers does not give the expected order in the atom list. Moreover it is a yet very simple way of joining two assemblies, even though the script, when joining, does not yet take care of site occupancy factors or U-values.

This script is called by the following script:	*This script calls the following other scripts:*
aGetParamsAtom.scp	a_isfreeR.scp a_isfreeA.scp a_isfreeP.scp

a_setUisoRAG.scp

When refining anisotropic displacement parameters too early, before the refinement of the positions has fully converged, the adp’s may become weird even though the refinement seems to be OK. The same effect can happen if the standard deviations on the restraints applied are released too early or too much. It is a pity and a loss of time in these cases to go back all the steps of refinement already done, as in most cases it is enough to restart the refinement at this point with isotropic U-values. This script reverts

all atoms of a residue, assembly or group to isotropic, and if the user wishes it calculates the mean value of all U-iso values and applies this mean value instead of the value calculated for each individual atom from its anisotropic displacement parameters.

This script is called by the following script:
aGetParamsAtom.scp

This script calls no other scripts.

a_changeGrp.scp

When splitting the atoms with a_splitJoin.scp the user has to decide which new position should be added to which group of the existing assembly. During refinement it can be necessary to change this if observing the ongoing refinement suggests that the initial choice might have been wrong. The script gives the possibility to flip the two atoms keeping the U-values and sof's consistent with the existing groups.

This script is called by the following script:
aGetParamsAtom.scp

This script calls the following other scripts:
a_xMessage.scp
a_getAtom.scp
a_checkAtom.scp
isfree.scp
nextfree.scp
a_selectAtoms.scp
a_setFlipParts.scp

a_exchangePos.scp

This is a very easy and elegant way to modify a part. Assuming the case the refinement would have gone well except for one position this could possibly be observed in the difference map. It is tedious to delete the old position, rename the Q peak to become the new atomic position, assure the assembly and group number are OK, and make sure the site occupancy factor is in agreement. The script helps the user by simply exchanging the positional parameters of the two positions in question. In favorable cases this will solve the problem in a satisfactory way.

This script is called by the following script:
aGetParamsSelection.scp

This script calls no other scripts.

a_setFlipParts.scp

Building the disorder assembly using the option "split and join" there is always the

decision which way round to take the two new positions that should be joined to the existing disorder assembly. This script gives the possibility to flip the two atoms belonging to the two disorder groups of the assembly so that the result can be visualized and evaluated, and the version which looks most promising can be retained for further refinement.

This script is called by the following scripts: *This script calls no other scripts.*

a_changeGrp.scp
a_splitJoin.scp

a_rotateAtomsMenu.scp

a_rotateAtomsMenu calls a_rotateAtoms after having got a rotation angle from the user. If the rotation angle zero is passed to the script a menu driven interface is invoked that allows modifying the atomic model interactively by calling a_rotateAtoms as many times as the corresponding button is hit while the rotation angle applied can be changed at the users choice. If a rotating angle other than zero is passed the script calls a_rotateAtoms once with this rotation angle and exits.

This script is called by the following scripts:	*This script calls the following other scripts:*
aGetParamsSelection.scp	a_addCentroid.scp
aGetParamsBond.scp	a_rotateAtoms.scp
aGetParamsAtom.scp	a_removeDummyAtoms.scp *

*The call is commented out as the delete command is done explicitly. Later this may be changed in favor of the call to have a uniform way to remove dummy atoms..

a_calcMolax.scp

Calculating best lines and planes can be very useful not only for analytical purposes at the end of a refinement, but it can be extremely helpful when preparing the model for refinement. It is possible to put atoms on best lines or planes in order to improve the starting geometry, but it is also possible to use the best line in order to shift atoms along this line, possibly equivalent with a chemical bond, thus permitting to achieve better agreement with the restraints used for subsequent refinement.

When discussing the steps needed to prepare a model for the refinement of a disordered structure there has been the suggestion if this step could not be executed in a better way using molecular mechanics. Very probably this would be a very valuable option, even though in practical work and by looking at what we can expect as a final result of such a refinement it would probably not show big differences. During the testing the option described earlier has been used to change the bond distances in

disordered aliphatic carbon chains to be 1.54 Angstrom for the starting model, and for that purpose the option worked well and gave the desired effect improving the agreement of the model with the restraints from the beginning of the refinement on. Extending the use of this option will reveal which cases can be treated successfully in this way.

Input to a_calcMolax.scp is possible in many different ways. The command that is executed in CRYSTALS is #GEOMETRY. A list of atoms is given from which the best plane or line is calculated. An optional list of atoms can be evaluated with respect to the plane or line definition. By executing the command on the command line the list of atoms to be evaluated is always given after the list of atoms defining the plane or line. In the case of the best line the GUI takes the selection of atoms as the list of atoms to be evaluated as long as the line is defined by a bond or by an atom plus the centroid of the selected atoms. Otherwise the evaluated atoms have to be stored beforehand in a scratch file. The same applies if atoms should be evaluated with respect to a plane definition.

This script is called by the following scripts:	*This script calls the following other scripts:*
aGetParamsSelection.scp	a_xMessage.scp
aGetParamsBond.scp	a_addCentroid.scp
aGetParamsAtom.scp	a_findCommonAtom.scp
	a_calcAnchorBond.scp
	a_calcNearest.scp
	a_bondSValMenu.scp
	a_refreshDefinition.scp
	a_refreshSelection.scp

a_transSelection.scp

This script enhances a function already present in the CRYSTALS interface. The ability to use the space group symmetry to put atoms in the nearest possible position is extended to work on a whole group of atoms. This can be practical if the model becomes complete only slowly and a group of positions is located in the Fourier map but results in being distant from the bulk of the structure. Moving the atoms one by one was always possible, but moving the whole group is faster and requires less user interaction.

Scripts to move residues, assemblies or groups have been written, but proved to be less useful as usually the positions to be moved are not entire residues, assemblies or parts. If for some reason the user wants to move a whole residue, assembly or group, the script a_selectAtoms.scp can be used to select that residue, assembly or group which then can be moved as a whole using the script described here.

This script is called by the following script: *This script calls no other scripts.*
aGetParamsBond.scp

8.5.7 Scripts that help to complete the user input

a_getAtom.scp

In some situations it is necessary that the user can specify some atom when automatic ways to determine all atoms needed for an operation may have failed. In these moments the script a_getAtom.scp presents a dialog that allows the user to specify the atom. This can be done by clicking on the atom or by typing its name. If typing the presence of the atom is checked before enabling the OK button.

This script is called by the following script: *This script calls the following other script:*
a_changeGrp.scp a_checkAtom.scp

a_addCentroid.scp

In the case of rotations or best line calculations where we need two atoms to build the reference axis this script completes the input when called from the atom context menu by adding the centroid of the selected atoms as the second set of coordinates that defines the rotation axis. The name of the dummy atom is given back to the calling script so that it is able to delete it automatically as soon as it will not be used anymore.

The script a_addCentroid.scp checks the input of a command that needs a rotation axis and determines if the centroid must be added in order to get a valid input. If it finds that the centroid is needed it calls a_createCentroid.scp to calculate the new coordinates and to add the dummy atom to list 5. This way of organizing makes it possible to call a_createCentroid.scp also directly from the interface.

This script is called by the following scripts: *This script calls the following other script:*
a_calcMolax.scp a_createCentroid.scp
a_createAssembly.scp
a_rotateAtomsMenu.scp

a_bondSValMenu.scp

This script can be called as a freestanding script from the bond context menu, or it can be called from other scripts. Depending on the input it can work in two modes. It basically tries to give the user hints about chemically sensible bond lengths. If one or both of the atoms near each other is a Q peak (i.e. a maximum of electron density that

has not yet been assigned a atom type), then the script works by looking up the range of the current distance giving hints about possible atoms types that could bond in this range of distances. If the atom types are already assigned then all bonds of these atom types are listed and the nearest match is chosen.

This script should help users with less experience to get a good feeling for what are sensible bond distances for different atom types. If called from other scripts this script may give back the distance chosen for further use. The file containing the definitions is a text file and can be easily adapted to local needs.

This script is called by the following scripts:	*This script calls the following other script:*
aGetParamsSelection.scp	ydisp.scp
aGetParamsBond.scp	
a_calcMolax.scp	

This script needs the text file a_bond.dat to be present in the script directory, as the bond distances are stored there. This file can be easily adapted and completed in order to meet local needs and preferences.

8.5.8 Scripts to prepare refinement directives and restraints

a_genL12singleAtoms.scp

When calling the script a_refineSelection.scp this script is called to generate the refinement directives according to the user input. The selection of atoms to refine is stored in a scratch file, and for each option the script reads the scratch file and adds a line for each atom.

This script is called by the following script:	*This script calls no other scripts.*
a_refineSelection.scp	

a_genL12Parts.scp

This script is setting up the list 12 that is used in the initial stages of the refinement of some assembly containing two groups. The limitation on two groups is maintained at the moment as the refinement of the site occupancy factor needs to be revised for more than two contributors. The work with the current version will show if this change will be of importance or not.

This script is called by the following script:	*This script calls no other scripts.*
a_refineAssembly.scp	

a_genL12Assembly.scp

This script generates the refinement directives for an assembly. The chosen options are passed as variables of type LOGICAL to this script, and according to whether they are true or false the list 12 input is generated.

This script is called by the following script:
a_refineAssMenu.scp

This script calls no other scripts.

a_l12partsSOF.scp

This script is used in the last stages of the refinement. The disordered assembly has been refined and needs now to be integrated in the refinement of the whole structure. There may be more than one assembly to be incorporated in the final refinement steps. The script reads the file Parts.dat and writes the instructions concerning the refinement of the site occupancy factors of all assemblies present in the structure to the file refsofass.12. The script xwrite5.scp then uses this bit together with others to build a consistent list 12.

Currently the EQUIVALENCE and WEIGHT -1 instructions are used that limit the sensible use of the script to assemblies with two groups. In order to cope with assemblies with more than two groups the script will be changed to use the SUMFIX instruction that does not have this restriction.

This script is called by the following script:
xwrite5.scp

This script calls no other scripts.

xwrite5.scp

This is the only script from the CRYSTALS distribution that needs to be changed in order to incorporate the new refinement strategies. The ability to generate instructions to refine site occupancy factors from disordered assemblies needed to be added to this script. This is done my calling the script a_l12partsSOF.scp that creates these instructions. These instructions are then incorporated in the new list 12. Like this the user can return to the known context of refinement once the disordered assemblies are resolved and prepared for the final refinement.

This script is not called by other scripts written in the scope of this project.

This script calls the following other scripts written in the scope of this project:
a_getResiduesAndParts.scp
a_l12partsSOF.scp

8.5.9 Scripts for the refinement of a disordered assembly or a group of atoms

a_refineSelection.scp

This script is a very simple help in all the cases where difference maps are not easy to interpret as residual electron density deriving from the anisotropy of some atoms are dominating the difference map. This is an easy way to try to flatten the map in these areas in order to get new difference maps that now show more features of the structure.

This script is called by the following script:	*This script calls the following other scripts:*
aGetParamsSelection.scp	a_restorePoint.scp a_genL12singleAtoms.scp a_refine5.scp a_reset1216.scp a_refreshselection.scp a_xMessage.scp

a_refineAssembly.scp

This is a first try to make easier the refinement of an assembly created by one of the scripts described earlier. It showed not to be flexible enough and has been substituted by a new script which gives access to all scripts dealing with the refinement of the disordered assembly. It will remain in place until the new script has proved to work well.

This script is called by the following script:	*This script calls the following other scripts:*
aGetParamsAtom.scp	a_restorePoint.scp a_getAssemblyAndParts.scp a_gelL12Parts.scp a_refine5.scp a_reset1216.scp * a_refreshselection.scp a_xMessage.scp

*The call is commented out. Please read about the motivation to do this in the description of the script a_reset1216.scp.

a_refineAssMenu.scp

This is the enhanced refinement script for assemblies which will replace a_refineAssembly.scp at a later stage. It lets the user do the following things.

