Durham E-Theses

Systematic computational analyses and novel search procedures for crystallographic information

Cole, Jason

How to cite:

Use policy

Systematic Computational Analyses and Novel Search Procedures for Crystallographic Information

by Jason Cole

Thesis submitted in part fulfilment of the requirements for the degree of

Doctor of Philosophy

at the

University of Durham

Department of Chemistry   September 1995

Thesis
1995/
COL

Systematic Computational Analyses and Novel Search Procedures for Crystallographic Information

Submitted for the degree of Doctor of Philosophy, September 1995, by Jason C. Cole, University of Durham

ABSTRACT

This thesis describes work in two distinct areas of crystallographic research. In chapters 3 and 4, a number of novel crystal structures are reported. Eight new potential dyotropic rearrangement precursors are discussed in chapter 3. A comparison with other known structures is given. Three conjugated zwitterionic TCNQ type derivatives are discussed in chapter 4. All three molecules show good second order hyperpolarisability and so have potential uses in non-linear optical applications.

Chapter 6 and chapter 7 use the Cambridge Structural Database in novel ways to derive new information. In chapter 6 the estimated standard deviations of the 3D crystallographic coordinates have been used to derive more accurate precision indicators, as functions of the R-factor and the chemical constitution by correlation and regression methods. This research has shown that it is possible to provide effective improve precision indicators within the CSD for the circa 50,000 entries that do not possess coordinate e.s.d's.

In chapter 7 a re-analysis of space group frequencies is given. New software has been written to search the CSD and derive the lowest symmetry point group utilised by the independent residues within each crystal structure. With this analysis it has been possible to produce matrices which rationalise space group frequency as a function of point group utilised within the crystal. The requirement of mirror plane occupation in mirror-symmetric space groups has been confirmed. Further trends are also noticed.

The work described in this Thesis was carried out in the Department of Chemistry, Durham University, from October 1991 until December 1994, under the supervision of Professor J. A. K. Howard. All the work is my own, unless stated otherwise, and it has not been submitted previously for a degree at this or any other university.

J. C. Cole. 1995

Acknowledgements

Within a thesis, the acknowledgements page is the only place where the writer has the opportunity to write whatever they like, without having to have the page proof read by one of their peers or supervisors. Phrases like "to boldly go" (split infinitive) or spellings, such as "idependant" will usually abound. For will start I.appologuse this

A number of people have been deeply involved in the progression of my PhD through the 3(4 really) years and I will try to thank them in a "pseudo-"chronological fashion. First on the list is Dr. Jeremy Cockcroft who gave me patient attention during my early years in Durham, particularly all those annoying questions about "how do I write a FORTRAN program to ...".

During those early years, our group was a friendly place, and my peers, Clare (when she wasn't gallivanting around the world) and Nigel were always available for stimulating research sessions that usually involved words like "lets go to the New Inn", all in the interest of the furtherment of academic science, of course. There lively discussion and friendship has always been appreciated.

Like all academic groups people come and go. A number of new people have been through the clutches of Durham University, and since this thesis is already 300 pages long, I feel it would be a bit verbose to mention them all, but Roy, Jussi and Christian all deserve thanks for making me eat humble pie in crystallographic arguments on many occasions, but all the remaining students and post-docs are all heartily thanked.

My Mum and Dad have always been great to me during my studies... well I guess its pay back time. The next round of golf is on me. Also they, along with my Uncle Allen proof read this thesis. Thanks for that.

My Girlfriend, Mary, has probably been the most tolerant person in the entire universe in the last two years. She's made my life much much easier, and made the act of writing a less painful experience.

I would also like to thank the CCDC for supporting me in the last seven months whilst I have been writing my thesis. Although doing a job and writing a thesis is quite taxing, the CCDC have been extremely supportive; even scheduling my thesis into my work plans.

In the classical style I've "left the best till last". I have to thank most of all Frank Allen and Judith Howard for supervising and advising me throughout my PhD. It never ceases to amaze me how taxing it must be to have to look after students, and their encouragement and enterprise have always inspired me.

Table of Contents

Tables Presented in this Thesis

Figures Presented in this Thesis

Chapter 1.

Introduction

1.1. Introduction.

Crystallographic studies have provided a bedrock of structural information which has allowed chemistry to develop considerably in the past seventy years, primarily due to the technique allowing direct measurement of bond lengths to a precision of a 0.01Å and of angular parameters to a precision of better than a degree. Such information can be compared and contrasted for many similar studies to reveal the nature of chemical changes that may occur as a result of structural modifications.

In this thesis two sets of related novel structures are presented. The first of these studies (Chapter 3) analyses a series of compounds, and precursors to compounds, that may undergo intra-molecular dyotropic rearrangement reactions. Eight new structures have been elucidated. This has added to an extensive number of structural studies that have already been carried out. It is hoped that these new studies, along with past and future studies, and in conjunction with kinetic data, will provide a detailed understanding of the rearrangement reaction mechanism.

The second set of structures elucidated (Chapter 4) are a series of TCNQ based zwiterionic molecules that have potential use in non-linear optical materials. Three novel structures are presented. All three can be regarded as being of a quinonoidal nature. The three dimensional structures are immediately useful to the synthetic chemists since they provide a means of easily estimating the molecular hyperpolarisabilities for these three molecules. Such data is not readily available experimentally, but can be estimated from precise atomic coordinates.

Since 1965, a database of crystallographic studies of organic and metallo-organic structures (The Cambridge Structural Database, (CSD), Allen *et al*, 1991) has been compiled at the Cambridge Crystallographic Data Centre. This compilation of structural information has allowed analyses of large numbers of similar structures by a variety of statistical methods. Any database of structural information relies on its ease of searchability. The complete CSD system contains a powerful software package to fulfil this rôle (*QUEST3D*, Allen et al, 1991). Chapters 6 and 7 of this thesis develop new methods of searching and analysing data contained in the CSD.

In Chapter 6, an analysis of precision of the stored structural coordinates is given. This study is based on an analytical derivation given by Cruickshank (1960). The crystallographic R-factor, (defined below in equation 1.1), and functions of the elemental constitution of the asymmetric unit have been used to derive estimates of the mean coordinate e.s.d's (defined below in equation 1.2) for any particular element. It is shown that reasonable precision estimates can be obtained using these variables. The derived functions will ultimately be used to derive precision indicators for a large number of structures on the CSD for which coordinate e.s.d's have not been stored. (The only precision variable currently available for all CSD structures is the crystallographic R-factor).

$$R = \frac{\sum_{hkl} (F_{o,hkl}^2 - F_{c,hkl}^2)^2}{\sum_{hkl} |F_{o,hkl}^2|} \qquad (1.1)$$

In equation 1.1, $F_{o,hkl}^2$, is the observed structure factor for the reflection with miller indices, h,k,l, $F_{c,hkl}^2$, is the calculated structure factor for the same reflection. The summations are over all sets of h,k,l indices in the measured data set.

The estimated standard deviation (e.s.d.) of the jth parameter p_j, in the least squares process, can be expressed as

2

$$\sigma_{p_j} = \left(\left(A^{-1} \right)_{jj} \frac{\sum\limits_{i=1}^{N_r} w_i \left(F_{o,i}^2 - F_{c,i}^2 \right)^2}{N_r - N_p} \right)^{\frac{1}{2}} \qquad (1.2)$$

Where A^{-1} is the variance-covariance matrix that is calculated during least squares, w_i is an appropriately derived weight, $F_{o,i}^2$ is the square of the observed structure factor for the ith reflection and $F_{c,i}^2$ is the calculated value. The number N_r is the number of reflections, and N_p is the number of refined parameters.

Chapter 7 gives a study of space group frequencies on the CSD. In this study, new software has been written to derive the crystallographic point groups of the individual moieties and ions in each structure. This software has been used to generate tables of space group frequency as a function of the symmetry utilised in the unit cell, and as a function of the number of moieties or ions in the structure. The intention is to include the derived information as a searchable field within CSD entries, so that further analysis of such data can be carried out in the future. The study has shown that certain space groups occur more frequently when specific symmetry elements are occupied. Individual symmetry elements do have more dramatic influence on space group symmetry than others. Mirror planes are always occupied, higher order rotation axes are usually occupied. Improper rotation axes seem to have less influence.

A more detailed introduction to each of these topics is given in their respective chapters.

1.2. References

Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M., Watson, D. G.; *J. Chem. Inf. Comput. Sci.* **31**, 187, 1991

Cruickshank, D. W. J., *Acta Cryst.* **13**, 774-777, 1960

Chapter 2.

Experimental Crystallography

2.1 Introduction

Since the discovery of X-ray diffraction by crystals in 1912 by Laue, Friedrich and Knipping, crystallography has revealed more information about three-dimensional (3D) chemical structure than any other single experimental technique, due to its ability to elucidate accurate 3D atomic positions from simple experimental measurements. These measurements, a record of the intensities of diffraction maxima, do however suffer from one problem: the relative phases of the diffracted beams cannot be observed in the experiment - the so called phase problem of crystallography. Direct methods of phase determination have now dramatically increased the application of crystallography. Prior to this development, only Patterson methods were available to crystallographers as a systematic means of establishing the phase information that is necessary to transform the structure factors into an electron density map. Any other attempt at solving structures relied on trial and error. The Patterson method relies on having at least one relatively heavy atom within the unit cell. Direct methods works in precisely the opposite situations and so allowed the structural determinations of unsubstituted organic molecular systems (i.e. non-heavy atom derivatives) for the first time.

The past three decades have seen a vast increase in computing power. In the mid 1960's the average crystal structure could take as much as a year's work from experiment to final publication, due to the time taken for data collection and for the complex calculations to be completed on the computers available then. Now computers can perform these calculations in very short times, even for compounds with large numbers of atoms. This has led to a much larger number of structures being solved and refined. The rate-limiting

step in structural determination is now often the crystal growth, rather than the structural study.

2.2 Experimental Details

2.2.1 Crystal selection & mounting

Any structural study firstly requires selection of a suitable single crystal from a batch of crystals provided. The selection procedure is usually as follows. Firstly, crystals are inspected under a microscope in order to investigate their size and morphology. The ideal crystal is fairly isotropic in shape with well defined crystal faces. The required dimensions depend on the nature of the experiment: in neutron crystallography samples with an average dimension of more than 2 mm are preferred due to the low incident neutron flux, but in conventional X-ray diffractometry dimensions between 0.2 mm and 0.4 mm are ideal, depending on the unit cell contents[†]. No one dimension should exceed 0.7 mm, since this induces certain systematic errors into data collection due to varying volumes of crystal lying in the beam at different crystal orientations. The crystal size also influences the decision on collimator size. Larger crystals require larger collimators to ensure that all of the crystal is bathed in the beam. A limit on collimator size is imposed by the size of the focal spot of the generated X-rays.

Once a crystal of this type is located it can sometimes be inspected by use of a polariser on a microscope. Polarised light is rotated within a crystal due to the variation of the refractive index in different directions within a crystal of non-cubic symmetry. This

---

[†] If strongly scattering atoms are present in the unit cell a small crystal is adequate, and sometimes even favourable, since in this case the small crystal will create less diffraction and hence be less susceptible to systematic errors such as counter saturation. A larger crystal is used for "light atom" stuctures. The size of crystal needed is also governed by the X-ray flux. Harding and co-workers (1994) have recently demonstrated that single crystal studies can be carried out on compounds with crystals of a few microns using monochromated synchrotron radiation

process identifies situations where two crystals are fused in different orientations. With a single crystal one expects clean extinction in crossed polarisers.

Having selected a crystal it has to be firmly mounted for the diffraction experiment. Crystals are mounted on the end of a fine glass fibre with epoxy glue, ideally with the longest dimension approximately parallel to the fibre axis. In certain cases it is not possible to mount crystals on glass fibres because exposure to the atmosphere causes the crystal to decay. In such cases two options are available. Firstly the crystal can be mounted in a Lindemann tube under a suitable dry inert atmosphere. The second, more recently developed method for this situation uses the inert nature of perfluorinated ethers to shield the crystal from the atmosphere. Such a method has been described recently (Stalke, 1993), although this is not the first occurrence of its use. The usual procedure here is to bathe the crystals in the oil within a large Schlenck tube under an N_2 atmosphere and then spoon crystals out of this tube onto a microscope slide. The perfluorinated oil is then also used as an adhesive, since it solidifies at low temperature. Studies of unstable compounds are routinely performed at low temperature within the author's laboratory. The benefit of using perfluorinated oils for unstable compounds is that it avoids the large amount of background scatter given by a Lindemann capillary.

2.2.2 The Diffractometers used in this thesis.

Three diffractometers were used in the crystallographic studies reported in this thesis. All three follow the 4-circle Eulerian geometry design (See, for example Arndt and Willis, 1966). This design consists of 4 concentric drive circles arranged as shown in Figure 2.1. Three of these circles are used to orient the crystal in space. The fourth circle (The 2θ circle) moves the detector to a particular Bragg position. The crystal is placed on a goniometer head at the centre point of all 4 circles such that movement of any one circle rotates the crystal without precession. The X-ray beam passes through this central point perpendicular to the "χ" circle when it is at its zero position. The detector is collinear with the source and the centre point of the circles.

6

Figure 2.1. Schematic drawing of a Eulerian 4-circle diffractometer.
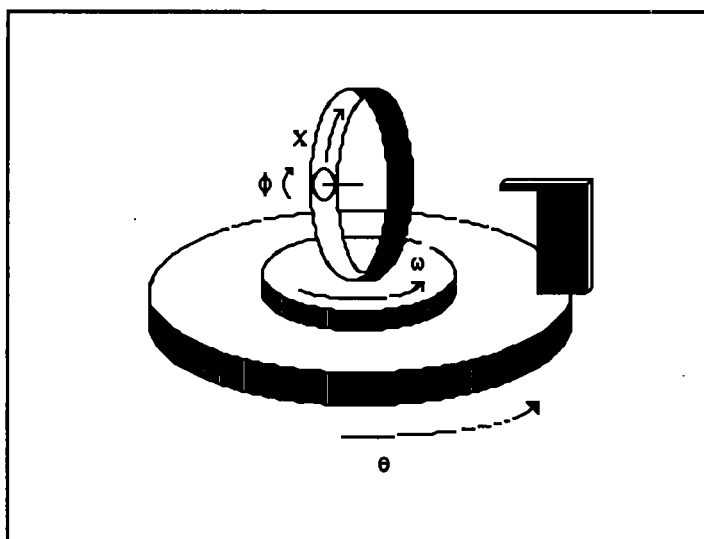
The three diffractometers used in the experiments described in this thesis were a Rigaku AFC6S and a SIEMENS P4, both with molybdenum sealed-tube sources, and an in-house designed diffractometer consisting of a set of HUBER circles mounted on a SIEMENS molybdenum rotating-anode source.

2.2.3 Unit cell determination; Identification of the Bravais Lattice

Once the crystal is on a diffractometer, the first process is the determination of the dimensions and symmetry of the unit cell. To do this a number of reflections have to be located. There are a number of ways in which this can be done. The usual method is by a search in reciprocal space. Initially the angles are rotated with the shutter open until counts are registered above the background level. This "peak" is then centred starting from this position, if possible, by one of two methods. The first is the half shutter method. In this there are two shutters, one of which obscures the left hand half of the detector aperture, the other the right. The diffractometer control program drives through each angle in turn and finds the point at which placing either shutter into the beam reduces the intensity by half; this is regarded as the centre of the peak at this angle. The other method scans through each rotation axis in turn and tries to find the two points on either side of the scan where the point intensity is half of the maximum point intensity in the profile. The average of these two points gives an estimate of the centre of the scan.

Both methods described above are iterative: the 4 circles are moved to the positions where the peak was first detected and then one axis is scanned. At the end of the scan a new position for this particular angle is derived by one of the processes described above (either by half-shutters or the half heights). The same routine is then performed for the next circle with the previous one updated. This process is repeated until no angle position moves by more than a pre-set threshold. In the diffractometers used only the AFC6S uses shutters, the other two use the half-height method.

By the above process, a number of reflections can be located and centred. The centring process also checks the sample quality, since it shows the reflection profile shape, and the integrated intensity.

The other method of initial reflection location is to use a rotation photograph. Here a Polaroid film is placed in front of the detector with all 4 circles set to zero. The ϕ circle is then rotated with the beam shutter open thus exposing the film. Diffraction occurs from all of the Bragg planes that are brought into a diffracting position. The diffraction vector can be in any direction, providing the Bragg condition is fulfilled, so spots are observed across all of the Polaroid film. This is a very efficient method of checking a crystal and obtaining the positions of a large number of reflections for centring but the cost of Polaroid film means that the method is only used sparingly.

Having optimised a set of angles for a number of reflections (Usually at least 10, and more commonly about 20) the diffractometer control software indexes the reflections.

The primary aim of indexing reflections is to define a transformation between the reflection position in real space and the Miller indices that represent the reflection position in reciprocal space, i.e. a coordinate system that is entirely governed by the crystal, but can be oriented anywhere in real space. The method first calculates the set of vectors between all reflections in real space. The three shortest non-coplanar vectors in this space

8

are then regarded as possible coordinate axes in reciprocal space (their lengths are the reciprocal axis lengths). Using the two axis sets it is possible to generate a transformation between them. This can then be tested with the remaining reflections to see if they give Miller indices that are integer values. In cases where fractional Miller indices are calculated, new basis vectors are assigned. New vectors can then be calculated which eliminate the fractional indices. Once all reflections have been assigned Miller indices, a non-linear least squares refinement of the cell parameters can be carried out using the refined angles of all the reflections as observations.

This process will usually derive a unit cell based on the three shortest non-coplanar lattice vectors [a "Buerger" (1957,1960) cell]. From this a Niggli (1982) reduced cell can be derived (Santoro & Mighell, 1970). This procedure usually fails in cases where there is more than one single crystal in the beam (i.e. when the crystal is twinned).

The cell thus found is then tested for symmetry. There are programs written and published in the literature to carry out higher symmetry searching (for example; *XPREP* Sheldrick, 1986) These programs take the cell dimensions and directions in real space (that can be calculated from the reciprocal cell parameters) and generate new cells which have basis vectors lying along combinations of the reduced cell vectors. If the new cell has a higher metric symmetry then this is reported as a higher symmetry cell. The derived cell can be further checked by using the Laue symmetry, as described below.

In strongly diffracting crystals, a better cell can usually be achieved by centring a set of higher angle reflections due to the better resolution of the d-spacings at higher 2θ. This can be explained quite simply by considering the sine curve. At low θ the d-spacing rapidly changes with $\sin\theta$ so a higher angular resolution is required to resolve a particular deviation δd in the d-spacing. Conversely at higher θ the change is slower and so the same deviation δd is more easily resolved. This advantage is offset by the fact that

stronger reflections are usually better centred than weaker ones and so the lower angle peaks are normally better resolved in θ[§].

Laue symmetry can be used to check the Bravais lattice that the above process predicts. A crystal has a specific geometric symmetry that is imposed as a result of the crystal lattice. As a result of this symmetry certain reflections become equivalent to one another. An example is the monoclinic lattice. This is shown below in Figure 2.2

It is apparent from this that the monoclinic lattice has 2/m symmetry. The drawing is shown with the respective name for the Miller indices in that direction. Due to the symmetry one can see that in monoclinic compounds the reflections +h, k, +l are equivalent to -h, k, -l. Similarly the mirror plane, in the plane of the paper, makes the reflections h, +k, l and h, -k, l equivalent so we can write:

$$h, k, l \equiv h, -k, l \equiv -h, k, -l \equiv -h, -k, -l$$

Figure 2.2 The Monoclinic Lattice.

_____
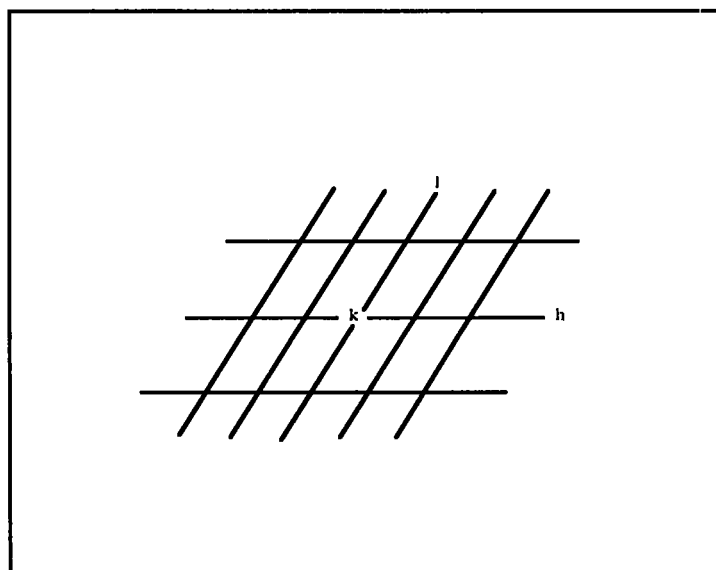
[§] Very intense reflections do not always centre as well as weaker ones due to counter saturation.

10

Modern software has available "Laue checks" that allow the user to measure the intensity of sets of equivalent reflections to check that they all have similar intensity. This is always a worthwhile procedure prior to data collection, since sometimes crystal cell parameters can show a higher metric symmetry than the symmetry of the diffraction intensities.

The Laue symmetry defines the unique set of data that is required for a structural study. If one set of equivalent reflections is collected then the structure can be fully defined, since the individual measurement can be regarded as representing its equivalent.[†]

2.2.4 Data Collection.

Data collection is the most important part of any single crystal experiment because ultimately it will yield the final results. The process of data collection is fully automated on modern diffractometers. Several parameters have to be set prior to data collection and many options are available. The first of these is the scan type with which data are to be collected. In the experiments described later either ω or $\omega/2\theta$ scans were used for data collection. The decision between these two scan types was always made on the peak width and on the individual knowledge of peak overlap for each structure: sometimes, if the cell volume is large, if one axis is particularly large, or if any angle is relatively far from 90°, then peaks can be close together in θ and thus the best policy is to scan in ω only. Since all three diffractometers use monochromatic radiation, filter streaking did not have to be considered. The widths of the scans were based on the peaks that were located during cell determination. These were allowed to vary during each experiment and in different ways depending on the machine used. The AFC6S diffractometer has many options available: $\omega/2\theta$ scans are available and can have a variable width based on the function, $\text{width}(\omega)° = A + B\tan\theta$ where A and B are parameters that are calculated experimentally from high angle peaks. The parameter, B, reflects the change in the diffraction angle of the $K_{\alpha 1}$ peak as compared to that of the $K_{\alpha 2}$ peak. The discrepancy

---

[†] In non-centrosymmetric space groups, Friedel opposites should be collected due to differences in the imaginary contributions to the structure factors.

11

between the two peaks increases with angle, and so the peak width broadens with θ. Alternatively the user can select Lehmann-Larsen fitted ω scans which are of a fixed width.

The SIEMENS P4 and the HUBER/SIEMENS rotating anode software both based their scan widths on the calculated positions of the $K_{\alpha 1}$ and $K_{\alpha 2}$ peaks for each reflection. The P4 also had an option for using ω scans with its width defined by the same function noted above.

Having decided the scan type and the scan width defining parameters, the unique set of reflections has to be considered. As already mentioned, the Laue symmetry, along with the lattice centring, defines the unique set of reflections. In order to solve and refine a structure as easily as possible, at least a unique data set should be collected[†]. Lattice centring, when present, reduces the number of reflections required. The centring causes certain reflections to have zero intensity, due to the added symmetry that the centring implies. Because the cell centring can be derived at the cell determination stage, the reflections affected can be omitted from data collection.

The AFC6S diffractometer has a repeat scan option available to the user. In this, a pre-set statistical target can be set by the user. The software calculates the ratio of intensity (I) to its estimated standard uncertainty [$\sigma(I)$] and repeats the data point. The two measurements are then merged to improve the statistics. This process is repeated until either a pre-set maximum number of scans is carried out or a pre-set "Target" $I/\sigma(I)$ is achieved. A similar routine exists on the P4 but it is unfortunately poorly conceived. The P4 equivalent routine measures a first scan and then calculates the speed of the second measurement based on the first. Unfortunately it does not use the first measurement at all in deriving the final reflection intensity and, thus, is a waste of time. The rotating anode

---

[†] Occasionally an experimental problem such as decay, or machine failure can curtail an experiment before completion.

can only measure reflections at a fixed scan speed. The data collection software on this machine is currently being extensively upgraded to provide a larger number of options.

The SIEMENS P4 and the rotating anode both collect and store profiles. The AFC6S can collect profiles in its Lehmann-Larsen mode but the profile is interpreted at the time of collection and not stored.

Another important parameter for data collection is the time that has to be spent on each reflection. This is based upon the strength of diffraction of a particular sample, although if a sample decomposes due to radiation damage a fast data collection becomes more favourable. A minimum time may be pre-set on all the machines described.

The final decision prior to data collection concerns the check reflections. A number of reflections are re-collected at pre-specified intervals during data collection in order to monitor the radiation damage that occurs to the crystal as well as whether it is moving during the experiment. The more sophisticated software checks the movement by monitoring its profile and its position. The P4 and the SIEMENS/HUBER machine only use intensity as a guide to crystal movement but the AFC6S checks the positioning of the check reflection to ensure that it has not moved. The average intensity of standard reflections indicates the amount of decay that has occurred as a result of X-ray exposure.

2.3 Data Reduction

2.3.1 Extracting the raw intensities

Having collected a data set, the raw intensities or profiles obtained need to be processed. This is a topic that has been extensively researched in the past two decades, and many methods have been developed for obtaining accurate values for the integrated intensity from the profiles measured. The AFC6S measures the integrated intensity directly, but the other two diffractometers both store profiles for each reflection. These are processed at the end of data collection. The P4 diffractometer software uses many parameters for

analysing its profiles. Firstly an "asymmetry" factor is used. The median of the profile is calculated and its position is compared to the position of the maximum peak point in the profile. If the absolute value of the difference between the two is greater than a pre-set number of profile steps then the reflection is rejected.

The P4 also analyses how far the peak median lies off the centre of the scan. If the median is too far away from the scan centre then the peak is rejected. A final test on the profile employs a measurement of the background asymmetry. If the left background is considerably different from the right background after analysis then the reflection is rejected. This form of analysis usually picks up most of the rogue peaks although there is no substitute for actually looking at bad profiles, which will usually have F values that fit poorly in least squares refinement.

The rotating anode data reduction software has no profile analysis (in one unpublished structure new software had to be written "on the fly" to remove certain rogue reflections that had been retained). It only uses the Lehmann-Larsen algorithm to estimate the value of the intensity. The current software upgrades will provide the necessary data format for using the DREAM system (Blessing, 1987) which should allow much more detailed analysis of data from this diffractometer.

2.3.2 Corrections for systematic effects

Crystal structure intensities are always affected by certain systematic effects during their collection, due to the way in which they are collected. Some of these corrections are applied automatically by data reduction programs, but it is worth understanding the basis of the corrections that are applied.

a) The Lorentz correction

This is a geometric correction that is applied to compensate for the speed at which the Bragg peak passes through its diffracting position, which varies with 2θ angle. In the case of the Eulerian geometry diffractometer the Lorentz correction can be written as:

$$L = \frac{1}{\sin 2\theta}$$
(2.1)

b) The polarisation correction

This is another geometric correction that accounts for the fact that X-ray scattering by electrons is not homogeneous in all direction, but depends on the angle at which it is observed, thus the relative intensity is θ dependant and so a correction has to be applied. A polarisation correction is required every time diffraction occurs and so involves a combination of two terms for monochromatic radiation. The reduced form for the Eulerian geometry with a monochromatic beam is:

$$P = \frac{\cos^2 2\theta \left| \cos^n 2\theta_m \right| + 1}{1 + \left| \cos^n 2\theta_m \right|}$$
(2.2)

Where θ is the diffraction angle from the sample, θ_m is the angle of diffraction from the monochromator crystal and n is a monochromator dependent power, that takes a value between 1 and 2, depending on the mosaicity of the monochromator.

These two corrections are applied almost invisibly in data reduction software. The parameters involved in their application do not change unless either the monochromator or the radiation source is changed.

Other corrections are usually applied with more user intervention. The details of the possible corrections are given below

c) The decay correction

Some crystals deteriorate during data collection as a result of the exposure to X-rays. This can cause a number of effects. The most serious is that raw intensity falls as a result of the crystalline material being destroyed. Decay occurs more rapidly along fault lines within crystals and so the mosaic spread also increases with exposure. In nearly perfect crystalline samples certain low angle reflections can increase in intensity as a result of this due to removal of extinction effects. Another direct result of the increase in mosaic spread is peak broadening.

There are various ways of dealing with decay. The simplest method is to fit a straight line function to describe the change of the standard intensities with time and then use the crystal exposure time for each reflection to correct the data point. More effective methods fit either a polynomial or an exponential function in the same way, since the decrease in standard intensities is rarely linear. One method (Ibers, 1969) refines a polynomial function during crystallographic least squares procedure.

d) The Absorption correction.

Absorption corrections, as the name would imply, correct for attenuation of the primary beam by the crystal. For a spherical crystal the effect can be disregarded since every reflection will be equally affected, but in the majority of crystals the shape is relatively anisotropic and so each reflection is attenuated to a greater or lesser degree depending on the path of the incident and reflected beams through the crystal. There are many methods for correcting structures for absorption. The most effective methods, for well defined crystals, require the crystallographer to establish the Miller indices of the faces which form the surfaces of the crystal. This data, along with the face dimensions can be used to calculate accurately the path length of a specific reflection in the crystal, either by analytical or numerical methods. This method should be used if the crystal has a high mass absorption coefficient.

The most common form of absorption correction uses the ψ scan. This is perfectly adequate for the majority of structures where a correction is necessary[†]. In a ψ scan the crystal is rotated around the diffraction vector and the reflection is measured at a variety of angle settings. The change in integrated intensity that is observed over the set of measurements is an absorption effect because, as the crystal is rotated, the path length through the crystal changes, hence this set of data can be used to empirically correct the data for absorption. Many methods exist that utilise this data: The AFC6S diffractometer uses the absolute scale of the variation in intensity of a number of ψ scans to correct the data (*AFC6S user's manual* ,1992).

Data collected on the P4 is usually corrected using XEMP (*SHELXTL* package, 1986). This approximates the crystal to an ellipsoid . It uses the ψ scan to adjust the ellipsoidal radii, to produce the best merging R-value for the ψ-scan data. This technique can also utilise equivalent reflections, although ψ scans are more advisable in this procedure. An equivalent reflection can be affected by other systematic unknowns within the data collection.

The original diffractometer software on the rotating anode machine only offers absorption corrections by indexing faces. Some new software has been written by the Author so that ψ scans can be collected and utilised via the program XEMP. This is particularly important for data measurement at low temperature where the sample is blind to the user on the rotating anode.

[†] This is a point often missed in many structural studies. With very low mass absorption coefficients a correction is probably a waste of time and may introduce a worse fit in the least squares model. One could compare the effect to adding a further random error to each reflection.

e) The extinction correction.

This is a correction that is usually only required in relatively perfect crystals. Two forms of extinction can be treated. The less significant of the two is primary extinction. Here a portion of the diffracted beam is diffracted a second time by the same set of Bragg planes. The exiting beam is thus attenuated by destructive interference by the phase lag that is induced. The second more common extinction effect, known as secondary extinction, occurs as a result of the crystal mosaic spread. In most crystals the mosaicity is too large for this to be of any significance, but when this is not the case some mosaic blocks are oriented identically relative to the incident beam. Because they are separate blocks, there is a phase difference in the diffracted beam from each block which causes destructive interference. Sometimes the direct effect of extinction can be observed if a suitably intense low angle standard is used. A diffraction experiment can increase the mosaic spread due to slight radiation damage. In this case the extinction is removed as the experiment progresses and so a steady increase in the standard intensity is observed as a function of time. The correction for extinction is usually applied during structural refinement. A parameter, x, is refined that fits a multiplying factor to the data set as below:

$$F_c^{2\prime} = k\left[1 + \frac{0 \cdot 001 x \lambda^3 F_c^2}{\sin(2\theta)}\right]^{-\frac{1}{4}} \qquad (2.3)$$

where λ is the wavelength of radiation used and k is the overall scale factor. This correction is empirical. It is a compromise that corrects approximately for the effect.

f) Thermal Diffuse Scattering

There is a portion of the measured reflection that does not result from coherent scattering. This component cannot be measured independently from the main Bragg peak and it cannot be modelled, since there is no direct relationship between the incident radiation and the scattered radiation, thus there is no way to predict the interference effects.

Correction for this is rarely applied in X-ray crystallography. This is because the relative quantity of incoherent scattering is very small compared to the coherent scattering. It is more important in neutron crystallography. Usually TDS corrections are only applied in very accurate structural determinations. The correction uses the background shape to estimate the total contribution to the intensity of each reflection (Blessing, 1987).

g) Multiple Reflection (The Renniger effect).

Multiple reflection is related to primary extinction. The difference is that multiple reflection implies that an independent set of Bragg planes scatter the diffracted beam as opposed to the same set of planes. This occurs when two reflections lie on the surface of the Ewald sphere at the same time. Usually Renniger effects can be detected when systematically absent reflections appear present, when certain strong reflections are weaker than they should be, or when weak reflections seem stronger. In most studies the effect is not considered. Reflections that manifest strong renniger effects can be measured accurately by changing the azimuthal angle at which they are measured.

2.4 Space group determination.

Following a diffraction experiment it is necessary to establish the symmetry properties of the crystalline lattice. Analysis of the "systematic absences" within the data set can reveal the required information. A systematic absence arises due to a specific symmetry element causing exact destructive interference between diffracted waves for specific reflections. For example a 2_1 screw axis along c can be described as making the following two positions equivalent within the lattice;

$$x, y, z \text{ and } -x, -y, z + \tfrac{1}{2} \qquad (2.4)$$

The effect on a particular reflection is easily calculated if we consider the Fourier expansion of the structure factor. For a given structure factor, $F_{(hkl)}$, we can write

$$F_{(hkl)} = \sum_{j=1}^{N} f_j e^{2\pi i(hx_j + ky_j + lz_j)} e^{-k(\sin^2 \theta)U_j} \qquad (2.5)$$

where N is the number of atoms, h, k and l are the Miller indices, θ is the Bragg angle, U is the isotropic thermal parameter and k is a constant which is equal to $8(\pi/\lambda)^2$.

If we consider the reflection 0,0,1 we can see that the summation simplifies to

$$F_{(100)} = \sum_{j=1}^{N} f_j e^{2\pi i(z_j)} e^{-k(\sin^2 \theta)U_j} \qquad (2.6)$$

which can be written as two components

$$F_{(100)} = \sum_{j=1}^{N} f_j \left\{ \cos(2\pi z_j) + i \sin(2\pi z_j) \right\} e^{-k(\sin^2 \theta)U_j} \qquad (2.7)$$

Expanding this in terms of the symmetry operators shown in (2.4) this becomes

$$F_{100} = \sum_{j=1}^{M} f_j \left\{ \cos\left(2\pi z_j\right) + \cos\left(2\pi\left[z_j + \tfrac{1}{2}\right]\right) + i\left[\sin\left(2\pi z_j\right) + \sin\left(2\pi\left[z_j + \tfrac{1}{2}\right]\right)\right] \right\} \qquad (2.8)$$

where the full summation is over M; all atoms that are not related by the 2_1 screw axis. Since $\cos(\theta) = -\cos(\theta + \pi)$ and $\sin(\theta) = -\sin(\theta + \pi)$, $F_{(001)} = 0$.

Using relationships such as these it is possible to determine the presence of some symmetry operators within the unit cell. (Systematic absences only occur as a result of translational symmetry elements). This is adequate for discerning the difference between space groups in most cases along with the Laue symmetry, except that some means is also required to determine whether a crystal is centrosymmetric or non-centrosymmetric.

The most widely used analysis in this respect is the utilisation of Wilson statistics (Wilson, 1942,1949). This is based upon the "normalised" structure factors, known as E-values. Wilson showed that the structure factors could be corrected for the effects of

thermal motion across the range of 2θ. He showed that to bring structure factors onto a relative scale one could use:

$$|E_{hkl}| = \frac{|F_{hkl}|}{\sqrt{\langle |F| \rangle^2_{\theta_{hkl}}}} \qquad (2.9)$$

where $\langle |F| \rangle^2_{\theta_{hkl}}$ is the mean value for all structure factors that have a scattering angle of θ_{hkl}. These "E-values" are independent of scattering angle, and thus distributions of intensities can be analysed.