Create and modify restraints for the assembly in question. The corresponding scripts can also be called from outside the refinement script. The functions used to create restraints will be described elsewhere. They are aimed to give the user the choice of using the restraints of her or his liking.

Create and modify refinement directives. The script uses the PART numbers in list 5 to address the groups of the assembly. This is practical and robust, in particular as in earlier stages of refinement the models still may change as atoms are added to or removed from the assembly. This is the reason why at this stage the user should avoid to add the hydrogen atoms as with the PART numbers the implicit exclusion of hydrogen atoms from refinement is not active. Adding the hydrogen atoms in the last step is the right strategy and lets the user optimize the model without being distracted by lots of partially occupied hydrogen atoms that obstruct the view of the structure by their usually high U-values.

The user can control the number of refinement steps to be executed in one go. Criteria can be defined to detect if the refinement is converging. Control can be passed to a batch file that updates the settings and changes them during the following refinement cycles automatically. This option is designed in particular for users with limited experience. Looking at the way the batch files operate the script can give hints of what to try another time in a similar situation.

In the last section titled “Model and controls” the user gets the possibility to go back in a controlled way. He may want to return to the beginning, or simply go back one step. (It is an option to give access to all steps of one run of this script.) Sometimes it is sufficient to restart with the isotropic model as described elsewhere. On completion the user can leave this script by clicking OK.

This script is very complex at the moment as it gives access to both lists 12 and 16. The creation of restraints needs to be separated as the window controlling everything is too large to be practical. Moreover it has been observed that it is rare to create new restraints once the generated ones have proved to work well. The changes are made mainly in the applied standard uncertainties, and for this purpose a separate script named a_replaceSU.scp has been written that will be accessible through the dialog window of a_refineAssMenu.scp.

This script is called by the following script:	*This script calls the following other scripts:*
aGetParamsAtom.scp *	a_restorePoint.scp a_getAssemblyAndParts.scp a_getFirstPart.scp a_xMessage.scp a_getOtherPart.scp a_genL12Assembly.scp edList12.scp ** a_refineN.scp a_getBondsAngles.scp edlist16.scp a_replaceSU.scp ** xshiftl.scp a_remShiftl.scp

*Should and will be aGetParamsAtom.scp, now, as still under development, called from the command line.
**The call is commented out, but will be made available in the future.

a_refine5.scp

Once lists 12 and 16 have been set up the model needs to be refined. This script simply does five refinement cycles, one after each other. So it also continues if singularities are detected. This is of course not optimal and will need to be enhanced.

This script is called by the following scripts:	*This script calls no other scripts.*
a_refineAssembly.scp a_refineSelection.scp	

a_refineN.scp

This is almost the same script as a_refine5.scp, only that the value of how many cycles should be performed is not hard coded and can be passed to the script using the integer variable ICYCLE. Apart from that the same problems need to be addressed as with a_refine5.scp.

This script is called by the following script:	*This script calls no other scripts.*
a_refineAssMenu.scp	

8.5.10 Scripts working in the background

The scripts presented here fulfill important tasks without ever producing output the user will notice. Nevertheless their contribution is essential for the successful interaction of all scripts involved in this project.

a_refreshSelection.scp and a_refreshDefinition.scp

These two scripts are related to the script a_resetSelection.scp but do exactly the contrary. They restore the selection stored in a scratch file to the model window. This can be the selection to modify (AtomsToWorkOn.dat) or the definition to use for a modification (MolaxDefinition.dat).

These two scripts are called by the following scripts:

a_calcMolax.scp
a_createAssembly.scp *
a_refineAssembly.scp *
a_refineSelection.scp *
a_rotateAtoms.scp *

These two scripts call no other scripts.

*Called only by a_refreshSelection.scp.

a_removeDummyAtoms.scp

For various operations the use of dummy atoms is required in order to obtain the desired results. Rotations can use a dummy atom, and if we wish to use a mirror plane for the model creation then this plane resides most probably between the atomic sites. The dummy atoms can be created for instance using the script a_createCentroid.scp which has been described earlier. The script a_addCentoid.scp produces the dummy atom it needs autonomously and deletes it also afterward. But this is not always possible. So the script is provided to have an easy possibility to remove all these dummy atoms with one click leaving the other Q peaks in place.

This script is called by the following scripts:

Direct call from popup-noatom.srt
a_createAssembly.scp *
a_rotateAtomsMenu.scp *

This script calls no other scripts.

*Planned calls.

a_setUiso.scp

During the creation of the assembly using the script a_createAssembly.scp the atoms need, in case they should already have been refined anisotropically, to be reset to have isotropic U-values before the refinement of the disordered assembly can be started. This script first calculates the mean value of all isotropic U-values in question and applies them to the group before being duplicated and modified.

This script is called by the following script:
a_createAssembly.scp

This script calls no other scripts.

a_getAssemblyAndParts.scp

With the help of this script the user can get hold of the assembly and group numbers involved in the current operation. The user gives one atom label chosen by a mouse click or in the context menus to the script which then gives the part numbers involved in the current operation back to the calling script in the scratch file PartsToWorkOn.dat.

This script is called by the following scripts:
a_refineAssembly.scp
a_refineAssMenu.scp *

This script calls the following other script:
a_getResiduesAndParts.scp

*Currently other scripts are used in this script to determine the part numbers and this call is therefore commented out. A testing phase will show which method proves to be more reliable.

a_getResiduesAndParts.scp

At many points we need to know the residues and parts that are defined for a given structure. This may be necessary when we want to create a new assembly and need to know the next free assembly number, but this is also important when generating new refinement directives as we need to give the instructions for each assembly present in the structure.

The script collects all residue numbers and part numbers used in the structure, sorts out all duplicates, sorts them in ascending order and writes the result to two scratch files called Residues.dat and Parts.dat. Other scripts can then read the scratch file and extract the information they need.

This script is called by the following scripts:
a_createAssembly.scp
a_splitJoin.scp
xwrite5.scp
a_getAssemblyAndParts.scp

This script calls the following other script:
a_restorePoint.scp

a_getFirstPart.scp and a_getOtherPart.scp

If the assembly number is already known there is some alternative way to get hold of the part numbers that is less complicated than the algorithm used in a_getAssemblyAndParts.scp where the whole list 5 is scanned for the information needed. Nevertheless this easier way may be abandoned in the future in order to have to maintain only one script doing this work.

The advantage of this pair of scripts is that they check if there are more than two groups present in the assembly very easily. Nevertheless this check can be performed in some other way, and extending the functionality to support more than two groups per assembly it will not be needed anymore.

These two scripts are called by the following scripts:
a_refineAssMenu.scp
a_getBondsAngles.scp

These two scripts call no other scripts.

a_checkAssembly.scp

This script checks if the assembly we work on is consistent, thus if the pairs of atoms can be determined. In the best case the groups can be checked by their atom names and the physical order in list 5. If this fails the pairs are searched for using their names and the offset of the serial number. If this fails too, the atoms are written pair wise to the file pairs.dat, which can then be used by other scripts.

This script is called by the following script:
a_getBondsAngles.scp

This script calls the following other script:
a_checkAtom.scp

a_getBondsAngles.scp

When generating restraints for the disordered assembly the connectivity needs to be determined. This script does this by analyzing the output of different distance and angle calculations executed by CRYSTALS stored in text files. The CRYSTALS command is:

```
#DISTANCES
PIVOT PART(XXXYYY)
OUTPUT MONITOR = OFF PUNCH = RESTRAIN
SELECT SYMMETRY = NONE ALLDISTANCES = YES/NO RANGE = L41
END
```

The directive ALLDISTANCES is used in dependence whether we are only looking for distances within the part or also want to examine bonds and angles to atoms outside the part. For bonds and angles within the part it is easier not to have to filter duplicates and ALLDISTANCES = NO is used. If one atom of the bond or angle is located outside the part, ALLDISTANCES = YES ensures that we find all contacts independent of the position of the atoms in the atom list.

A first preliminary step is to check is to see if the assembly is consistent, and if the offset in the serial numbers can be determined. This check is done by the script a_checkAssembly.scp described later.

The second step is to extract the relevant information from the listings created by the bond and angle calculations. Then there is a division made between bonds that link the disorder assembly to the ordered bulk structure and the bonds inside the two groups of the assembly. They are referred to as external and internal bonds and angles. An external bond is characterized by the fact that one of the two bonded atoms is not part of the assembly, and the external angle is present if any of the three atoms that form the bond is lying outside the assembly.

Now the information about the bonds and angles and about the pairs of atoms in the assembly are merged together and templates for the different commands are generated. Such a template could look as follows:

```
CONTINUE C(56) TO O(1) C(66) TO O(11)
```

Now these templates can be used to generate different restraints. The line preceding this restraint could be one of the following:

```
DISTANCES  0.0 , 0.0010 = MEAN
```

or

```
VIBRATIONS 0.0, 0.00200  =
```

No matter which one is used, the combination will be a valid restraint. The values to be achieved are provided in the form filled by the user.

The information in the file pairs.dat can also be used to generate SAME, DELU and SIMU restraints. They can be combined with the restraints for external bonds and

angles, and like that there is a large choice of combinations that can be used in order to automatically create a consistent set of restraints.

In order to be able to accumulate information from different sources in list 16 the already existing script zAddLi16.scp is used to update list 16.

This script is called by the following script:	*This script calls the following other scripts:*
a_refineAssMenu.scp	a_getFirstPart.scp a_xMessage.scp a_getOtherPart.scp a_listPart.scp a_checkAssembly.scp a_calcDistPivot.scp a_writePairs.scp a_checkAtom.scp a_getPairs.scp zAddLi16.scp edlist16.scp *

A direct call to generate restraints independently from the script a_refineAssMenu.scp would be most welcome.
*Commented out, will be available as an option.

a_calcDistPivot.scp

This script produces a listing of all distances and angles around a given pivot atom. This listing is needed by the script a_splitJoin.scp to explore the neighborhood of the atom to split, and the atoms found in this listing are later analyzed to see if they are part of an assembly.

Using the same script with a part instead of a single atom produces a listing of all bonds and angles that are necessary for the generation of restraints in a script that is still to be written.

This script is called by the following scripts:	*This script calls no other scripts.*
a_splitJoin.scp a_getBondsAngles.scp	

a_isfreeR.scp, a_isfreeA.scp and a_isfreeP.scp

These three scripts work in a similar way to the script provided already by the CRYSTALS distribution that is called isfree.scp. While isfree.scp gives back the logical value TRUE or FALSE depending on whether an atom label is free to be used or already in use causing a clash, the three scripts do the same for residue, assembly or part numbers. The check for part numbers instead of group numbers is done as there may be different groups with the same group number in different assemblies.

These three scripts are called by the following script:
a_changeRAG.scp

These three scripts call no other scripts.

a_checkAtom.scp

In some moments it is necessary to know if a given atom name can be found in the atom list. This script does this, and if it finds the atom it selects it and deselects it a second later to show the user that the input was correct.

This script is called by the following scripts:
a_changeGrp.scp
a_checkAssembly.scp
a_getAtom.scp
a_getBondsAngles.scp

This script calls no other scripts.

a_restorePoint.scp

As refinement of disordered assemblies does not always work on the first try, a concept has been devised to give the user a safe way to step back. The idea is to save consistent sets of lists 5 (coordinates), 12 (refinement directives) and 16 (restraints) so that it is easy to go back to a defined point before some other action was taken. This is safer than stepping back in a whole series of coordinate sets (list 5) in order to find the one that represents the state before the refinement deteriorated the model. A call to this script at the start of any action gives the possibility to have a safe “undo” option for everything changed during the execution of the script.

The development of this script is not finished. Restore points can be saved easily now, but there needs to be a mechanism developed to give the possibility to step back more than one step. A proper way of storing the whole history needs to be found. As soon as this will be available this script will be called very frequently from different places.

This script is called by the following scripts: *This script calls no other scripts.*
a_refineAssembly.scp
a_refineAssMenu.scp
a_refineSelection.scp
a_regularise.scp
a_xInfoAboutStructure.scp
a_getResiduesAndParts.scp

a_resetToRP.scp

This script is the complementary script to a_restorePoint.scp. At the moment it can be used to step back to the last restore point saved. The serial numbers of the lists to restore are passed as variables defined in both scripts, the usual way of passing variables from one script to the other.