If the distribution of atomic positions within the unit cell can be regarded as random the central limit theorem can be applied. The probability of a particular value of $|E|$ is different for centrosymmetric structures as compared to non-centrosymmetric structures, due to the differing properties of the imaginary component in each cases. The two distributions that arise from the central limit theorem can be written as

$$P_{\bar{1}}(|E|) = \sqrt{\tfrac{2}{\pi}} e^{(|E|^2/2)} \qquad \text{(centrosymmetric structures) (2.10a)}$$

$$P_1(|E|) = 2|E| e^{(|E|^2)} \qquad \text{(non-centrosymmetric structures) (2.10b)}$$

These distributions can be used to calculate theoretical mean values of functions of $|E|$ and can be used to predict the number of E-values that should theoretically be greater than a particular threshold. Some theoretically predicted values are shown in Table 2.1

Table 2.1. Theoretical values for the means of various functions of |E|. (Karle, Dragonette and Brenner, 1965)

	centrosymmetric	non-centrosymmetric
$\langle |E|^2 \rangle$	1.000	1.000
$\langle |E^2 - 1| \rangle$	0.968	0.736
$\langle |E| \rangle$	0.798	0.886
$\% |E| > 1.0$	31.7	36.8
$\% |E| > 2.0$	4.6	1.8

21

$\%|E| > 3.0$	0.3	0.01

With this information space groups can be determined unequivocally. Caution has to be taken with weak data sets. Both the above techniques rely on differences in intensity and so weak data can yield incorrect answers. In this respect space group frequency tables are of some value. Energetic considerations suggest that certain space groups are more favourable than others and so the distribution of space groups is non-random. Using this information, the most likely candidates can be established.

2.5 Structure Solution

In a conventional diffraction experiment[†] raw values of intensity are measured. Wave mechanics tells us that the raw intensity is related to the square of the wave amplitude. The electron density pattern can be revealed by taking the Fourier transform of these amplitudes for all *hkl* planes. Because of this there is an inherent problem in a diffraction experiment. Since we only have intensity measurements there is no experimental information on the relative phases of the wave amplitudes, hence we are unable to calculate the electron density distribution directly from the experimental data. This is the crystallographic phase problem and two methods have been proposed for its solution.

Patterson methods use the Fourier transform of the square of the modulus of the structure factor. This function corresponds to the self-convolution of the electron density. The strength of this method is that a Patterson function can be calculated without any phase information. However peaks in the Patterson map do not correspond directly to atomic positions but to inter-atomic vectors. The peak height on an absolute scale corresponds to the product of the atomic numbers of the elements that are linked by the vector. The most common application of this technique is in cases where a small number of atoms dominate the scattering. In such cases a few vectors between the heavier atoms are very prominent

---

[†] there are some examples in the recent literature of dual beam experiments where some phase information can be gained.

and so can be used to find the atomic positions of these atoms. A set of phases, based on the heavy atoms only, can be calculated and combined with the measured intensities to calculate an electron density map. which usually yields further atomic positions. Repeating this process of recalculating the phases from the developing model and then recalculating the electron density distribution should eventually reveal all non-hydrogen atomic positions.

Nowadays direct phasing methods are most commonly used. These use statistical relationships between reflections to establish an approximation to the phases of reflections by probabilistic methods. The presence of certain strong reflections in a data set suggests relationships between their phases (Sayre, 1952). In effect, if a small number of reflections have their phases established then the remaining phases can be predicted, with a "reliability" for each reflection based on its intensity. Provided an origin can be defined it is possible to calculate full sets of trial phases. This reduces the number of possible solutions tremendously and means that computers are able to try a small number of starting phases and generate probable solutions to the phase problem. A rigorous treatment of either of these methods would be beyond the scope of this thesis; both are already well documented in crystallographic texts (see for example Giacovazzo, 1992 and references therein).

2.6 Structural refinement.

2.6.1 The process of least squares

Structure solution provides an approximation to the reflection phases which in turn usually yield approximate atomic positions for most of the atoms within the lattice. In order to get a full structural picture it is necessary to create a more precise atomic description of the unit cell that not only models the positions of atoms but how they vibrate as well. In order to do this the atomic positions are refined by either Fourier or least squares methods. The method used for refinement in this thesis was exclusively the "least squares method". In this, the fitted parameters are refined to minimise the weighted sum of the squares of the deviations between the observed and calculated structure

factors. To assess the quality of the final overall fit, two residuals are used, the Goodness-of-fit (*Goof*) and the weighted R-factor (wR_2), both of which are shown below:

$$wR_2 = \left[\frac{\sum \left\{ w \left(F_o^2 - F_c^2 \right)^2 \right\}}{\sum \left\{ w \left(F_o^2 \right)^2 \right\}} \right]^{\frac{1}{2}} \qquad (2.11)$$

$$Goof = S = \left[\frac{\sum \left\{ w \left(F_o^2 - F_c^2 \right)^2 \right\}}{(N - P)} \right]^{\frac{1}{2}} \qquad (2.12)$$

where N is the number of reflections and P is the number of parameters refined. In a perfect fit wR_2 will in theory tend approximately towards a parameter known as R_σ, which is the sum of the reflection sigmas divided by the sum of the observed structure factors. The *Goof* will tend towards unity.

$$R_\sigma = \frac{\sum \sigma \left(F_o^2 \right)}{\sum \left(F_o^2 \right)} \qquad (2.13)$$

The principle of least squares is based around the postulate of a Gaussian distribution of errors within the data set. In order to ensure that this is true, a reasonable weighting scheme has to be chosen to correct for inadequacies in the data reduction process. The variance in the weighted residual across regions of data separated both in 2θ angle and intensity indicate the quality of a weighting scheme. The more consistent across all regions of data, the more reasonable is the weighting scheme.

In the past few years there has been some debate in the crystallographic community as to the way in which structures should be refined. Some schools of thought encourage discarding of the weak reflections due to their unreliability, whereas others believe that this approach introduces an unnecessary statistical bias into refinement. In this thesis all refinements have been conducted using the SHELX93 program (Sheldrick, 1993) which

uses all data in refinement, and refines on F^2 as opposed to F. The authors decision here is not for any particular statistical reason, however. The reason for the approach is two-fold. Firstly, general familiarity with the old program SHELX76 leads logically on to using the new version and secondly, there are now available within SHELX93 a wide variety of useful restraints and constraints, as well as more sophisticated methods for refining hydrogen positions.

2.6.2 Interpretation of results: precision.

Least squares refinement provides values for the refined parameters, together with estimates of their precision, which come directly from the variance-covariance matrix. This added information arises due to the over-determinacy within a structural study: we have many observational equations compared to the number of parameters that we refine, so effectively a distribution of possible answers exists, since in an ideal data set, where all reflections fitted the structural model perfectly, only "N_p" reflections would be needed to derive the correct values for "N_p" parameters of the structural model. Since the number of reflections measured is much greater than the number of parameters in the structural model, the result can, in principle, be derived many times over, giving a distribution of possible values for the derived parameters.

This idea is directly analogous to the statistical theories that govern distributions: the very foundation of the least squares method relies on the uncertainties in reflections following a Gaussian distribution and the same postulate now exists here. Because of this it is worth outlining a few concepts of statistics.

If we measure a quantity (say x) on N separate occasions we have a "distribution". As the population of this distribution becomes large the variation in the quantity can be plotted and the values will follow a normal, or Gaussian distribution. It is possible to write a function that describes this distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(-(x-\bar{x})^2/2\sigma^2\right)} \qquad (2.14)$$

where $\left\{\bar{x} = \sum\limits_{i=1}^{N} x_i\right\} \Big/ N,$ i.e. the mean of the values within the distribution, and σ^2 is the variance of the distribution. This is governed by the shape of the distribution. If a wide range of values is observed in the distribution then this value becomes large, so the size of the variance directly relates to the range of values that have been observed and, as such, is a measure of the confidence that we can have in stating the mean value of all of the observations as the true value of the quantity being measured.

In the process of least squares the variance-covariance matrix of the parameter estimates is revealed within a scale factor. It has to be estimated before the variances of the parameters can be derived correctly. This, fortunately, is available in the least squares process. The estimated standard deviation[†] (e.s.d.) of the jth parameter p_j, can be expressed as

$$\sigma_{p_j} = \left(\left(A^{-1}\right)_{jj} \frac{\sum\limits_{i=1}^{N_r} w_i \left(F_{o,i}^2 - F_{c,i}^2\right)^2}{N_r - N_p} \right)^{\frac{1}{2}} \qquad (2.15)$$

Where A^{-1} is the variance-covariance matrix that is calculated during least squares, w_i is an appropriately derived weight, $F_{o,i}^2$ is the square of the observed structure factor for the ith reflection and $F_{c,i}^2$ is the calculated value. The number N_r is the number of reflections and N_p is the number of parameters refined. (The above equation is shown for refinement on F^2 rather than F). Note here that the calculated scale factor is shown in full in equation 2.15: it corresponds to the goodness-of-fit given above (*Goof*).

It is important, when estimating the precision of a parameter, that the full matrix is used. A more fundamental point is that, when calculating the e.s.d. for a derived quantity (such

---

[†] The IUCr commission on crystallographic nomenclature (1994) has suggested that a more appropriate term for the standard deviation would be the "standard uncertainty" however this has not come into common usage so in this thesis the original terminology has been used.

as a bond length), the correlation between the parameters used in deriving the quantity have to be taken into account. To do this the full variance-covariance matrix is required. In situations where the correlation is not accounted for, the e.s.d's tend to be systematically underestimated. The equation for the σ of a derived parameter is shown below:

$$\sigma^2\left(f(p_1 \cdots p_n)\right) = \sum_{i,j=1}^{n}\left(\frac{\partial f}{\partial p_i} \cdot \frac{\partial f}{\partial p_j} \cdot \left(A^{-1}\right)_{ij} \cdot \frac{\sum_{k=1}^{N_r} w_k (F_{o,k}^2 - F_{r,k}^2)}{N_r - N_p}\right) \qquad (2.16)$$

Having derived the e.s.d. in a parameter correctly it is important that we should understand its meaning. In a real sense the e.s.d. allows us to specify a degree of confidence in our results; a commonly heard figure within crystallography is the magical "3σ". The interpretation of this number is that a measurement of a quantity q will lie within the boundaries of q$\pm$ 3σ in over 99% of occasions of its being measured. This tends to be regarded as a reasonable level of confidence, although in some cases higher cut-offs are imposed. One question often posed in the chemical interpretation of a particular crystallographic quantity is: "Are these two values of a particular quantity (e.g. two bond lengths) significantly different?". The simple answer to this question is that if the two ranges of possible values defined by the σ cut-off for each measurement overlap then the two quantities should not be regarded as significantly different.

2.7 Presentation of result in Chapters 3 and 4.

In the analysis of the results of the majority of the structural studies the "OFIT" option in the SHELXTL XP package (Sheldrick, 1986) has been used to study the structural changes induced by chemical change. This program takes two independent sets of orthogonal coordinates and superposes them. The relative positioning of the two models is derived by a least squares fit between a number of pairs of atoms, one atom in each pair coming from each model. The XP package allows the user to generate *crystal* coordinates for this superposition and so relative changes can be quantified. An example of the use of

this features would be a fit of the cage atoms of the dyotropic reactant and product. as described in section 3.3.

Figure 2.3 Angles defined for inter-molecular bonding tables
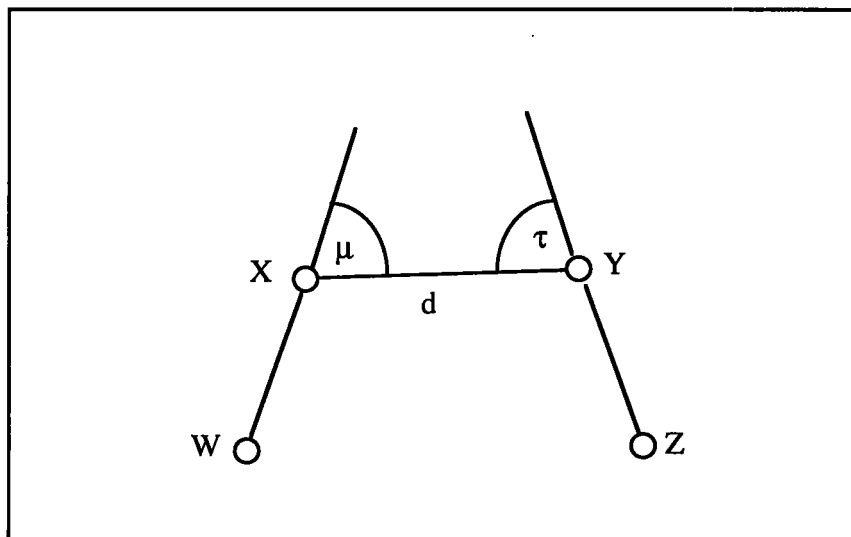
Tables of relevant inter-molecular contacts for each molecule are all presented in a fixed format for contacts which do not involve hydrogen. The first column is the non-bonded distance (d in Figure 2.3), the second two being the two "polar" angles (τ and μ in Figure 2.3)

Here, μ is defined as the highest polar angle and τ as the smaller one. This definition is the same as that used by Nyburg and Faerman (1985). It should be noted that in this definition small angles correspond to linear interactions. The torsion angle defined by the two bonds is not recorded; Nyburg and Faerman found that this had no influence on the non-bonded patterns that they observed in systems where like-like close contacts exist.

Close contacts involving hydrogens are shown in separate tables. In most cases two distances are shown, one using the refined position and one where the H atom is placed in an ideal geometric position at a fixed distance from its parent atom. This is referred to as the "idealised" distance. The reason for this approach is because the H positions were of relatively low precision.

For each compound a short comment is given that compares and contrasts the individual compound with other similar examples. In the discussion at the end of the chapters any general trends across all of the compounds are discussed.

2.8 References

AFC6S user's manual , Distributed by the MSC-Rigaku Corporation, 1992

Arndt, U. W.; Willis, B. T. M.; *Single Crystal Diffractometry*, Cambridge University Press, 1966

Blessing, R. H.; *J. Appl. Cryst.*, **194**, 412, 1986

Buerger, M. J.; *Zeitschrift für Kristallographie*, **109**, 42, 1957

Buerger, M. J.; *Zeitschrift für Kristallographie*, **113**, 52, 1960

Fundamentals of Crystallography, (IUCr text on crystallography No 2). Published by Oxford University Publications, edited by Giacovazzo, C., 1992

Harding, M. M.; Kariuki, B. M.; Cernik, R.; Cressey, G.; *Acta Cryst.*, **B50**, 673-676, 1994.

Ibers, J. A. ; *Acta Cryst.*, **B25**, 1667 1969

Karle, I. L.; Dragonette, K. S.; Brenner, S. A.; *Acta Cryst*, **19**, 713, 1965

Niggli, P.; *Handbuch der Experimentalphysik*, **Vol. 7(1)**, Academische verlagsgesellschatt, Leipzig, 1982

Nyburg, S.C.;Faerman, C.H, *Acta Cryst.*, **B41**, 274-279, 1985

Report of the IUCr Subcommission on Statistical Descriptors, *Acta Cryst.*, **A45**, 63-75, 1989

Santoro, A.; Mighell, A. D.; *Acta Cryst.*, **A26**, 124, 1970

Sayre, D.; *Acta Cryst.*, **5**, 60, 1952

Stalke. D.; *J. Appl. Cryst.* **26**,615,1993

Wilson, A. J. C.; *Nature*, **150**,152,1942

Wilson, A. J. C.; *Acta Cryst.*, **2**,318,1949

XPREP Part of the Shelxtl package distributed by SIEMENS. Written by Sheldrick, G.M. 1986

Chapter 3.

Structural studies of a number of potential rearrangement complexes

3.1 Introduction.

The majority of chemical reactions studied are of an intermolecular nature. Two or more compounds, usually solvated, are forced, by manipulation of physical conditions, to react with one another when they meet. The kinetics of the process are dependent upon concentration terms in the solution of the reactants. To describe this process we use the rate law. In the reaction:

$$A + B \xrightarrow{\ k\ } AB$$
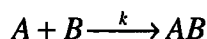
the rate law is expressed as:

$$\frac{d[AB]}{dt} = k[A][B] \qquad (3.1)$$

where $[A]$, $[B]$ and $[AB]$ are the concentrations of the reactants and product and k is the rate constant. This dependency on concentration suggests a relationship between reaction rate and the likelihood of two molecules approaching one another, a factor that is tentatively accounted for in thermodynamics by the entropy. In intramolecular reactions the problem of molecular approach is significantly reduced; the reacting parts within the molecule are already in relatively close proximity, and so the entropic term is less significant. However there is evidence that the rate enhancement which is seen in intramolecular reactions, compared to their intermolecular analogues, cannot be accounted for by the reduced entropy term alone.

As a result of the considerable enhancement of reaction rate in intramolecular reactions, a great deal of potential exists for their application in new synthetic routes (see for example Craig, 1987; Fallis, 1984; Ciganek, 1984). Another synthetic benefit of such reactions is that they are particularly "clean": they do not show side reactions. This is probably due to the orientational control that is inherent in the molecule itself; there is no possibility for ambiguity in reaction due to different modes of approach. Also, it has been suggested (Menger, 1985) that enzyme catalysed reactions in biological systems may well be mimicking intramolecular reactions, by applying orientational control to reactants in an A+B type of reaction. Thus, detailed analysis of structural effects on reaction kinetics for intramolecular reactions is of considerable interest.

Many theories exist that attempt to explain fully the rate enhancement which is usually seen in enzyme catalysed reactions. This also relates directly to intramolecular systems The earliest ideas tended to be subtly different ways of interpreting the influence of molecular orientation on reaction rate, (see for example Storm, 1970; Milstein, 1972; Menger, 1985; Dorigo, 1987), although a great deal of uncertainty exists as to the correct reason for rate enhancement, since the independent studies described in the literature above seem to contradict one another. A principal reason for this situation is probably a lack of understanding of the form of the transition state, which, of course, will be different in each different reaction.

To resolve this situation, the most reasonable approach is to study a series of chemically and structurally related compounds which all undergo a specific reaction. The varying rates of reaction can then be rationalised against the changes in 3D structure from one reactant to the next and hopefully this can be used to interpret the mechanism of a specific reaction.

Figure 3.1. Dyotropic Rearrangement in isodrin type derivatives.

(a) → heat → (b)
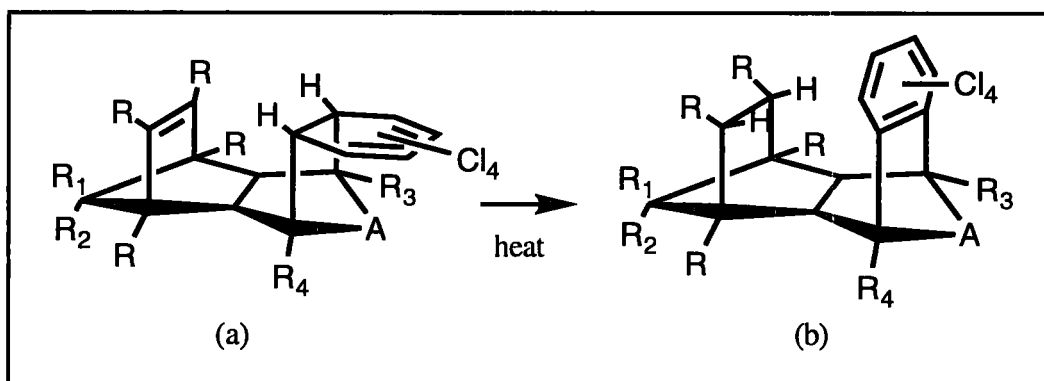
In Durham we have been studying, in collaboration with Dr K. Mackenzie at the University of Bristol, a series of compounds of the type shown in Figure 3.1(a), which are derived from the isodrin moiety.

These molecules undergo a "dyotropic" rearrangement to a triene structure (Figure 3.1b). Dyotropy can be defined as follows:

"An uncatalysed process in which two σ-bonds simultaneously, but not necessarily via a fully synchronous mode, migrate intramolecularly" (Reetz, 1973)

The reaction here is best regarded as a $[4\sigma + 2\pi]$ rearrangement (Mackenzie 1965, 1969) through a "pericyclic" intermediate. Similar rearrangements have been reported for $[4\sigma + 2\pi]$ reactions in similar isodrin complexes. (Vogel et al, 1981, Paquette et al, 1990, 1991). Recently $[4\sigma + 6\pi]$ reactions have also been reported (Grimme, 1992). Kinetic studies have revealed that a wide variety of reaction rates can exist in these systems as a result of fairly minor structural changes in the substituent groups, which seem fairly remote from the reaction cavity. Mackenzie and co-workers (1993) reported an example of the system (Figure 3.1, R, R_1, R_2, R_3, R_4=H, A=CH_2) where the reaction progressed so fast that the diene, 3.1(a), could not be isolated. (The rate constant was estimated at $4 \times 10^{-2}s^{-1}$). They have since noted that remarkable attenuation in this reaction rate can be achieved by manipulation of the substituent groups. Chlorine atoms in the "R" positions in Figure 3.1(a) particularly

retard the reaction rate, and thus chlorine substituted derivatives provide useful model systems in which both isomers 3.1(a) and 3.1(b) can be isolated in order to study this reaction.

In this chapter the structures of a number of novel examples of substituted isodrin derivatives are reported. The purpose of this work was to establish the shape of the reaction cavities, with particular reference to the proximity of the bridging carbon atoms in these compounds.

Structural diagrams are included with each molecule in the respective table of experimental details. Also a bookmark card is provided with structural diagrams and chemical names for the readers' reference.

3.2 The structures of two precursors to dyotropic reactants.

There is some interest in the precursors to the fully described dyotropomers, since intimate structural detail can be elucidated which yields information on the structural cavity, and thus can suggest whether further synthetic efforts are worthwhile. The first two compounds are examples of crystalline precursors to potential dyotropomers.
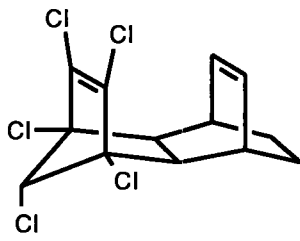
3.2.1 Structure (1).

Structure (1), shown in Figure 3.2 (chemical diagram and experimental details given in Table 3.1), is a penta-chloro analogue of homo-isodrin, a known compound that is a potential insecticide. Structure (1) crystallizes with two molecules in the asymmetric unit. The structure of homo-isodrin has already been elucidated by two independent groups (Dong *et al*, 1984; Kennard and Smith, 1984). A comparison of homo-isodrin with structure (1) is clearly worthwhile. Kennard and Smith's study of homo-isodrin appears to be more precise than Dong's since the R-factor and "*sigma flag*[†] " are both

---

[†] The "*sigma flag*" indicates the mean e.s.d of C-C bonds for a particular structure on the CSD

Table 3.1. Experimental details for structure (1)

Chemical Details	
Empirical formula: $C_{13}H_{11}Cl_5$	Chemical Diagram
Formula weight: 344.5	
Diffractometer Experimental details	
Unit Cell	Data collection parameters
a = 12.147(2) Å	Crystal Size:0.43 x 0.45 x 0.48 mm
b = 13.934(2) Å	Collection method: $\omega/2\theta$ scans
c = 8.830(2) Å	Speed of collection: 2-8°/min.
α = 91.16(2)°	Scan width (2θ): 1.47 + 0.3$\tan\theta$
β = 92.78(2)°	θ range for collection: 2.74 to 29.38°
γ = 110.41(2)°	Index ranges :
Volume = 1397.9(4)Å^3	$0 \leq h \leq 15$
Based on 23 reflections in the range	$-19 \leq k \leq 17$
32.5-35.1° 2θ	$-12 \leq l \leq 12$
	Temperature: 293(2)K
Refinement details	
Space group: P$\overline{1}$	Z = 4
ρ_c= 1.637 g/cm^3	Absorption coefficient: 1.015 mm^{-1}
F(000): 696	Goodness-of-fit on F^2: 1.025
Reflections collected: 6872	Final R indices [I>2σI)]
Independent reflections: 6531	R1 = 0.0359 | wR2 = 0.0874
R_{int} = 0.0421	R indices (all data)
Data: 6528	R1 = 0.0610 | wR2 = 0.1037
Restraints: 0	Largest diff. peak and hole
Parameters : 344	0.361 and -0.363 eÅ^{-3}

Hydrogen Treatment: Hydrogens were held fixed at idealised positions relative to their parent carbon atom during refinement. Their isotropic temperature factors were held fixed at 1.2 × their parent carbon atom.

Figure 3.2, Atomic representation of molecule (1) of structure (1).
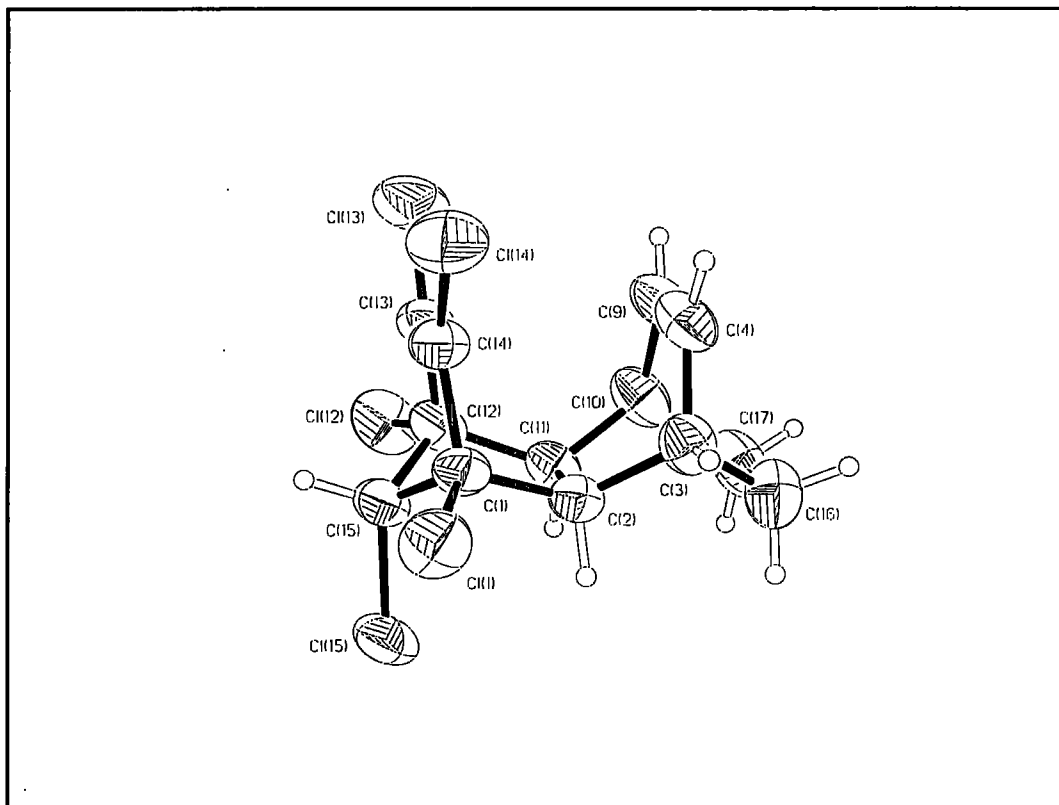
lower than in the first study. The cell parameters also differ from one study to the other: a difference of 0.01Å exists in the reported length of the **a** axis

The more precise structural study of homo-isodrin was used for comparison with structure (1). The structure of homo-isodrin is pseudo-tetragonal, $\mathbf{a} \approx \mathbf{c}$. The average inter-cavity C..C contact in this system is 2.947Å, some 6σ different from the mean of $2.962(2)\text{Å}$ observed in structure (1). (See Table 3.2)

This may not be significantly different, for reasons that are explained in Chapter 2, but unfortunately no estimate of the e.s.d. for the homo-isodrin contact is easily available. The change is consistent with the chemistry of the system; one would expect the presence of the extra chlorine atom in homo-isodrin to force the two ethene bridges fractionally close together, for steric reasons.

35

Table 3.2 Intra-cavity distances within structure (1)

Contact	Distance (Å)
Molecule 1	
C(4)-C(14)	2.958(4)
C(9)-C(13)	2.976(4)
Molecule 2	
C(54)-C(64)	2.957(3)
C(59)-C(63)	2.956(3)
Mean	2.962(2)

Figure 3.3 OFIT plot of homo-isodrin fitted with molecule (1) of structure (1)

All Carbon atoms are fitted, homo-isodrin is dotted
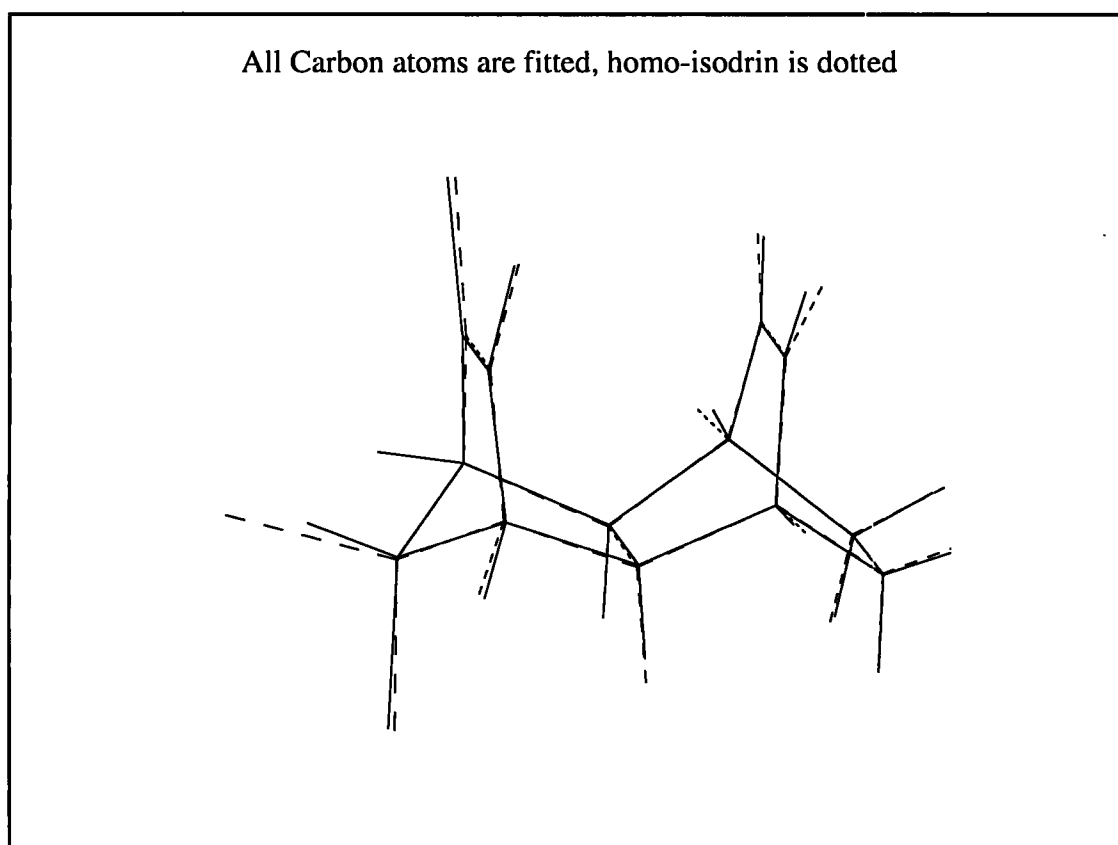
Figure 3.4. Definition of the Chlorine-ethene and Bridgehead-ethene planes.

Bridgehead-ethene plane

Chlorine-ethene plane
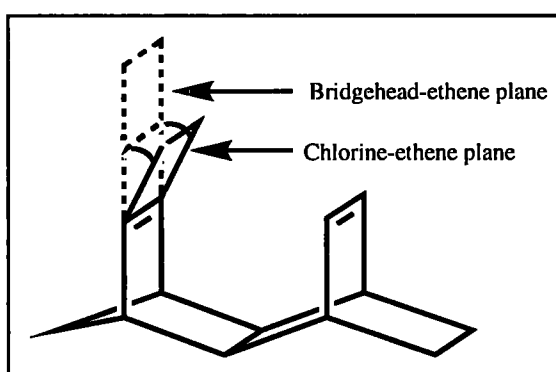
One puzzling effect in structure (1) is the asymmetric distribution of the intramolecular distances across the cavity in one of the two independent molecules present in the asymmetric unit, (see Table 3.2) although this may be a statistical quirk. (See Chapter 2). There is very little difference between molecule (1) and molecule (2), as can be seen from the bond length and angle tables given in Appendix 3A. The slight differences in the bond lengths and angles are probably insignificant.

In both molecules the ethene bridged chlorine atoms are pointing in towards the second ethene bridge. This is easily seen by considering the inter-planar angle as defined in Figure 3.4 The inter-planar angle is the angle between the "bridgehead-ethene" plane normal and the "chlorine-ethene" plane normal.

The mean value of the inter-planar angle over the two independent molecules is 6.2(2)°. This is identical to that observed in homo-isodrin. This shows that the added chlorine atom has very little effect on the inter-planar angle described above, a logical conclusion since the distance between this chlorine and the ethene bridged chlorine atoms is 4.238Å, well above the range of short interactions that one infers from van der Waals radii. (V(Cl) = 1.80Å from *The handbook of chemistry and physics*, Weast, 1984). This confirms the conclusions that one can make from Figure 3.3.

The inference that can be made from all the above evidence is that increasing the substitution at the methylene bridge has little steric effect on the structural cavity. If any rate enhancement is seen in dyotropomer derivatives of these compounds then this rate enhancement probably cannot be attributed to changes in the steric conditions within the skeleton. The individual dyotropomers would need analysis for conclusive proof.

The packing in structure (1) is interesting. A number of contacts exist that are close to the sums of the van der Waals radii as described by Nyburg and Faerman (1985). The packing diagram shown in Figure 3.5(a) and (b) shows these interactions. They are tabulated in Tables 3.3 and 3.4. The two molecules in the asymmetric unit are clearly

Figure 3.5(a). Herring bone patterns in structure (1)
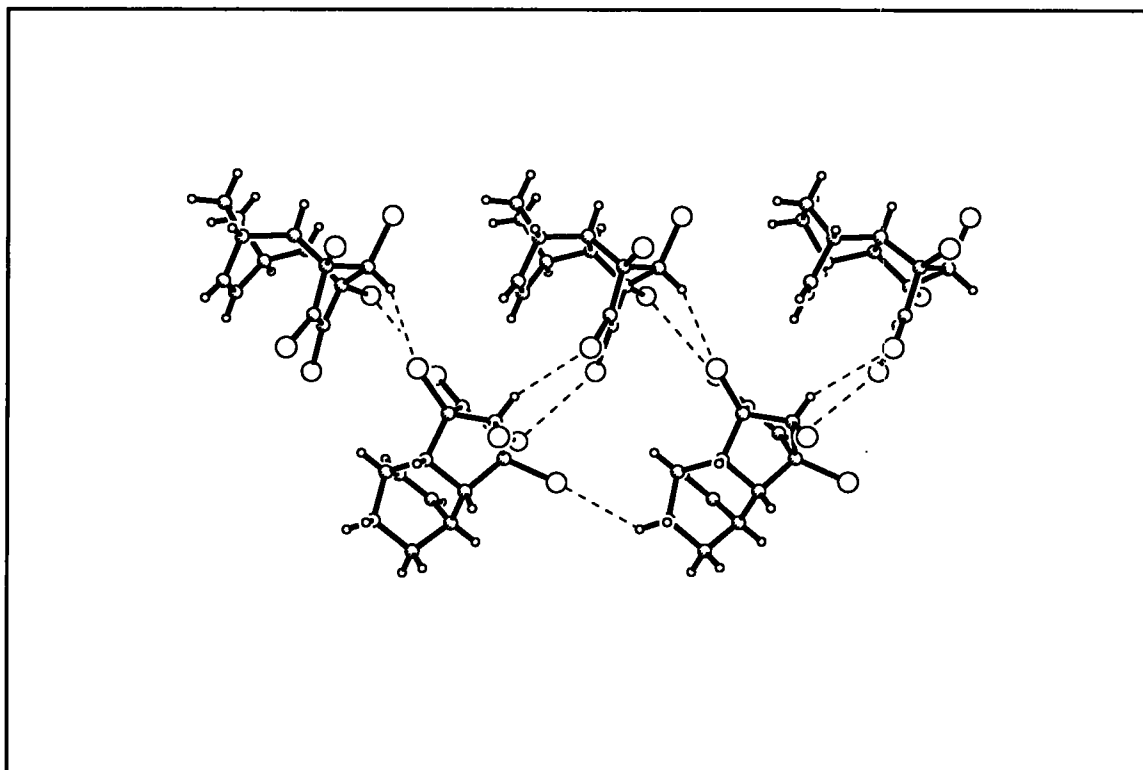
Figure 3.5(b). Layer packing in (1)
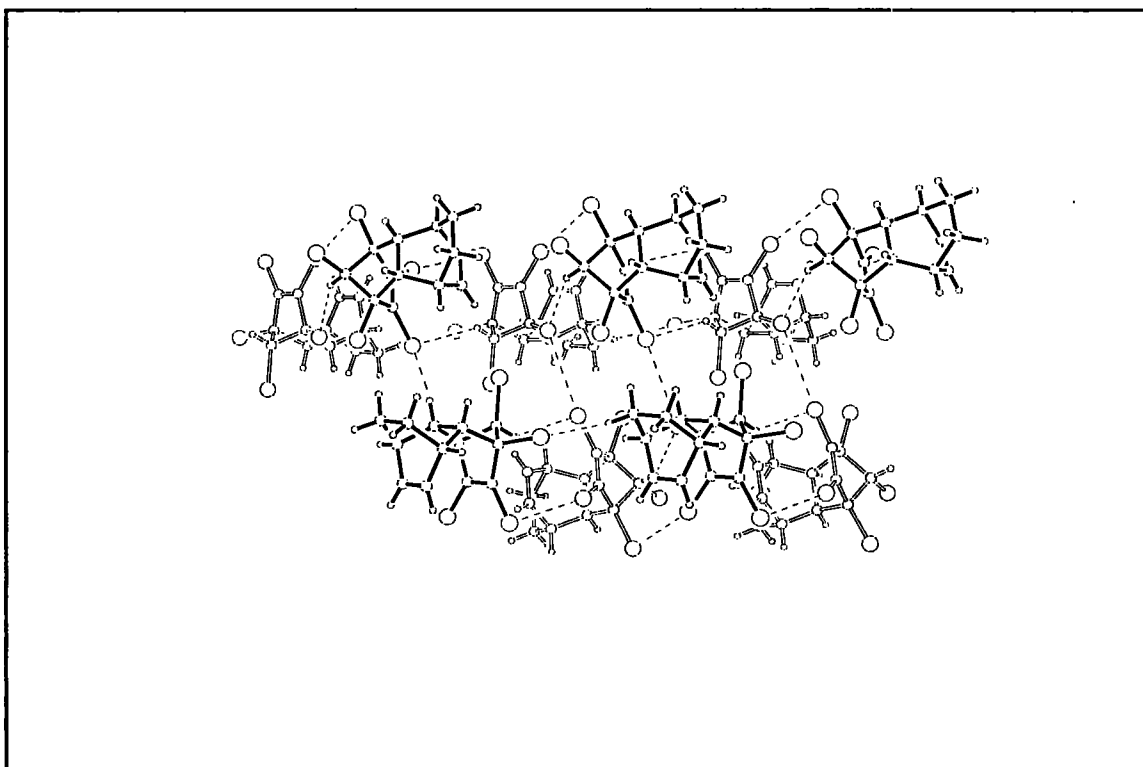
Table 3.3. Selected close intermolecular contacts in structure (1)

Contact	Distance (Å)	μ (°)	τ (°)
Cl(1) (2-x,-y,1-z)...Cl(63)	3.472(1)	82.36(9)	25.17(10)
Cl(13)...Cl(62)	3.537(1)	60.22(9)	33.05(11)
Cl(14) (2-x,-y,2-z)...Cl(64)	3.571(1)	99.88(8)	27.13(10)

Table 3.4. Contacts involving hydrogen for structure (1)

Contact	Distance	Idealised distance[†]
Cl(13)...H(65) (2-x,-y,2-z)	2.89(17)	2.89
Cl(51)...H(67B) (x,y,z+1)	2.89(15)	2.90
Cl(62)...H(15) (2-x,-y,1-z)	2.88(19)	2.91
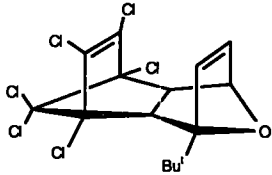
shown to be maximising certain contacts in the lattice in Figure 3.5(a). They form a herring bone like pattern, forming a pseudo-2_1 axis. 2_1 translation axes have been shown to be more energetically favourable than molecular translation alone (Filippini and Gavezzotti, 1992). These layers are stacked as shown in Figure 3.5(b) with one close inter-layer contact.