More versatile options to use these two scripts are under development. The improvements will encourage wider use. This option is important for the general use, but in the current implementation it is too primitive to be used widely.

This script is called by the following script: *This script calls no other scripts.*
a_regularise.scp

a_reset1216.scp

It is sometimes useful to reset lists 12 and 16, while the coordinates should remain the refined ones. This script does this by leaving the list 5 with the refined coordinates unchanged while resetting lists 12 and 16 to the version with a serial number one lower than the actual one.

Most of the time to the best choice would be to let the user decide whether she or he wishes to reset the refinement directives and restraints or not. In tests it has been seen that in the case of refinement of single atoms (a_refineSelection.scp) the reset of lists 12 and 16 is practical, while in all other situations it seems to be less useful. Further developments will take into account these observations.

This script is called by the following scripts: *This script calls no other scripts.*
a_refineAssembly.scp *
a_refineSelection.scp

*Call currently commented out.

a_sortRing.scp

The idea behind this script is to facilitate the user input for the options CP-RING, HEXAGON and PHENYL of the REGULARISE command. Clicking in some order on the atoms on the screen suggests to the user that the atoms will be stored in this order. As REGULARISE expects the input for all options in some ordered way this may produce unexpected output.

There are two possibilities to overcome this problem. One is to keep track of the order in which the user clicked on the atoms assuming she or he knows the right order. The second possibility is to order the input automatically, and this approach has been used in this script. The distance from each atom to all other atoms in the ring is calculated in order to find 1-2, 1-3 and, in the case of six membered rings, 1-4 relationships between the atoms. Doing this by taking one atom after the other as the pivot atom and exchanging atoms in the case of mismatch between the expected and the calculated distances gives in the end the ordered input for these calculations. In the case of the options CP-RING and HEXAGON the mean distance is calculated and proposed to the user alongside with a literature value.

It is planned to write a similar script for the other options of REGULARISE, in particular for TETRAHEDRON and OCTAHEDRON. As they are not crucial in the context of the refinement of disorders the realization of these additional features has been postponed.

This script is called by the following script:
a_regularise.scp

This script calls no other scripts.

a_rotateAtoms.scp

a_rotateAtoms makes use of the ROTATE function in #EDIT. For the moment only the mode using a vector defined by two atoms as the rotation axis is used. For the work with disorders this seems adequate, and if later use will show the necessity other options can be implemented.

This script is called by the following script:
a_rotateAtomsMenu.scp

This script calls the following other scripts:
a_refreshSelection.scp
a_xMessage.scp

a_orthoAtoms.scp

In some situations it is practical to have the coordinates of an atom or a group of atoms at hand in orthogonal space. This script takes the selected group of atoms, extracts the corresponding matrix to transform the coordinates to orthogonal space from list 1 holding the unit cell parameters and other information about the unit cell, locates all

atoms in list 5 and writes the transformed coordinates to the file AtomsOrtho.dat which is then available to other scripts.

This script is called by the following script: *This script calls no other scripts.*
a_regularise.scp

a_findCommonAtom.scp

Best line and plane calculations have two sets of atoms as input. There are the atoms defining the line or plane, and there are the atoms evaluated in respect to that line or plane. This script checks for the case of two atoms defining a line if one atom is also present in the group of evaluated atoms.

This is done by calculating the distance between the two atoms defining the line and all atoms to be evaluated. If the distance is less than 0.25 Angstrom it is regarded as a match and the atom from the group to be evaluated is taken to calculate the line. This feature can be useful in disorder assemblies where the proximity of positions of pairs of atoms makes it difficult to click on the correct bond.

This script is called by the following script: *This script calls no other scripts.*
a_calcMolax.scp

a_calcAnchorBond.scp

If there is a common atom between the line definition and the group of evaluated atoms then this line definition acts like an anchor for the evaluated atoms. The atom opposite to the common atom is the anchor in respect to which the evaluated atoms can be shifted along the line. This feature of a_calcMolax.scp is useful in very early stages of refining disordered chains of atoms as it gives the possibility to start with sensible bond distances. The script described here calculates the value that is displayed by a_calcMolax.scp to the user.

The script is called later than a_findCommonAtom.scp, thus if there is a common atom it will be labeled identically and can be identified by matching the labels. When both the anchor atom and the common atom are found their distance is calculated. In case of failure the script gives back a value of zero.

This script is called by the following script: *This script calls no other scripts.*
a_calcMolax.scp

a_calcNearest.scp

The script a_calcAnchorBond.scp is only called if it is a best line that is calculated and if this line is defined by two atoms. This is typically the case when calling a_calcMolax.scp from the bond context menu.

In all other cases of calling a_calcMolax.scp the nearest distance between some atom from the group defining the line of plane and some other atom from the group of evaluated atoms is calculated and displayed.

This script is called by the following script: *This script calls no other scripts.*
a_calcMolax.scp

a_remShiftl.scp

Some cases of disorder refinement require the use of shift limiting restraints to be successful. The script xshiftl.scp from the CRYSTALS distribution helps the user to set them up. Towards the end of the refinement it is desirable to remove those restraints as their presence may slow down the last cycles of refinement considerably. In the GUI interface it seemed to be simplest to write a dummy input for this part of list 16 and to use again zAddLi16.scp to add the input containing only a comment to list 16 in order to remove the shift limiting restraints from list 16.

This script is called by the following script: *This script calls the following other script:*
a_refineAssMenu.scp zAddLi16.scp

a_listPart.scp

When determining the connectivity of the disordered assembly it is necessary to get hold of the parts that form the assembly under investigation. This script gets the part number from the calling script and writes all atoms that have this part number to the punch file that needs to have been opened by the calling script. As a side effect the number of atoms being members of that part is counted.

This script is called by the following script: *This script calls no other scripts.*
a_getBondsAngles.scp

a_writePairs.scp

During the process of checking the consistency of the disorder assembly before generating restraints the matching pairs of atoms of both groups are written to the file

pairs.dat by the script a_checkAssembly.scp. This information will later be used to derive the restraints for the second group from those for the first group by substitution of all atom labels from the first group by those of the second group.

The merging process uses two files to read from. In order to be used with older versions of CRYSTALS where only one file is available for reading a workaround has been found where the information from the two files is read from one file, and the script a_writePairs.scp prepares the data file accordingly. This script will be obsolete as soon as the current distribution of CRYSTALS will support two files to read from in the scripting environment.

This script is called by the following script: *This script calls no other scripts.*
a_getBondsAngles.scp

a_getPairs.scp

The restraints are created in pairs. This means to have at hand the label of one atom in the first group together with the label of the equivalent atom in the second group. This script gets an atom label as input and scans the file pairs.dat in order to find its equivalent which is then given back to the calling script.

This script is called by the following script: *This script calls no other scripts.*
a_getBondsAngles.scp

a_tidyup.scp

This script is called every time at the end of the execution of other scripts. The easiest way of doing this is to place it at the end of the interface scripts. No matter of how the individual scripts have terminated the control is in any case given back to these scripts making it easy to get rid of scratch files not needed anymore. The only moment when this is not the case is if the user had the possibility and the knowledge to interrupt the executing by entering DIRECT at the prompt.

All scratch files with the exception of the file containing a selection of atoms for later use are deleted. This ensures that first the directory the user is working in is not flooded by lots of files with unknown content and function. And the presence of an old file could also lead to inconsistent input to the scripts actually running, even though the design of the script is made in a way that the running script only uses files it has written itself. The exception is the file atomsToWorkOn.dat that might contain the selection of atoms used for further evaluation.

It is planned to use this script at the end of all calls to the three interfacing scripts to eliminate the individual deletion calls from single scripts that are difficult to maintain.

This script is called by the following script: *This script calls no other scripts.*
aGetParamsAtom.scp

a_xMessage.scp

This is a simple script to give the user some information that has to be acknowledged before continuing the work. This is useful before stopping execution of a script when an error condition is detected.

This script is called in many places, too many to be listed one by one.

a_doNothing.scp

This script is a pure placeholder. It can be used to keep the place occupied until the true script will be finished, or it can stay at a place where a call should never happen, but where we need the line of code for syntactical reasons.

Currently this script is not in use. On completion of the project this script will become obsolete.

8.6 Limits and ways to improve reliability

Again I would like to state very clearly that these scripts are intended to be helpers for model building and refinement of disordered structures. If the reason of the problem is other than disorder, they will probably not give the expected results. There are presumably lots of structures in the databases that have been refined as being disordered while the true reason for the problem was not disorder, but the equipment used did not give the possibility to observe the additional data needed to solve the problem in the way it should have been done.

For all these cases the development of new equipment is likely to improve the situation. More sensitive detectors and brighter, more focused X-ray sources will improve the situation from two sides so that a few “false” disorders can be excluded. But already now there are possibilities to be more cautious.

In the Crystallography Laboratory at the University of Basel the integration of the raw frames is usually done at least twice. The first integration is done when about two thirds of the data collection is done. With careful examination of the Laue class the attempt is made to prevent to lose time working on data sets with wrong unit cells and cell settings. The structure is solved, and if this is successful it is visible in a short time if there are problems like disorder around. This then gives the possibility to think before the crystal is taken from the diffractometer if a more careful look is needed, for instance

to see if there could be satellites or weak spots indicating a bigger unit cell, or if there are indications suggesting twinning. By the end of the data collection important information is already present, and this allows taking the crystal from the diffractometer without the fear of having eventually lost important information.

Nevertheless a considerable part of the structures requiring more time and effort will remain the disordered ones. New equipment will also improve the situation for disordered structures as refinements will not so quickly run out of reflections and will be more stable. The new scripts will help the crystallographer to treat a few of them with bigger ease and hopefully comparable or better results than with the traditional methods. Finding the best model for a disorder is rewarding as a nicely resolved disorder reveals details of the structure that otherwise would remain hidden.

9 Literature and structures

9.1 Literature

The literature cited in this manuscript is ordered in the order of appearance in the manuscript

1 Betteridge, P. W., Carruthers, J. R., Cooper, R. I., Prout, K., Watkin, D. J. (2003). J. Appl. Cryst. **36**, 1487

2 F. H. Allen, *Acta Cryst.*, B**58**, 380-388, 2002

3 Palatinus L., Chapuis G. (2007): Superflip - a computer program for the solution of crystal structures by charge flipping in arbitrary dimensions. *J. Appl. Cryst.* **40**, 786-790

4 Burla M.C., Camalli, M., Carrozzini, B., Cascarano, G., Giacovazzo, C., Polidori, G. Spagna, R. (2000). Acta Cryst., **A56**, 451-457

5 Sheldrick, G.M. (2008). *Acta Cryst*. **A64**, 112-122

6 P.T. Beurskens, G. Beurskens, R. de Gelder, S. Garcia-Granda, R.O. Gould, and J.M.M. Smits (2008). The DIRDIF2008 program system, Crystallography Laboratory, University of Nijmegen, The Netherlands

7 M. Zehnder, M. Neuburger, P. von Matt, A. Pfaltz (1995). Acta Cryst. **C51**, 1109-1112

8 Berlepsch, P. (1996). Schweiz. Mineral. Petrogr. Mitt., 76, 147–157

9 Alain Schlatter, Mrinal K. Kundu Dr., Wolf-D. Woggon Prof. (2004). Angewandte Chemie, **116**, Issue 48, 6899–6902

10 Wolff, P. M. de (1974). Acta Cryst. **A30**, 777–785

11 Wagner, T., Schönleber, A. (2009). Acta Cryst. **B65**, 249–268

12 R. I. Cooper, R. O. Gould, S. Parsons and D. J. Watkin (2002). J. Appl. Cryst. **35**, 168-174

13 H.-G. Liu, Y.-C. Qui, J.-Z. Wu, *Acta Cryst.* (2007). E**63**, o4876

14 Petricek,V., Dusek,M. & Palatinus,L.(2006). Jana2006. The crystallographic computing system. Institute of Physics, Praha, Czech Republic.

15 Vikram S. Bajaj, Patrick C.A. van der Wel, Robert G. Griffin (2009). J. Am. Chem. Soc., **131**, 118-128

16 Ludger Schröder, David John Watkin, Alain Cousson, Richard Ian Cooper, Werner Paulus (2004). *J. Appl. Cryst.*, **37**, 545-550

17 Johnson, C. K. and Levy, H. A. (1974). *International Tables for X-ray Crystallography*, Vol IV, Kynoch Press, Birmingham, England, p. 311

18 P. R. Mallinson, T. Koritsanszky, E. Elkaim, N. Li and P. Coppens (1988). Acta Cryst., **A44**, 336-343

19 V. Schomaker and K. N. Trueblood (1968). Acta Crystallogr. **B24**, 63-76

20 Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.*, **91**, 201-228

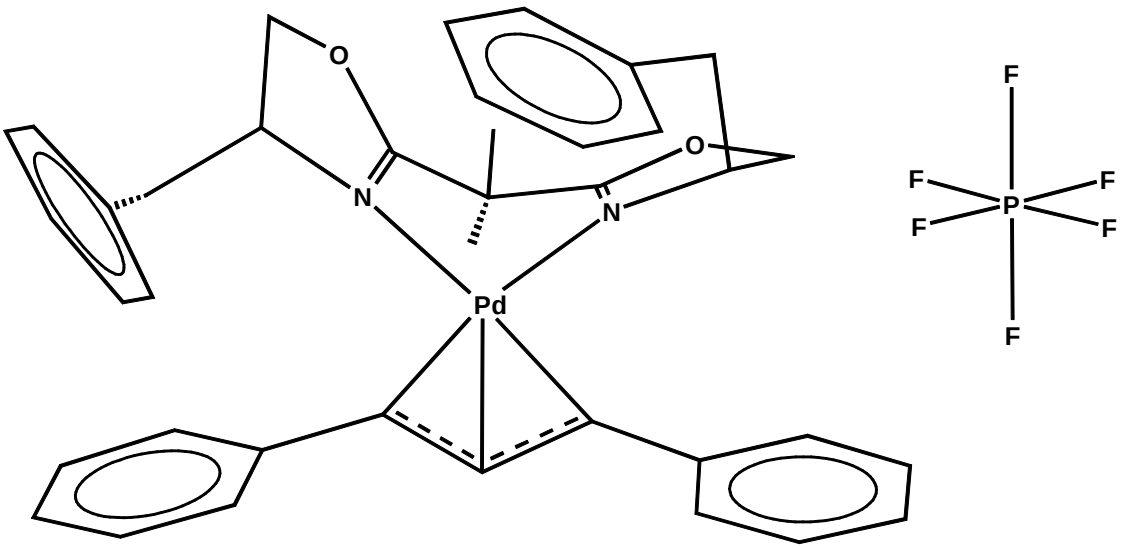
21 A.L.Spek (2002). PLATON, A Multipurpose Crystallographic Tool, Utrecht University, Utrecht, The Netherlands

9.2 Structures

1 pm362 (dynamic disorder)

Table 1. Crystal data for **pm362**

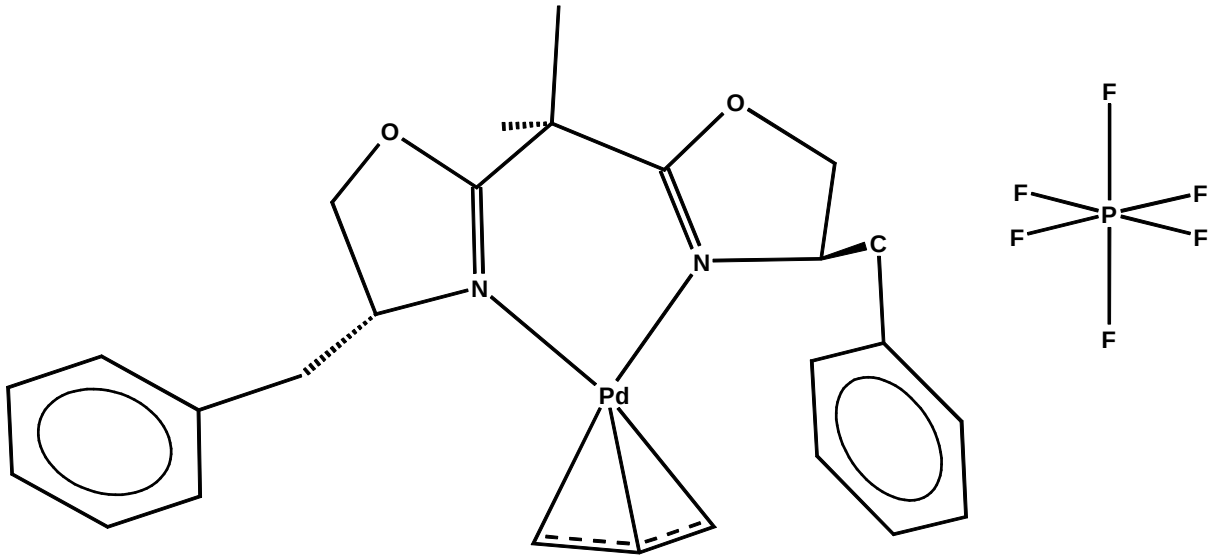
formula	$C_{38}H_{39}F_6N_2O_2P_1Pd_1$
formula weight	807.10
Z, calculated density	4, 1.508 Mg · m^{-3}
F(000)	1648
description and size of crystal	colourless block, 0.21 · 0.24 · 0.35 mm^3
absorption coefficient	0.635 mm^{-1}
min/max transmission	0.88 / 0.88
temperature	250K
radiation(wavelength)	Mo K_α (λ = 0.71069 Å)
Crystal system, space group	orthorhombic, P 2_1 2_1 2_1
a	12.609(3) Å
b	15.226(19) Å
c	18.521(7) Å
α	90°
β	90°
γ	90°
V	3555.6(4.8) $Å^3$
min/max Θ	2.097° / 28.002°
number of collected reflections	4746
number of independent refections	4746 (merging r = 0.000)
number of observed reflections	4251 (I>2.0σ(I))
number of refined parameters	452
r	0.0359
rW	0.0488
goodness of fit	0.9193



2 pm335 (static disorder)

Table 2. Crystal data for **pm335**

formula	$C_{26}H_{31}F_6N_2O_2P_1Pd_1$
formula weight	654.91
Z, calculated density	1, 1.620 Mg · m^{-3}
F(000)	332.000
description and size of crystal	yellow prism, 0.20 · 0.20 · 0.40 mm^3
absorption coefficient	0.819 mm^{-1}
min/max transmission	0.8488 / 0.8488
temperature	250K
radiation(wavelength)	Mo K_α (λ = 0.71069 Å)
Crystal system, space group	triclinic, P 1
a	9.171(4) Å
b	9.387(7) Å
c	9.558(3) Å
α	112.99(6)°
β	113.55(2)°
γ	93.40(5)°
V	671.3(8) $Å^3$
min/max Θ	2.437° / 27.983°
number of collected reflections	3424
number of independent refections	3418 (merging r = 0.000)
number of observed reflections	3424 (I>2.0σ(I))
number of refined parameters	355
r	0.0296
rW	0.0250
goodness of fit	1.1270



3 jentschite (substitutional disorder)

Table 1. Crystal data for **global**

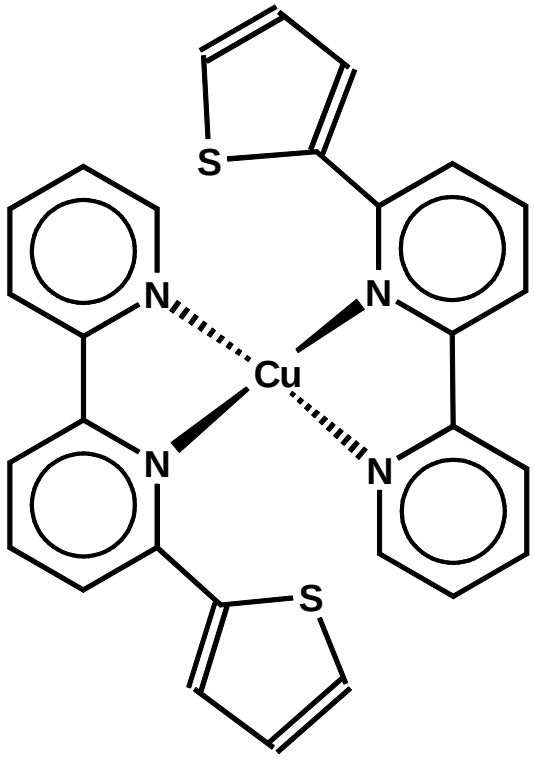
formula	$TlPbSb_{.55}As_{2.45}S_6$
calculated density	5.237 Mg · m^{-3}
Crystal system, space group	monoclinic, P $2_1/n$
a	8.0958 Å
b	23.917 Å
c	5.8876 Å
α	90°
β	108.063°
γ	90°
V	1083.816 $Å^3$

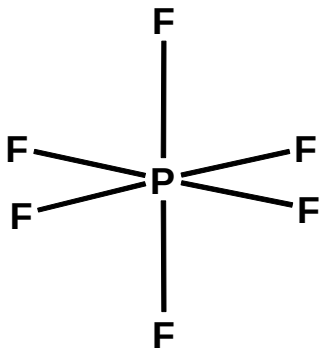
(A more complete data set is not available from databases of the internet.)

4 pka60 (thiophene ring with two orientations)

Table 4. Crystal data for **pka60**

formula	$C_{28}H_{20}Cu_1F_6N_4P_1S_2$
formula weight	685.13
Z, calculated density	2, 1.624 Mg · m^{-3}
F(000)	692
description and size of crystal	red block, 0.08 · 0.19 · 0.31 mm^3
absorption coefficient	1.053 mm^{-1}
min/max transmission	0.82 / 0.92
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	triclinic, P -1
a	7.1898(2) Å
b	14.0902(3) Å
c	14.6598(4) Å
α	103.270(2)°
β	99.206(2)°
γ	98.373(2)°
V	1400.92(7) $Å^3$
min/max Θ	1.808° / 29.574°
number of collected reflections	22905
number of independent refections	7740 (merging r = 0.046)
number of observed reflections	4821 (I>2.0σ(I))
number of refined parameters	404
r	0.0434
rW	0.0696
goodness of fit	1.0610

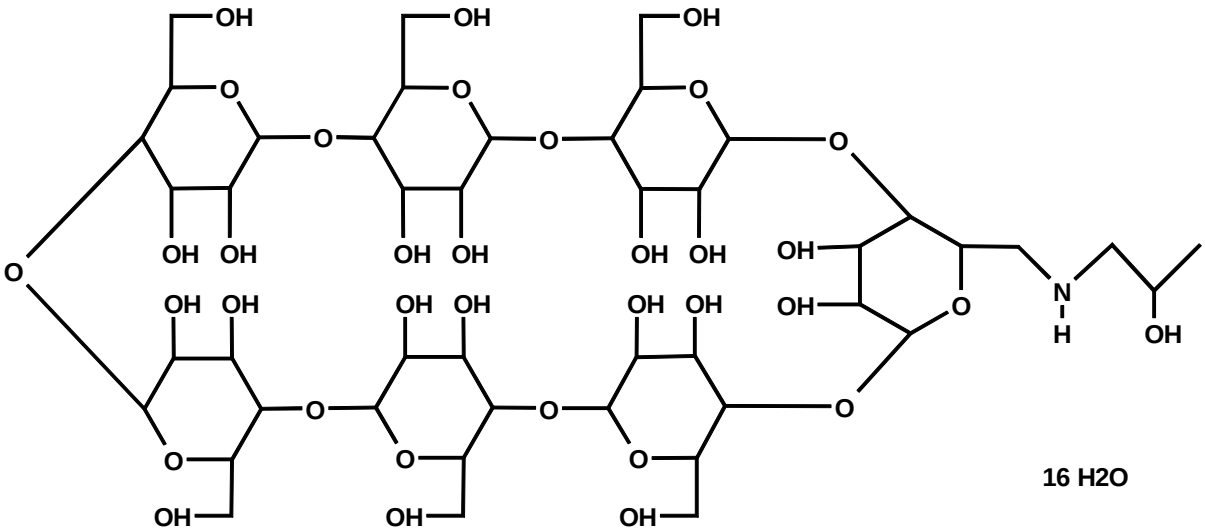




5 cyclodextrine (solvent disorder)