3.2.2 Structure (2)

Molecule (1) of structure (2) is shown in Figure 3.6. Like structure (1) this compound crystallizes with two molecules in the asymmetric unit. (A chemical diagram and experimental details are given in Table 3.5). The change in cavity shape as a result of the chemical changes to the molecule is of interest here: one might expect the ^{t}Bu group to induce asymmetry within the cavity as a result of its location. The cavity separations are shown in Table 3.6. The distribution is similar to that observed in structure (1), and does not show the marked asymmetry that might have been expected. Only one of the independent molecules has distinctly different separations, and even these are within 6σ of one another, in the same way as observed in (1). The mean difference is significantly smaller than that observed in (1) and approximately the same as the mean distance

---

[†] The idealised distance is defined as the intermolecular Cl-H distance that is measured when the hydrogen atom is placed at an idealised distance and geometry from its parent atom.

Table 3.5 Experimental details for structure (2)

Chemical Details	
Empirical formula: $C_{15}H_{14}Cl_6O$	Chemical diagram
Formula weight: 422.96	

Diffractometer Experimental details	
Unit Cell	Data collection parameters
$a = 10.621(2)$Å	Crystal Size: $0.25 \times 0.25 \times 0.4$ mm
$b = 13.730(3)$Å	Collection method: $\omega/2\theta$ scans
$c = 13.885(3)$Å	Speed of collection: 2-16°/min
$\alpha = 111.57(3)°$	Scan width: $1.26 + 0.3\tan\theta$
$\beta = 110.07(3)°$	θ range for collection: 2.77 to 26.99 °
$\gamma = 90.53(3)°$	Index ranges :
Volume=$1747.3(6)$Å^3	$-13 \leq h \leq 13$
Based on 19 reflections in the range	$0 \leq k \leq 17$
20.9-32.8° 2θ	$-17 \leq 1 \leq 16$
	Temperature: 150(2) K

Refinement details		
Space group P$\bar{1}$	Z=4	
ρ_c=1.608 g/cm^3	Absorption coefficient: 0.980 mm^{-1}	
F(000): 856	Goodness-of-fit on F^2: 1.039	
Reflections collected: 7918	Final R indices [I>2σI)]	
Independent reflections:7594	R1 = 0.0373	wR2 = 0.0814
R_{int} = 0.0249	R indices (all data)	
Data: 7592	R1 = 0.0819	wR2 = 0.0958
Restraints 0	Largest diff. peak and hole	
Parameters 419	0.410 and -0.356 e.Å^{-3}	

Hydrogen Treatment: C-H bond lengths were refined. X-C-H angles were constrained to ideal values. Non-methyl hydrogen temperature factors were constrained to 1.2 times the value of U_{iso} of their parent atom. Methyl hydrogens had a value of 1.5 times the U_{iso} of their parent atoms. The methyl hydrogen positions were determined from maxima in the Fourier map of a ring at an idealised distance from the parent carbon atom

Figure 3.6. A computer generated representation of structure(2)
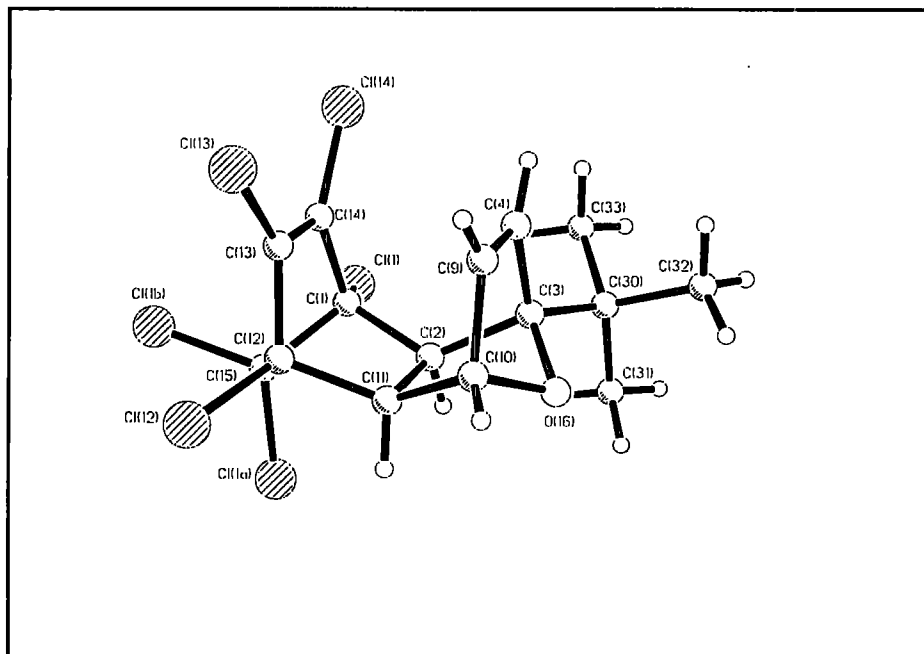
Figure 3.7. "OFIT" of the relative coordinates of Isodrin with structure.(2)

All skeletal C atoms are fitted. Isodrin is represented by a dotted line

41
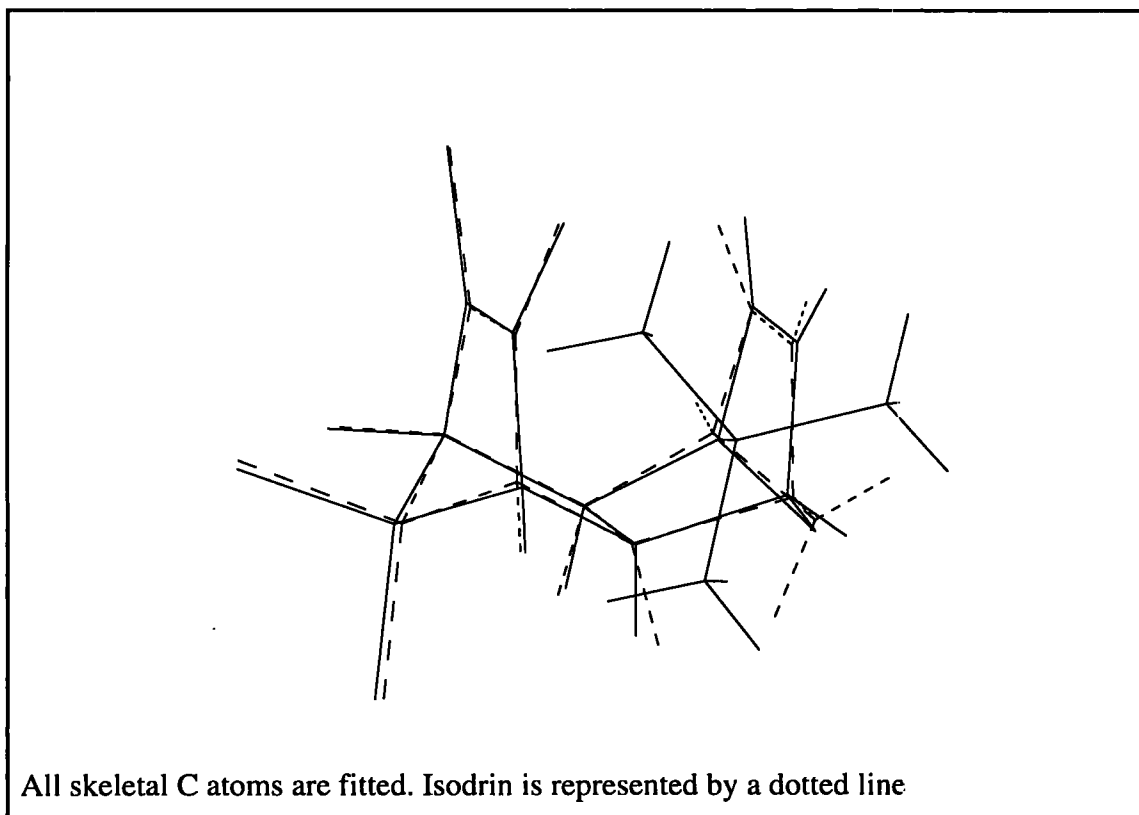
Table 3.6 Intra-cavity distances in structure (2)

Contact	Distance
Molecule 1	
C(4)...C(14)	2.951(4)
C(9)...C(13)	2.936(4)
Molecule 2	
C(54)...C(64)	2.965(5)
C(59)...C(63)	2.954(4)
Mean	2.952(2)

Table 3.7(a). Selected intermolecular contacts for structure (2).

Contact	Distance	μ	τ
Cl(1B)...Cl(6A) (1-x,1-y,-z)	3.488(1)	53.6(1)	21.7(1)
Cl(1B)...Cl(14) (1-x,2-y,1-z)	3.489(1)	57.3(1)	37.2(1)

Table 3.7(b) Selected intermolecular: contacts involving hydrogen atoms for structure (2)

Contact	Distance	Idealised distance
Cl(62)...H(81C) (x-1,y,z)	2.8(2)	2.74
O(16)...H(11) (-x,2-y,-z)	2.45(10)	2.36
O(66)...H(61) (1-x,1-y,-z)	2.50(9)	2.42

in homo-isodrin (2.947Å). The structure of isodrin is also known (Kennard *et al*, 1979). In this compound the mean distance is 2.87Å. The original literature suggests that the precision of this measurement is relatively low; the mean σ(C-C) for the structure is over 0.02Å, although this does not explain the general difference.

An OFIT superposition of isodrin with structure (2) is shown Figure 3.7. This shows that the oxygen atom in (2) is folded away from the methylene bridge position, which may explain the slightly longer intramolecular distances in (2) relative to isodrin. The mean angle between the chlorine-ethene and bridgehead-ethene planes (as defined

Figure 3.8. Packing in structure (2)
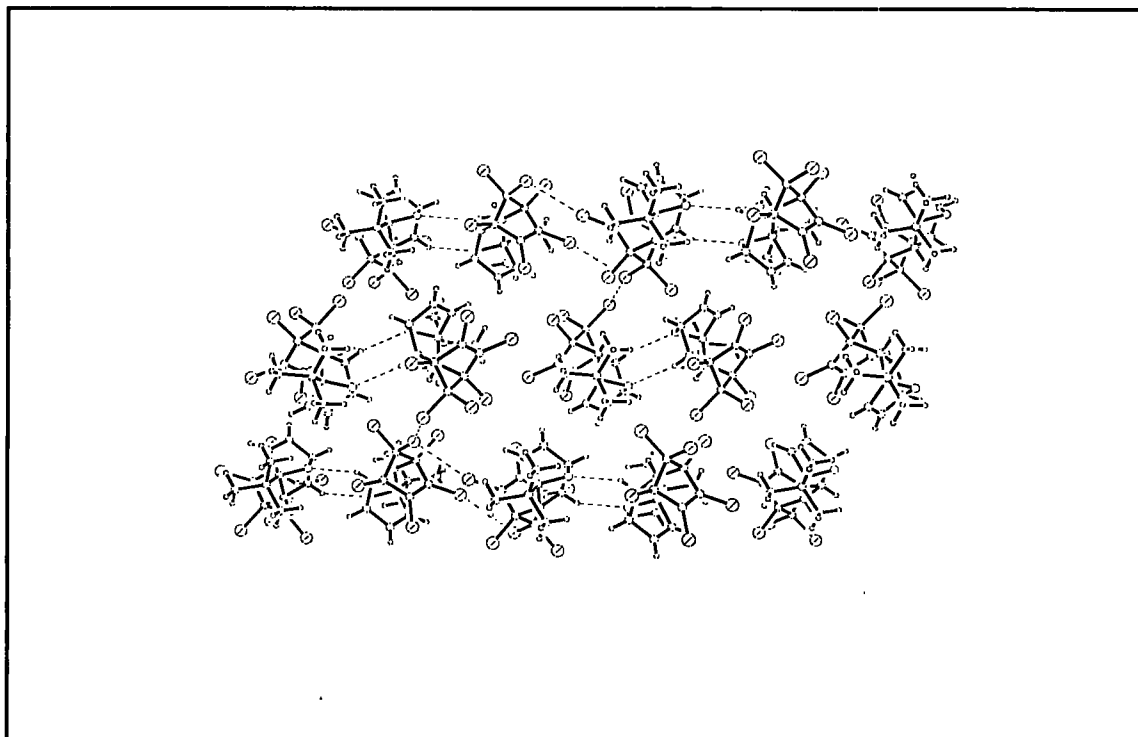
in Figure 3.4) is smaller than in structure (1) at 3.8(2)°. The calculated value in the structure of isodrin is 1.1° but this will be subject to a relatively large uncertainty.

The packing in structure (2) is shown in Figure 3.8. The closest contacts that are highlighted are given in Tables 3.7(a) and (b). The packing is somewhat different from that observed in (1). The ether oxygen atoms are utilised; both independent molecules form independent dimer pairs around inversion centres. These independent dimers then form a close-packed arrangement.

3.3 Studies on dyotropic systems

3.3.1 The dyotropic pair (3) and (4).

Structures (3) and (4) (see Figures 3.9(a) and (b) and Tables 3.8 and 3.9) form a dyotropic pair and are considered together. The intramolecular C-C distances observed for these two compounds are similar to those seen in other examples of these systems

Figures 3.9(a) and (b). Dyotropomers (3) and (4)

(a)

(b)
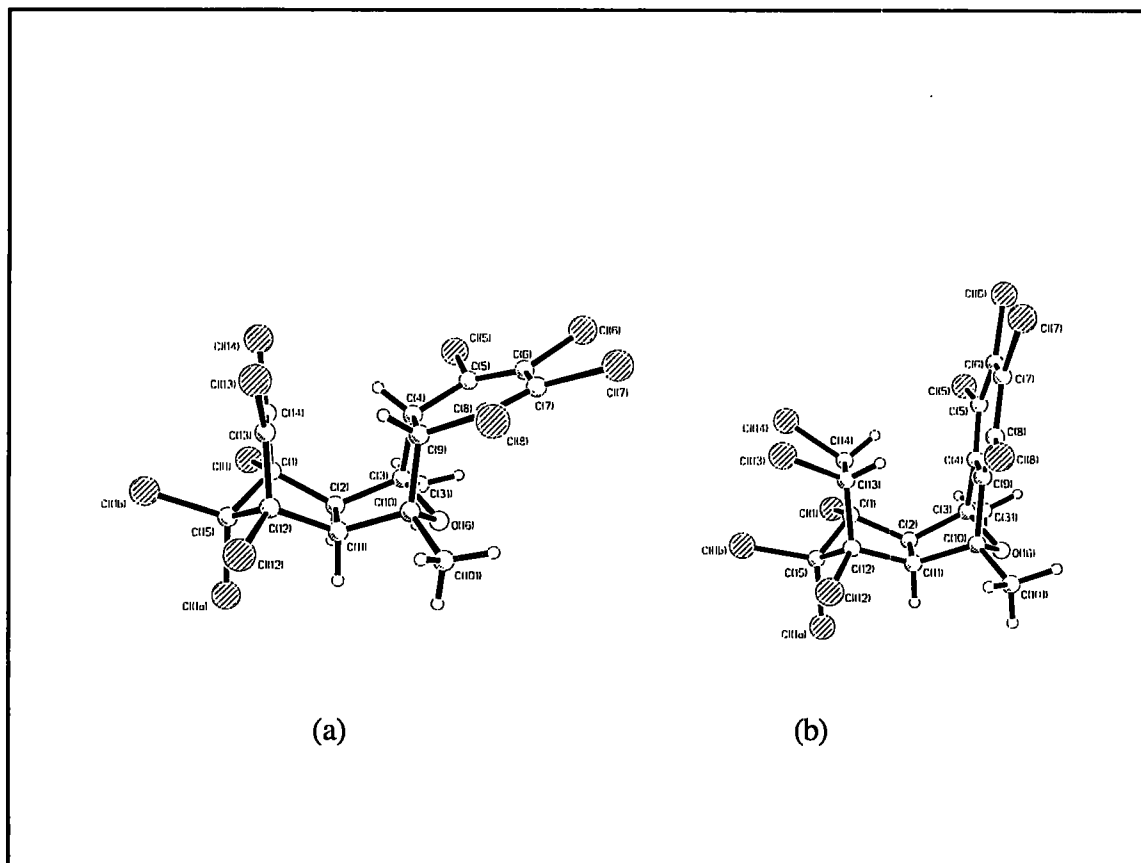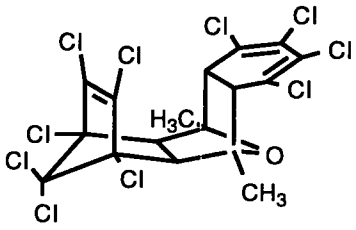
Figure 3.10. Structural representations of HMW1(a) and (b).

(a)

(b)
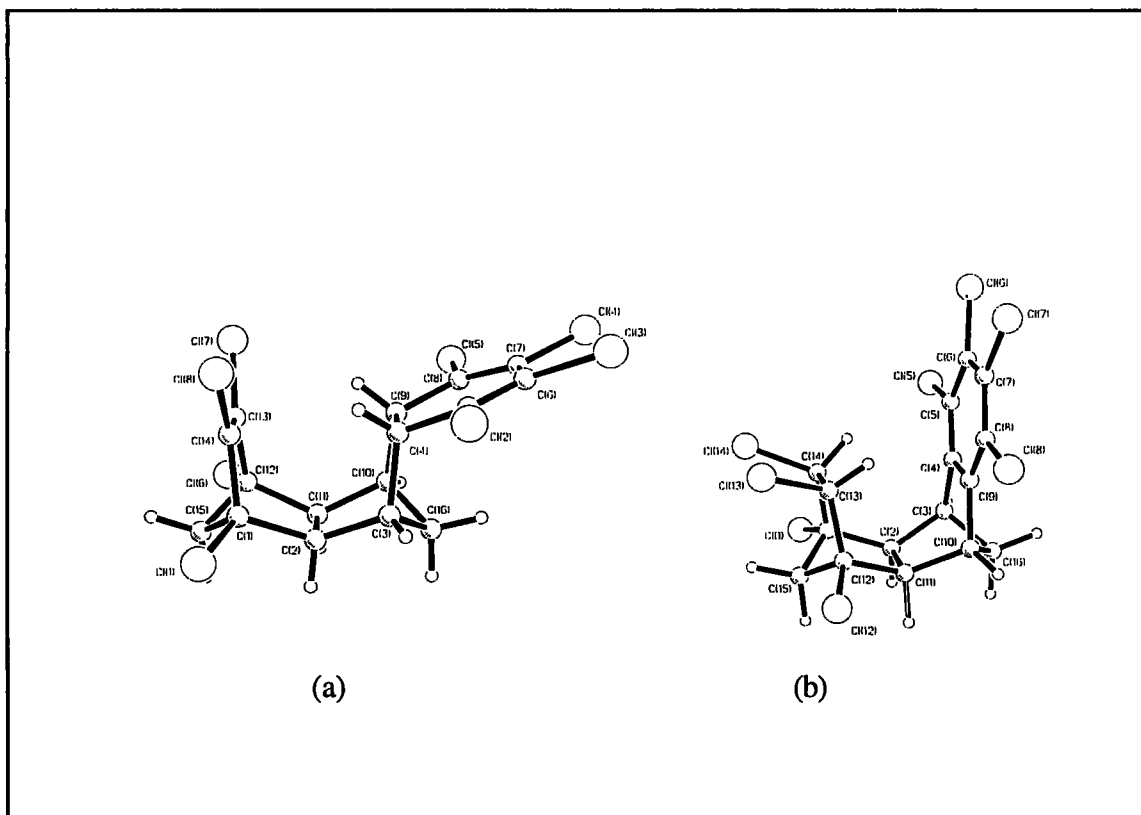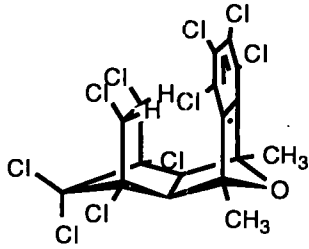
Table 3.8. Experimental details for structure (3)

Chemical Details	
Empirical formula: $C_{17}H_{10}Cl_{10}O$	Chemical diagram
Formula weight: 584.75	

Diffractometer Experimental details	
Unit Cell	Data collection parameters
$a = 15.300(3)$Å	Crystal Size:0.5 × 0.35 × 0.32 mm
$b = 8.364(2)$Å	Collection method: $\omega/2\theta$ scans.
$c = 17.015(3)$Å	Speed of collection: 6-1.5°/min.
	Scan width: $(K_{\alpha 1} + 1°)$ to $(K_{\alpha 2} - 1°)$ †
$\beta = 93.34(3)°$	θ range for collection: 1.33 to 30.00°
	Index ranges :
Volume = $2173.7(8)$Å^3	$0 \leq h \leq 18$
Based on 26 reflections in the range	$0 \leq k \leq 11$
10-39.4° 2θ	$-23 \leq l \leq 23$
	Temperature: 293(2)K

Refinement details		
Space group: $P2_1/c$	Z = 4	
$\rho_c = 1.787$ g/cm^3	Absorption coefficient: 1.291 mm^{-1}	
F(000): 584.75	Goodness-of-fit on F^2: 1.032	
Reflections collected: 5064	Final R indices [I>2σI)]	
Independent reflections: 4850	R1 = 0.0358	wR2 = 0.0849
$R_{int} = 0.0175$	R indices (all data)	
Data: 4840	R1 = 0.0503	wR2 = 0.1006
Restraints: 0	Largest diff. peak and hole	
Parameters: 293	0.423 and -0.376 eÅ^{-3}	
Hydrogen Treatment: Hydrogen positions were allowed to refine without restraint. Isotropic temperature factors were also refined independantly.		
Remark: No Absorption correction was applied to this structure		

---

† From the angle of diffraction of $K_{\alpha 1}$ radiation + 1° to the angle of diffraction of $K_{\alpha 2}$ radiation - 1°.

45

Table 3.9. Experimental details for structure (4)

Chemical Details	
Empirical formula: $C_{17}H_{10}Cl_{10}O$	Chemical diagram
Formula weight: 584.75	

Diffractometer Experimental details	
Unit Cell:	Data collection parameters
$a = 17.216(3)$Å	Crystal Size:0.60 x 0.40 x 0.20 mm
$b = 13.462(3)$Å	Collection method: $\omega/2\theta$ scans
$c = 17.947(4)$Å	Speed of collection: 2°/min
	Scan width (2θ): 1.3 + 0.3tanθ
	θ range for collection: 2.81 to 24.98°
	Index ranges :
Volume = 4159(2)Å^3	$0 \leq h \leq 16$
Based on 25 reflections	$0 \leq k \leq 16$
	$-21 \leq l \leq 0$
	Temperature: 293(2)K

Refinement details		
Space group: Pbca	Z = 8	
$\rho_c= 1.868$ g/cm^3	Absorption coefficient: 1.349 mm^{-1}	
F(000): 2320.0	Goodness-of-fit on F^2: 1.090	
Reflections collected: 2528	Final R indices [I>2σI)]	
Independent reflections: 2528	R1 = 0.0277	wR2 = 0.0492
Data: 2521	R indices (all data)	
Restraints: 0	R1 = 0.0461	wR2 = 0.0628
Parameters: 284	Largest diff. peak and hole	
Extinction Coefficient: 0.00175(10)	0.238 and -0.286 e Å^{-3}	
Hydrogen Treatment: Hydrogen positions were allowed to refine without restraint. Isotropic temperature factors were also refined independantly.		

46

(Mackenzie, Howard, *et al*, 1993). For example, in the diene shown in Figure 3.10(a) (from now on referred to as diene HMW1a) the mean intramolecular cavity C-C distance is 3.097Å (3.090Å from neutron diffraction work), as compared to 3.091Å in structure (3). This suggests that the system should be able to undergo dyotropy in the same way as comparable examples. The mean intramolecular cavity distance observed for structure (4) was found to be 3.089Å. This compares to a value of 3.075Å (3.080Å from neutron data) for the dyotropomer shown in Figure 3.10(b) (referred to here as HMW1b).

Previous studies of similar systems have suggested that the rate of intramolecular dyotropy can be increased by a factor of 10^4 by changes of the order of 0.2Å in the C..H intra-cavity distance. Unfortunately, X-rays rarely provide particularly reliable hydrogen positions, but in good quality datasets a C-H bond length can be determined to a reliability of better than 0.1Å. With such data some approximation to this distance can be gained, provided some care is taken in the interpretation of the distances. The C-H bond lengths and intra-cavity C..H distances are given below for the dyotropic pair (3) and (4) in Table 3.10. The comparable values for idealised hydrogen atom positions are also given. The estimated standard uncertainties on these distances are relatively high, but small enough to suggest that the distances are meaningful.

In HMW1a the measured intra-cavity neutron diffraction C..H distances are 2.361 and 2.435Å. These distances are comparable with those observed for (3). One interesting effect that is apparent is the asymmetry of the C..H contacts across the cavity. A comparable effect is noted by Paquette (1990) in other cage-type compounds.

Figure 3.11 shows an OFIT superposition of the X-ray models for (3) and (4). What is apparent from this figure is that the carbon cage is not strongly perturbed as a result of intramolecular dyotropy. Using this, we can estimate the distance that the two protons travel during dyotropy in this system. These values are shown in Table 3.11, for the final model and for idealised proton positions.

Table 3.10. Intra-cavity distances for structures (3) and (4).

Structure (3)			Structure (4)		
Contact	*Refined*	*Idealised*	*Contact*	*Refined*	*Idealised*
C(13)...H(9)	2.38(2)	2.34	C(9)...H(13)	2.42(3)	2.33
C(14)...H(4)	2.44(2)	2.38	C(4)...H(14)	2.45(3)	2.35

Table 3.11. Inter-model hydrogen distances between (3) and (4).

Contact	*Distance [Å]*	*Idealised distance [Å]*
H(4)...H(14)	1.55	1.44
H(9)...H(13)	1.47	1.39

Figure 3.11. A least-squares fit (OFIT) of structure (3) and (4)
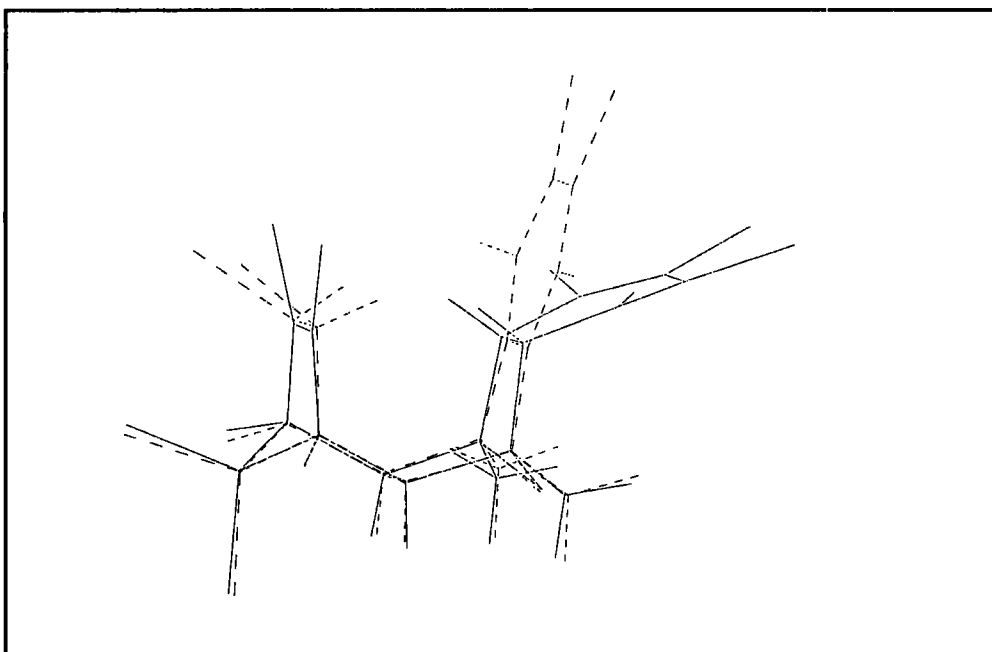
These distances are comparable to those for HMW1a and b, where the mean distance that the two protons have to travel during dyotropy is 1.455Å. This parameter is important because it discredits possible mechanisms for reaction where the proton movement has to be small during the change from reactant to product.

The distance here shows that the protons have to move from their ground state positions prior to reaction. It is possible that the thermal vibration modes have a large enough amplitude to fulfil this function. This would be consistent with the temperature dependence of the reaction rate. (Mackenzie, Howard *et al*, 1993).

Angular data can also be abstracted from the OFIT superposition of (3) with (4). Details of this are as is shown below in Figure 3.12. Specific planes are defined in the table attached. Values for the angles defined in Figure 3.12 are given in Table 3.12 for the (3)/(4) and HMW1a/b super-positions.

The largest differences in Table 3.12 are C-D and D-E. The additional chlorine atoms in this system have a steric influence on the geometry in this position, forcing the chlorine atoms in (4) into a more linear conformation. The geometry of the protons in the product molecules (4) and HMW1b are similar. Again the slight differences can be accounted for by the steric influence of the added substituent chlorine atoms. In this case there is a "knock on" effect due to the movement of plane C, as described above.

The steric influence of the CH_3 groups on plane B is apparent from the difference in the inter-plane angle A-B. This movement of plane B in turn has a small effect on the plane involving the reactant cavity protons, H(4) and H(9), Plane G. However, a simple geometrical calculation shows that a small change in this angle is possibly significant, due to the wide range of reaction rates that are observed as a result of small differences in cavity geometry.

Figure 3.12 The definition of planes A-H in the OFIT super-position of (3) and (4)
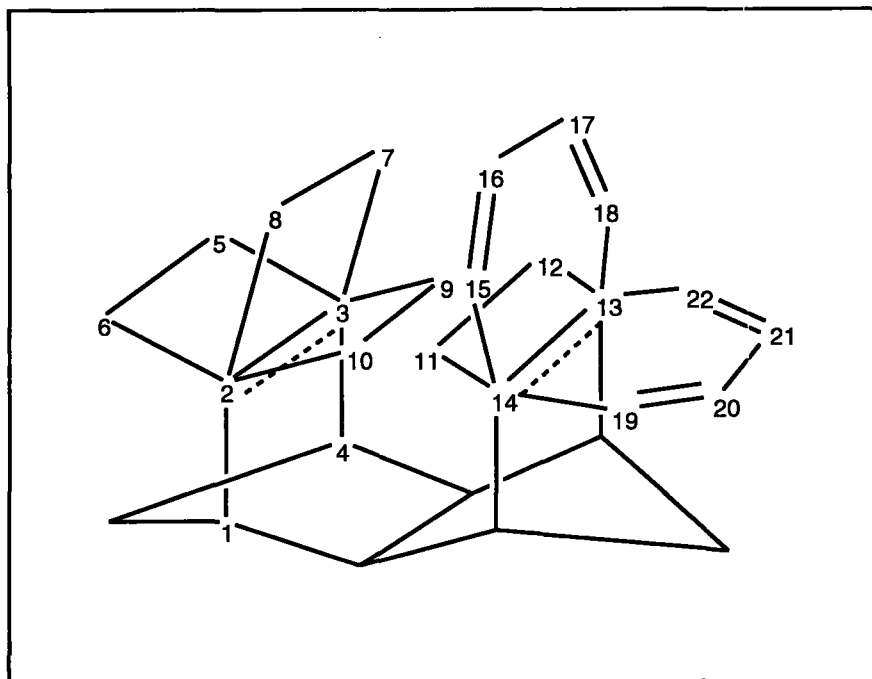
Plane A: 13-14-15-16-17-18	Plane B: 13-14-19-20-21-22
Plane C: 2-3-5-6	Plane D: 2-3-7-8
Plane E: 2-3-9-10	Plane F: 1-2-3-4
Plane G: 11-12-13-14	Plane H: 9-10-14-13
(Plane H is not shown explicitly in Figure 3.12 to preserve the figure's clarity)	

Table 3.12. Selected Inter-planar angles for the OFIT superpositions of (3) and (4), and for HMW1(a) and (b)

Planes	(3)/(4)	HMW1a/b
A-B	54.0°	58.2°
C-D	48.9°	62.6°
D-E	65.6°	56.8
E-F	62.1°	58.8
G-H	19.4°	20.7

Figure 3.13: A schematic representation showing the angle between planes G and H (defined in Figure 3.12),.

The lower case letters correspond to mean inter-atomic distances between the atoms at the periphery which define the planes. Plane H, defined as the best plane through 9,10,13 and 14 is again omiited for clarity.
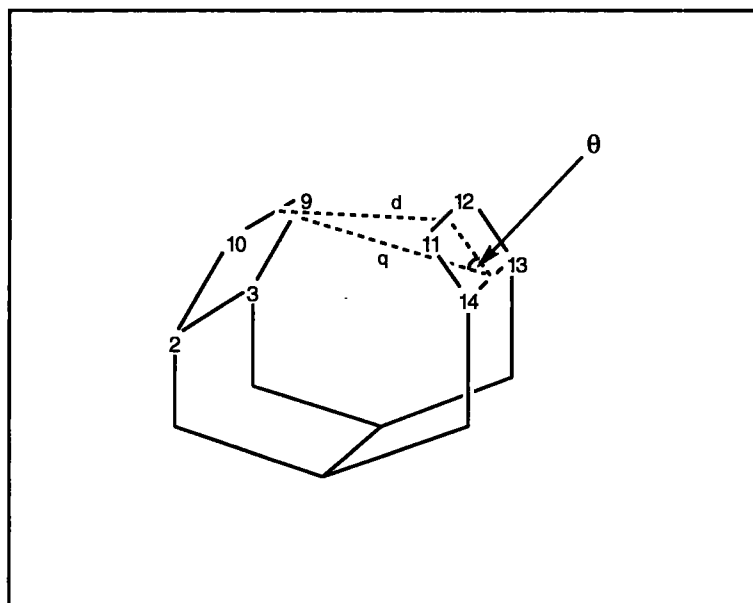
Consider Figure 3.13. By the cosine rule:

$$d = \sqrt{(r^2 + q^2 - 2rq\cos(\theta)}$$

The first derivative of this with respect to θ is

$$\frac{\partial d}{\partial \theta} = rq\sin(\theta).\sqrt{(r^2 + q^2 - 2rq\cos(\theta))}$$

Using this derivative we can get an approximation to the rate of change in the distance d at the approximate observed values of θ, r and q. For $\theta = 20°$, r = 1.0Å and q = 2.34Å, the derivative is approximately 0.01Å deg.$^{-1}$. This is only a small change, but Paquette (1991) has reported rate changes of $\sim10^4$ for similar systems where intra-cavity distances change by only 0.1Å, thus some effect would be expected as a result of a small change such as 0.01Å.

The rate of reaction in this dyotropic pair has been measured (Mackenzie, 1995) and is found to be approximately the same as the rate observed in the unmethylated analogue,

51

so either a small motion of the proton positions does not contribute greatly to the reaction rate in this system, or the methyl groups retard the reaction rate electronically, cancelling this effect out.

One unusual feature present in these two molecules is an out of plane bending of the chlorine atoms from the dienyl fragment in structure 3 and, more interestingly, from the benzene ring plane in molecule 4. The angle subtended between the Cl5-C5 bond and the plane defined by atoms C(4), C(5), C(6), C(7), C(8) and C(9) is 10.2°. The angle between the same plane and the Cl(8)-C(8) bond was observed as 9.3° in structure 3. In structure 4 the equivalent angles between the same chlorine atoms and the benzene ring were observed as 5.0 and 4.2° respectively. This may be primarily an effect of the molecular packing in the crystal.

Table 3.13(a). Close contacts in structure (3)

Contact	Distance	μ	τ
Cl(7)...Cl(8) (2-x, y+1/2, 3/2-z)	3.395(1)	47.67(9)	32.38(10)
Cl(5)...Cl(7) (2-x, 1-y, 2-z)	3.577(1)	88.76(9)	87.95(9)
Cl(5)...Cl(13) (x, 1/2-y, z+1/2)	3.569(1)	11.86(9)	57.77(10)
Cl(1B)...Cl(1) (1-x, -y, 2-z)	3.553(1)	16.75(8)	57.77(10)
Cl(1B)...O(16) (x, y-1, z)	3.182(2)	41.71(10)	-

Table 3.13(b). Close contacts in structure (4)

Contact	Distance	μ	τ
Cl(5)...Cl(12) (x+1/2 1/2-y 1-z)	3.252(1)	11.51(12)	64.48(11)
Cl(6)...Cl(1B) (1-x 1-y 1-z)	3.416(1)	63.01(11)	58.80(3)
Cl(13)...Cl(14) (1-x 1-y 1-z)	3.571(1)	82.65(12)	80.92(12)
Cl(13)...Cl(1A) (1/2-x y+1/2 z)	3.128(2)	40.76(13)	51.87(12)

Figure 3.14(a). Packing in structure (3).
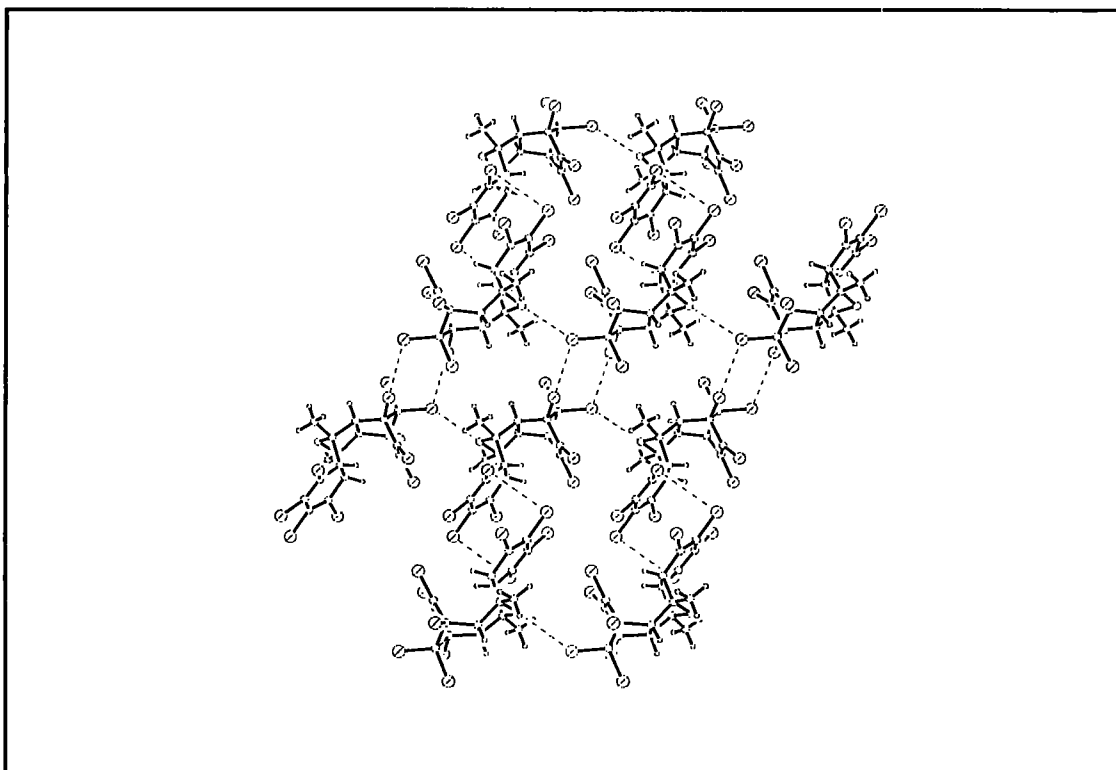
Figure 3.14(b). Dimeric packing in structure (3)
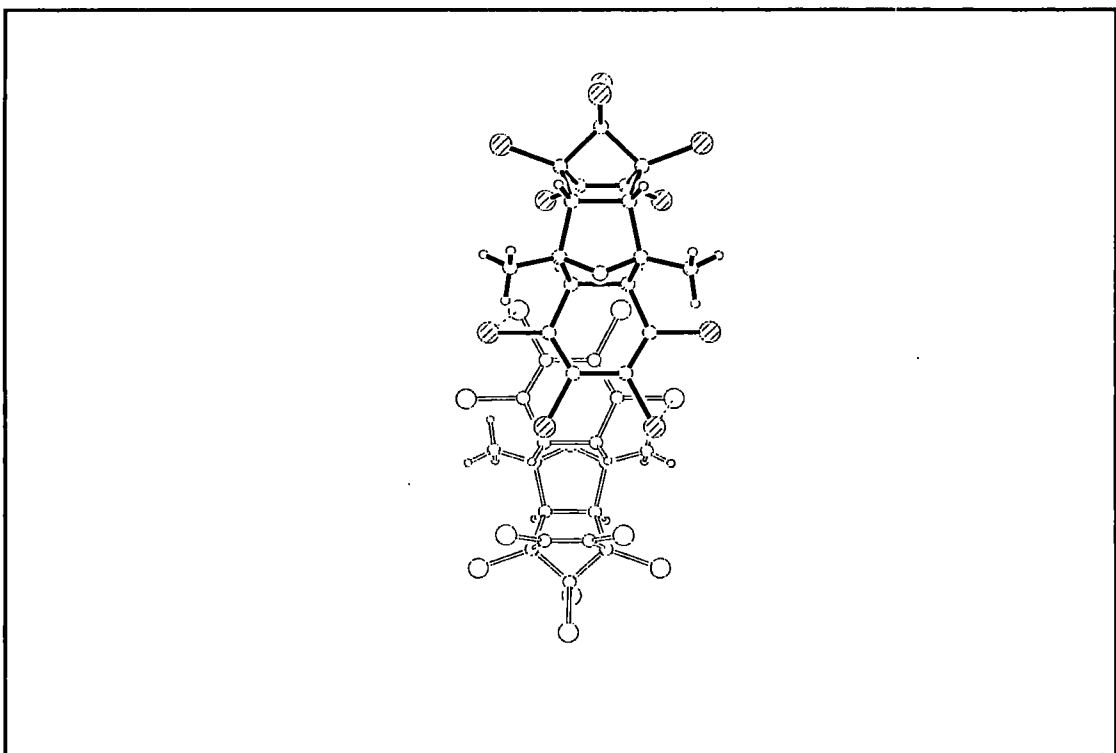
Figure 3.15 Packing in structure (4)
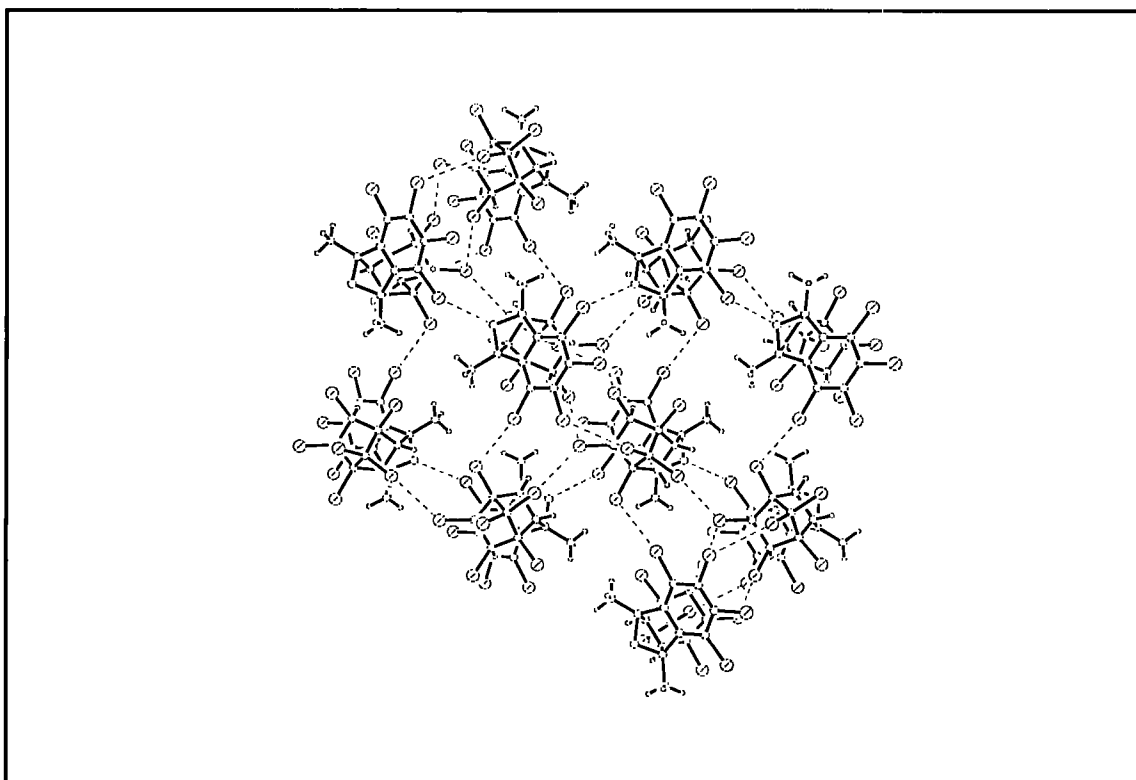
The packing of (3) and (4) shows a rich array of non-bonded contacts. These are listed in Tables 3.13(a) and (b). The packing arrangements for structures (3) and (4) are given in Figures 3.14(a,b) and 3.15 respectively. In structure (3) five contacts are listed, none of which involve hydrogen atoms. The distances observed are again within the limits that were suggested by Nyburg and Faerman (1985). The closest contacts only utilise six of the ten available chlorine atoms. A search with wider tolerances revealed that the remaining four chlorine atoms also have near neighbours, although not as close as those shown in the tables. This is to be expected considering the closed packing requirements of the crystalline state.

In (3) the dimeric pairs pack with their dienyl groups facing one another as shown by Figure 3.14(b)

Structure (4) also has an array of non-bonding contacts. The closest Cl-Cl contact is quite short at 3.252Å. The packing is again a simple close-packed arrangement. As one usually sees in centrosymmetric structures the molecules pack in dimeric pairs with

interactions between these pairs. The close packed arrangement is clearly seen in Figure 3.15. The closest contact is still within the bounds observed by Nyburg and Faerman (1985).

3.4 The Structure of (5)
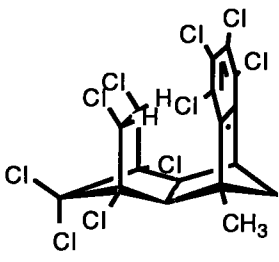
This compound is another example of a dyotropic product, similar to structure (4) (See Figure 3.17 and Table 3.14), and can be analysed in the same way. Unfortunately structural data is not available for the dyotropic isomer of (5) and so similar comparisons cannot be drawn. Some information is available, however.

The data quality in this structure was not as high as in (3) and (4) and so the hydrogen atoms were fixed at idealised X-ray positions during refinement.

Figure 3.16. A computer generated representation of structure (5).
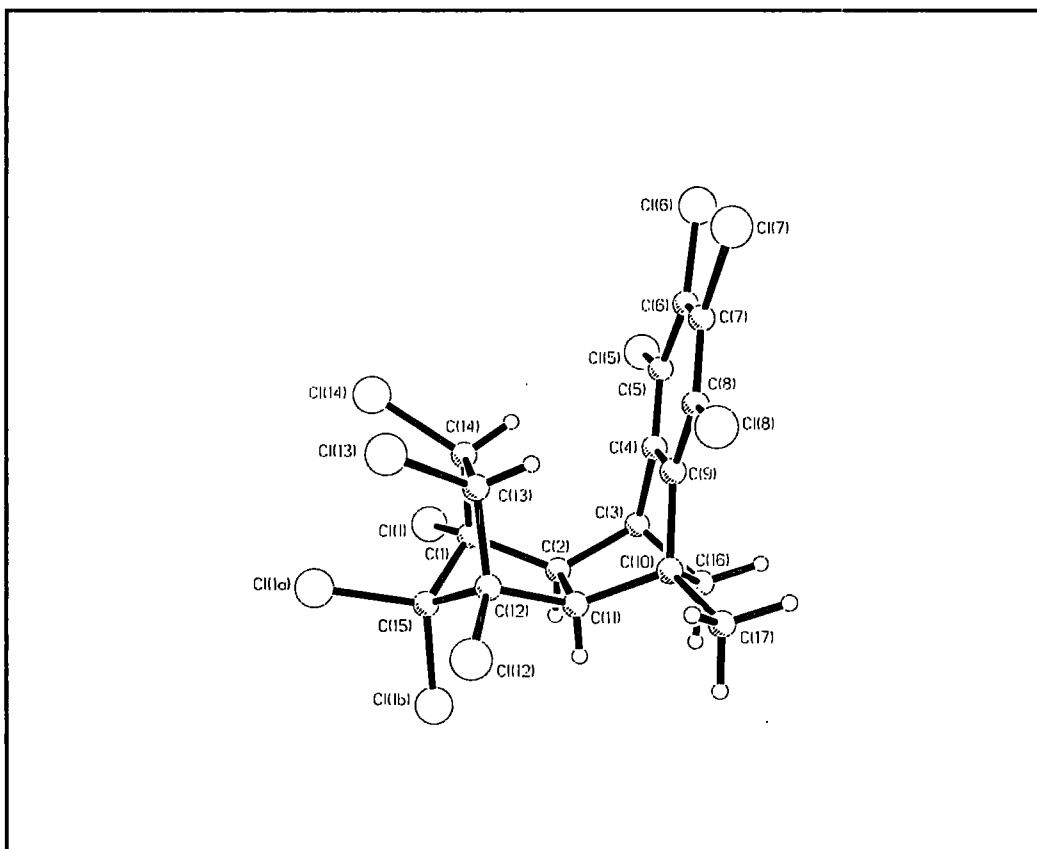
Table 3.14. Experimental details for structure (5)

Chemical Details	
Empirical formula: $C_{17}H_{10}Cl_{10}$	Chemical diagram
Formula weight: 568.75	

Diffractometer Experimental details	
Unit Cell	Data collection parameters
a = 11.317(1)Å	Crystal Size:0.5 mm cube.
b = 14.555(1)Å	Collection method: ω scans
c = 12.925(1)Å	Speed of collection: 3-30°/min
	Scan width:1.0 + 0.5tanθ
β = 101.11(1)°	θ range for collection: 2.13-25°
	Index ranges :
Volume = 2089.1(3)Å^3	-1 $\leq$ h $\leq$ 13
Based on 42 reflections between 10 and	-1 $\leq$ k $\leq$ 17
25° in 2θ	-15 $\leq$ l $\leq$ 15
	Temperature: 293(2)K

Refinement details		
Space group P2$_1$/n	Z = 4	
ρ_c= 1.808 g/cm^3	Absorption coefficient: 1.337 mm^{-1}	
F(000): 1128	Goodness-of-fit on F^2: 1.028	
Reflections collected: 4631	Final R indices [I>2σI)]	
Independent reflections: 3661	R1 = 0.0464	wR2 = 0.1084
R$_{int}$ = 0.0281	R indices (all data)	
Data: 3661	R1 = 0.0630	wR2 = 0.1184
Restraints 0	Largest diff. peak and hole	
Parameters 246	0.520 and -0.390 eÅ^{-3}	
Hydrogen Treatment: Hydrogens were placed at geometrically predicted positions. Methyl group hydrogens were placed at points of the points of maximum electron density† an ideal distance from the parent carbon atom. Temperature factors were fixed at 1.2 × the parent atom's for non-methyl hydrogens, and at 1.5 × their parent carbon atom for methyls.		

---

† Derived by calculating a fourier map in the ring perpendicular to the C-X bond at an idealised distance

Figure 3.17 OFIT of (5) with (4)
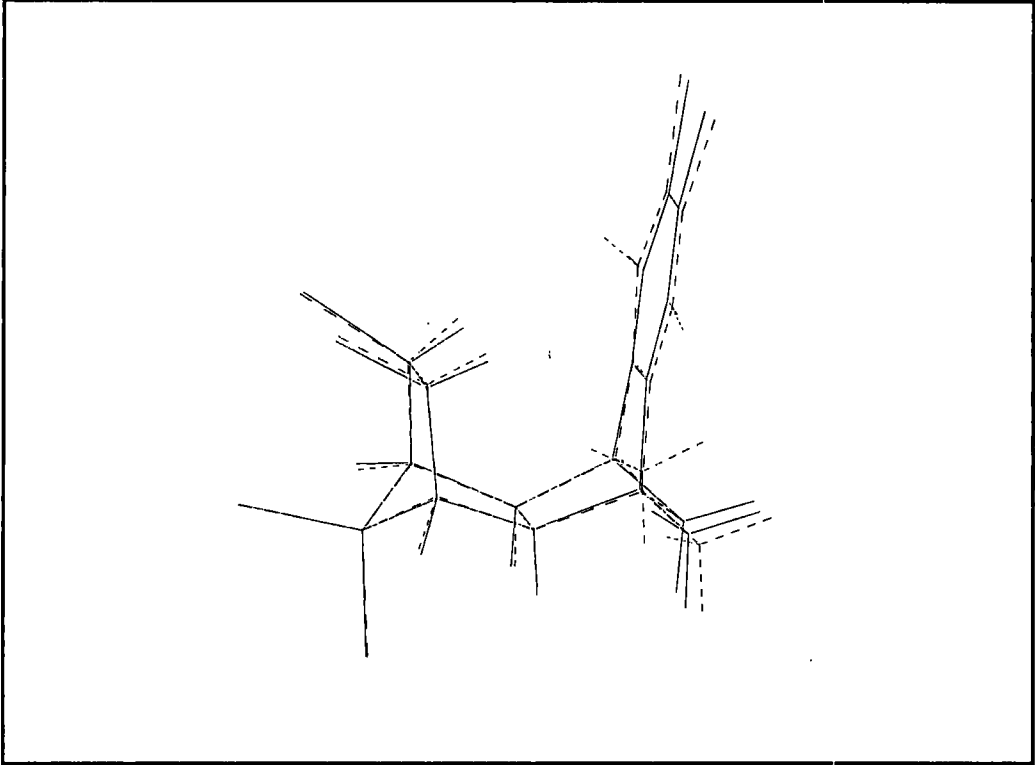
Figure 3.18. OFIT of (5) with HMW1(b)
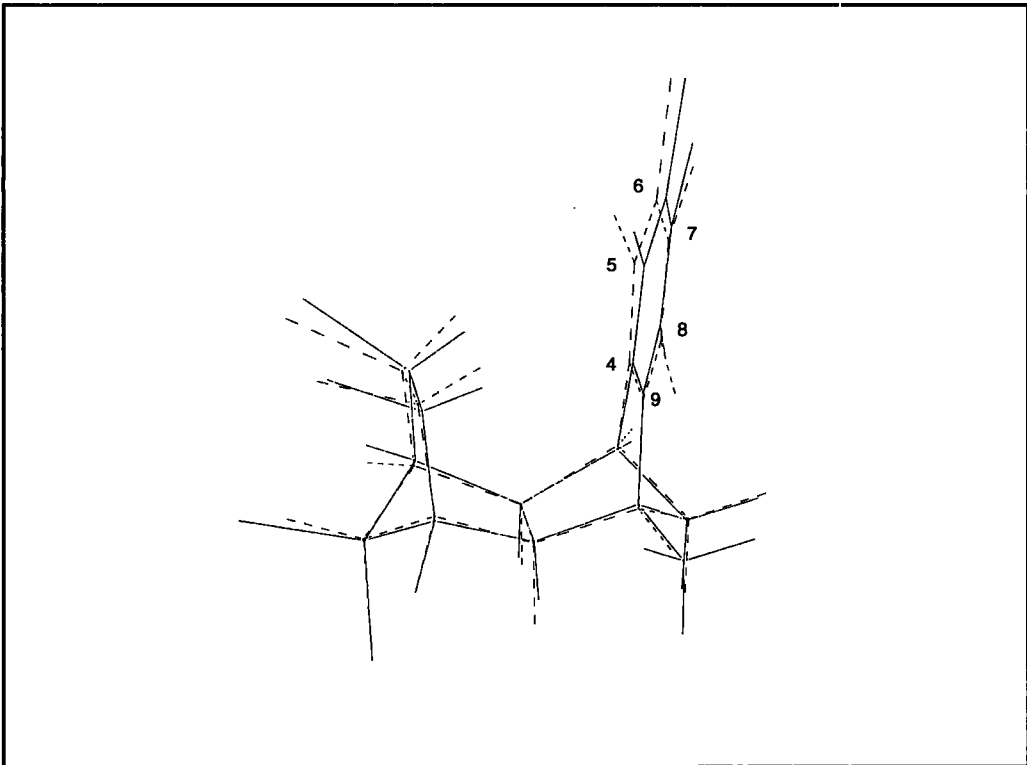
The intra-cavity C-C distances in this structure, C(9)...C13 and C(4)...C(14), were found to be 3.062(5)Å and 3.070(5)Å respectively. The less reliable C-H intra-cavity distances, C(9)...H(13) and C(4)...H(14) were found to be 2.33 and 2.31Å respectively. These values compare closely to those found in HMW1b and in structure (4). As expected in these systems, the isodrin skeleton is not seriously affected by tertiary substitution. The effect is only seen in the peripheral groups. This is shown very clearly in Figures 3.17 and 3.18 where structure (5) is fitted with (4) and HMW1b. It is again apparent from these figures that the substitution of the added chlorine atoms on the methylene bridge forces the ethyl bridge chlorine atoms into a more planar conformation.

The addition of a CH_3 group forces a slight twisting of the molecule. The overall effect of this is most pronounced in the benzene ring in Figure 3.18. The shift distances of the benzene carbon atom positions are shown below in Table 3.15.

The angle between the two ring mean planes was calculated from the OFIT superposition to be 4.2°, which explains some of the above shift. Using simple trigonometry and the "across ring" distance of 2.75Å one can calculate a maximum shift of approximately 0.2Å due to change of planar angle at the furthest C atom, thus a translational component is present within the distances shown above which is of the order of 1.5Å. The methyl group induces a small twist into the cage which is magnified in the periphery.

Table 3.15. Shift distances abstracted from the least-squares fit of (5) with HMW1(b)

Atom	Shift
C(4)	0.140
C(5)	0.307
C(6)	0.402
C(7)	0.385
C(8)	0.283
C(9)	0.137

The rate of reaction in this system is six times faster than that for the comparable unmethylated compound (Mackenzie,1995). This is probably a factor that is induced by the slightly lower intra-cavity distances in the mono-methylated compound, although the precision of the final results is not high enough to be certain.
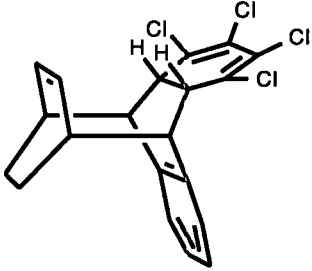
3.5 The structure of (6).

In this compound (experimental data and chemical diagram in Table 3.16) a different framework (shown in Figure 3.19) is investigated for dyotropy. The rate of reaction is dependent on the strain energy of the reactant. It is probably also dependent on the transition state strain energy as well, and so a direct comparison of intramolecular cavity distances does not explain differences in the dyotropic rearrangement rate. One useful aspect of this compound is that part of it represents the equivalent of a dyotropic product, and part a dyotropic reactant for this particular system. Thus some idea as to the change in geometry required for dyotropic rearrangement can be gained by fitting one side of the molecule to the other using OFIT. Such a fit is shown in Figure 3.20.

Figure 3.19. A computer generated representation of structure (6)
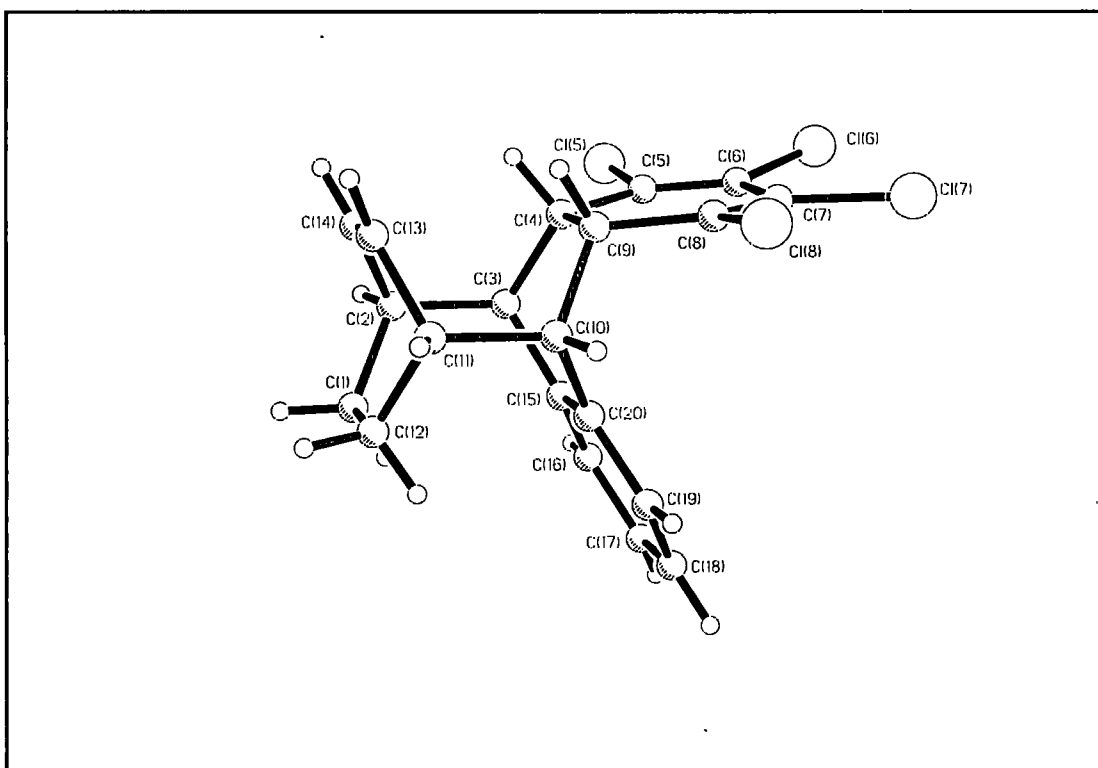
Table 3.16. Experimental details for structure (6)

Chemical Details	
Empirical formula: $C_{20}H_{14}Cl_4$	Chemical diagram
Formula weight: 396.11	

Diffractometer Experimental details	
Unit Cell	Data collection parameters
$a = 8.276(2)$Å	Crystal Size:$0.51 \times 0.46 \times 0.48$ mm
$b = 9.127(2)$Å	Collection method $\omega/2\theta$ scans
$c = 12.369(2)$Å	Speed of collection: 4-8°/min
$\alpha = 83.10(3)°$	Scan width: $1.68 + 0.3\tan\theta$
$\beta = 72.39(3)°$	θ range for collection: 2.67 to 25.0°
$\gamma = 73.70(3)°$	Index ranges :
Volume = 854.1(3)Å^3	$-9 \leq h \leq 9$
Based on 25 reflections in the range 18.4	$-9 \leq k \leq 0$
to 30.18° in 2θ	$-14 \leq l \leq 14$
	Temperature: 150(2)K

Refinement details		
Space group $P\bar{1}$	Z = 2	
$\rho_c = 1.540$ g/cm^3	Absorption coefficient: 0.691 mm^{-1}	
F_{000}: 404	Goodness-of-fit on F^2: 1.075	
Reflections collected: 2955	Final R indices [I>2σI)]	
Independent reflections: 2748	R1 = 0.0274	wR2 = 0.0687
R_{int} = 0.0153	R indices (all data)	
Data: 2745	R1 = 0.0345	wR2 = 0.0745
Restraints: 0	Largest diff. peak and hole	
Parameters: 281	0.241 and -0.275 e.Å^{-3}	
Hydrogen Treatment: Hydrogen positions were allowed to refine without restraint. Isotropic temperature factors were also refined independently.		

Figure 3.20. OFIT of top of molecule with bottom
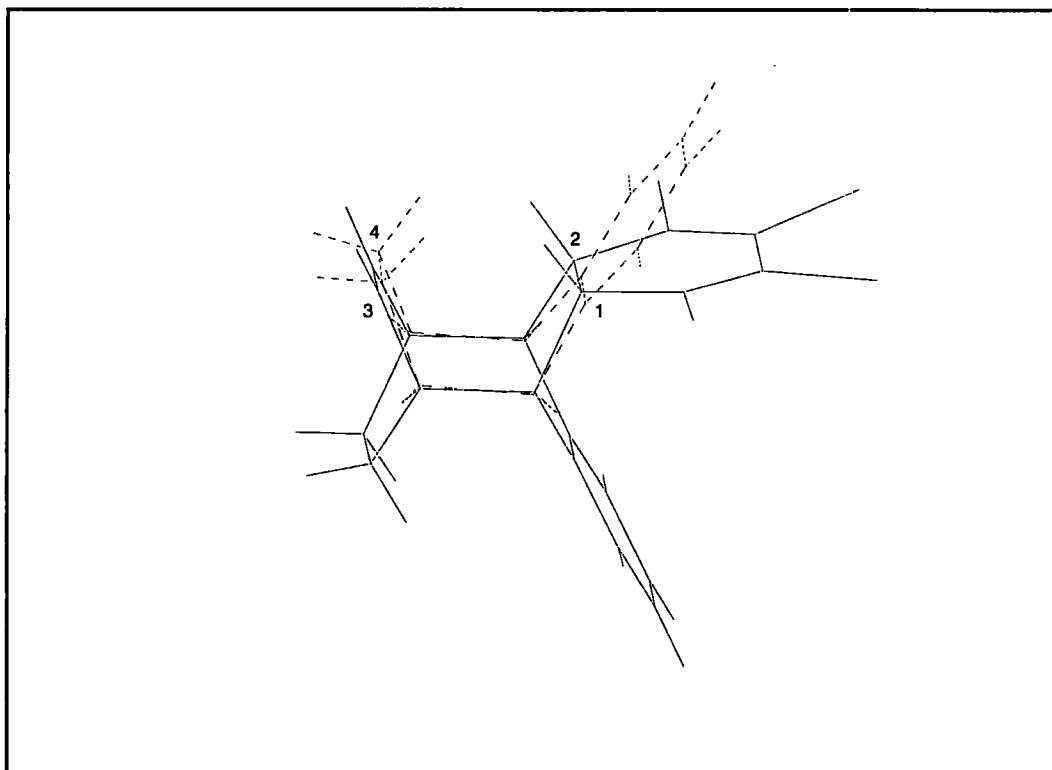
This tentative form of analysis gave some approximate data for atomic motion required during dyotropic rearrangement. The separation of the two atoms at points 1 and 2 in Figure 3.20 have to change by 0.23Å and 0.24Å respectively to go from reactant to product. The atoms at sites 3 and 4 both have to move by 0.22Å. Clearly this is only an order of magnitude estimate, since we are not sure of the geometry of the rearrangement product, but even so it is worth comparing these values with the true carbon movements observed in the change from structure (3) to structure (4).

In going from structure (3) to structure (4) the equivalent atomic sites to 1, 2, 3 and 4 shown in Figure 3.20 (i.e. C(4), C(9), C(13), C(14) in (3) and (4)) have to move by only an average of 0.14Å each. In going from a C-C single bond to a C-C double bond the bond length changes by ~0.25Å, so two sites changing by 0.14Å in one bond can acccounted to the change of bond type. The larger changes of each site observed for structure (6) suggest that some other change of the cage geometry is also present.

Figure 3.21 Dienyl conformation in structure (6). Projection is on to the dienyl Carbon atom plane.
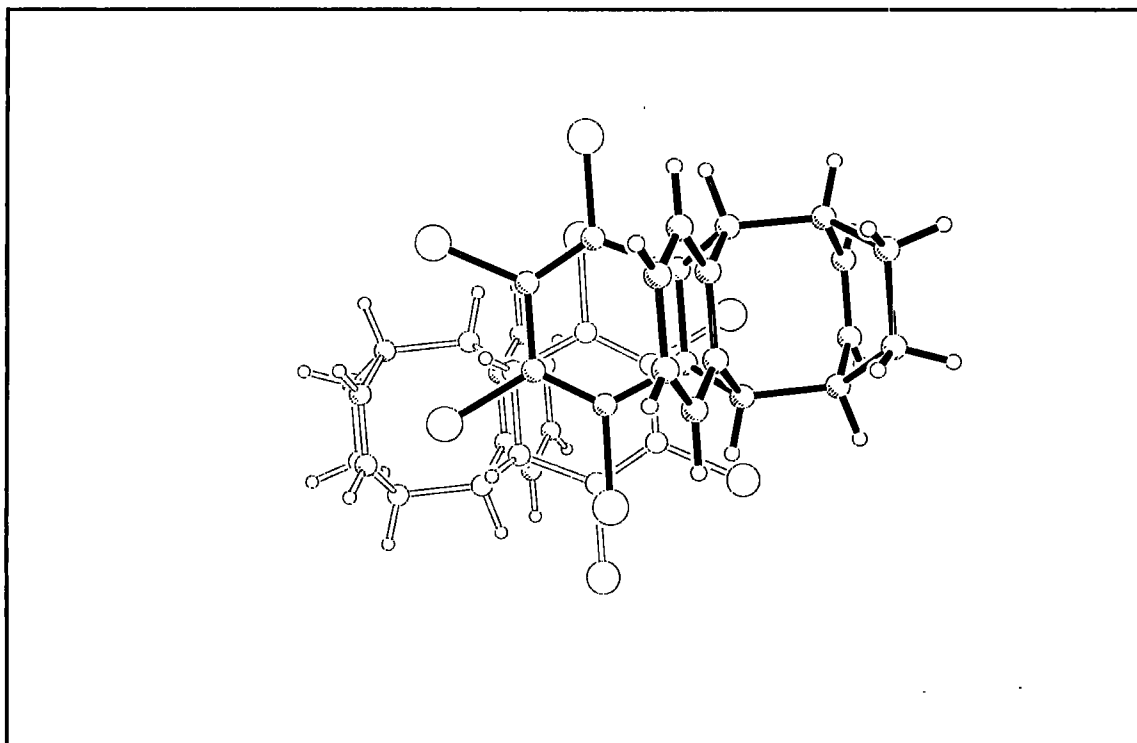
The intra-cavity separations, shown below in Table 3.17 is certainly well within the range required for rearrangement, They are shorter than their equivalent distances for the isodrin and homo-isodrin type compounds, as would be expected for this different type of cage. Unfortunately the presence or absence of dyotropy in these systems has not been tested.