Table 5. Crystal data for **cyclodextrine**

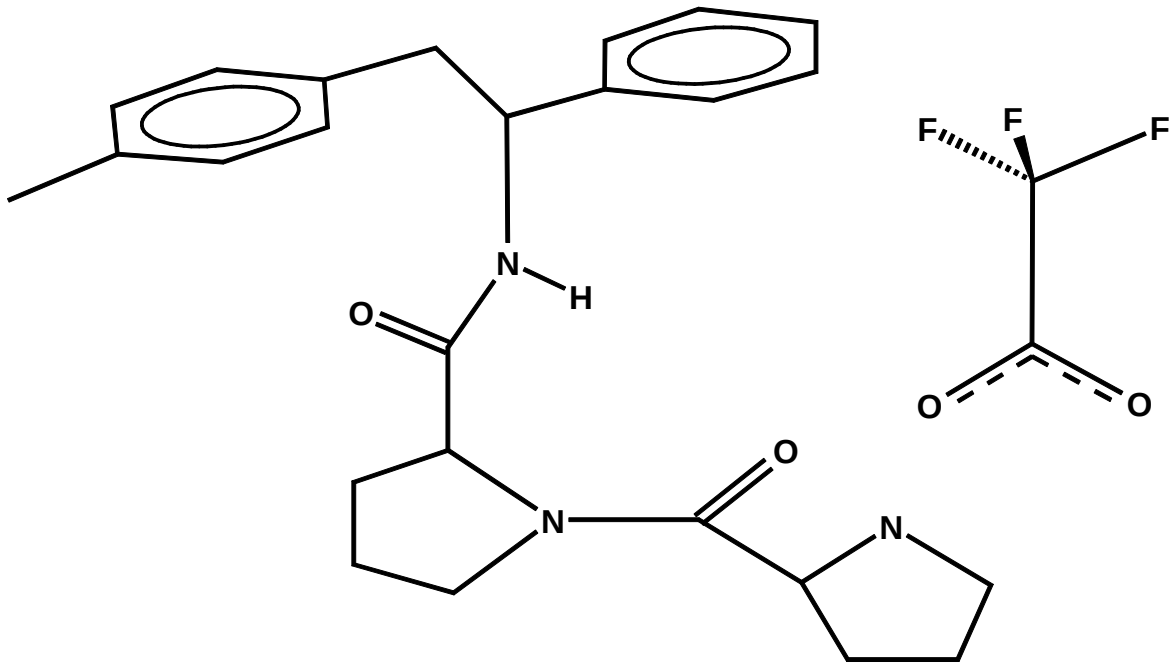
formula	$C_{45}H_{109}N_1O_{51}$
formula weight	1480.34
Z, calculated density	4, 1.468 Mg · m^{-3}
F(000)	3178.471
description and size of crystal	colorless needle, 0.10 · 0.10 · 0.32 mm^3
absorption coefficient	0.135 mm^{-1}
min/max transmission	0.99 / 0.99
temperature	173K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	orthorhombic , P 2_1 2_1 2_1
a	12.8092(4) Å
b	19.6407(5) Å
c	26.6247(5) Å
α	90°
β	90°
γ	90°
V	6698.3 $Å^3$
min/max Θ	4.10° / 28.01°
number of collected reflections	85399
number of independent refections	16082 (merging r = 0.07)
number of observed reflections	9379 (>3.00σ(I))
number of refined parameters	1090
r	0.0339
rW	0.0283
goodness of fit	1.0501



6 rk237 (disorder in the center of the molecule)

Table 6. Crystal data for **rk237**

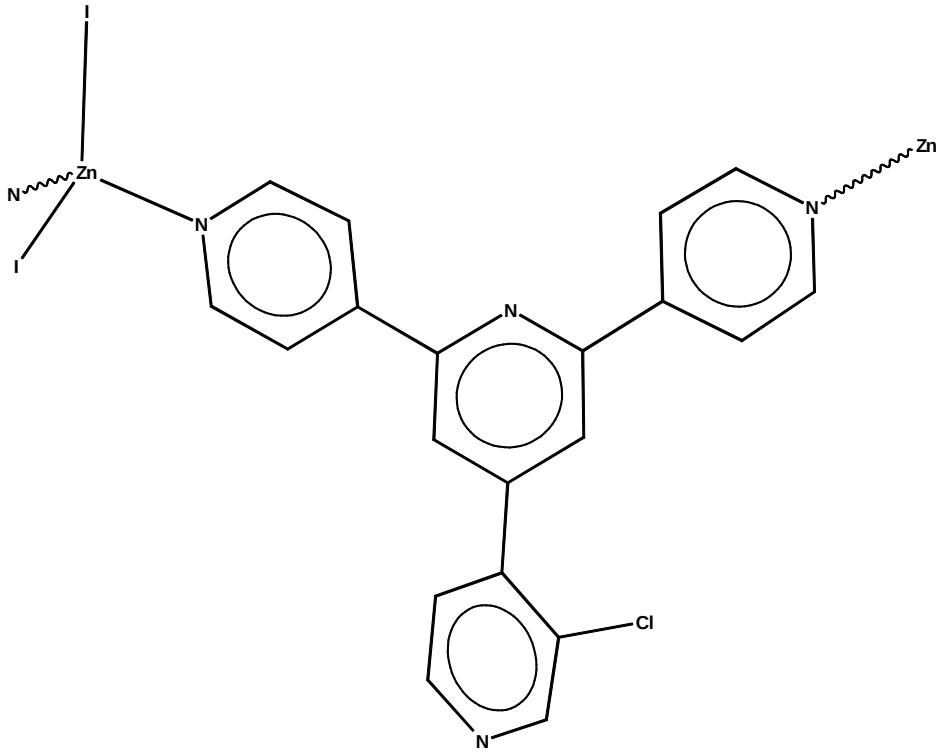
formula	$C_{27}H_{32}F_3N_3O_4$
formula weight	519.56
Z, calculated density	4, 1.314 Mg · m^{-3}
F(000)	1096
description and size of crystal	colourless prism, 0.07 · 0.13 · 0.21 mm^3
absorption coefficient	0.103 mm^{-1}
min/max transmission	0.99 / 0.99
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	orthorhombic, P 2_1 2_1 2_1
a	11.8295(7) Å
b	12.6083(7) Å
c	17.6036(9) Å
α	90°
β	90°
γ	90°
V	2625.6(3) $Å^3$
min/max Θ	1.987° / 33.205°
number of collected reflections	108716
number of independent refections	5540 (merging r = 0.039)
number of observed reflections	5270 (I>2.0σ(I))
number of refined parameters	380
r	0.0416
rW	0.0576
goodness of fit	1.0849



7 gq213 (whole ligand is disordered)

Table 7. Crystal data for **gq213**

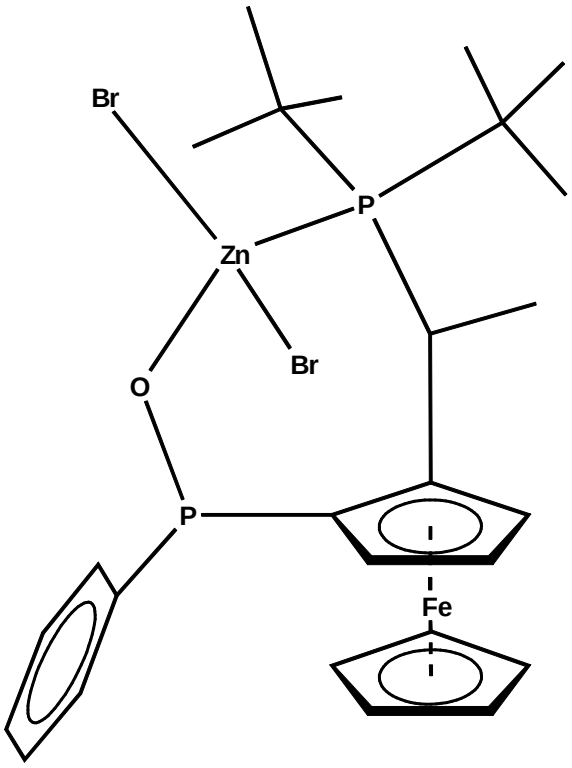
formula	$C_{20}H_{13}Cl_1I_2N_4Zn_1$
formula weight	663.99
Z, calculated density	4, 2.064 Mg · m^{-3}
F(000)	1256.000
description and size of crystal	colourless block, 0.06 · 0.17 · 0.24 mm^3
absorption coefficient	4.177 mm^{-1}
min/max transmission	0.49 / 0.78
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	monoclinic, P $2_1/n$
a	9.8811(5) Å
b	10.5421(5) Å
c	20.9451(11) Å
α	90°
β	101.637(2)°
γ	90°
V	2136.95(19) $Å^3$
min/max Θ	1.985° / 45.307°
number of collected reflections	125341
number of independent refections	17699 (merging r = 0.033)
number of observed reflections	10861 (I>2.0σ(I))
number of refined parameters	479
r	0.0426
rW	0.0775
goodness of fit	1.0918



8 awy025 (wrong unit cell)

Table 8. Crystal data for **awy025**

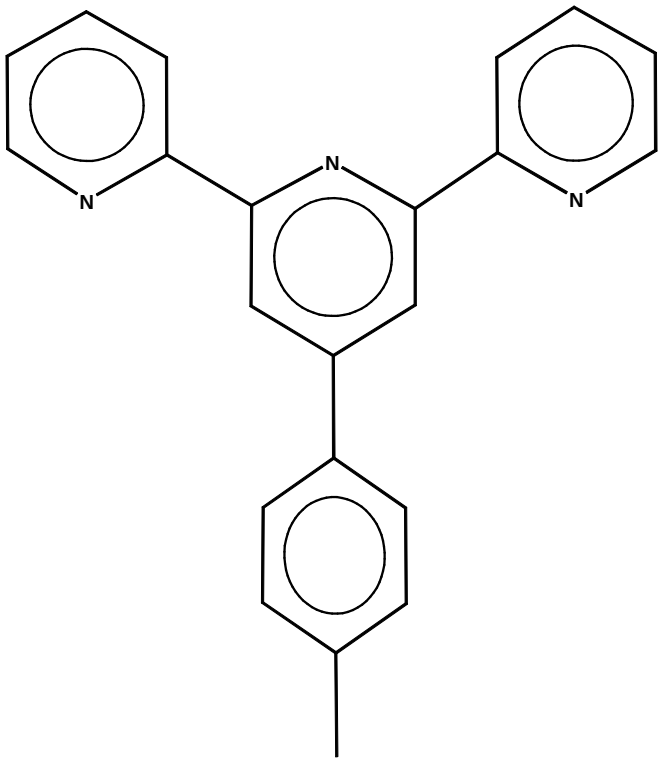
formula	$C_{26}H_{35}Br_2Fe_1O_1P_2Zn_1$
formula weight	706.55
Z, calculated density	12, 1.644 Mg · m^{-3}
F(000)	4260
description and size of crystal	orange needle, 0.03 · 0.08 · 0.22 mm^3
absorption coefficient	4.277 mm^{-1}
min/max transmission	0.71 / 0.88
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	orthorhombic, P 2_1 2_1 2_1
a	9.7252(6) Å
b	24.2001(14) Å
c	36.388(2) Å
α	90°
β	90°
γ	90°
V	8563.9(9) $Å^3$
min/max Θ	1.683° / 29.575°
number of collected reflections	217019
number of independent refections	24026 (merging r = 0.098)
number of observed reflections	14263 (I>2.0σ(I))
number of refined parameters	893
r	0.0435
rW	0.0508
goodness of fit	0.9846



9 bb66 (wrong unit cell or bad crystal quality)