There is only one notably close Cl-Cl contact in the packing of this compound between Cl(7) and its symmetry related equivalent (by the operator 2-x,2-y,1-z), with a value of.3.460(2)Å The angular terms, μ, and τ, as defined in chapter 2, were both 63.86(8)°, since the molecules are packed around an inversion centre. This is again consistent with expected values for these systems. The most interesting feature of the packing is the similarity of the stacking as compared to structure (3). the dienyl rings are in similar packing conformation to one another in both structures (See Figure 3.21), but the rings are more overlapped in (6) than in (3).

Table 3.17. Intra-cavity distances for structure (6)

Contact	Distance
C(9)...C(13)	2.818
C(4)...C(14)	2.803
C(1)...C(15)	2.848
C(12)...C(20)	2.834

3.6 Compounds for potential new directions of research

The final 2 compounds in this chapter represent potential new systems that could be investigated for dyotropy.

3.6.1 The structure of compound (7).

Figure 3.22. Postulated dyotropic rearrangement in (7)
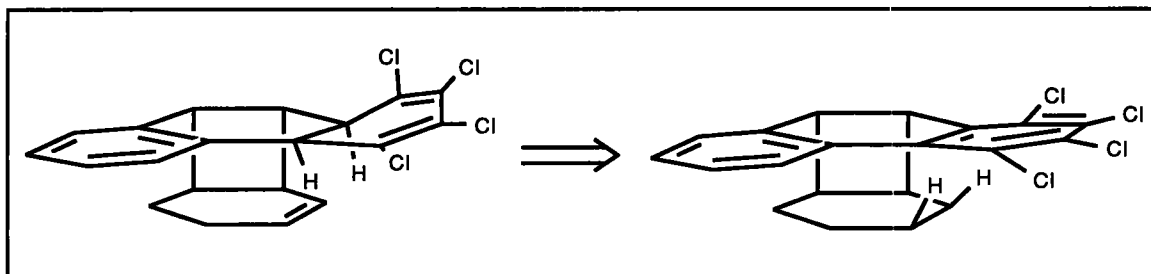
In this molecule the possibility of dyotropic rearrangement was considered as shown in Figure 3.22. The structural study (Table 3.18, Figure 3.23) shows that this is not possible due to the molecular conformation adopted by this compound. The proposed rearrangement involves a dyotropic 2H transfer from C atoms 15 and 20 to 12 and 11 respectively. The intramolecular C-C distances C(12)...C(15) and C(11)...C(20) are found to be 3.149(3) and 5.293(3)Å. This second value is much longer than in the isodrin cage compounds and thus an intramolecular dyotropic rearrangement is not plausible (see Figure 3.23).
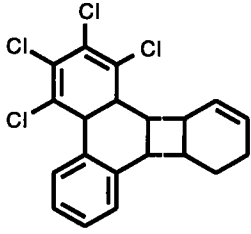
Figure 3.23. structure (7)
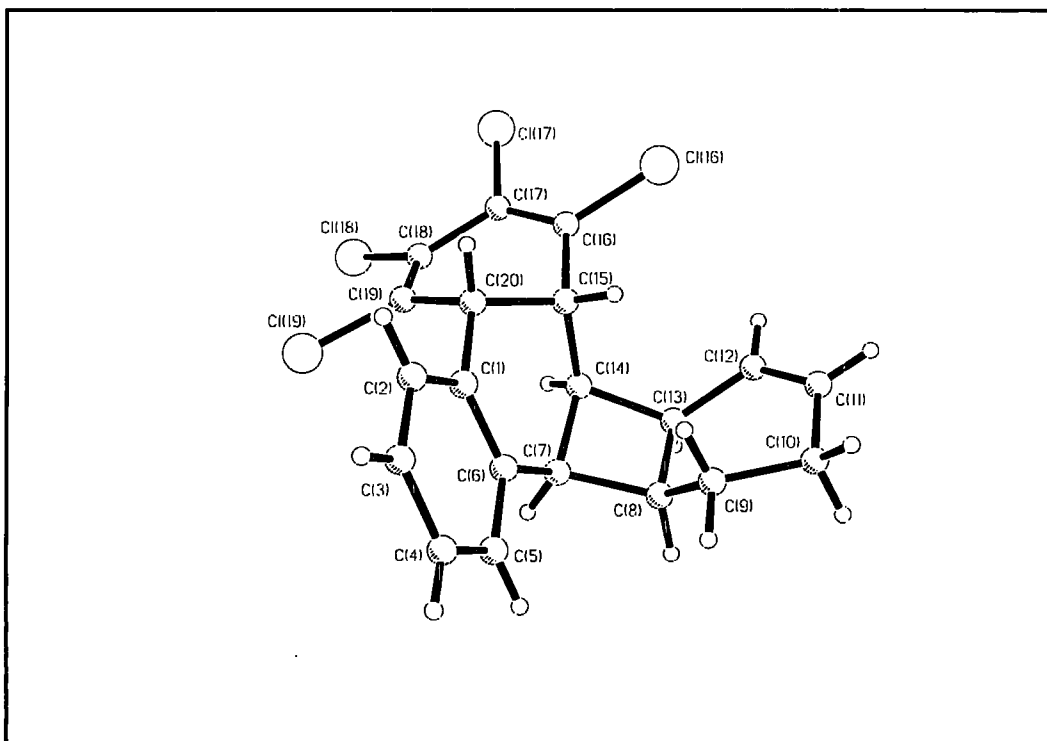
Figure 3.24. Herring bone packing in structure (7)
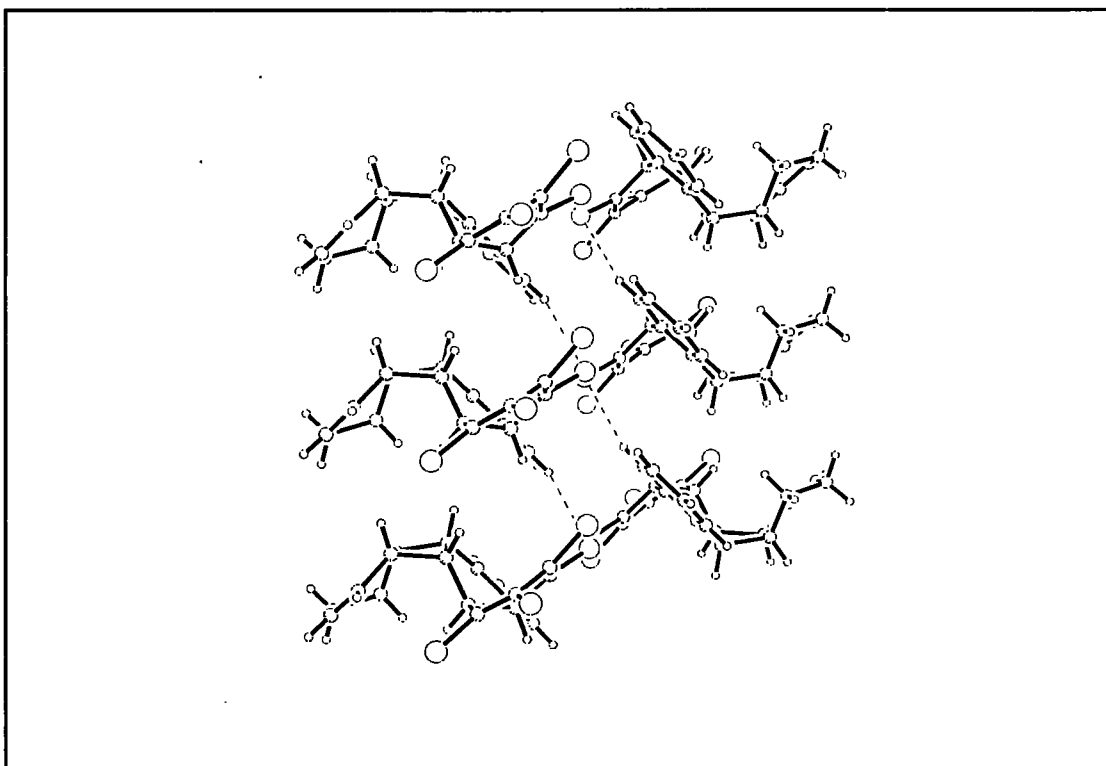
Table 3.18. Experimental details for structure (7)

Experimental Details	
Empirical formula: $C_{20}H_{16}Cl_4$	Chemical diagram
Formula weight: 398.13	
Diffractometer Experimental details	
Unit Cell	Data collection parameters
a = 5.2160(5)Å	Crystal Size: 0.44 x 0.32 x 0.24 mm
b = 16.5800(10)Å	Collection method: ω scans
c = 20.2510(10)Å	Speed of collection: 2-40°/min
	Scan width: $1.0 + 0.5\tan\theta$
β = 91.42°	θ range for collection: 2.0 to 25.0°
	Index ranges :
Volume = 1750.79(14)Å^3	$-1 \leq h \leq 6$
Based on 69 reflections in the range 8.5 to	$-1 \leq k \leq 19$
25.0° in 2θ	$-24 \leq l \leq 24$
	Temperature: 293(2)K
Refinement details	
Space group P2$_1$/c	Z = 4
ρ_c= 1.510 g/cm^3	Absorption coefficient: 0.674 mm^{-1}
F(000): 816	Goodness-of-fit on F^2: 1.027
Reflections collected: 4467	Final R indices [I>2σI)]
Independent reflections: 3079	R1 = 0.0409	wR2 = 0.0868
R_{int} = 0.0199	R indices (all data)	
Data: 3079	R1 = 0.0668	wR2 = 0.0988
Restraints: 0	Largest diff. peak and hole	
Parameters: 231	0.186 and -0.242 eÅ^{-3}	

Hydrogen Treatment: C-H bond lengths were refined. X-C-H angles were constrained to ideal values. Hydrogen temperature factors were constrained to 1.2 times the value of U$_{iso}$ of their parent atom.

Table 3 19. Relative close contacts in (7).

Contact	Distance (Å)
Cl(18)...H(10A) (x-1,0.5-y,z-0.5)	2.883(3)
Cl(19)...H(2) (x-1,y,z)	2.865(3)

The packing in this structure has no striking features. No markedly close Cl-Cl contacts are present in the packing. The closest contacts between any elemennts, relative to van der Waals radii, are given in Table 3.19. The compound close-packs in a classical herring bone fashion with bumps filling hollows as suggested by Pauling (1940), and Kitiagorodskii (1961). (See Figure 3.24)

3.6.2 The structure of compound (8)
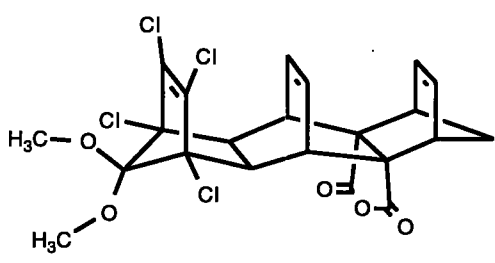
Structure (8) (See Table 3.19) represents a new potential precursor for a dyotropic system that involves a tricyclic system. It is hoped that a rearrangement as shown below may occur if the dienyl fragment can be added at the appropriate site.

Figure 3.25. Possible rearrangements in molecules based on structure (8)
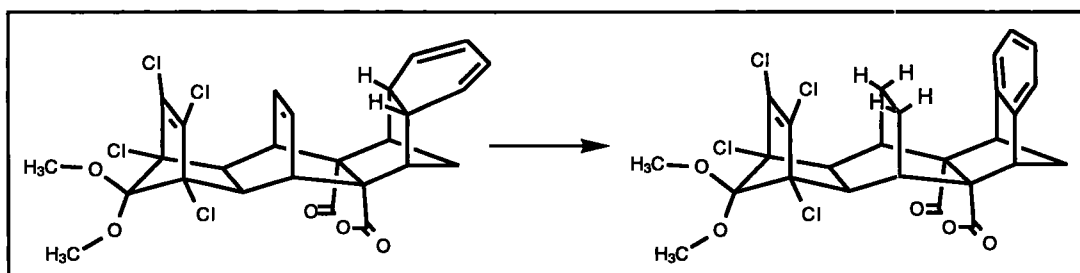
This study allows us to analyse the feasibility of such a reaction scheme by studying the geometry of the potential precursor, (8), primarily in terms of the critical inter-atomic cavity distances, C(21) to C(41), C(22) to C(42) and C(41) to C(61), C(42) to C(62), which would have an effect on the rate of intra-molecular dyotropy. The observed distances in this molecule are tabulated below in Table 3.20. A computer generated representation of structure (8) is shown in Figure 3.26

Table 3.19: Experimental details for structure (8)

Chemical Details	
Empirical formula: $C_{22}H_{18}Cl_4O_5$	Chemical diagram
Formula weight: 504.16	
Diffractometer Experimental details	
Unit Cell	Data collection parameters
a = 14.622(1)Å	Crystal Size:0.35 × 0.38 × 0.40 mm
b = 10.992(1)Å	Collection method: ω scans
c = 15.220(1)Å	Speed of collection: 3-30°/min
	Scan width: $1.0 + 0.5\tan\theta$
β = 118.39°	θ range for collection: 2.40 to 25.0 °
	Index ranges :
Volume = 2152.0(3) $Å^3$	$-1 \leq h \leq 17$
Based on 44 reflections between 9.6 and	$-1 \leq k \leq 13$
24.36° 2θ	$-18 \leq l \leq 16$
	Temperature: 293(2) K
Refinement details	
Space group $P2_1/c$	Z = 4
ρ_c= 1.556 g/cm^3	Absorption coefficient: 0.583 mm^{-1}
F(000): 1032	Goodness-of-fit on F^2: 1.032
Reflections collected: 4753	Final R indices [I>2σI)]	
Independent reflections: 3791	R1 = 0.0406	wR2 = 0.0808
R_{int} = 0.0283	R indices (all data)	
Data: 3790	R1 = 0.0722	wR2 = 0.0954
Restraints 0	Largest diff. peak and hole	
Parameters 296	0.214 and -0.242 e.$Å^{-3}$	

Hydrogen Treatment: C-H bond lengths were refined. X-C-H angles were constrained to ideal values. Non-methyl hydrogen temperature factors were constrained to 1.2 times the value of U_{iso} of their parent atom. Methyl hydrogens had a value of 1.5 times the U_{iso} of their parent atoms. The methyl hydrogen positions were determined from maxima in the Fourier map of a ring at an idealised distance from the parent carbon atom

Figure 3.26. Computer generated representation of structure (8)
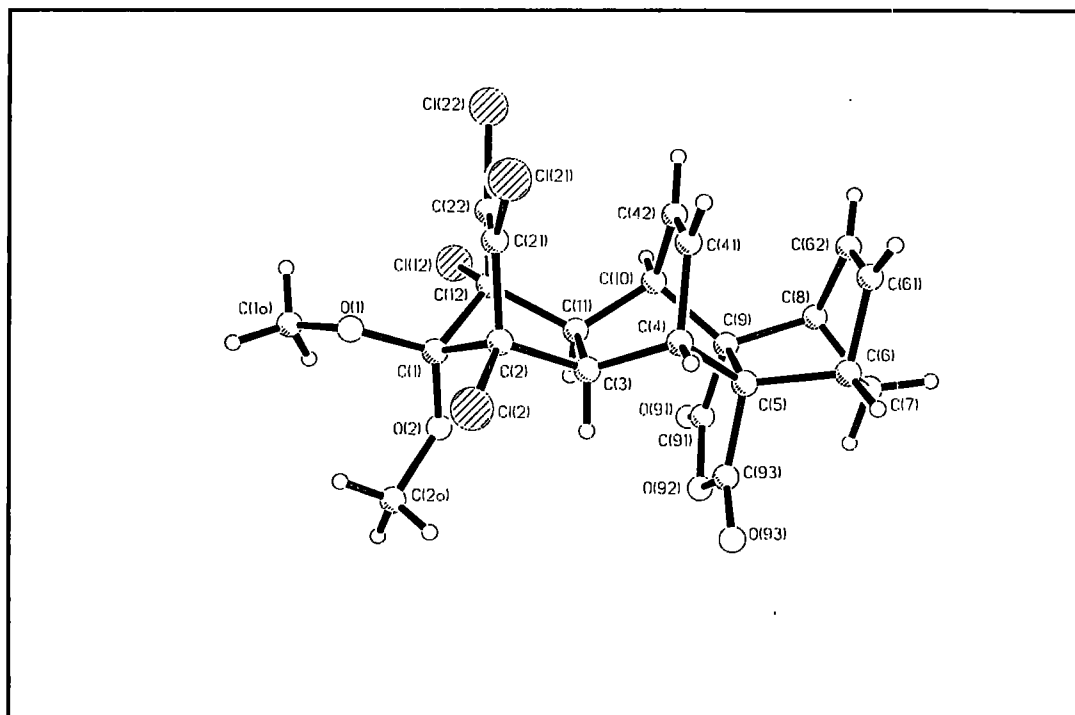
Figure 3.27. Packing in structure (8)
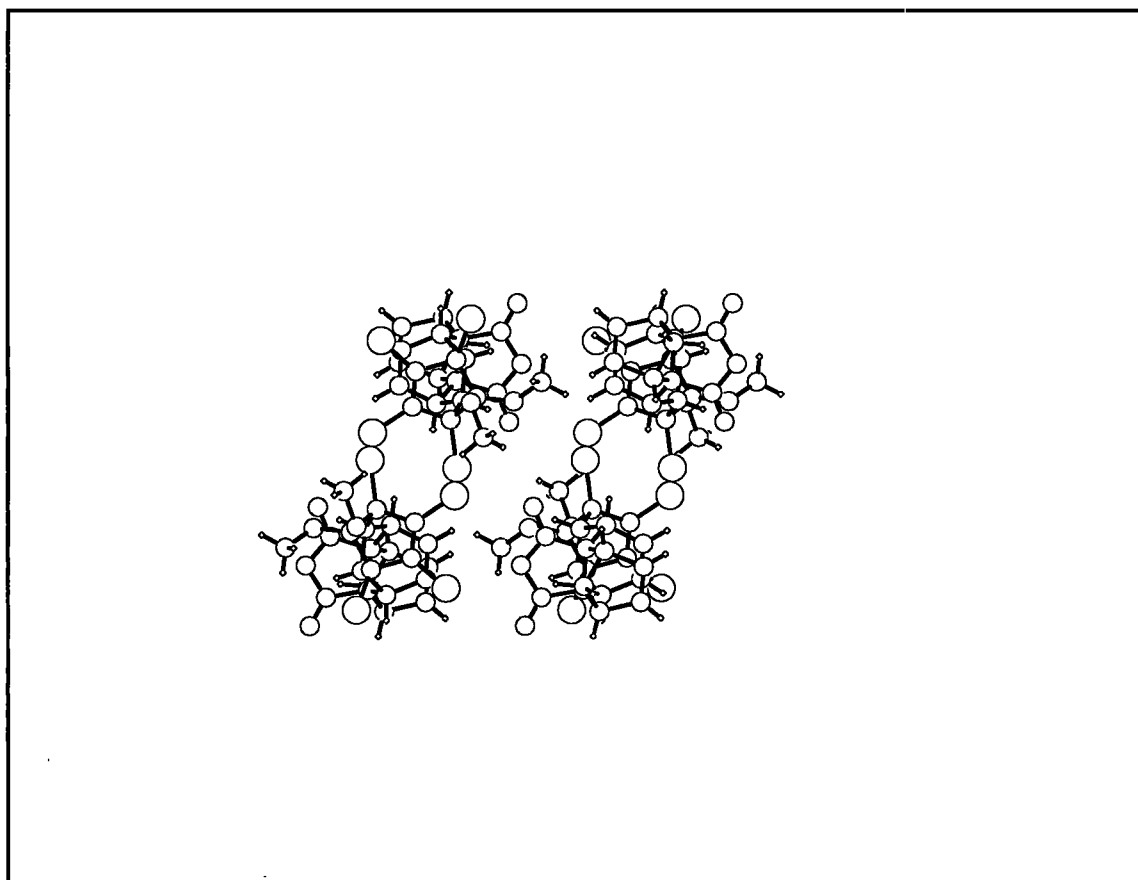
Table 3.20. Intramolecular distances in structure (8).

Contact	Distance
C(21)...C(41)	2.905(4)
C(22)...C(42)	2.929(4)
C(41)...C(61)	2.879(4)
C(42)...C(62)	2.869(4)

The intra-molecular distances given in Table 3.20 are comparable to those observed in isodrin (Kennard *et al*, 1979) and homo-isodrin (Kennard and Smith, 1984), the mean distance in these being 2.870 and 2.946Å respectively, and so the reaction scheme detailed above is plausible if the appropriate reactant can be synthesised.

The asymmetry that is noted by Paquette (1991) and others is also present in one of the cavities in this compound although the other cavity appears relatively symmetric, there being only a difference of 0.01Å between C(41)...C(61) and C(42)...C(62). The remainder of the geometry in this molecule is unremarkable, the chlorine atoms, Cl21 and Cl(22) are bent by 6.3(2)° out of the cis-ethenyl bridge plane. A similar value (6.2°) is observed for homo-isodrin itself. In parent isodrin this angle has a small negative value. This value, however, is subject to a large uncertainty, due to the low precision of the isodrin structure.

There are no notable close contacts in this structure. The packing is a classical example of a van der Waals close-packed array. The calculated density is relatively high at 1.57g/cm^3. A view down the b axis is shown in Figure 3.27.

3.7. Conclusions

The studies described in this Chapter has been carried out as a part of a wider research effort, and as such it does not attempt to answer an individual question, rather the work attempts to elaborate the amount of structural information that is available to researchers in this area. A structural study on a novel dyotropic system has been

presented. Evidence has been presented on new structural systems (Structures (7) and (8)) which could be used as dyotropomers, or precursors to dyotropic products. The evidence discredits the possibility of rearrangement in (7), but suggests that a dyotropic reactant could be formed from (8).

3.8 References

Allen, F. H.; Brammer, L.; Kennard, O. ; Orpen, A. G.; Taylor, R.; Watson, D.G.; *J. Chem. Soc. Perkin. Trans. II*, S1-S19, 1987.

Ciganek, E.; *Org. React.,* **32** ,1,1984,

Craig, D.; *Chem. Soc. Rev.,,* **16**, 187, 1987

Dong, D.C.; Wong-Ng, W.;Nyburg, S.C.; Siew, P.Y.; Edward, J.T., *Can. J. Chem.,* **62**, 452, 1984

Dorigo, A.E.; Houk, K.N.; *J. Am. Chem. Soc.,* **109**, 3698-3708, 1987

Fallis, A. G.; *Can. J. Chem.,* **62**, 183,1984

Filipinni, G.; Gavezzotti, A., *Acta Cryst. ,* **B48**, 230-234, 1992

Grimme, W. ; Geich, H.; Proske, K., J. Am. Chem. Soc. , **114**, 1492-1493, 1992

Hagenbuch, J-P; Stamfli, B.; Vogel, P., *J. Am. Chem. Soc.,* **103**, 3934, 1981

Weast, R. C. (ed.), *The Handbook of chemistry and physics,* CRC press, Boca Raton, 1984

Kennard, C.H.L.; Smith, G.; *J. Agric. Food Chem.,* **32**, 395, 1984

Kitiaigorodskii, A. I.; *Organic Chemical Crystallography;* Consultant's Bureau: New York, 1961; (English translation of the Russian original published by Press of the Academy of Sciences of the USSR, Moscow, 1955).

Mackenzie, K.; Adams, C. H. M.; *J. Chem. Soc.* , 4646, 1965

Mackenzie, K.; Howard, J.A.K.; Mason, S.; Gravett, E.C.; Astin, K. B.; Shi-Xiong, Liu; Batsanov, A.S., Djordje, V.; Maher, J. P.; Murray, M.; Kendrew, D.; Wilson, C.; Johnson, R. E. ;Preiß, T.; Gregory, R. J.; *J. Chem. Soc Perkin Trans. II*, 1211-1228, 1993

Mackenzie, K; Private communication

Menger, F.M., *Acc. Chem. Res.,* **18**, 128, 1985

Milstein, S.; Cohen, L.A. , *J. Am. Chem. Soc.,* **94**, 9158-9165, 1972

Nyburg, S.C.;Faerman, C.H, *Acta Cryst.,* **B41**, 274-279, 1985

Paquette, L.A.; Kesselmayer, M. A. ; Rogers, R. D. ; *J. Am .Chem. Soc.,* **112**, 284, 1990

Paquette, L. A. ; O'Doherty, G. A.; Rogers, R. D., *J. Am. Chem. Soc.*, **113**, 7761, 1991

Pauling, L; Delbrueck, M, *Science*, **92**, 77-79, 1940

Reetz, M. T.; *Tetrahedron,* **29**, 2189-2194, 1973

Sheldrick, G. M.; SHELXTL *Program Rev.* , 1990

Storm, D.R.; Koshland, D.E., *Proc. Natl. Acad. Sci. U.S.A.*, **66**, 445-452, 1970

Vogel, P.; Hagenbuch, J-P; Stampfli, B, *J. Am. Chem. Soc.*, **103**, 3934, 1981

Chapter 4

Novel structures of three new compounds for use in non-linear optics

4.1 Introduction

In the past three decades a large amount of industrial and academic effort has been invested in the development of new materials with enhanced optical properties. The invention of high powered laser sources has caused more attention to be paid to the harmonic effects known as "non-linear" applications. These harmonic effects are observable when a highly polarisable state of matter is subjected to an intense electric field. Under these conditions the polarisation can be written, in scalar form, as a Taylor series expansion:

$$P = P_0 + \varepsilon_0 \sum_{i=1}^{\infty} \tfrac{1}{i!} \chi^{(i)} E^i = P_0 + \varepsilon_0 \sum_{i=1}^{\infty} \chi'^{(i)} E^i \qquad (4.1)$$

In this equation, P is the overall polarisation, P_0 is the permanent polarisation of the medium, E is the applied electric field and $\chi'^{(i)}$ is the ith order susceptibility.

The susceptibility, $\chi'^{(i)}$, rapidly decreases with i due to the incorporation of the factorial term and so usually the higher order terms are neglected and only the linear term, where $i = 1$ is considered in linear optics. These higher order terms can, however, be relatively large. Any optical effect associated with cases where $i = 2$ is known as a second order effect. Equation 4.1 shows that the magnitude of the polarisation terms is dependant on the applied field. In non-constant fields the polarisation will also vary as a function of time. Electromagnetic theory tells us that this creates a field of its own in response to the moving charge. Consider an applied field, E, where:

$$E = E_0 \cos(\omega t) \qquad (4.2)$$

The polarisation, by equation 4.1 can then be written as:

$$P = P_0 + \varepsilon_0 \chi'^{(1)} E_0 \cos(\omega t) + \varepsilon_0 \chi'^{(2)} E_0^2 \cos^2(2\omega t) \qquad (4.3)$$

The second induced term here, $\varepsilon_0 \chi'^{(2)} E_0^2 \cos^2(\omega t)$ can be re-written as:

$$P^{(2)} = \tfrac{1}{2} \varepsilon_0 \chi'^{(2)} E_0^2 (1 + \cos(2\omega t)) \qquad (4.4)$$

Which implies that the application of a field E of frequency ω, induces two new fields, one D.C., and one of frequency 2ω. This phenomenon is known as second harmonic generation and has a number of lucrative industrial applications. In laser optics such an effect can be utilised to produce blue laser light from red. Different effects (such as the electro-optic effect) are achievable by applying different fields to a system. These, in turn, have their own uses. Some attention is also now being paid to the third harmonic terms that are present in equation 4.1. These also may well have their uses, particularly since third order responses are not dependent upon the medium being non-centrosymmetric. This requirement for a non-centrosymmetric medium[†] is a major obstacle for the design of crystalline materials with enhanced second order non-linear optical properties. In centric media the term $\chi'^{(2)}$ is zero.

The primary objective of any attempt to gain media with favourable second order non-linear properties is thus to gain a stable non-centric medium that is highly polarisable. The polarisability of a medium is governed primarily by the ease with which charge can move within the molecules in the medium. The molecular polarisability takes a similar form to that of the polarisability of the medium:

$$P_{mol} = P_{0,mol} + \varepsilon_0 \sum_{i=1}^{\infty} \zeta'^{(i)} E^i \qquad (4.5)$$

[†] Modern approaches tend not to utilise crystals. A common approach incorporates highly polarizable molecules into Langmuir-Blodgett films.

The second order term in this expansion, $\zeta'^{(2)}$, has been derived (Oudar, 1977) as:

$$\zeta'^{(2)} = C.\frac{\omega_0 \Delta\mu_{gn}}{\left[(\hbar\omega_0)^2 - (\hbar\omega)^2\right]\left[(\hbar\omega_0)^2 - (2\hbar\omega)^2\right]} \qquad (4.6)$$

In this expression, ω_0 is the resonant frequency associated with the molecule, as derived from the anharmonic oscillator model, $\Delta\mu_{gn}$ is the change of dipole moment from the ground to excited state, and C is a constant associated with the system.

Equation 4.6 shows that $\zeta'^{(2)}$ has two poles, one where $\omega = \omega_0$, and a second where $2\omega = \omega_0$. Thus, in principle, high molecular polarisabilities are attained near these frequencies, but they both coincide with high absorption. This creates a problem. When a high non-linear response to a field is observed, a large amount of the fundamental or second harmonic can be absorbed by the medium, thus the high response is severely attenuated. An ideal sample has a very sharp absorption profile, such that absorption near the pole is low, but at the pole it is high, thus a relatively high $\zeta'^{(2)}$ can be obtained with only relatively low losses to absorption of the generated second harmonic or fundamental fields.

Due to the obvious applications associated with the physical effects described above, a large number of potential compounds have been studied, and new methods for utilising their bulk properties have been proposed. A recent publication (Ashwell, Bryce et al, 1990) has discussed a number of new molecules based on TCNQ-like moieties, as shown in Figures 4.1(a), (b) and (c)

In these systems the zwitterionic fragments have been spliced onto long alkyl chains (the R-groups) and the resulting compounds have been formulated into Langmuir-Blodgett films by described techniques (Ashwell, Bryce *et al*, 1990 and references therein).

Figure 4.1(a), (b), (c) Schematic representation of compounds studied previously using Langmuir-Blodgett films.

a

b

c

X and Y in these molecules represent halogen atoms.
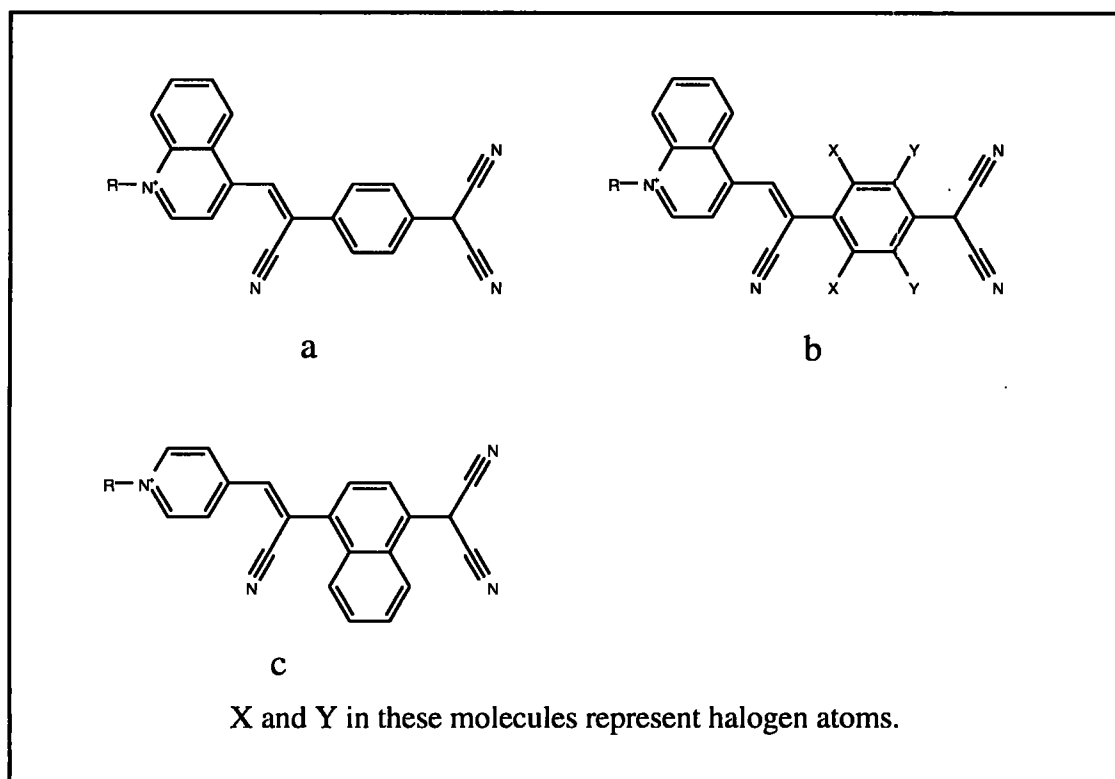
Figure 4.2. Resonance forms for the carbon-nitrogen backbone of the systems studied.
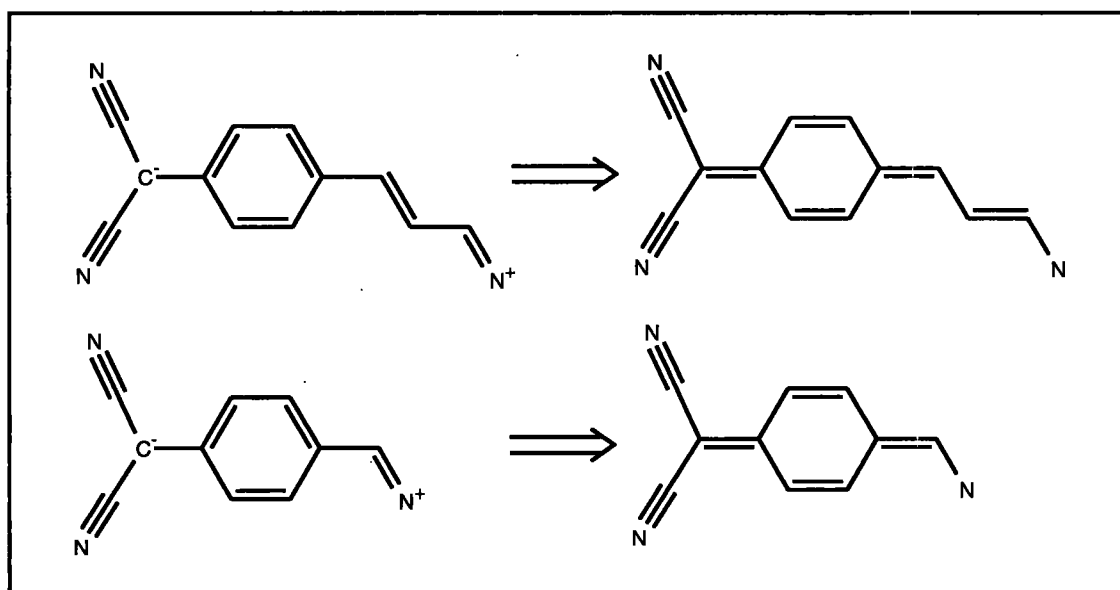
Studies of these films has proved that they show encouraging second harmonic properties.

In this chapter, the crystal structures of three similar novel compounds that are based on the TCNQ type derivatives are elucidated (see Figure 4.4(a), (b) and (c)). The compounds have very high polarisabilities due to the possible quinonoidal form, which represents an excited molecular state of these systems (see Figure 4.2).

These compounds are ideal for second harmonic generation not only because of their high polarisabilities, but also because their absorption is low in the blue part of the visible spectrum.
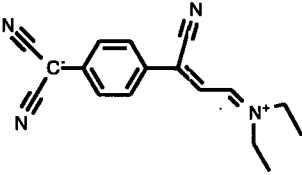
4.2 Results

The structures of three compounds have been elucidated by the author. Schematic diagrams of these three compounds, (1), (2) and (4), are shown below in Figure 4.3, along with two other compounds, (3) and (5). Structures of (3) and (5) have been elucidated by others (Metzger, 1984; Cole, 1995).
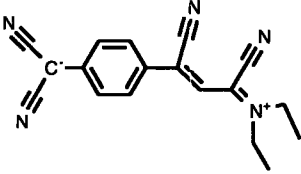
Figure 4.3. Schematic representations of compounds (1)-(5)

76
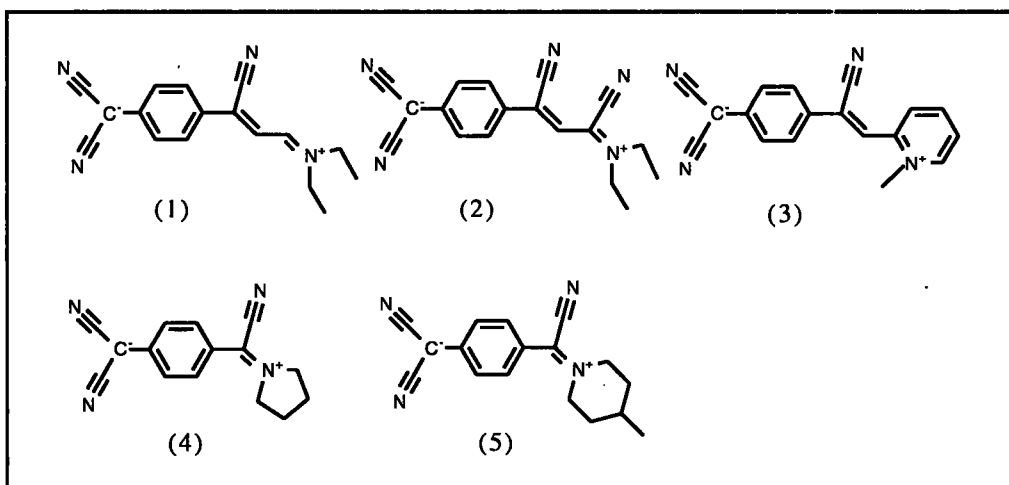
Table 4.1. Experimental details for structure (1)
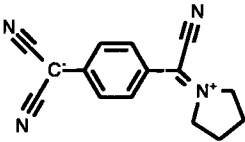
Chemical Details	
Empirical formula: $C_{17}H_{16}N_4$	Chemical diagram
Formula weight: 276.34	
Diffractometer Experimental details	
Unit Cell	Data collection parameters
a = 7.221(2)Å	Crystal size: $0.10 \times 0.20 \times 0.50$mm
b = 7.2620(10)Å	Collection method: $\omega/2\theta$
c = 14.670(2)Å	Speed of collection: 1-8°/min
α = 76.97(1)°	Scan width: $1.418 + 0.3\tan\theta$
β = 82.28(1)°	θ range for collection: 2.7 to 26.0°
γ = 79.40(1)°	Index ranges :
Volume = 733.2(2)Å^3	$-8 \leq h \leq 8$
Based on 25 reflections between 18.6	$-8 \leq k \leq 0$
to 23.05° 2θ	$-17 \leq l \leq 17$
	Temperature: 150(2) K
Refinement details	
Space group: P$\bar{1}$	Z = 2
ρ_c= 1.252 g/cm^3	Absorption coefficient: 0.077 mm^{-1}
F(000): 292.0	Goodness-of-fit on F^2: 1.056
Reflections collected: 2816	Final R indices [I>2σI)]	
Independent reflections: 2592	R1 = 0.0414	wR2 = 0.0926
R_{int} = 0.0301	R indices (all data)	
Data: 2591	R1 = 0.0948	wR2 = 0.1131
Restraints: 0	Largest diff. peak and hole	
Parameters: 255	0.193 and -0.181 eÅ^{-3}	
Hydrogen Treatment: The hydrogen positions and temperature factors were refined without constraint.	

Table 4.2 Experimental details for structure (2)

Chemical Details	
Empirical formula: $C_{18}H_{15}N_5$	Chemical diagram
Formula weight: 301.35	

Diffractometer Experimental details	
Unit Cell	Data collection parameters
$\mathbf{a} = 7.615(4)$ Å	Crystal size: $0.2 \times 0.5 \times 0.5$
$\mathbf{b} = 7.962(4)$ Å	Collection method: $\omega/2\theta$ scans
$\mathbf{c} = 14.740(4)$ Å	Speed of collection 2-16°/min
$\alpha = 91.45(4)°$	Scan width: $1.47 + 0.3\tan\theta$
$\beta = 96.93(4)°$	θ range for collection: 2.80 to 27.50°
$\gamma = 118.03(4)°$	Index ranges :
Volume = 779.7(6) Å^3	$-9 \leq h \leq 8$
Based on 25 reflections in the range	$0 \leq k \leq 10$
37.3 and 43.7° 2θ	$-19 \leq l \leq 19$
	Temperature: 150(2) K

Refinement details		
Space group: P$\bar{1}$	Z = 2	
$\rho_c = 1.284$ g/cm^3	Absorption coefficient: 0.081 mm^{-1}	
F(000): 316	Goodness-of-fit on F^2: 1.034	
Reflections collected: 3847	Final R indices [I>2σI)]	
Independent reflections: 3590	R1 = 0.0381	wR2 = 0.1009
$R_{int} = 0.0090$	R indices (all data)	
Data: 3589	R1 = 0.0706	wR2 = 0.1158
Restraints: 0	Largest diff. peak and hole	
Parameters: 269	0.280 and -0.183 eÅ^{-3}	
Hydrogen Treatment: Hydrogen positions and isotropic temperature factors were refined without restraint or constraint.		

Table 4.3. Experimental details for structure (4)

Chemical Details	
Empirical formula: $C_{15}H_{12}N_4$	Chemical diagram
Formula weight: 248.29	
Diffractometer Experimental details	
Unit Cell	Data collection parameters
a = 6.8370(10)	Crystal size: $0.05 \times 0.15 \times 0.3$ mm
b = 25.080(5)	Collection method: ω scans
c = 7.3860(10)	Speed of collection 1.3-8°/min
	Scan width: $1.313 + 0.3\tan\theta$
β = 92.70(3)°	θ range for collection: 2.88-25°
	Index ranges :
Volume = 1265.1(4)Å^3	$-8 \le h \le 8$
Based on 15 reflections in the range 20.6 -	$-29 \le k \le 0$
25.0° 2θ	$-7 \le l \le 0$
	Temperature: 293(2)K
Refinement details	
Space group: $P2_1/n$	Z = 4
ρ_c= 1.304 g/cm^3	Absorption coefficient: 0.082 mm^{-1}
F(000): 520	Goodness-of-fit on F^2: 0.920
Reflections collected: 2233	Final R indices [I>2σI)]
Independent reflections: 2057	R1 = 0.0586 $\quad$ wR2 = 0.1137
R_{int} = 0.0945	R indices (all data)
Data: 2056	R1 = 0.3495 $\quad$ wR2 = 0.1943
Restraints: 3	Largest diff. peak and hole
Parameters: 173	0.232 and -0.216 eÅ^{-3}

Hydrogen Treatment: hydrogen atoms were placed at idealised positions relative to their parent atom. Their isotropic temperature factors were fixed at a value 1.2 times that of the parent atom's isotropic equivalent temperature factor.

Figure 4.4(a), (b) and (c). Computer simulated representations of (1), (2) and (4)

(a) Molecule (1)
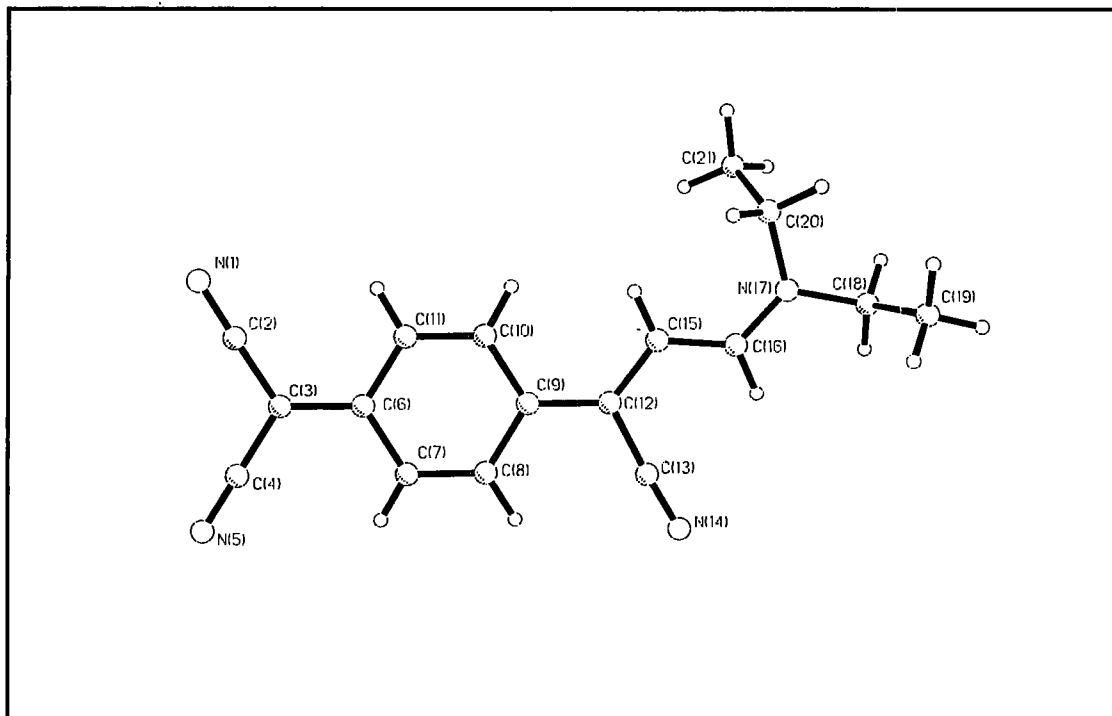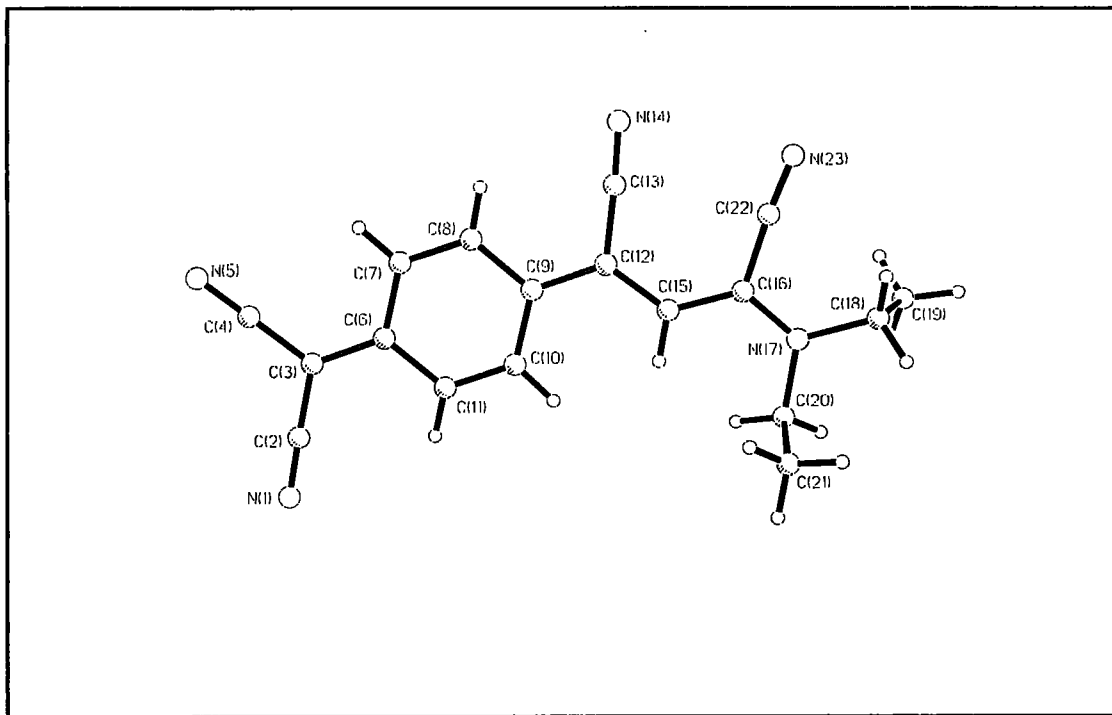
(b) Molecule (2)

(c). Molecule (4)
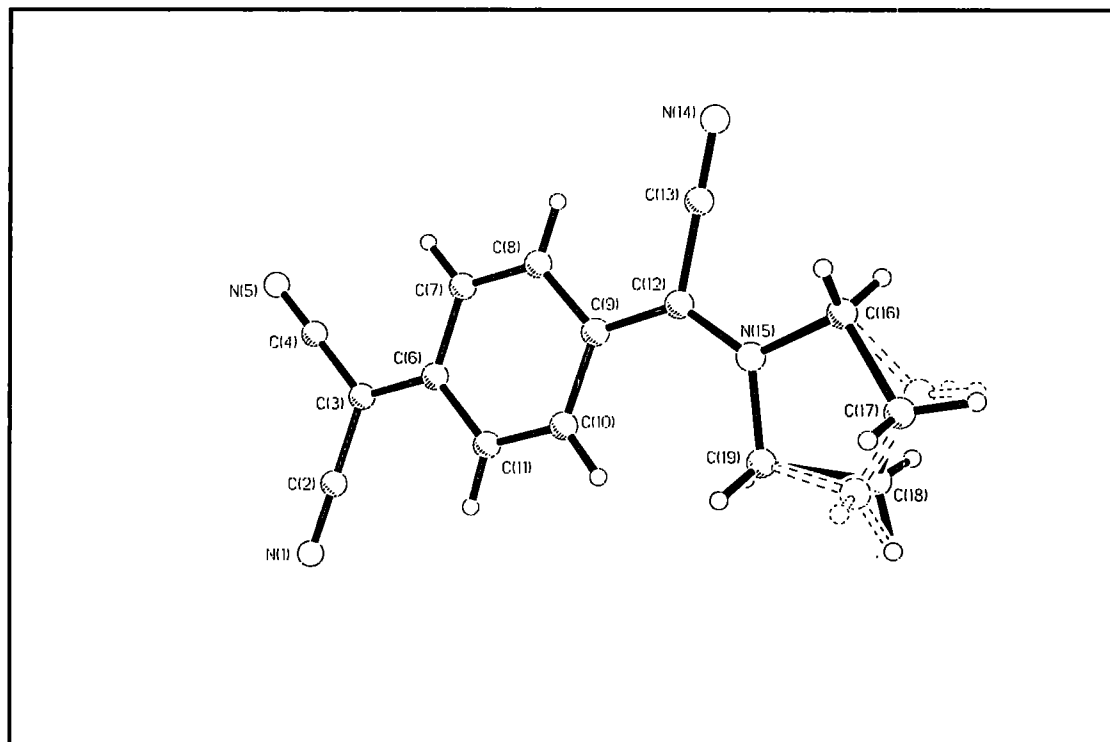
Bond lengths and angles, are given in a tabular format in Appendix 4A. Experimental details for the three structural studies presented in this Chapter are given below in Tables 4.1 to 4.3. No absorption corrections were applied. One special feature is worth mentioning separately from the Tables: the five membered ring in structure (4) is disordered. Two alternative envelope conformations exist and both occur in the crystal structure, as shown in Figure 4.4(c). The minor conformation is shown by the dotted lines. The occupancies of the major and minor conformations are 73.3% and 26.7%. Computer generated representations of molecules 1, 2 and 4 are shown in Figures 4.4(a), (b) and (c).

The primary area of structural interest in these molecules is the extent of conjugative interactions within the molecule. Above, in Table 4.4(a) and below in Table 4.4(b), the bond lengths along the chain are shown for all five compounds. The chain length varies from molecule to molecule.

Compounds (1) and (2) show a progression towards the more quinonoidal resonance form as a result of the addition of a second cyano substituent on the dienyl chain. Compound (1) is half way between zwitterionic and quinonoidal. The 6-membered ring from C(6) to C(11) has two short bonds, C(7)-C(8) and C(11)-C(10), which are comparable to the mean value of 1.353(10)Å that has been observed from studies using the Cambridge Structural Database (Allen *et al*, 1987). This would suggest that the best description would be the quinonoidal form, but for this to be the case one would expect a single bond between C(16) and N(17). The distance is more comparable, however, to a C=N bond. (For example, in furoxan the observed mean cited is 1.316(9)Å (Allen et al, 1987)). The C-C bonds within the molecular chain are also a little confusing. They are all comparable to the mean exocyclic C=C bond lengths observed in TCNQ of 1.392(17)Å (Allen et al 1987). The pattern of bond lengths observed in the chain is consistent with a zwitterion, contrary to the conclusions one would draw from the 6-membered ring.

Table 4.4(a). Conjugation in (1),(2) and (3)

Bond	Structure		
	(1)	(2)	(3)
C(3)-C(6)	1.419(3)	1.404(2)	1.440
C(6)-C(11)	1.424(3)	1.438(2)	1.411
C(6)-C(7)	1.419(3)	1.428(2)	1.398
C(7)-C(8)	1.364(3)	1.361(2)	1.374
C(11)-C(10)	1.361(3)	1.356(2)	1.366
C(8)-C(9)	1.422(3)	1.435(2)	1.404
C(10)-C(9)	1.417(3)	1.431(2)	1.396
C(9)-C(12)	1.427(3)	1.418(2)	1.467
C(12)-C(15)	1.389(3)	1.411(2)	1.355
C(15)-C(16)	1.397(3)	1.393(2)	1.458
C(16)-N(17)	1.315(2)	1.347(2)	1.383[†]

[†] In compound 4 the C-N bond is part of a pyridinium ring, rather than a conjugated chain bond hence this bond is considerably longer than its counterparts in structures (1) and (2).

Table 4.4(b) Conjugation in (4) and (5)

Bond	Structure	
	(4)	(5)
C(3)-C(6)	1.405(9)	1.427(4)
C(6)-C(7)	1.426(8)	1.415(4)
C(6)-C(11)	1.421(8)	1.419(4)
C(7)-C(8)	1.351(8)	1.363(4)
C(11)-C(10)	1.363(8)	1.361(4)
C(8)-C(9)	1.435(9)	1.430(4)
C(10)-C(9)	1.424(9)	1.418(4)
C(9)-C(12)	1.397(9)	1.428(4)
C(12)-N(15)	1.333(9)	1.321(4)

Molecule (2) is more obviously quinonoidal in nature. The backbone C-N bond with a length of 1.347(2)Å is comparable to the mean value for single bonds between formally Csp^2 and tertiary Nsp^2 which is 1.358(14)Å (Allen et al, 1987). The conjugated backbone bond lengths are also more consistent with the quinonoidal form in this molecule.

The presence of the cyano groups substantially increases the conjugation along the chain. The mean C-C single and double bond lengths in trienyl fragments in the CSD have been calculated as 1.445(13) and 1.345(12)Å respectively (Allen et al, 1987). This is important in molecules that are for use in non-linear optical applications. Studies on polymeric systems have shown that hyperpolarisability is improved as a function of chain length, but the additional benefit of extra carbon atoms depreciates the longer the chain length, because the system is not fully conjugated (see, for example, Prasad, Williams, 1991).

Figure 4.5(a), (b), (c) UV-spectra for molecules (1),(2) and (4).

(a). Molecule (1)
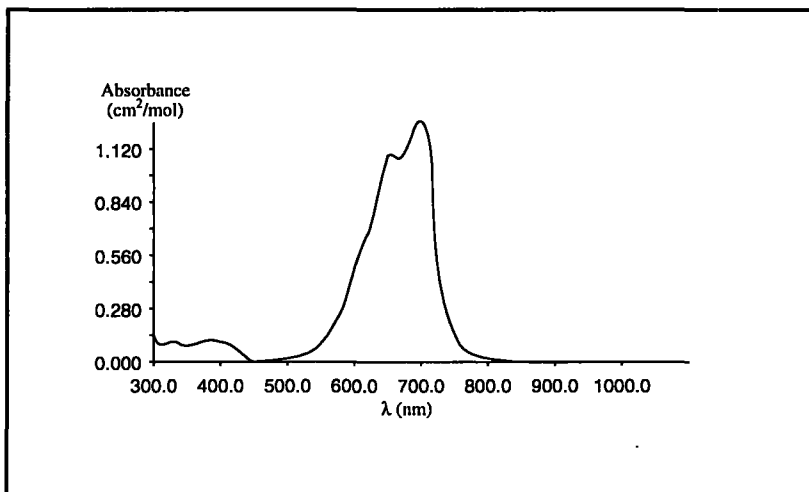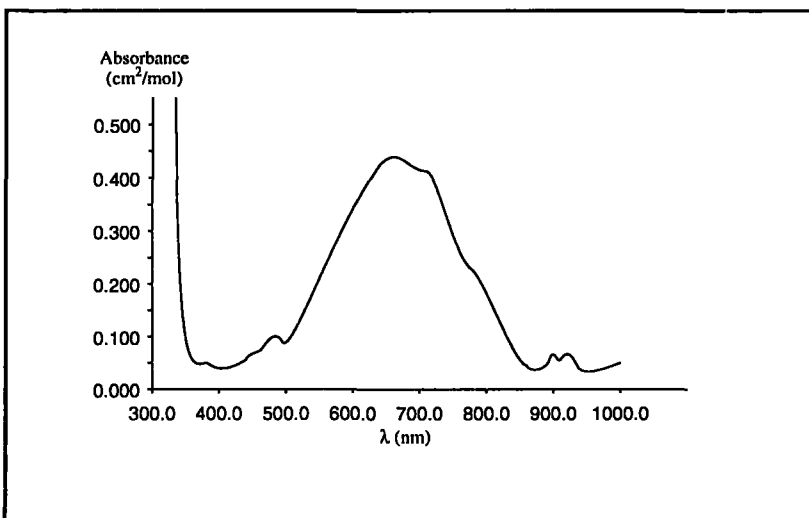
(b). Molecule (2)
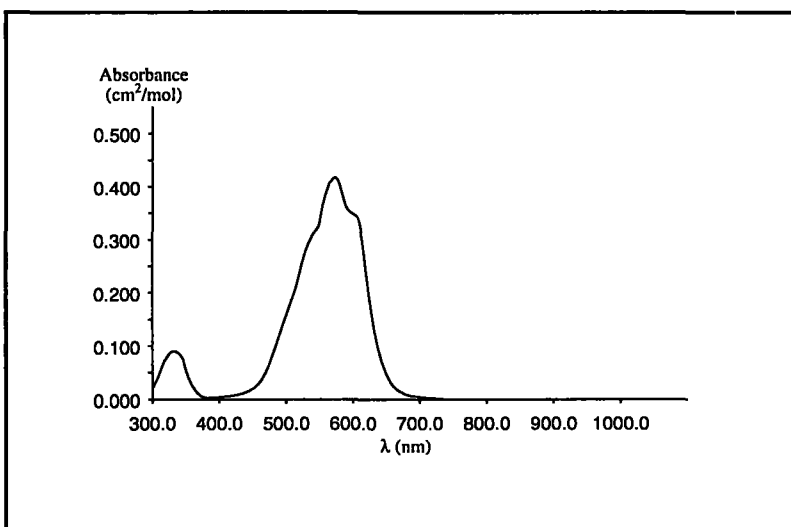
(c). Molecule (4)

Molecule (3) is zwitterionic rather than quinonoidal. This is probably due to the additional aromaticity present in molecule (4) due to the pyridinium ring. The substitution of the pyridinium ring with an α methyl group also forces the pyridinium ring out of plane relative to the carbon backbone of the molecule. This combination of effects stabilises the zwitterionic form.

Molecules (4) and (5) appear to be more quinonoidal in nature rather than zwitterionic based on the 6-membered ring bond lengths between atoms C(7)-C(8) and C(10)-C(11) in each molecule (see Table 4.4(b)). They are again comparable to the mean endocyclic bond distances observed in TCNQ. The backbone C-N bonds in the two compounds, C(12)-N(15) are relatively short in a similar way to compounds (1) and (2).

The benefit of a highly conjugated system is that it provides a high hyperpolarisability due to the density of electronic states at a low energy. One drawback of this is that the molecules efficiently absorb light at a frequency where the hyperpolarisability is at its highest. Equation 4.1 shows the dependency of $\zeta^{(2)}$ on frequency; it is desirable to have an incident field as close to the fundamental frequency as possible, exactly where absorption is worst. The compromise is to find a compound that has a narrow absorption band, and thus has low absorption at a frequency close to the resonance frequency. The UV spectra of molecules (1),(2) and (4) are shown in Figure 4.5. As is clear from these three diagrams two of the three molecules have low absorption. In molecule (1), the window is between 450 and 470 nm. In (4), a similar window is seen between 380 and 420 nm. In (2), absorption occurs across the whole range of the visible spectrum, although the absorption is lower between 350 and 500 nm. This "blue window" in compounds (1), (2) and (4) makes them ideally suited to non-linear optical applications. The molecular hyperpolarisability has been calculated for molecule (1). The value obtained is in the range of 400 to 500×10^{-30} esu (Cole, Cross et al, 1995), a very high value as compared to urea, which has a hyperpolarisability of 0.45×10^{-30} esu.

The packing of these zwitterionic molecules is obviously of key importance with respect to second order non-linear optical applications if the proposed bulk medium is the crystal, due to the key requirement that the crystalline lattice should be non-centrosymmetric. Also of importance are the directions of the components of the molecular hyperpolarisability tensor within the unit cell. In most conjugated molecules this tensor is primarily one dimensional: it only has a large component in the direction of the conjugation. If a molecular symmetry element aligns this component in the wrong way, such that the principal components are oppositely aligned in adjacent molecules, then the overall bulk second order hyperpolarisability will be low, due to the localised cancelling effects. Optimum molecular orientations for all suitable crystal point groups have been proposed (Oudar and Zyss, 1982) The primary aim, however, is to break the inversion symmetry.

Unfortunately all the molecules elucidated here are in centrosymmetric space groups, but the latest compound elucidated within the laboratory, (molecule (5): Cole, 1995.) crystallises in space group Pn.

Figure 4.6(a), (b), (c), (d) Dimeric pairs in structures (1), (2), (4) and (5)

(a) Structure (1)
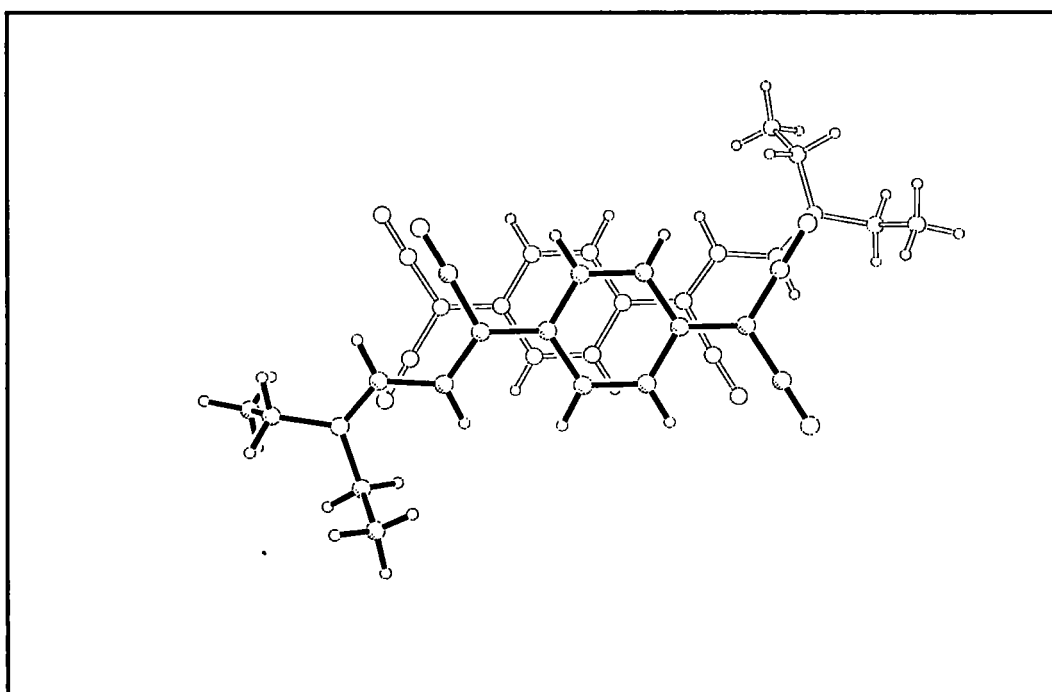
(b). Structure (2)
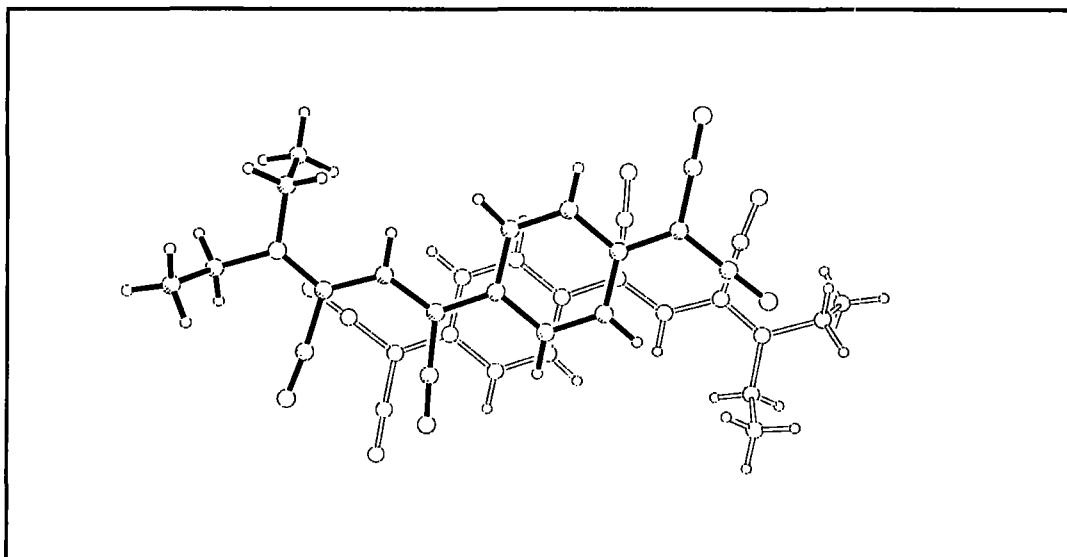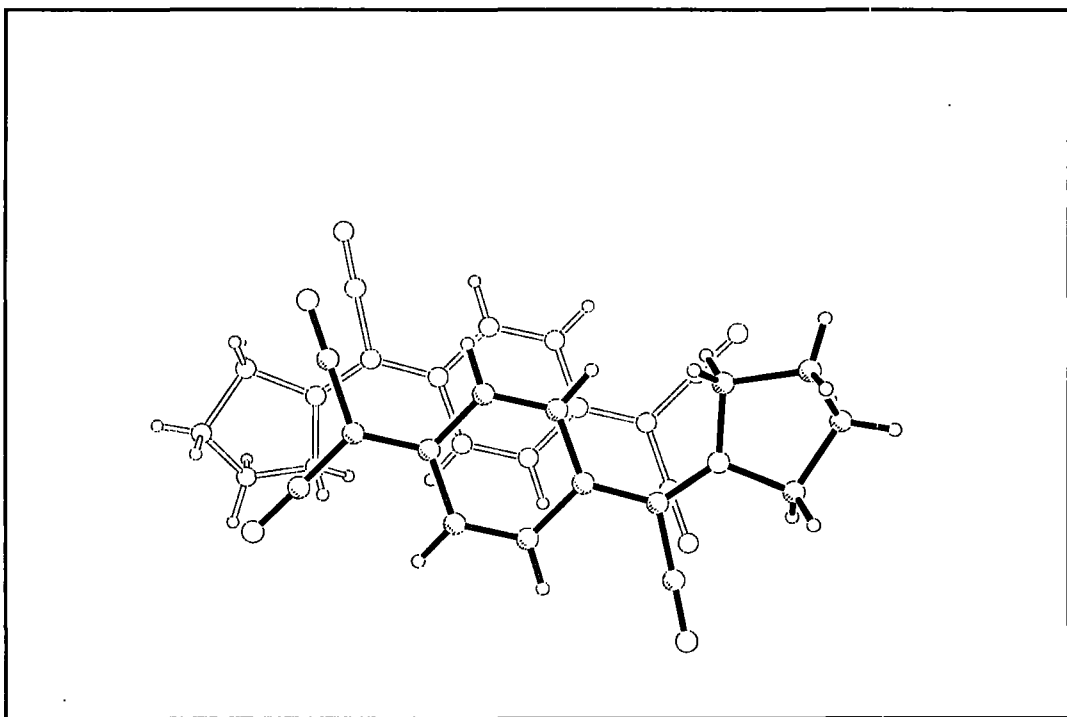
(c). Structure (4)

(d) Structure (5)
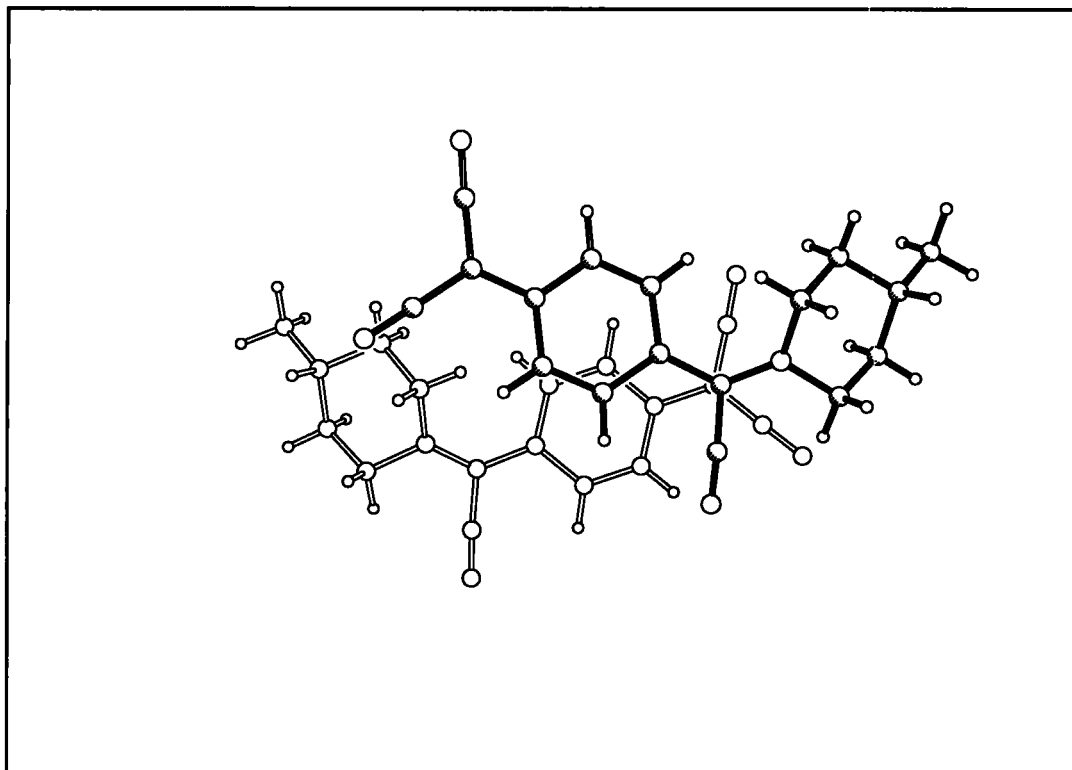
The predominance of centrosymmetric structures in these zwitterionic molecules is probably due to the energetic benefit of the proximity of opposite charges within the lattice, although it is inappropriate to write a formal charge on any one. The molecules form dimeric pairs within the lattice. These pairs are shown for molecules (1),(2) and (4) in Figure 4.6(a), (b) and (c). Structure (5) also forms a dimeric pair, but not around an inversion centre. This dimeric pair is shown in Figure 4.6(d); again, the proximity of the charged centres is probably the driving force in this system. The inversion dimer is not formed due to steric influences of the substituted 6-membered ring. The packing utilised minimises the effect of this. Also notable in this particular stack is that the quinonoidal 6-membered rings are twisted with respect to one another.