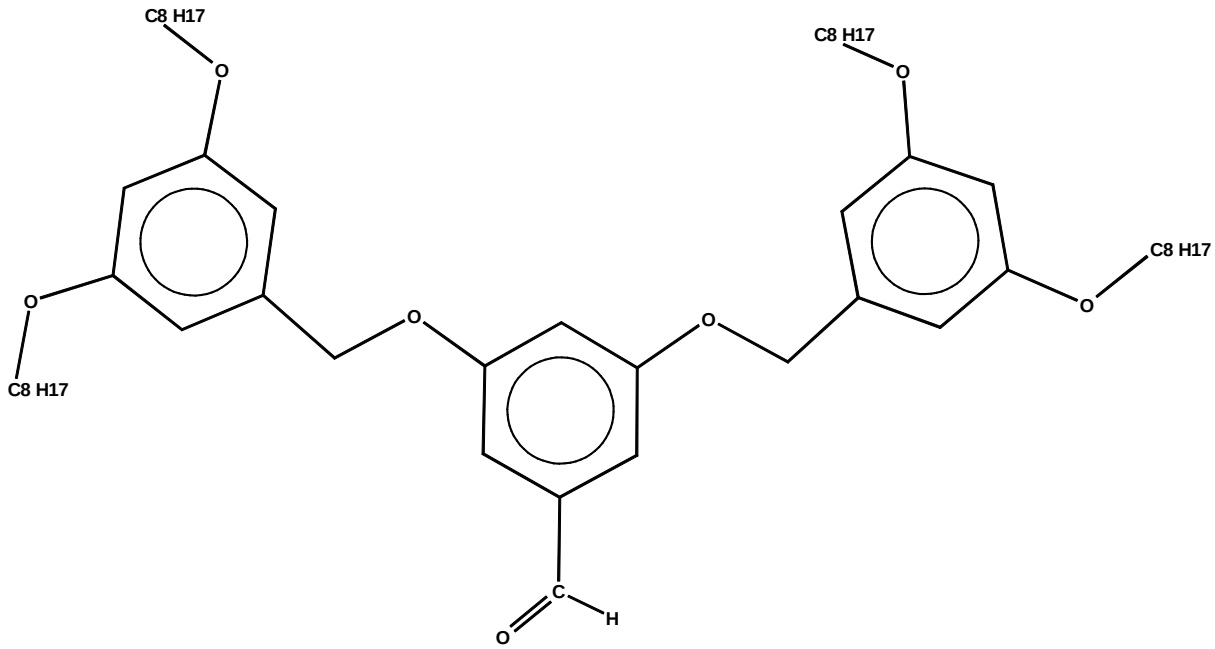
Table 9. Crystal data for **bb66**

formula	$C_{22}H_{17}N_3$
formula weight	323.40
Z, calculated density	8, 1.308 Mg · m^{-3}
F(000)	1360
description and size of crystal	colourless plate, 0.06 · 0.20 · 0.53 mm^3
absorption coefficient	0.079 mm^{-1}
min/max transmission	0.98 / 1.00
temperature	173K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	monoclinic, P $2_1/c$
a	9.3566(4) Å
b	33.3794(13) Å
c	11.3619(4) Å
α	90°
β	112.2436(11)°
γ	90°
V	3284.5(2) $Å^3$
min/max Θ	1.220° / 27.943°
number of collected reflections	23796
number of independent refections	7844 (merging r = 0.046)
number of observed reflections	4838 (I>2.0σ(I))
number of refined parameters	451
r	0.0422
rW	0.0715
goodness of fit	1.0264



10 mh356 (phase transition)

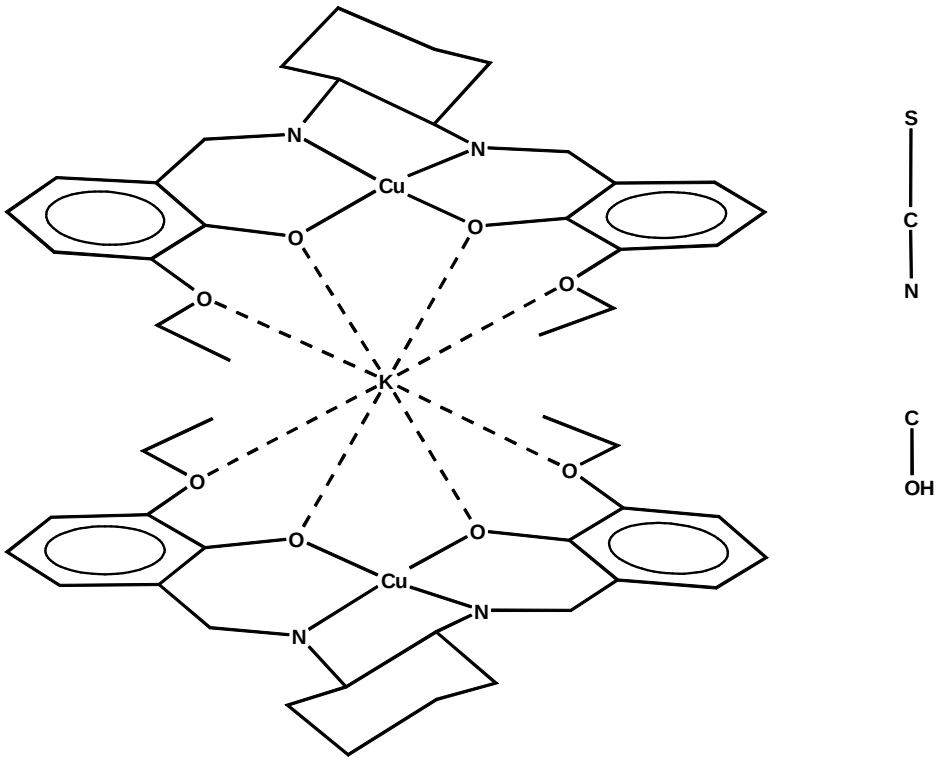
Table 10. Crystal data for **mh356**	**high temp**	**low temp**
formula	$C_{53}H_{82}O_7$	
formula weight	831.23	
Z, calculated density	2, 1.103 Mg · m^{-3}	1.132 Mg · m^{-3}
F(000)	912	
description and size of crystal	colourless block, 0.07 · 0.11 · 0.23 mm^3	
absorption coefficient	0.071 mm^{-1}	0.073 mm^{-1}
min/max transmission	0.99 / 1.00	
temperature	223K	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)	
Crystal system, space group	triclinic, P -1	
a	10.3926(5) Å	10.4362(7) Å
b	16.1928(8) Å	15.4916(11) Å
c	16.3589(7) Å	16.3189(12) Å
α	103.076(2)°	100.110(4)°
β	102.323(2)°	103.655(4)°
γ	103.409(3)°	101.785(4)°
V	2503.7(2) $Å^3$	2439.3(3) $Å^3$
min/max Θ	1.598° / 29.574°	2.078° / 30.034°
number of collected reflections	67547	62053
number of independent refections	14037 (r = 0.033)	14198 (r = 0.045)
number of observed reflections	7646 (I>2.0σ(I))	6603 (I>2.0σ(I))
number of refined parameters	614	550
r	0.0745	0.0520
rW	0.1574	0.1672
goodness of fit	1.1766	1.1074



11 gq68 (wrong spacegroup)

Table 11. Crystal data for **gq68 (contains partially occupied methanol)**

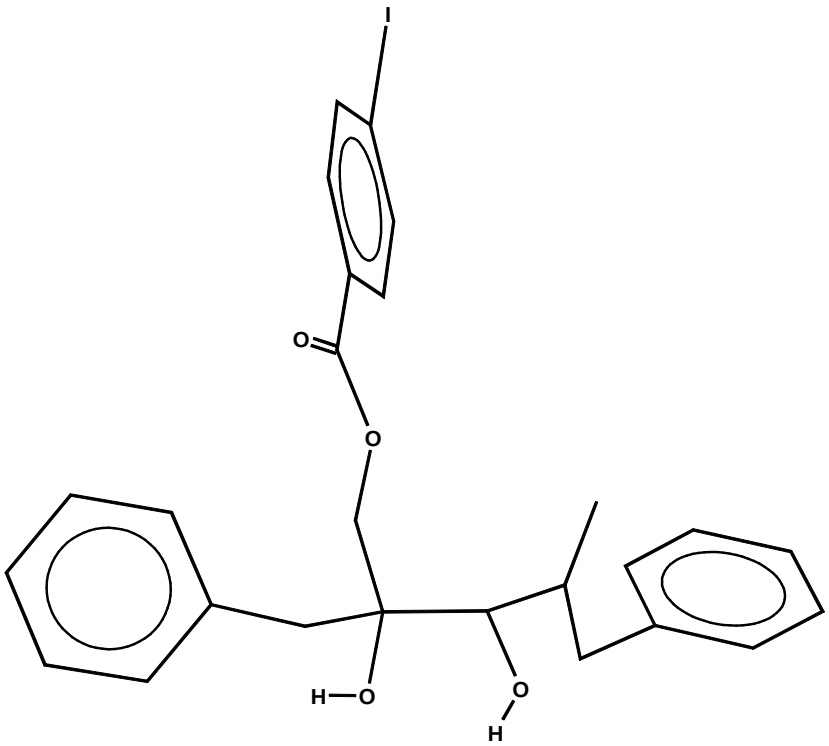
formula	$C_{49.80}H_{59.20}Cu_2K_1N_5O_{8.80}S_1$
formula weight	1066.90
Z, calculated density	1, 1.459 Mg · m^{-3}
F(000)	556.400
description and size of crystal	red needle, 0.02 · 0.06 · 0.60 mm^3
absorption coefficient	1.066 mm^{-1}
min/max transmission	0.94 / 0.98
temperature	173K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	triclinic, P 1
a	9.830(2) Å
b	11.531(2) Å
c	12.209(2) Å
α	105.53(3)°
β	103.18(3)°
γ	105.74(3)°
V	1214.0(6) $Å^3$
min/max Θ	2.276° / 32.015°
number of collected reflections	56546
number of independent refections	15546 (merging r = 0.081)
number of observed reflections	13799 (I>2.0σ(I))
number of refined parameters	614
r	0.0420
rW	0.0543
goodness of fit	1.0761



12 sb45-10 (space group ambiguity, h-bonds)

Table 12. Crystal data for **sb45-10**

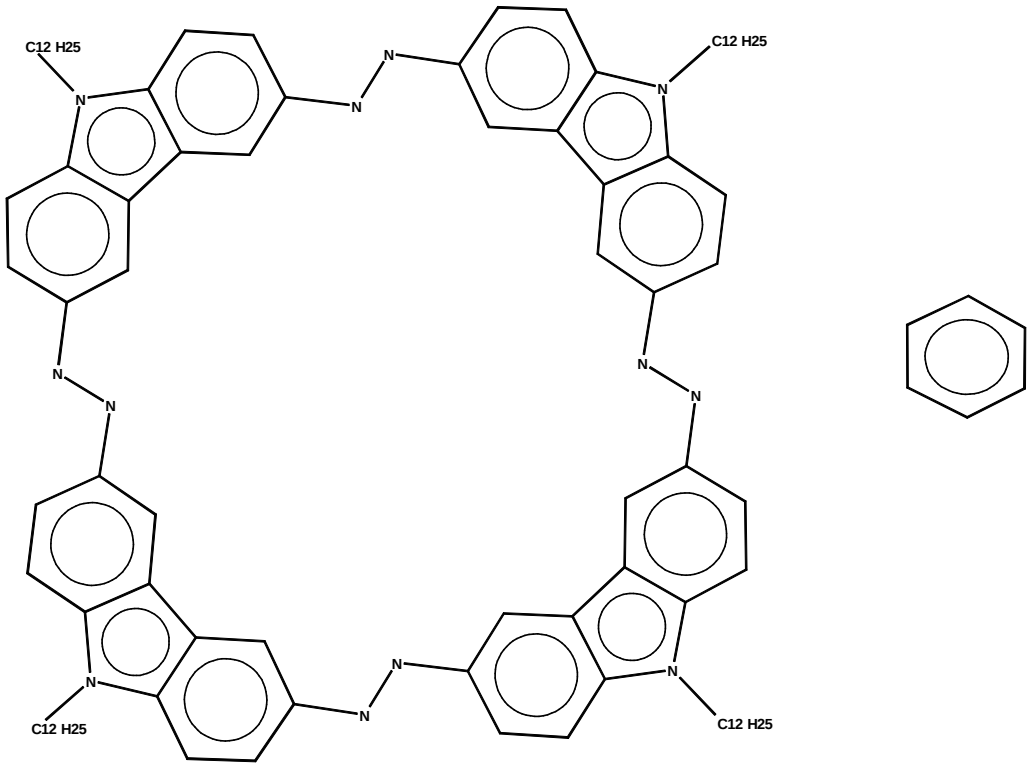
formula	$C_{26}H_{25}I_1O_4$
formula weight	528.39
Z, calculated density	2, 1.544 Mg · m^{-3}
F(000)	532
description and size of crystal	colourless needle, 0.03 · 0.04 · 0.31 mm^3
absorption coefficient	1.439 mm^{-1}
min/max transmission	0.94 / 0.96
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	triclinic, P 1
a	5.6359(7) Å
b	13.7031(17) Å
c	15.0995(17) Å
α	82.314(8)°
β	79.604(5)°
γ	89.600(6)°
V	1136.5(2) $Å^3$
min/max Θ	1.899° / 27.484°
number of collected reflections	19016
number of independent refections	10123 (merging r = 0.036)
number of observed reflections	7325 (I>2.0σ(I))
number of refined parameters	560
r	0.0441
rW	0.0501
goodness of fit	1.1145