Figure 4.7(a), (b), (c). Packing of layers in structures (1), (2) and (4)

(a) Structure (1)
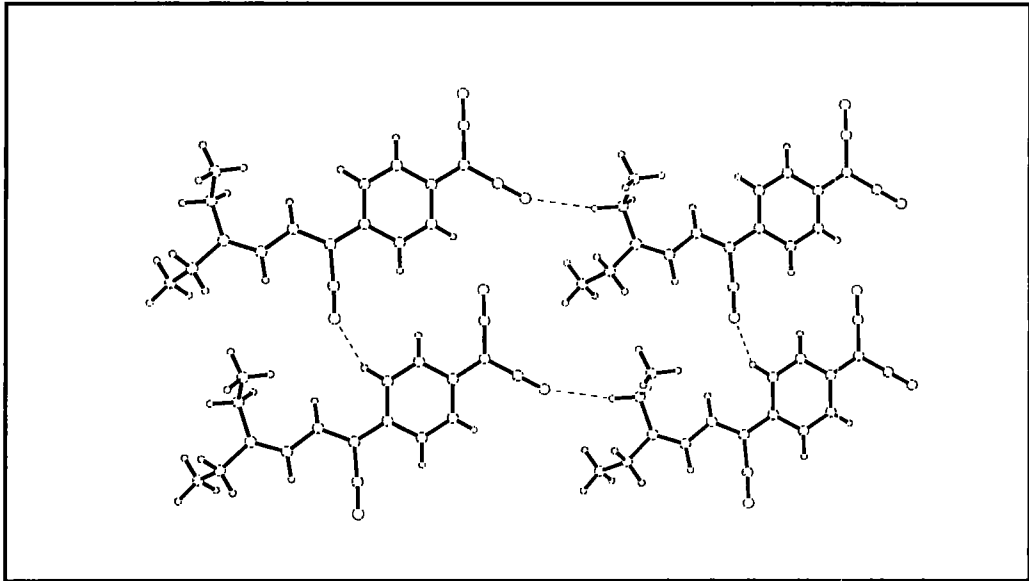
(b). Structure (2)
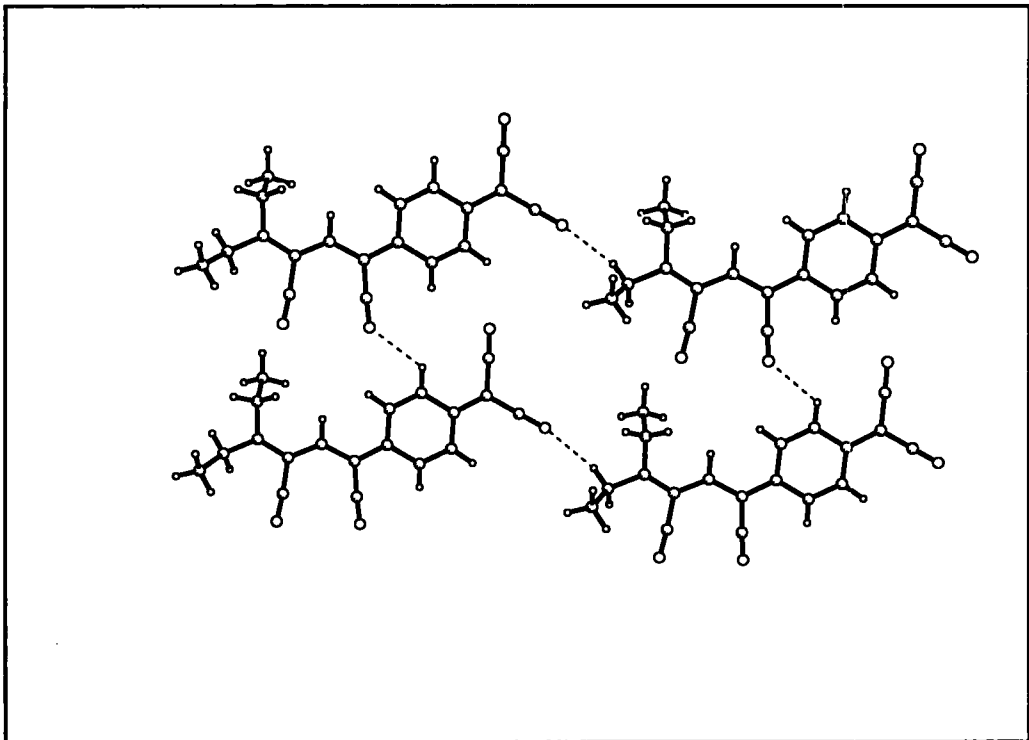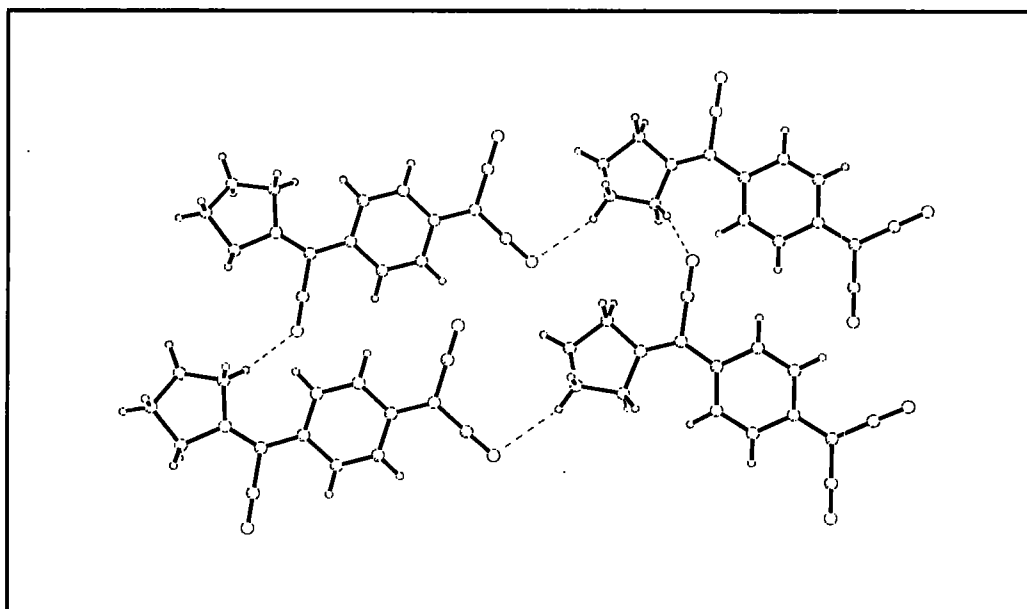
(c). Structure (4)

Table 4.5. Close contacts in structures (1), (2) and (4). Only contacts of ~2.6Å and smaller are given.

Contact	Distance [Å]	μ [°]	τ [°]
Structure (1)			
N(5)...H(201) (x, y, z+1)	2.560(24)	21.9(5)	16.5(17)
N(14)...H(10) (x+1, y, z)	2.587(20)	46.4(5)	37.1(15)
N(1)...H(202) (-x-1,1-y,1-z)	2.511(24)	52.4(5)	23.6(17)
Structure (2)			
N(14)...H(11) (x, y-1, z)	2.593(17)	53.4(13)	44.3(4)
N(5)...H(181) (x, y, z-1)	2.526(18)	32.3(4)	14.2(14)
Structure (4)			
N(14)...H(19A) (x-1, y, z)	2.542(9)[†]	58.3(7)[†]	13.0(2)[†]
N(5)...H(18A) (3/2 -x, y -1/2, 1/2 - z)	2.608(10)[†]	76.6(6)[†]	31.3(3)[†]

[†] The e.s.d is smaller in this structure, because the hydrogen atoms were placed at geometricaly idealised positions.

The crystal structures of (1),(2) and (4) are held together by arrays of hydrogen bonds. The precision of the hydrogen positions in structures (1) and (2) is high enough for confidence in the lengths of the C≡N..H contacts that are noted in Table 4.5. Some tentative contacts were noted in structure (4), but the hydrogen atoms were added at idealised positions, and the disordering in this compound means that little confidence can be ascribed to the distances observed.

The distances given in Table 4.5 are all considerably shorter than the sum of the associated van der Waals radii, (1.2Å for hydrogen, 1.55Å for nitrogen: Bondi, 1964). Structures (1), (2) and (4) all pack in layers, as is seen in Figure 4.7, maximising these interactions. The third interaction given for structure (1) (between N(1) and H(202)) is not shown in Figure 4.7(a). This acts between the layers of molecules, rather than within the layer.

Figure 4.8. Possible layer dipole arrangements.
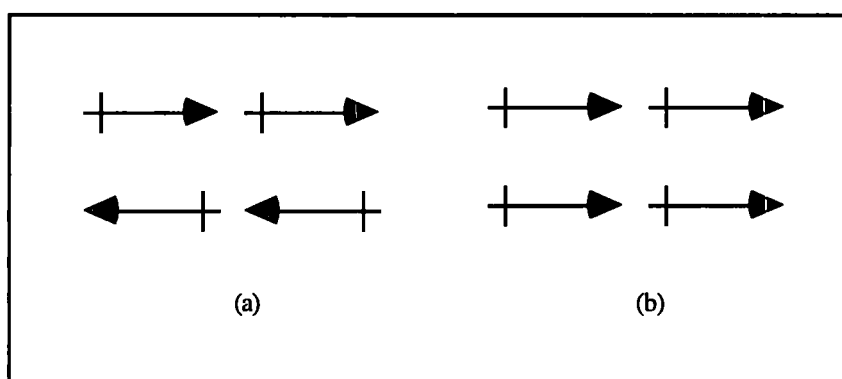
The intermolecular interactions noted for structures (1), (2) and (4) are classical examples of C≡N..H hydrogen bonds. Systems with C≡N..X intermolecular contacts in the CSD (where X is Cl, Br or I) have been studied by Desiraju and Harlow (1989). The contacts in structures (1), (2) and (4) could be considered to be structure determining. If one only considered potential zwitterionic nature of the molecules in this study, one would expect the molecular dipoles to be aligned in an *anti-parallel* fashion (see Figure 4.8(a)), whereas in reality the dipoles are *parallel* to one another (see Figure 4.8(b)), showing that the hydrogen bonds are strongly influencing the structure. (This also suggests that the

molecules are more quinonoidal than zwitterionic, in keeping with earlier arguments based on bond lengths). The anti-parallel arrangement is seen from one layer to the next, as can be inferred from Figure 4.6.

4.3 Conclusions

This study of three highly conjugated molecules has provided some explanation as to the nature of their interesting ultra-violet spectra. Most importantly, the three-dimensional structure has been revealed to the workers studying these systems. These co-ordinates are now being used in theoretical calculations to predict the molecular hyperpolarisabilities that would be observed. Such data provides an insight into the optical properties of Langmuir-Blodgett films that incorporate such structural moieties.

The work also confirms the mode in which these zwitterionic molecules tend to pack: primarily as close packed sheets of molecules stacked via an inversion centre such that pairs of molecules can form "inversion dimers" as has been suggested by Wilson (1993).

4.4 References

Allen, F. H.; Brammer, L.; Kennard, O. ; Orpen, A. G.; Taylor, R.; Watson, D.G.; *J. Chem. Soc. Perkin. Trans. II*, **S1**, 1987

Aswell, G. J.; Bryce, M. R.; Dawnay, Grainger, A. M.; Hasan, M.; E. J. C.; Kuczynski, A. P.; Szablewski, M.; Sandy, I. M.; *J. Chem. Soc. Faraday Trans.* 1117-1121, **86**, 1990

Bondi, A.; *J. Phys. Chem.* **68**, 441, 1964

Cole, J. C.; Cross, G.H.; Howard, J. A. K.; Szablewski, M.; *Acta Cryst.*, **C50**, 715, 1995

Cole, J.M. *Personnal communication.* University of Durham, 1st year report, 1995.

Desiraju, G. R.; Harlow, R. L.; *J. Am. Chem. Soc.*, **111**,6757-6764, 1989

Metzger, R. M., Heimer, N. E.; Ashwell, G. J.; *J. Mol. Cryst. Liq. Cryst.*, **107**, 133-149, 1984

Oudar, J. L.; Chemla, D. S.; *J. Chem Phys*, **66**, 2664, 1977

Oudar, J. L.; Zyss, J.; *Physical Review A*, **26**(4), 2028, 1982

Prasad, P. N., Williams, D. J.; *Introduction to Non-linear Optical Effects in Molecules and Polymers*, Wiley Press, 1991

Wilson, A. J. C. *Acta Cryst.* **A46**, 795-806, 1993

Chapter 5.

The Cambridge Structural Database.

5.1. Introduction

By the 1940's crystallography was well established as a primary tool for detailed analysis of structure at atomic resolution. In that time a great deal of chemical knowledge was gained through systematic analyses of related structures (see for example, Pauling, 1940; Pimental & McClellan, 1960; Sutton, 1958,1965). Chemists were able to find and explain trends in fundamental chemical quantities, such as bond lengths and angles and molecular constitution which gave rise to a more detailed understanding of chemistry. Such analyses were only possible due to the easy accessibility of structural data.

Over the following two decades, the quantity of structural studies increased. This increase occurred for a number of reasons. Firstly, experimental equipment became more widely available, due to companies independently producing machines for the specialist market that developed. Secondly many experimental techniques were automated, which meant that data was more easily collected for a particular structure.

To efficiently record the ever-increasing volume of structural data available, a number of institutions produced independent collations of crystallographic data. *"Wyckoff's Crystal Structures "* and *"Structure Reports"* were examples of secondary publications, which began in the late 1930's, that summarized the primary literature and provided crystallographers with classified sources of structural information in printed form.

By the mid to late 1960's computers began to be more generally accessible to most independent laboratories. Also, the arrival of direct methods meant that many more compounds could be solved routinely and refined in shorter periods of time. The

Figure 5.1 The Growth of the CSD since 1965.
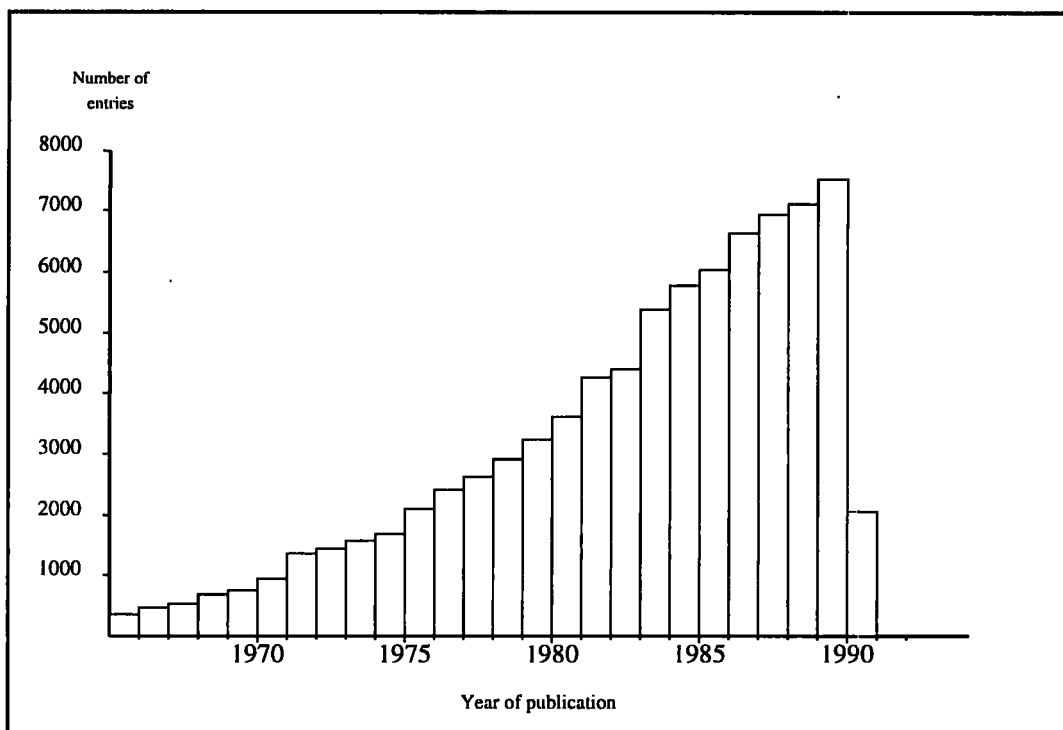
explosion of structural information that resulted (see Figure 5.1) meant that it was a cumbersome job, even with specialised crystallographic journals, to retrieve sets of related chemical structures, hence the data began to be collated into specialised data bases which could be searched computationally.

The Cambridge Structural Database (CSD, Allen *et al*, 1991) is one of four fully retrospective databases which fulfils this role. Since 1965 structural details have been stored in the CSD for all compounds of an organic and metallo-organic nature, along with information which describes the chemistry of the molecular system. Also stored are details of the quality of the structural data, such as the R-factor. The ability to automatically locate and retrieve structural data *via* the CSD has caused a resurgence of interest in systematic studies of chemical structure, which reveal both novel and fascinating aspects of molecular and solid state chemistry.

Figure 5.2: A histogram showing the increase in the number of publications utilising the CSD since 1965.
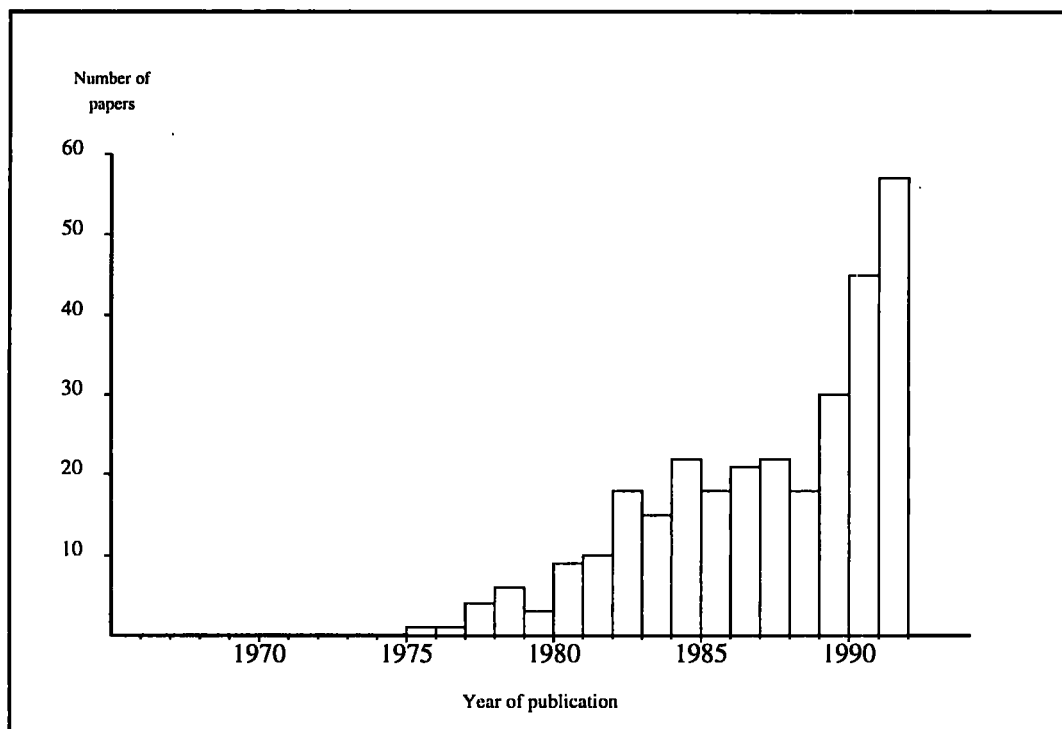
Three other databases exist to store information on other types of compounds. The inorganic chemical structural database (ICSD), as the name would suggest, stores compounds of an inorganic nature (Bergerhoff et al, 1983). Large biological molecules are stored on the protein data bank (PDB,). Metals and alloys are stored in the metals data file (MDF). In total some 250,000 structures are now stored between these four primary databases, representing a vast amount of academic and industrial research.

The re-kindling of interest in systematic studies of structural data can clearly be seen in Figure 5.2, where the research usage of the CSD is shown as a histogram. This histogram has been derived from a new component of the CSD recently introduced: a database of database use, i.e. of references to publications that use the CSD as their data source. (Allen, Kennard and Watson, 1995). Figure 5.2 shows that there was a steady increase in the number of database-related publications until a plateau of ~20 papers per

year was reached in 1985. Since the introduction of more interactive software, the number of CSD related publications has sharply increased, reaching 56 papers published in 1991. This upward trend might, perhaps, be compared to that of the early years of structure determination, shown in Figure 5.1.

5.2 Information Content of the CSD.

The CSD contains structural information from over 700 different journals. Over 20 major journals utilise the CSD as their official depository for unpublished coordinate data. The structures within the database are either "organic" or "organometallic" in nature. The definitions of organic and organometallic are fairly wide ranging: compounds are included if they contain C-C or C-H bonds.

The CSD store three distinct regions of data. These are defined as the one, two and three dimensional data respectively (see Figure 5.3). These are summarised in Sections 5.2.1-3. One important concept is the "*field*". A field is a region of data which contains one explicit data type, for example the bond properties field in the two dimensional information would contain the bond orders of all the bonds in the structure.

5.2.1. One dimensional information content

The one dimensional information content is summarised in Tables 5.1(a) and (b). For each entry a number of bibliographic, numerical and text-based data fields are stored which allow the user quick access to important crystallographic and chemical details, as well as the obvious benefit of providing the citation to the primary literature.

Table 5.1(a). Written one dimensional data content on the CSD

REFCODE	Any previous REFCODE
Journal of publication	Journal volume
Journal Page Number	Space Group Symbol
Formula Sort Key	Abstracted Formula
Compound name	Compound name qualifier
Compound name synonym	Authors' names
Comment on connectivity	"Remarks" text
"Disorder" comment	"Error" comment
Properties text	

Figure 5.3. Information content of an entry in the Cambridge Structural Database.

1D information

ABUCUP
bis(1,3-Diaminobutane) copper(ii)
perchlorate
blue-violet form
C8 H24 Cu1 N4 2+,2(Cl1 O4 1-)
A.Pajunen,K.Smolander,
I.Belinskij
Suom.Kemistil.B, 45, 317,1972

2D information

CSD

3D molecular information

3D crystallographic information
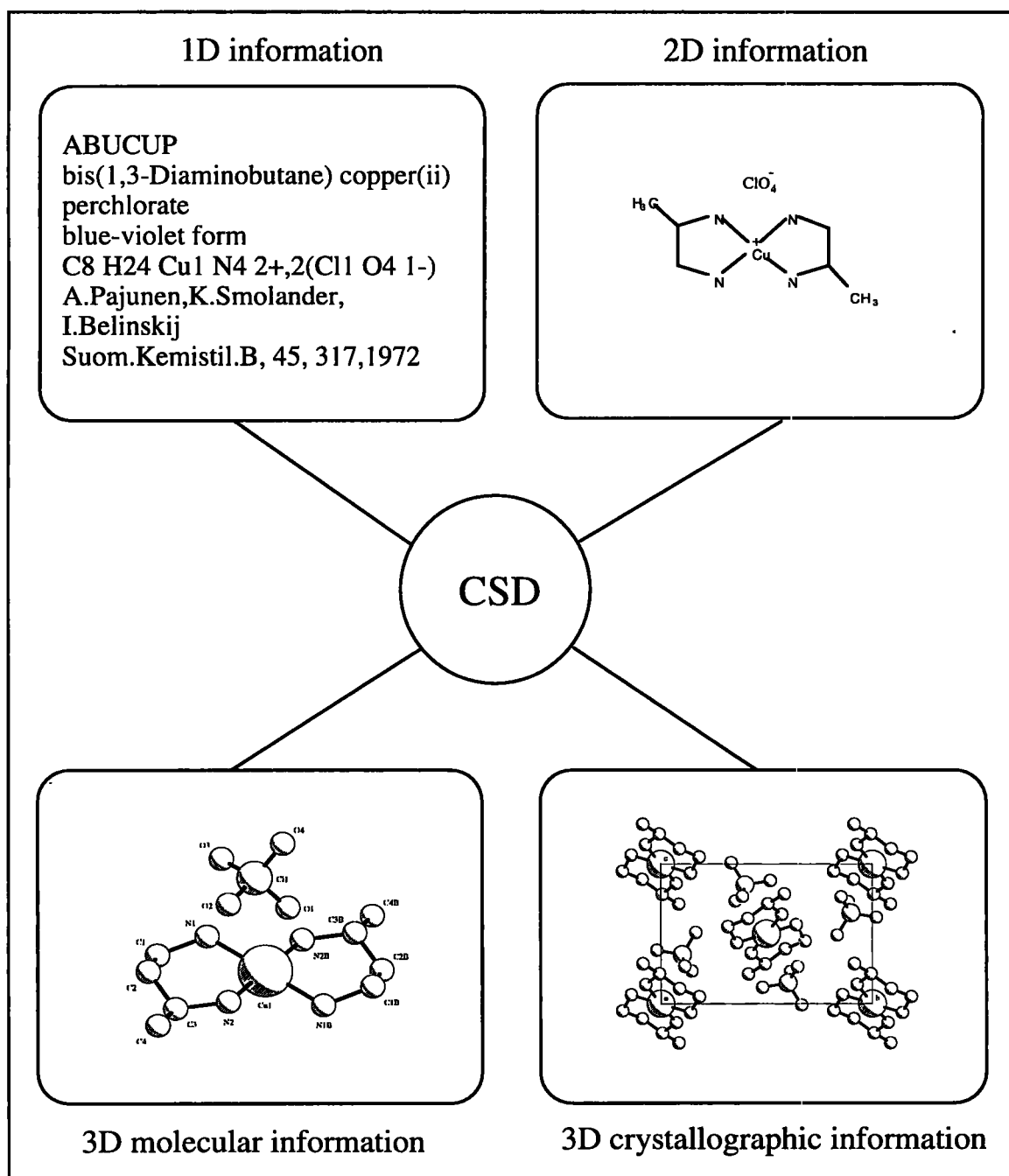
Table 5.1(b). Numerical one dimensional data content on the CSD

MSD* bibliography number
CAS§ number
NBS† Id Number:
NBS† Crystal Data number
Journal Coden: A local database journal identification code
Publication Year
Chemical Class
The Number of atoms having 3D coordinates
Space Group Number
R-factor
Temperature of study
Maximum atomic number in the structure
Z-value and Z′, the number of molecules in the unit cell and asymmetric unit respectively
Measured density
Author-calculated density
Database-calculated density
Cell parameters, **a b c** α β γ
Reciprocal cell parameters
"*AS*" or "*sigma*" flag: A flag that is set to 1 for structures with a mean C-C bond length e.s.d.
   ($\overline{\sigma}$(C-C)) ≤ 0.005Å, to 2 for structures with 0.005Å ≤ $\overline{\sigma}$(C-C) < 0.010Å, to 3 for structures
   with 0.010Å ≤ $\overline{\sigma}$(C-C) < 0.030Å and to 4 for structures with $\overline{\sigma}$(C-C) > 0.030Å. The AS flag
   is set to zero for structures where bond length e.s.d's are unavailable
Connectivity matching flag: Details whether the 3D coordinates coincide with the stored chemical connectivity
"Intensity Measurement flag": set to 1 for diffractometer data
Entry category flag -> set to either 1,2,3, or 4, depending on the type of structural study
The metric class of the reduced cell
Bravais lattice: Author's Cell
Bravais lattice: Reduced Cell
Reciprocal cell volume
The numbers of characters/I*4 integers in the TEXT CONN and DATA records in total
"*Element group*" symbol.
The chemical classification: A database assigned classification, for example "Boron compounds"
A number of *bit-screen* integers.

5.2.2 Two dimensional information content

Two dimensional connectivity information, which consists of a 2D diagram and information on atom and bond properties, is stored for each entry as a "*graph*" (Harary, 1972). A graph is a network of connections between a number of points. The points are regarded as "*nodes*" connected by "*edges*". In a chemical graph the nodes are atoms, the edges are bonds. The CSD stores atom and bond properties explicitly for each entry in two separate fields. (The information content is summarised in Figure 5.4)

---

* "Molecular Structures and Dimensions": Published by the IUCr in conjunction with the Cambridge Crystallographic Data Centre from 1970-1984

§ Chemical Abstracts Service

† National Bureau of Standards

Figure 5.4. A connectivity table of the 2D structure of 1-alanine hydrochloride

Atom Properties

Atom Number[†] :	1	2	3	4	5	6	7
Element Symbol:	C	C	N	C	O	O	Cl
Number of connected non-hydrogen atoms:	3	3	1	1	1	1	0
Number of terminal hydrogen atoms:	0	1	3	3	0	1	0
Net charge:	0	0	1	0	0	0	-1

Bond Properties

Bound atom pairs:	1-2	1-5	1-6	2-3	2-4
Bond type:	1	2	1	1	1

Together these form the chemical graph, provided the ordering of the two sets of data in the two fields are consistent with one another (i.e. atom i in the properties field is refered to as atom i in the bond fields). A third connectivity field is also stored. This is used to describe the connectivity based on interatomic separations. An example of a graph is shown for 1-alanine hydrochloride in Figure 5.4. The first atom in the graph is a carbon atom which is attached to three other non-hydrogen atoms and so on for all atoms. Note that the element type, charge, and number of terminal hygrogen atoms are stored together in the CSD as atom properties.

[†] These numbers are derived by a modification of the Morgan algorithm (Morgan, 1965) that accounts for elemental constitution and bond order.

100

The most important aspect of a graph that represents the connectivity is that the atoms within the graph are in a canonical order. An unambiguous graph can usually be defined by methods derived from the connectivity by utilising the Morgan algorithm (Morgan, 1965). The first step of the algorithm is to calculate an extended property (EP) number for each node, (in this case an atom), which are initially in an arbitrary order. The EP number is quite simply the number of contacts that the node makes with other nodes in the graph. If all nodes have different EP numbers, then the graph can be assigned a unique set of labels. More usually, the graph will have nodes which have the same number of contacts (and hence the same EP numbers), and so a unique set of labels can not be set at this point. The second step is to calculate new EP numbers by again looking at the connectivity. The sums for the second and subsequent steps use the EP numbers from the previous step. The node EP number of a particular node is recalculated as the sum of the old EP numbers of all connected nodes. The conventional algorithm continues in this way until either all the nodes are uniquely numbered, or until the number of unique assignments decreases as a result of a further step. The algorithm does not provide a unique assignment for molecules that possess molecular symmetry: clearly a molecule with symmetry usually has sets of nodes with identical extended properties. The algorithm can be modified to account for atomic constitution and bond order to make unique assignment more frequent.

To label the atoms, the atom with the highest EP number is labelled as atom 1. The next atom is the atom with the highest EP number that is connected to atom 1. A similar process is then performed for this atom, until a terminal atom is reached, or all atoms are assigned an order. If a terminal atom is reached, the process is re-started for the unassigned atoms. The unassigned atom with the highest EP number is used as the starting point until all atoms are given a label. This provides a *non-arbitrary* order for the atoms which means that it is easier to match a chemical fragment to the two dimensional connectivity.

5.2.3 Three dimensional information content

For each individual entry, the CSD stores a number of data items. Fractional crystallographic coordinates (which can be interpreted with the cell parameters, stored in the one dimensional data fields) and the space group symmetry are both stored explicitly, hence the CSD also stores *crystallographic* information as well as the molecular structure. Another key information field stored here is the 2D-3D matching integers field which link the atoms having 3D coordinates with the 2D graph described above in Section 5.2.2.

5.3 The CSD software system

The CSD software system provides users with a suite of software packages that permit both retrieval and analysis of the data. The system is summarised in Figure 5.5.

Figure 5.5. The CSD software system.
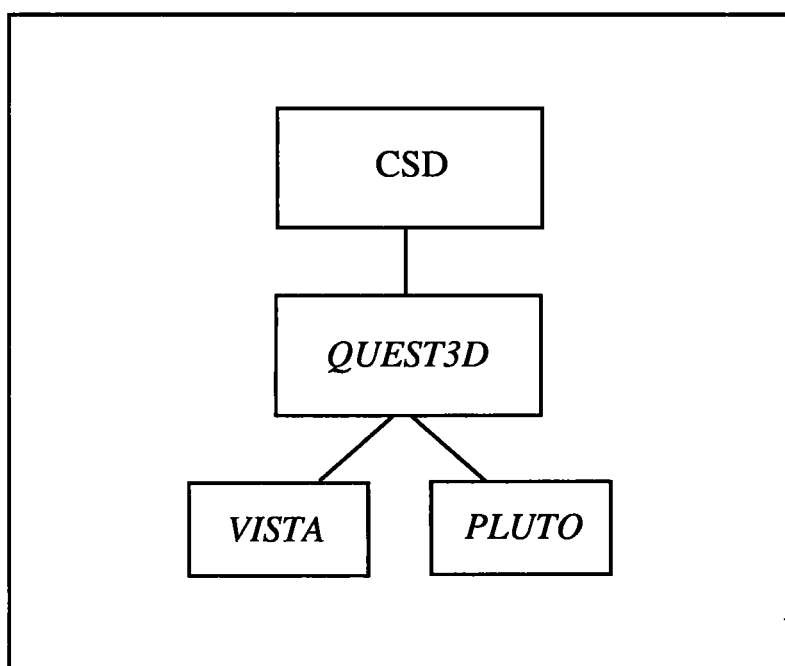
The quality of a structural database is not only governed by the information contained within the database, but also by the ease with which this data can be searched. The CSD has developed a comprehensive package (*QUEST3D*, Allen *et al*, 1991) that allows users to search for specific structures which fulfil certain criteria.

These criteria can be derived from the one dimensional, two dimensional and three dimensional information on the CSD. Normally a user will define a particular chemical "*sub-structure*": A sub-structure is a molecular fragment which may occur as part of molecules within the CSD. An example is shown in Figure 5.6.

QUEST3D is then used to search the database. Users of the CSD will usually also impose secondary criteria within a search which may define the tolerable levels of precision or other factors within a retrieved data set. The CSD stores an integer (the "*AS*" or "*sigma*" flag) for each entry that denotes the approximate mean structural precision for the C-C bonds. This is defined in Table 5.1(b). Users may limit their searches to structures with only an AS flag of 1 or 2. Another precision indicator is the R-factor, a parameter which is recorded for over 99% of CSD structures. More oblique precision criteria include limiting the search to only organic structures, or to structures that had data collected using diffractometers. Another is to use only those structures that were solved and refined after a certain date, due to the higher relative precision of more recent structures. In cases where accurate hydrogen positions are required, users may limit the data to neutron structures alone.

Figure 5.6. An example of a ketone sub-structure search and hit
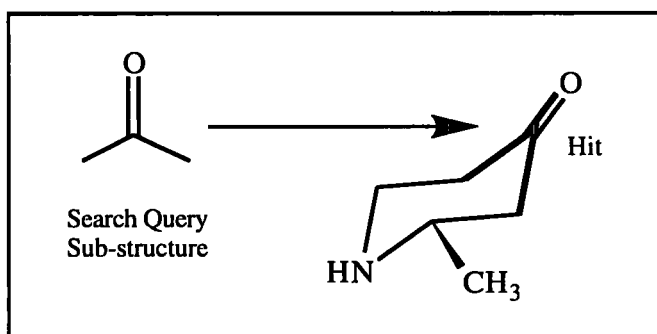
More modern searching techniques are now being utilised on the CSD which operate on the three dimensional structure. Users can now specify particular geometrical constraints within a search, which are calculated prior to retrieval of an entry (rather than being stored within the explicit information content of the entry). An example would be to return only the boat conformations of cyclohexane derivatives present in the CSD using torsion angle constraints. The same software also allows users to collate specific derived geometrical parameters (such as torsion angles) during data retrieval, rather than at a later stage during data analysis. This process also allows users to inspect values of the derived parameters for entries at search time.

QUEST3D allows users to define one dimensional numerical and text-based "queries", as well as tests on the 2D and 3D sub-structure using a fully interactive menu driven system. Individual tests can be combined using Boolean logic to formulate more complex questions (for example, one might devise a search for five membered rings but not a combination of four and three membered rings: see Figure 5.7)

Any primary sub-structure search uses the stored 2D connectivity for its location, by matching the graph (Harary, 1972) of the sub-structure with components of the stored graph for a each complete structure. Within this graph theoretical method the connectivity and bond order can be fully encapsulated in one search. This method is fully described by Ash and co-workers (1985).

Figure 5.7. An example of the use of Boolean logic within a 2D substructure search.

104
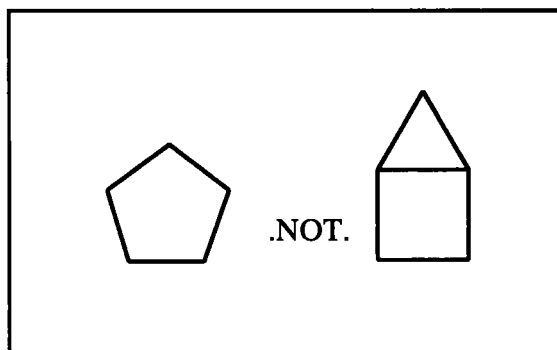
Graphical searching methods have been found to be slow due to the need to match all atoms and bonds within each structure with all atoms and bonds within the search fragment. Because of this each data entry contains bit-encoded information automatically derived from the 1D, 2D and 3D information fields which can be scanned independently to speed up searching. More detail is given in Section 5.6.1.

The individual hits can then be analysed using the *VISTA* (*VI*sual *ST*atistical *A*nalysis) (*CSD User's Manual*, 1991) program, to study the mathematical inter-relationships between the defined parameters, and the *PLUTO* program (see below) to visualize individual hits within the data set.

The *VISTA* package allows users to apply a number of statistical methods to individual data sets. Histograms of particular parameters can be obtained, as well as scatterplots and linear regressions of one parameter against another. A strength of *VISTA* is that it allows interactive analysis: if a particular entry is a clear outlier in a histogram (or scatterplot), the user can click on the point in the histogram and the CSD 1D fields, the 2D diagram and a fully rotatable structural representation are presented on screen for inspection. Any reasons for the structure being an outlier can then be investigated.

VISTA is being developed further to allow users similar interactive access to multivariate statistical analysis methods, such as principal component analysis and cluster analysis techniques.

PLUTO is a crystallographic graphics package that allows detailed analysis of individual structures within a retrieved data set. The program allows the user to manipulate the view of the molecular structure interactively. *PLUTO* can be used to view molecules in a number of ways. Ball-and-stick, space filling and stereo view options are available. Packing plots can be generated.

PLUTO can also be used for investigating non-bonding contacts within the lattice. Facilities have been introduced that allow users to easily investigate intermolecular contacts by visual means - users can click on an atom, a_1, input a radius, r_{tol}, and *PLUTO* will show all contacts from a_1 to any other atom, a_i, that fulfils the requirement that the distance $a_1...a_i < r_{vdw}(a_1) + r_{vdw}(a_i) + r_{tol}$. Hydrogen bonding interactions can easily be investigated by this method.