13 lus65 (disordered side chain near inversion center)

Table 13. Crystal data for **lus65**

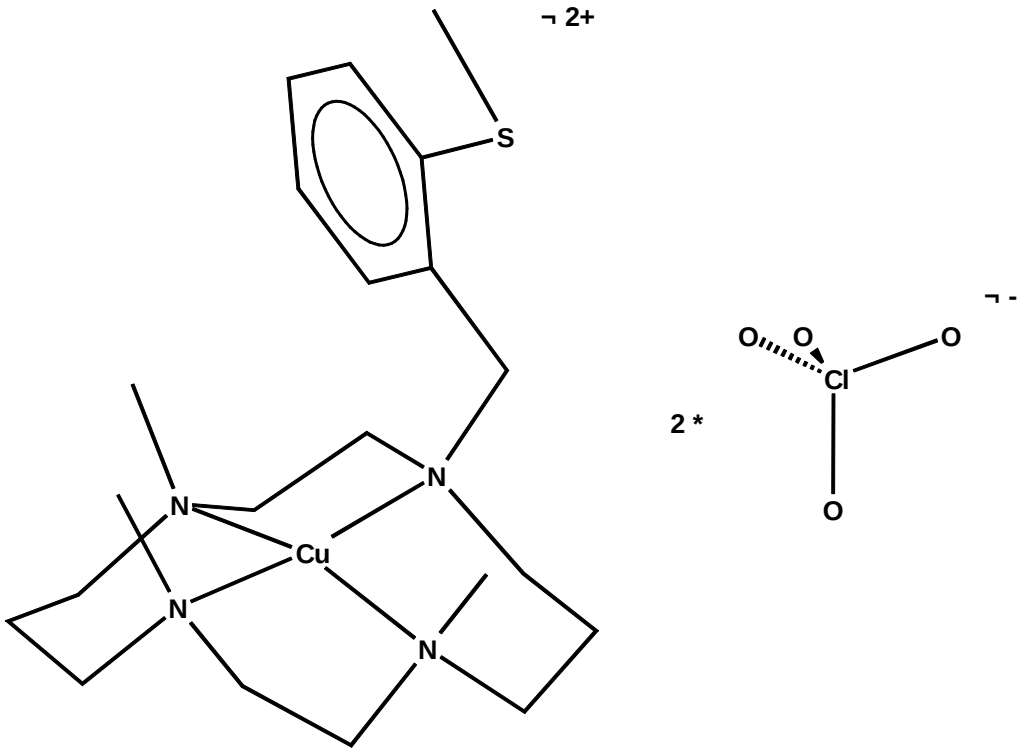
formula	$C_{108}H_{136}N_{12}$
formula weight	1602.35
Z, calculated density	2, 1.148 Mg · m^{-3}
F(000)	1736
description and size of crystal	colourless block, 0.06 · 0.13 · 0.27 mm^3
absorption coefficient	0.067 mm^{-1}
min/max transmission	0.99 / 1.00
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	monoclinic, P $2_1/n$
a	21.828(4) Å
b	9.4747(16) Å
c	22.495(4) Å
α	90°
β	95.114(10)°
γ	90°
V	4633.8(13) $Å^3$
min/max Θ	1.818° / 26.391°
number of collected reflections	53279
number of independent refections	9461 (merging r = 0.148)
number of observed reflections	4724 (I>2.0σ(I))
number of refined parameters	596
r	0.0633
rW	0.1986
goodness of fit	1.0530



14 cs-Cu-sme (two orientations of a whole group)

Table 14. Crystal data for **cs-Cu-sme**

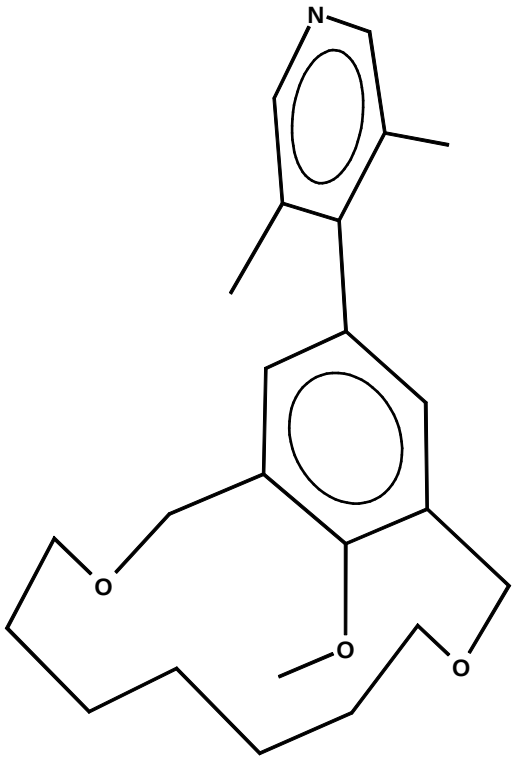
formula	$C_{21}H_{38}Cl_2Cu_1N_4O_8S_1$
formula weight	641.07
Z, calculated density	2, 1.550 Mg · m^{-3}
F(000)	670
description and size of crystal	blue plate, 0.10 · 0.27 · 0.31 mm^3
absorption coefficient	4.070 mm^{-1}
min/max transmission	0.34 / 0.67
temperature	293K
radiation(wavelength)	Cu K_α (λ = 1.54180 Å)
Crystal system, space group	triclinic, P -1
a	9.450(2) Å
b	12.143(2) Å
c	12.594(5) Å
α	82.43(2)°
β	77.72(2)°
γ	77.62(2)°
V	1373.7(7) $Å^3$
min/max Θ	3.606° / 77.375°
number of collected reflections	6112
number of independent refections	5829 (merging r = 0.013)
number of observed reflections	4848 (I>2.0σ(I))
number of refined parameters	508
r	0.0746
rW	0.0629
goodness of fit	1.2630



15 ts-a242-3 (strategy when splitting atoms)

Table 15. Crystal data for **ts-a242-3**

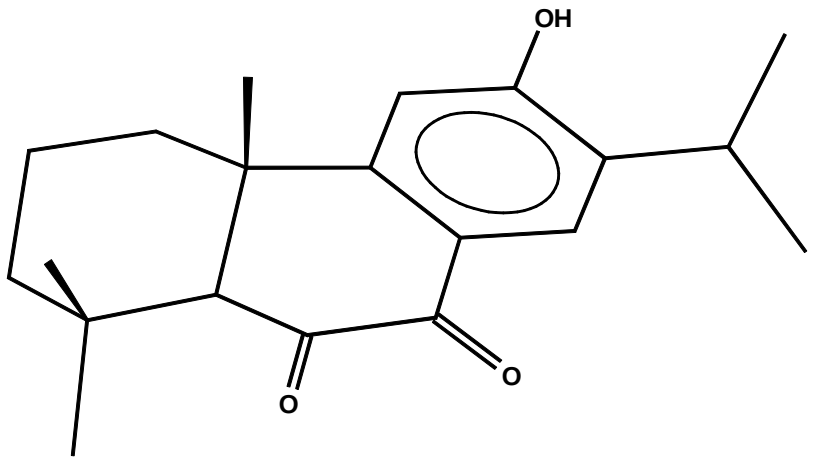
formula	$C_{23}H_{31}N_1O_3$
formula weight	369.50
Z, calculated density	2, 1.193 Mg · m^{-3}
F(000)	400
description and size of crystal	colourless block, 0.09 · 0.17 · 0.26 mm^3
absorption coefficient	0.078 mm^{-1}
min/max transmission	0.99 / 0.99
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	triclinic, P -1
a	8.2574(6) Å
b	9.7222(8) Å
c	13.4278(10) Å
α	84.159(5)°
β	73.626(4)°
γ	89.789(5)°
V	1028.54(14) $Å^3$
min/max Θ	1.589° / 30.033°
number of collected reflections	23593
number of independent refections	5991 (merging r = 0.036)
number of observed reflections	3904 (I>2.0σ(I))
number of refined parameters	263
r	0.0419
rW	0.0719
goodness of fit	1.0943



16 ckj_2_145b (disorder between methyl and ethyl group)

Table 16. Crystal data for **ckj_2_145b**

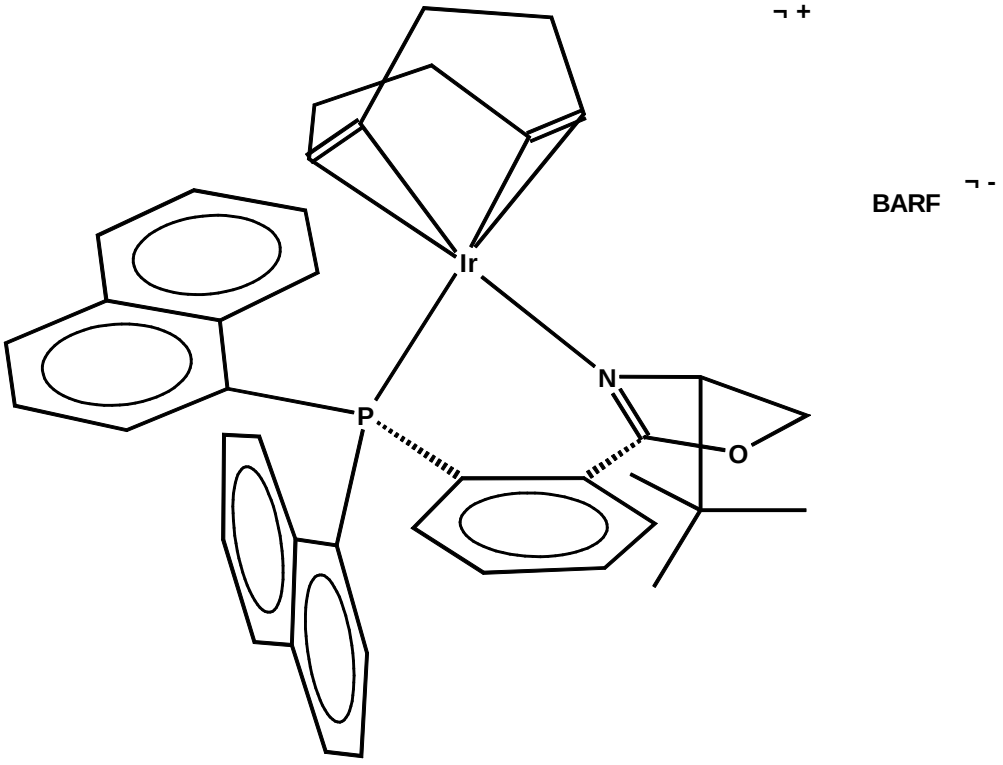
formula	$C_{20.36}H_{26.72}O_3$
formula weight	319.46
Z, calculated density	4, 1.214 Mg · m^{-3}
F(000)	691.489
description and size of crystal	colourless needle, 0.04 · 0.06 · 0.33 mm^3
absorption coefficient	0.080 mm^{-1}
min/max transmission	1.00 / 1.00
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	orthorhombic, P 2_1 2_1 2_1
a	9.7586(4) Å
b	12.6301(6) Å
c	14.1851(7) Å
α	90°
β	90°
γ	90°
V	1748.34(14) $Å^3$
min/max Θ	2.159° / 30.578°
number of collected reflections	43906
number of independent refections	3022 (merging r = 0.101)
number of observed reflections	2025 (I>2.0σ(I))
number of refined parameters	227
r	0.0347
rW	0.0778
goodness of fit	1.1429



17 ek006 (various disordered parts)

Table 17. Crystal data for **ek006**

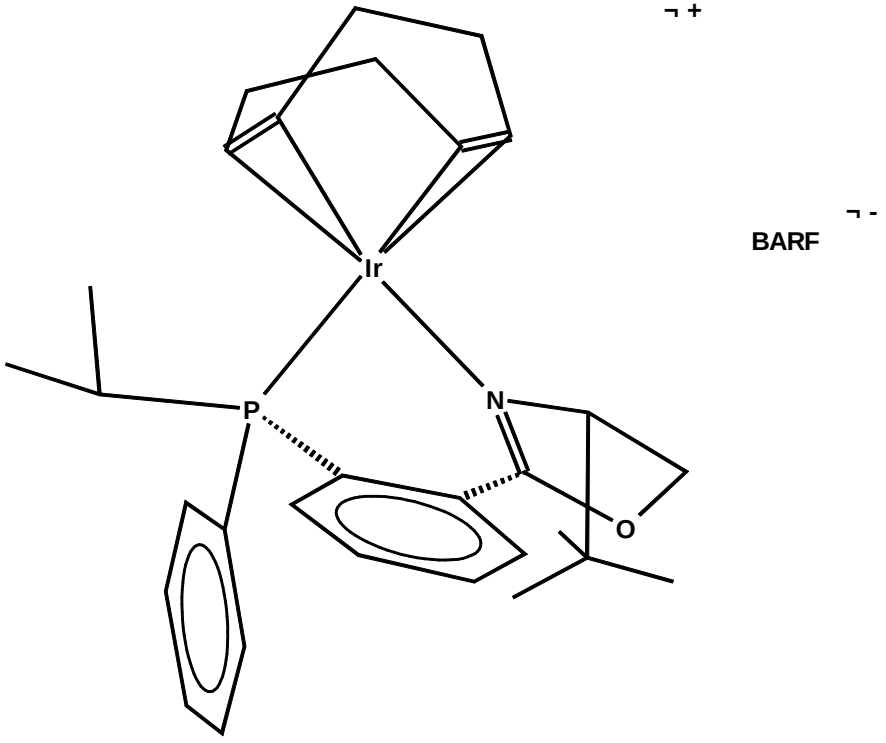
formula	$C_{73}H_{54}B_1F_{24}Ir_1N_1O_1P_1$
formula weight	1651.19
Z, calculated density	2, 1.622 Mg · m^{-3}
F(000)	1640
description and size of crystal	orange block, 0.06 · 0.11 · 0.18 mm^3
absorption coefficient	2.111 mm^{-1}
min/max transmission	0.79 / 0.88
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	monoclinic, P 2_1
a	13.7400(5) Å
b	12.5827(5) Å
c	19.9469(7) Å
α	90°
β	101.467(2)°
γ	90°
V	3379.7(2) $Å^3$
min/max Θ	1.657° / 33.728°
number of collected reflections	173257
number of independent refections	26771 (merging r = 0.036)
number of observed reflections	22662 (I>2.0σ(I))
number of refined parameters	1212
r	0.0255
rW	0.0381
goodness of fit	0.9905



18 sps080b (mirror plane as local symmetry of disorder)

Table 18. Crystal data for **sps080b**

formula	$C_{62}H_{51}B_1F_{24}Ir_1N_1O_1P_1$
formula weight	1516.06
Z, calculated density	4, 1.589 Mg · m^{-3}
F(000)	3004
description and size of crystal	red-brown needle, 0.08 · 0.10 · 0.42 mm^3
absorption coefficient	2.244 mm^{-1}
min/max transmission	0.80 / 0.84
temperature	173K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	monoclinic, P 1 2_1 1
a	21.6718(4) Å
b	13.5386(2) Å
c	22.8722(4) Å
α	90°
β	109.1860(8)°
γ	90°
V	6338.09(19) $Å^3$
min/max Θ	1.580° / 25.010°
number of collected reflections	21522
number of independent refections	21507 (merging r = 0.000)
number of observed reflections	16437 (I>2.0σ(I))
number of refined parameters	1668
r	0.0462
rW	0.0345
goodness of fit	0.9368

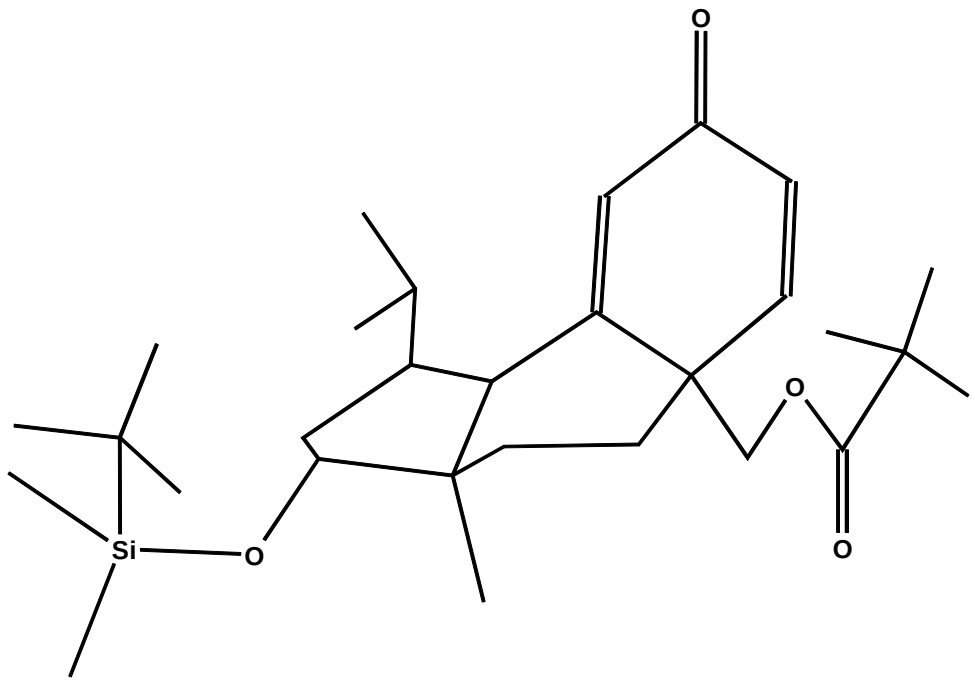


BARF

19 ep117b (again the mirror plane as local symmetry of disorder)

Table 19. Crystal data for **ep117b**

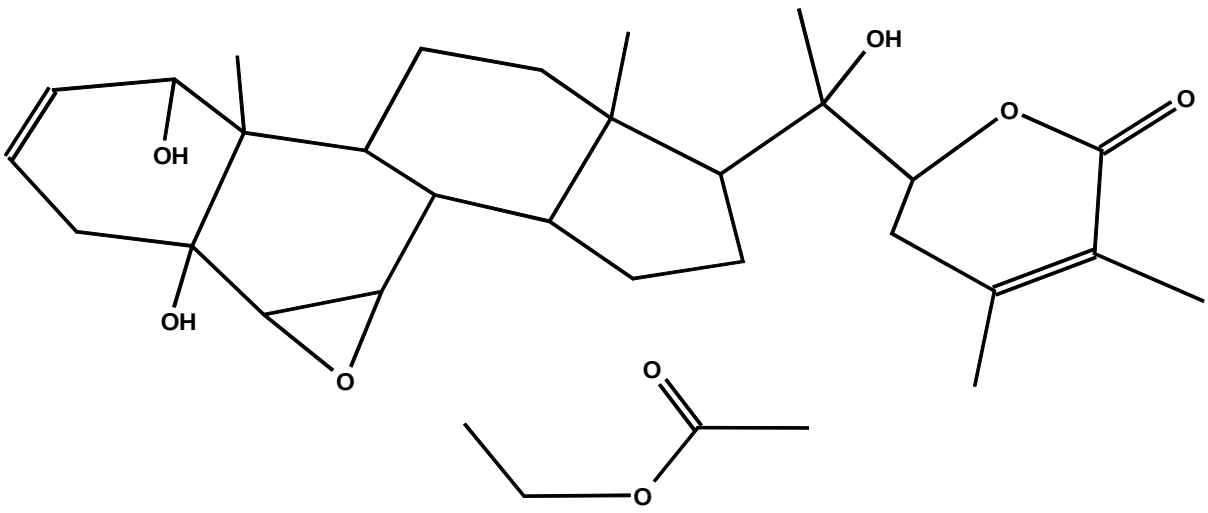
formula	$C_{29}H_{46}O_4Si_1$
formula weight	486.77
Z, calculated density	4, 1.119 Mg · m^{-3}
F(000)	1064
description and size of crystal	colourless block, 0.04 · 0.11 · 0.26 mm^3
absorption coefficient	0.111 mm^{-1}
min/max transmission	0.99 / 1.00
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	orthorhombic, P 2_1 2_1 2_1
a	9.7412(4) Å
b	11.7340(5) Å
c	25.2718(9) Å
α	90°
β	90°
γ	90°
V	2888.6(2) $Å^3$
min/max Θ	1.612° / 32.668°
number of collected reflections	77051
number of independent refections	10587 (merging r = 0.046)
number of observed reflections	7886 (I>2.0σ(I))
number of refined parameters	381
r	0.0345
rW	0.0561
goodness of fit	1.0790



20 jh120 (a disordered solvent molecule)

Table 20. Crystal data for **jh120**

formula	$C_{32}H_{48}O_8$
formula weight	560.73
Z, calculated density	4, 1.273 Mg · m^{-3}
F(000)	1216
description and size of crystal	colourless block, 0.09 · 0.14 · 0.25 mm^3
absorption coefficient	0.090 mm^{-1}
min/max transmission	0.99 / 0.99
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	orthorhombic, P 2_1 2_1 2_1
a	11.8379(5) Å
b	13.3459(6) Å
c	18.5218(9) Å
α	90°
β	90°
γ	90°
V	2926.2(2) $Å^3$
min/max Θ	1.881° / 32.358°
number of collected reflections	65338
number of independent refections	5762 (merging r = 0.045)
number of observed reflections	4785 (I>2.0σ(I))
number of refined parameters	416
r	0.0311
rW	0.0479
goodness of fit	1.1101



21 gq214 (ordered and disordered ligands side by side)

Table 21. Crystal data for **gq214**

formula	$C_{40}H_{34}Co_1N_{12}O_2S_2$
formula weight	837.85
Z, calculated density	4, 1.401 Mg · m^{-3}
F(000)	1732
description and size of crystal	brown needle, 0.04 · 0.06 · 0.23 mm^3
absorption coefficient	0.590 mm^{-1}
min/max transmission	0.97 / 0.98
temperature	123K
radiation(wavelength)	Mo K_α (λ = 0.71073 Å)
Crystal system, space group	monoclinic, P $2_1/c$
a	23.3334(7) Å
b	10.3868(3) Å
c	17.3827(6) Å
α	90°
β	109.477(2)°
γ	90°
V	3971.8(2) $Å^3$
min/max Θ	1.851° / 35.018°
number of collected reflections	102487
number of independent refections	17451 (merging r = 0.056)
number of observed reflections	9242 (I>2.0σ(I))
number of refined parameters	678
r	0.0697
rW	0.1443
goodness of fit	1.0536