5.4 Structural Research Applications of the CSD

Uses of the database are varied. Most structural crystallographers will use it to check for previous studies of a particular compound or compound type. It is also frequently used to identify unknown compounds from their cell parameters. Obviously, every additional structure makes the database a more effective tool for chemical identification. Other uses have been more detailed, and have applied statistical analysis to specific data sets to elucidate chemical information. Here, a number of geometrical parameters are defined within the chemical fragment and the values of these parameters are examined, individually or together, to establish trends or correlations of chemical significance.

A preliminary search is used to retrieve examples of compounds which contain a particular 2D fragment. The three dimensional geometric structure can then be computed for each *"hit"*. These examples are then analysed using various statistical methods which can be broadly summarised as follows. The techniques described have been applied extensively by workers such as Bürgi, Murray-Rust and Dunitz (1975). A recent book (*"Structure Correlation"*, Burgi and Dunitz, 1994) has dealt with the topics described in Sections 5.4.1 and 5.4.2 in detail.

(a) Univariate methods

As the name would suggest, these methods analyse a single variable within a system. A classical example would be an analysis of the distribution of a particular bond length, angle, or torsion angle within a particular class of structures. Univariate methods analyse

the shape and metrical chracteristics of the distribution of values observed within a data set visualised *via* a simple histogram. One benefit of such analysis is that it provides reasonable global estimates of the mean or median of a particular parameter, along with an estimate of the variance across a data set. Different shapes of distribution indicate particular properties, which in turn influence the choice of analytical tool (For example, an unweighted mean value tends to be more sensitive towards outliers: in such cases a more reasonable estimate is given by the median, which by definition is unparametric, i.e. each individual entry has an equal bearing in the final result).

Univariate methods provide descriptive statistics for a paticular data set. They also allow users to interpret unusual distributions, such as the multi-modal distribution. This usually indicates that more than one class of parameter is present within a data set. A trivial example would be an analysis of all C-O bonds (of any order) for organic compounds on the CSD. The distribution would not be normal: at least two major maxima would be observed, one for C-O single bonds and one for C=O double bonds.

(b) Bivariate methods

Bivariate methods compare two parameters with one another. An example of a visual tool of analysis in bivariate statistics is the scatterplot where the values of two parameters are plotted against one another. Relationships are suggested by specific point distributions. A more mathematical method for analysing such systems is the regression method where a model is fitted which describes the relationship between the two parameters - this method can, of course, be extended to higher dimensionality. Another important technique in bivariate statistics that is employed for establishing dependancies between parameters is the analysis of correlation. The correlation, c, between two parameters, x and y is defined as:

$$c = \frac{\sum_{i=1}^{n}\left[(x_i - \bar{x})(y_i - \bar{y})\right]}{\left\{\left[\sum_{i=1}^{n}(x_i - \bar{x})\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})\right]\right\}^{\frac{1}{2}}}$$

where $\bar{x}$ and $\bar{y}$ are the mean values of the parameters x and y. This ratio is high if both x and y deviate from their respective means for the same members of the distribution.

(c) Multivariate methods

These methods deal with systems which have a high number of parameter variables, such as the conformation of a cyclopentane which has five deformable bonds. To visualise such a system initially one has to consider 5-dimensional space. Most multivariate methods attempt to reduce the dimensionality of the problem from a high number of dimensions to a more manageable number, which can be more easily visualized. An example of such a method is principal component analysis (PCA), where the correlation matrix between the Np parameters is analysed to find a set of axes which fully describe the original system in a lower dimensionality space. The aim of PCA is to find a set of axes that are mutually orthogonal to one another that can be used to fully describe the parameter variations within a data set. In other words, the correlations between the values of each principal component (PC) for each data set entry are zero.

5.4.1 Intramolecular applications

(a) Mean molecular dimensions

A great strength of crystallography is that it gives direct access to accurate interatomic separations *via* a routine method. A database dramatically increases the use of this data by giving numerous examples of similar separations in different chemical environments. This not only provides information about the influence of each particular environment, but also gives mean values for individual bonds in chemical environments that can be classified as "similar".

Numerous examples of studies of mean interatomic distances have been carried out. The most recent version of such tables were produced by Allen and co-workers (1987). Orpen and co-workers have produced similar tables for organometallic compounds

(Orpen *et al*, 1989). A study of bond angles and conformational angles within peptide chain structures also exists (Ashida *et al*, 1987). These tables can only be improved by the presence of larger quantities of data, due to the extended scope and precision that the added data delivers.

(b) Conformational analysis

The database has been used extensively to analyse molecular conformation. It is reasonable to assume that molecules in crystal structures exist in an energetically accessible form. This form corresponds to either the global energy minimum or to some low energy local minimum.

Studies of conformation have great potential for use in chemical modelling (e.g. in rational drug design), where starting models that are close to an energy minimum are required. Low energy starting points greatly reduce the complexity of minimisation by reducing the amount of computer processing unit (CPU) time required for a particular calculation. The database provides *experimental* evidence on the relative probability that a particular conformation will occur within a particular chemical environments.

Examples of conformational studies are numerous: a search of the "CSD use" database for the text string "conformation" yielded 131 hits. An early example of such a study is given by Murray-Rust and Motherwell (1978). They used principal component analysis (PCA) of the torsion angles in 110 beta-ribofuranoside fragments to show that the variability of the conformation is described by two internal ring puckering parameters and one side-chain torsional parameter. They used this analysis to describe the interchange pathway between conformations. Allen (1980) has analysed the conformation of small rings retrieved from the CSD. More recently, Allen, Howard and Pitchford (1993) have performed a conformational analysis of seven membered rings, and mapped the two pseudo-rotation pathways that exist for fragments of this type.

5.4.2 Intermolecular applications

Another new innovation in the modern database searching software is the ability to search for specific *non-bonded* contacts within the extended three-dimensional lattice. This aspect of database use is still in its infancy. It is only now that the full potential of information in extended crystal stryuctures is becoming apparent.

(a) Hydrogen bonding

A key interaction in chemistry is the hydrogen bond, which, for N and O acceptors, is the strongest of all so-called "weak" non-bonded interactions. Frequently it can be shown that such interactions stabilise molecular conformation, particularly within proteins. They are also crucially important in molecular packing within crystals. ax exemplified in Chapter 4 of this thesis. The addition of non-bonding searching on the CSD makes it easier to perform routine analysis of such interactions. Desiraju and co-workers have produced a number of papers on this subject (see for example Sarma and Desiraju, 1985). Rules regarding the formulation of hydrogen bond networks have been proposed by Etter (1990).

One of the earliest CSD based studies of hydrogen bonding analysed, and proved the existence of, C-H...O, N and Cl interactions (Taylor & Kennard, 1982). This work analysed data from 113 neutron diffraction studies that were stored on the CSD at that time. Their analysis showed that many C...O distances in C-H...O systems were significantly shorter than the van der Waals sums. They also showed that the proton in these systems lies within 30° of the plane of the oxygen lone pairs. They also noted that a number of short contacts exist between other heteroatoms and C-H protons. The subject of C-H...O interactions has recently been reviewed by Desiraju (1991)

This area of chemical crystallography has received more attention in the past decade since the need has arisen to design solids with specific macroscopic properties (a classical example being non-linear optics, see Chapter 4). Clearly the ability to engineer crystals by

molecular design is both academically appealing and of potential commercial value. It is also hoped that the information regarding non-bonding interactions abstracted for small molecular systems may be useful for designing inhibitor drugs that block specific enzyme receptors.

(b) General non-bonded interactions and reaction pathways

Many other types of intermolecular contacts have been studied using the CSD. For example, Nyburg & Faerman (1985) studied the relationship between C-X...X-C contact distances (where X represented any halogen) and the C-X...X angles subtended. They found that near linear interactions tended to have the shortest distances.

It is also possible to analyse reaction pathways using a combination of non-bonded distances and molecular conformations. This idea was initially postulated by Bürgi and Dunitz (1973), in a paper which studied the pyramidalization of the C=O group as a function of intramolecular N...C=O non-bonded interactions in six natural products. They later formalized the concept of studying reaction pathways through related functional groups.

Effectively, the molecular conformation of a molecule will change as a near neighbour approaches it. One could regard such a situation as a mapping along the reaction pathway. This principle has been used to study pseudo-rotation in five membered complexes. (Auf der Heyde and Bürgi, 1989,1,2,3)

In their study, coordinates for five coordinate d8 complexes were retrieved from the CSD. This data set was analysed using cluster analysis[†] so that the achetypal geometries within the data set could be described. The reaction coordinate of such complexes could be

---

[†] Cluster analysis is an empirical technique which groups data points that lie in close proximity in parameter space.

Figure 5.8. Schematic reaction pathway for the pseudo-rotation of ligands in five coordinate metal complexes.
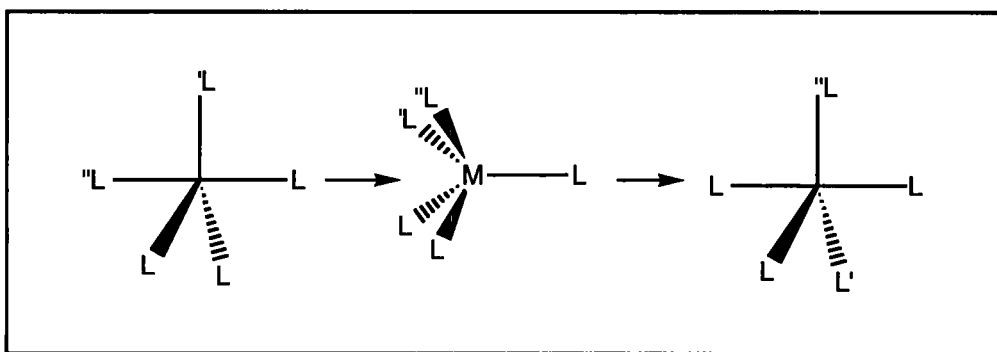
mapped by studying the deformations from the derived "idealised" geometries that occured within the data set.

Studies of reaction pathways have been pioneered by Dunitz and Bürgi. (See, for example, Britton and Dunitz, 1981; Bye, Dunitz and Schweizer, 1982; Bürgi and Chandresekhar, 1983).

5.5 Crystallographic Research Applications of the CSD

The majority of studies using the CSD have been of a chemical nature. Even studies of essentially crystallographic features, such as crystal packing, have naturally tended to try to rationalise the packing used against chemical features (such as interaction energy). Very little research has been carried out on crystallographic "systematics." There are some examples of such studies however. One study has looked into the precision of structural studies utilising a small subset of data from the CSD (Kennard & Taylor, 1986). Other studies have investigated the populations of space groups and rationalised them against crystallographic parameters (such as the symmetry elements present or number of molecules in the asymmetric unit: Wilson, 1988, 1990, 1991, 1993; Brock & Dunitz, 1994, Scaringe, 1991).

(a) Previous studies on structural precision

As noted above, Taylor and Kennard (1986) have used data abstracted from the CSD to analyse the validity of structural precision estimates. In their work they analysed the estimated standard deviations for a number of studies where the structure had been determined independently by two different research groups *by accident* in some 100 cases. They analysed the observed difference in parameter estimates between the two studies and compared these to the e.s.d's cited within the studies. Through this analysis they were able to conclude that

i) Parameter e.s.d's tend to be underestimated by a factor of 1.4-1.5, particularly in the more precise structural studies

ii) Structures tend to differ *systematically* by small amounts.

iii) The e.s.d's in some studies are underestimated by a greater degree than others.

iv) E.s.d's of atoms with high atomic masses tend to be underestimated by a greater degree than those of atoms with low atomic masses.

v) Parameter e.s.d's tend to be underestimated equally in all directions

vi) Cell parameter e.s.d's are grossly underestimated: Cell length e.s.d's may be underestimated by as much as a factor of 5, cell angles by as much as 2.5.

Their study highlights the deficiencies in the e.s.d's provided from the least squares refinement. One area it did not cover is the inter-relationship between coordinate e.s.d's and other precision indicators, such as the R-factor.

(b) Space group statistics

Since the earliest crystallographic analyses, it has been noted that certain space groups tend to occur more frequently than others. To quantify this observation there have been numerous compilations of space group frequency tables, the earliest examples being

compiled by Nowacki (1942, 1943, 1951,). The first compilation that utilised a database of information was published by Mighell, Himes and Rodgers (1983). They showed that organic compounds tend to crystallise in a few space groups such as $P2_1/c$, $P\bar{1}$ and $P2_12_12_1$, an observation that was noted by earlier researchers such as Kitiagorodskii (1962) and Nowacki (1942, 1943, 1951, 1954).

More recently workers have tended to try to correlate the occurence of certain space groups with crystallographic and chemical parameters. Recent studies by Wilson (1988, 1990, 1991, 1993), Scaringe (1991), and Brock and Dunitz (1994) have all utilised the CSD. In all these studies the distributions of Z', the number of molecules in the asymmetric unit, has been used to subdivide the distribution. The reason for these three independent analyses is that the writers wished to use the results for different purposes. Wilson has been developing a mathematical model of the space group distribution based on the crystal class, Scaringe wishes to incorporate this data into crystal structure prediction programs whereas Brock and Dunitz are interested in rationalising crystal packing. They hope that a detailed understanding of the crystal packing will assist in "crystal engineering". A more detailed review of space group frequencies is given in Chapter 7.

(c) Studies utilising the CSD within this thesis

Two studies are presented within this thesis, both of which analyse the more crystallographic aspects of the CSD, and extend the work described above in Sections 5.5(a) and (b). In Chapter 6 an effort has been made to derive a suitable function to estimate the precision of a particular structure from information stored in the CSD. This study has utilised theoretical work by Cruickshank (1960). The purpose of the study was to provide a suitable precision estimate for ~50000 entries on the CSD which have no coordinate e.s.d's stored. If the e.s.d's could be predicted accurately enough it was hoped that the derived information could be used to approximate weighted means for geometric parameters (Taylor and Kennard, 1983) and, more generally, to improve the precision criteria available in database searches.

The second study, presented in Chapter 7, is a more detailed analysis of space group frequency than any available so far. New software has been written to establish the crystallographic symmetry that a particular molecule or ion adopts within the lattice for each structure on the CSD. Using this data, tables showing space group frequency as a function of crystallographic point group have been derived. Other parameters have also been analysed, such as the influence of charge and the effect of having more than one chemical moiety in the asymmetric unit (as opposed to having $Z' > 1$).

Both studies required the development of novel software to act on the CSD master file. For this reason, the data structures used in the CSD are described in more detail in Section 5.6.

5.6 Data structures used in the CSD

5.6.1 One dimensional structure

Each entry within the CSD consists of four separate regions of data storage. The first level of data is stored in integer format, the so-called MASK array. The MASK array contains the one dimensional numerical data. Details of the data fields have already been summarised in Section 5.3. The one dimensional data includes a number of integers that contain binary information about the chemical sub-structures and the crystallographic details of the particular entry. The so-called *bit-screens* are stored in the following way. A conventional integer is stored in 32 bits. 31 of these are used to store the numerical value of a particular number. The 32nd bit is used to store the sign of the integer. Each bit can be used to store the answer to a question, provided the answer is simply, "yes" or "no": if the answer is "yes" the bit is set to one, otherwise it is set to zero. For example consider the question "Does this entry contain fused rings?". In entries where this is true a specific bit in a specific integer is set to 1.

The bit-screens contain information that is derivable from other information stored for each entry. The redundancy is present because the bit-screens provide information rapidly: it is much quicker to ask a single question "Does this entry contain fused rings" *via* a single bit indicator than *via* software that abstracts the information from the two-dimensional connectivity.

The second level of one dimensional information is stored in an array of single characters (the *TEXT* array) In conjunction with this a number of integers are stored that indicate the point in the array where a text field ends (i.e. a "*directory*"). The first eight fields of text are of a fixed length and contain information common to all entries, such as refcode, the journal volume and page. The remaining fields are of variable length: each one contains one specific piece of information, such as the authors' names or the chemical name of the compound studied.

5.6.2 Two dimensional structure.

The database stores the two-dimensional connectivity explicitly (see the example in Section 5.2.2.) in the so-called CONN field. This area of data is very important within the searching structure since the bond type is stored along with connectivity. A single one dimensional array of integers is stored for each entry. Again a separate "*directory*" (as defined in Section 5.6.1) defines the position and size of the individual fields within the array.

Each CSD entry has both an associated "*crystal*" connectivity and a "*chemical*" connectivity. The crystal connectivity describes the calculated connectivity as based on the measured interatomic separations: for a bond to be present in this field, the two atoms have to be closer together than the sum of their covalent radii plus a specified tolerance. The chemical connectivity is the connectivity of the formally described chemical structure from the original literature. Included within it are the orders of bonds.

The crystal connectivity is stored as two separate fields. The first field contains a set of $N_{atom} + 2(N_{bond} - N_{atom})$ integers where N_{atom} is the number of atoms and N_{bond} is the number of bonds. The first N_{atom} integers represent each atom; the integer stored points to a second atom to which the first is bound. The remaining pairs of integers point to separate atoms which are also bound to one another. A second field is also stored that contains the atomic properties. This stores each atom's element type, charge and the number of terminal hydrogens connected to it.

The chemical connectivity is described by combining the atom properties field (described above) with a "bond properties" field. For each bond, an integer is stored that describes the formal bond between two atoms and the bond's formal order.

5.6.3 Three dimensional structure.

The extensive numeric data defining the 3D structure is stored in the *DATA* array. Again the ends of different fields are indicated by the use of a "*directory*", as described above for the *TEXT* array in Section 5.6.1

The CSD entries are based around the concept of *crystal chemical unit*. This is the group of atoms and symmetry related atoms that describes in full the chemical moieties represented within the crystal structure. The crystal chemical unit can be broken down into "*residues.*" A "*residue*" is defined as a discrete bonded moiety or ion within the crystal chemical unit . For example, a solvent molecule within the crystalline lattice would be defined as an independent residue from the main structure. Another example is when more than one independent molecule is present within the asymmetric unit.

Figure 5.9. The CSD entry ABUCUP.

The figure demonstrates (a) that it is possible to have more than one residue in the asymmetric unit and (b) that it is chemically convenient to store the symmetry related atoms explicitly: in this structure, for example, N2 is related to N2B by a centre of inversion. The full molecular structure is only revealed when the symmetry related atomic positions are included.
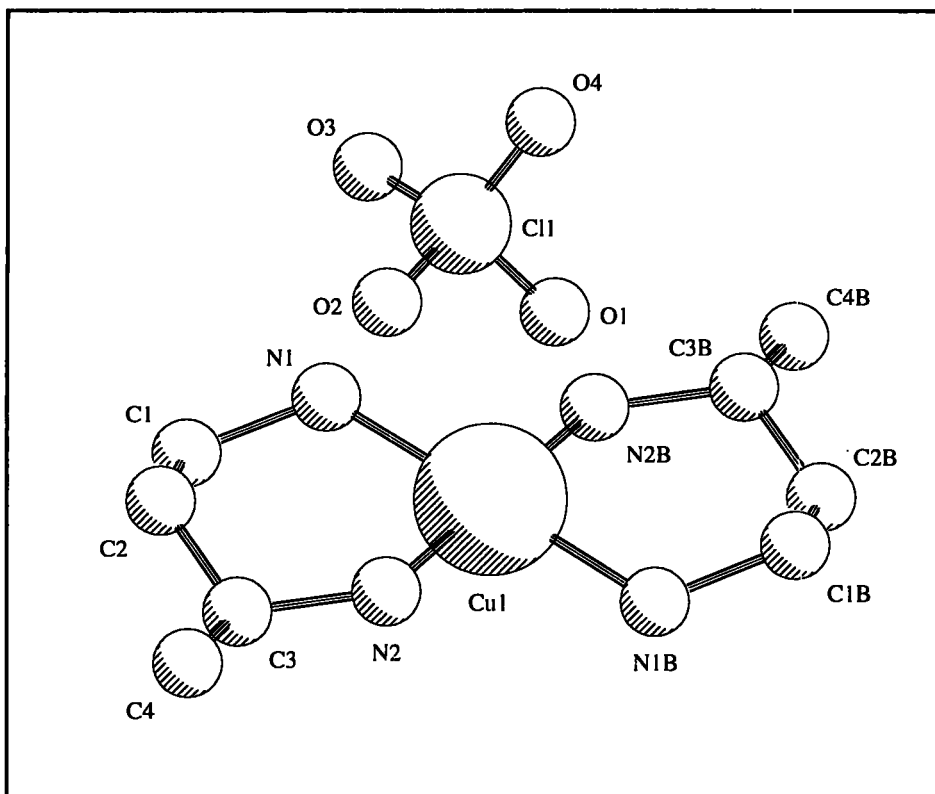
Each residue has to describe the chemistry of the moiety. When a molecule lies on a symmetry element, the molecule will consist of both *"atoms"* and *"symmetry related atoms"*. Because the CSD stores the chemical motif, both the atoms and these symmetry related atoms are stored explicitly in two different formats, as described below.

Both of the above concepts can be seen for the CSD entry ABUCUP shown in Figure 5.9. There are two independent residues stored: a perchlorate counter ion and a charged copper complex. The copper complex sits on a centre of symmetry and so only half the molecule lies within the asymmetric unit.

The atomic coordinates are stored as integers within the *DATA* array. Three integers are stored for each atom within the asymmetric unit, one for each coordinate. The coordinate and its attached e.s.d. are stored on the database compactly as a single 32 bit integer. In decimal this can comprise a maximum of nine digits, or as is the case in the CSD format,

a variable length coordinate integer of up to six digits, e.s.d's of up to two digits and a final single integer that indicates the number of significant digits in the coordinate field.

Also stored for each entry are the symmetry operators, including any space group centrings. Each operator is stored as 12 integers; 9 form components of a 3×3 rotational matrix, the other stores a 3×1 translational component; each component in the translation having been multiplied by 12. In non-centrosymmetric space-groups all operators are stored. In centrosymmetric space groups only the non-inversion related set of operators are stored, if the unit cell origin coincides with a centre of symmetry. When this is not the case, all operators are stored. There are very few examples of this in the CSD.

Symmetry-related atoms, ("*satoms*"), are stored independently from the atom field. Each "*satom*" is stored as a single integer. The first three digits are used to point to the number of the atom from which the satom is generated in the atom list. The second three digits indicate the number of the stored symmetry operator. The final three digits indicate unit cell length translations that are needed to bring the atom into its respective "*residue*"

5.7 References

Allen, F. H.; *Acta Cryst*, **B36**, 81, 1980

Allen, F.H.; Kennard, O.; Watson, D. G.; In preparation, 1995

Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R.; *J.Chem. Soc. Perkin Trans. 2*, **S1**, 1987

Allen, F.H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M., Watson, D. G.; *J. Chem. Inf. Comput. Sci.* **31**, 187, 1991

Allen, F. H.; Howard, J. A. K.; Pitchford, N. A.; *Acta Cryst.*, **B49**, 910, 1993

Ashida, T.; Tsunogae, Y.; Tanaka, I.; Yamane, T.; *Acta Cryst.*, **B43**, 212, 1987

Ash, J. E.; Chubb, P. A.; Ward, S. E.; Welford, S. M.; Willett, P.; *Communication, Storage and Retrieval of Chemical Information*, Wiley, Chichester, UK, 1985

Auf der Heyde, T. P. E.; Bürgi, H.-B.; *Inorg. Chem.*, **28**, 3960,1989,1

Auf der Heyde, T. P. E.; Bürgi, H.-B.; *Inorg. Chem.*, **28**, 3970,1989,2

Auf der Heyde, T. P. E.; Bürgi, H.-B.; *Inorg. Chem.*, **28**, 3982,1989,3

Bergerhoff, G. ;Brown, I. D.; Hundt, R., Sievers, R.; *J. Chem. Inf. Comput. Sci.*, **23**, 66-69, 1983

Britton, D.; Dunitz, J. D.; *J. Am. Chem. Soc.*, **103**, 2971, 1981

Brock, C.P; Dunitz, J.D. *Chem. Mater.*, **6**, 1118, 1994

Burgi, H.-B.; Dunitz, J. D.; Shefter, E.; *J. Am. Chem. Soc.*, **95**, 5065, 1973

Bye, E. Dunitz, J.D.; Schweizer, W. B.; *J. Am. Chem. Soc.*, **104**, 5893, 1982

Chandrasekhar, K.; Bürgi, H.-B.; *J. Am. Chem. Soc.*, **105**,7081, 1983

Cruickshank, D. W. J.; *Acta Cryst.*, **13**, 774, 1960

Desiraju, G. R.; Acc. Chem. Res, **24**, 290, 1991

Etter, M. C.; *Acc. Chem. Res.*, **23**, 120, 1990

Harary, F; *Graph Theory*, Addison-Wesley Publishing Co., 1972

Kennard, O.; Taylor, R.; Acta Cryst., **B42**, 112, 1986

Kitiagorodskii, A.I. *Organic Chemical Crystallography*; Consultant's Bureau: New York,

1961. (English translation of the Russian original published by Press of the Academy of

Sciences of the USSR, Moscow, 1955).

Mighell, A.D.; Himes, V.L.; Rogers, J.D. *Acta. Cryst.* , **A39**, 737, 1983

Morgan, H. L.; J. Chem. Doc. **5**, 107, 1965

Motherwell, W. D. S.; Murray-Rust, P.; *Acta Cryst.*, **B34**, 2534, 1978

Murray-Rust, P., Bürgi, H.B., Dunitz, J.D.; *J. Am Chem. Soc.* **96**, 921, 1975

Nowacki, W. *Helv. Chim. Acta.*, **25**, 863, 1942

Nowacki, W. *Helv. Chim. Acta.*, **26**, 459, 1943

Nowacki, W. *Helv. Chim. Acta.* , **34**, 1957, 1951

Nowacki, W. *Crystal Data*; J.D.H.Donnay and W.Nowacki, Ed. Geol. Soc. Amer.; pp. 85, 1954

Nyburg, S. C.; Faerman, C. H.; Acta Cryst., **B41**, 274, 1985

Orpen, A. G.; Brammer, L.; Allen, F. H.; Kennard, O.; Watson, D. G.; Taylor, R.; *J. Chem. Soc., Dalton Trans.*, **S1**, 1989

Pauling, L.; *The Nature of the Chemical Bond*, Ithaca, New York, Cornell Univ. Press, 1940

Pimental, G. C.; McClellan, A. L.; *The Hydrogen Bond*, San Francisco, 1960

Sarma, J. A. R. P.; Desiraju, G. R.; Chem. Phys. Lett., **117**,160, 1985

Scaringe, R.P. In *Electron crystallography of Organic Molecules*; Fryer, J.R. and Dorset, D.L., Ed.; Kluwer: Dordrecht; **Vol.328**; pa. 85, 1991

Structure Correlation, Ed. Bürgi, H.-B.; Dunitz, J.D.; Published by VCH Publishers Inc. New York, USA, 1994

Struture reports, Ed. IUCr,Utrecht: Osthoek, 1956

Sutton, L. E.; Editor. *Tables of Interatomic distances and Configuration in Molecules and Ions*, Special Publication No. 11, London: The Chemical Society, 1958

Sutton, L. E.; Editor. *Tables of Interatomic distances and Configuration in Molecules and Ions (Supplement)*, Special Publication No. 18, London: The Chemical Society, 1965

Taylor, R; Kennard, O.; *J. Am. Chem. Soc.*, **104**, 5063, 1982

Taylor, R.; Kennard, O.; *Acta Cryst.*, **B39**, 517, 1983

Wilson, A.J.C. *Acta Cryst.*, **A44**, 715, 1988

Wilson, A.J.C. *Acta Cryst.*, **A46**, 742, 1990

Wilson, A.J.C. *Z. Krystallogr.*, **197**, 85, 1991

Wilson, A.J.C. *Acta Cryst.* , **A46**, 795, 1993

Wykcoff's Crystal Structures; Interscience, New York, 1948

Chapter 6

An analysis of structural precision in the CSD.

6.1 Introduction

6.1.1 Precision on the Cambridge Structural Database (CSD)

The precision of experiments is a cause for frequent debate in all of the physical sciences. Every experimental result has to be interpreted individually or compared with other results in order for conclusions to be drawn, hence estimates of precision are important quantities. In the earliest days of crystallography systematic studies of geometrical information from structural work played key roles in the earliest understanding of chemical bonding (see for example, Pauling 1940; Pimental & McClellan 1960; Sutton 1958, 1965). The frequency of such studies fell away during the 1970's probably due to the difficulty in locating the necessary data from the literature. Databases, such as the CSD, have largely eliminated these problems due to the availability of user friendly software to extract the necessary information and, as already detailed, a large increase has been seen in the number of studies that utilise the CSD. This increased interest in the chemical content of the CSD has presented a new problem: that of data selection for a particular study. Data selection will primarily be governed by the chemical substructure that is under investigation; a particular compound will be selected based on its 2D chemical structures, its 3D chemical shape, or on how it interacts with other molecules within the lattice. Secondary criteria are, however, required in both selection of structure and in interpretation of the final results. These, of course, usually relate to the precision and the accuracy of the data that satisfy the primary search criteria. The accuracy of a study is hard to quantify. The CSD does store some information on this. A number of bit-screens are set for structures in which numerical inconsistencies were detected during database building, but usually gross errors (due to a poor experiment) are detected during interpretation of results rather than

during data acquisition. Inaccurate structures will often occur as outliers in a geometrical comparison. The precision of a crystallographic study can, however, be approximately derived during the structure determination. The least squares process yields estimates as to the structural precision in the form of estimated standard deviations (e.s.d's) for the refined parameters. Unfortunately these data are not stored on the database for entries prior to 1985[†], and so in terms of the CSD they do not constitute suitable precision parameters for all structures.

Normally a user of the CSD will apply secondary constraints using parameters that are stored for nearly all structures. The most commonly used "precision" parameter is the R-factor. (In the database this is recorded as the lowest of R, Rw, etc.). The R-factor is ubiquitous in crystallography, and is available for more than 99% of all CSD entries. This parameter is not ideal but it does at least give some indication of the precision of a structure, due to its inherent link to the expression for the estimated standard deviation: the numerator of the weighted R-factor is the same as the numerator of the goodness-of-fit. The second parameter stored for precision indication in the CSD is the "AS flag". This is a parameter which encodes the mean C-C bond length e.s.d within a structure. Although this parameter is more directly linked to precision, it is less than satisfactory. Until 1985, the AS flag was assigned manually by the CSD editorial staff only in cases where bond *length* e.s.d's for C-C bonds ($\sigma(C-C)$) were available. Even when atomic e.s.d's were recorded the AS flag remained unassigned and so some 16% of all structures do not possess an AS flag. The AS flag assigned is unfortunately also inappropriate because the banding of the parameter was conceived using quantiles

---

[†] Due to the constraints on computer storage space at the inception of the CSD only a limited quantity of information was stored for each structure. Unfortunately a new data field included at a later date leaves a large backlog of structures that do not possess the new information. The standard uncertainties are an example of this but no updating is realistically possible to the backlog due to the vast number of structures, and the ever increasing acquisition rate for new structures.

Table 6.1 AS flag bandings

$\overline{\sigma}(C - C)$ (in Å) is the mean of the e.s.d's for all C-C bonds in a given structure

AS=0	$\overline{\sigma}$(C-C) not available
AS=1	$0.000 < \overline{\sigma}$(C-C) ≤ 0.005
AS=2	$0.005 < \overline{\sigma}$(C-C) ≤ 0.010
AS=3	$0.010 < \overline{\sigma}$(C-C) ≤ 0.030
AS=4	$0.030 < \overline{\sigma}$(C-C)

appropriate for the structural studies of the 1960's and the early 1970's. Advancement in technology has meant that the precision of most studies has vastly improved and so the AS bands are now inappropriate for the structures of today. The bandings are shown in Table 6.1.

The AS = 3 band is too broad, since it subsumes many organic structures that have bond length e.s.d's in the range 0.011-0.015Å. Such structures may well be considered precise enough for use in many studies. Further, the AS flag is discontinuous, and concentrates entirely on light atoms (particularly carbon, although occasionally C-O or C-N bonds are used in the assignment). The AS flag does not convey explicit information on any particular atom. This information is necessary if fully weighted means of geometrical parameters are to be generated from structures within the CSD. Currently the best alternative to this is either to use a "semi-weighted" mean as suggested by Taylor & Kennard (1983) or to use unweighted means over subsets of the CSD for which we may assume a relatively narrow precision range, such as R-factor $\leq 7\%$ and AS ≤ 2.

A CSD user will normally also apply other restraints on data selection. For example, in organic geometrical studies, a user will frequently apply a constraint on the maximum atomic number of any element in the structures accepted. Another criterion sometimes applied is a limit on the year in which the study was carried out. In both these cases, experience is being used to impose indirect precision constraints: Figure 6.1 shows the variation of mean R-factor with year of publication. The mean R-factor is clearly seen

Figure 6.1: R-factor vs year
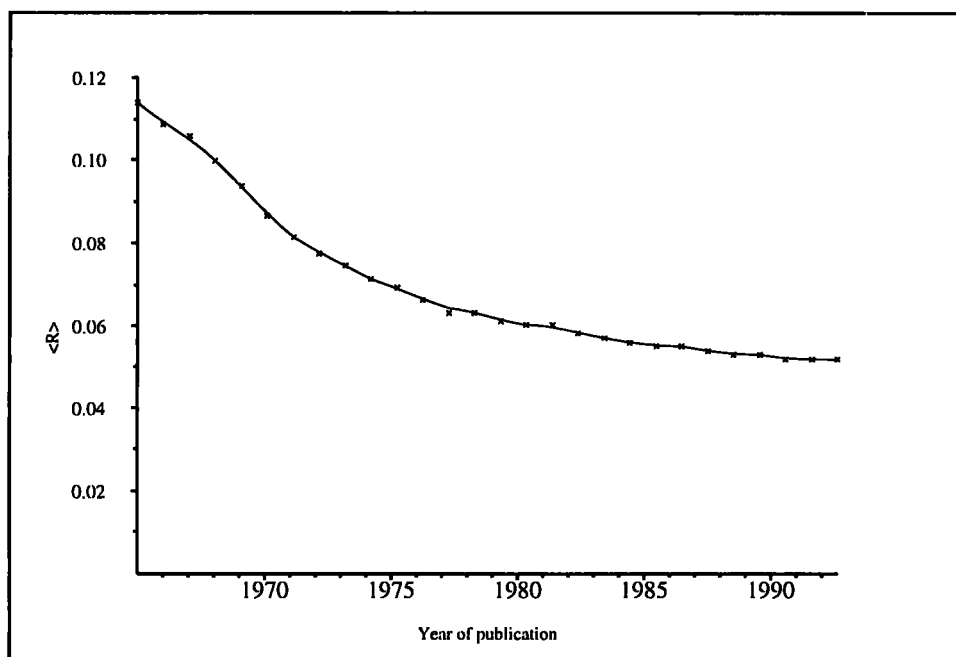
to fall with year. One probable reason for the fall is the increased use of anisotropic refinement over isotropic, coupled with better tools for measuring intensities to a higher precision, which makes anisotropic refinement worthwhile. This, is usually a good way of improving the final results.

The "light-atom" constraint again arises primarily through experience, since generally when heavier elements are present they tend to dominate the scattering. As such, a very good fit can be obtained provided that the heavier atoms are well modelled and so a "light-atom" constraint in database searching tends to increase the relative precision of a data set within a specific R-factor banding.

As already mentioned, since 1985, the atomic e.s.d's have been incorporated into the CSD in order to improve the information content on precision, but there still exists a large backlog of information of some 50000 entries. This study attempts to use the 45763 entries (up until January 1992) that do have coordinate e.s.d's assigned to try to examine the possibility that reasonable estimates of the mean C-C e.s.d. for a particular

structure ($\overline{\sigma}(C-C)$) can be predicted using some function of variables that are stored in a vast majority of the CSD entries;

$$\overline{\sigma}(C-C)_p \approx f(x_1, x_2, \ldots x_n) \qquad 6.1$$

where f is the predicting function involving variables x_1 to x_n. If such a function can be found then it could potentially be used as a single, real-valued, and continuous parameter representing precision for the large backlog of structures, complementing the AS flag in cases where it is present and substituting for it in cases where it is not. A preliminary study of this kind was reported by Allen and Doyle (1987) based on the 4187 entries that contained coordinate e.s.d's at that time. This showed that the approach was feasible. A more comprehensive study is reported here.

6.1.2 Theoretical background.

An analysis of the required precision of X-ray data that is required to yield a mean isotropic coordinate e.s.d, $\overline{\sigma}(A)$ for any element A in a particular structure that may also contain other elements B,C *etc.* is given by Cruickshank (1960). He derived a simple approximate formula that relates $\overline{\sigma}(A)$ to the residual (R), the chemical constitution of the asymmetric unit and the scattering limits during an experiment.

$$\overline{\sigma}(A) = \frac{R\overline{N}_A^{\frac{1}{2}}}{\overline{s}(mp)^{\frac{1}{2}}} \qquad 6.2$$

In this equation $\overline{s}$ is the root mean square reciprocal radius for the "observed" reflections, R is the conventional R-factor, $\overline{N}_A$ is the number of atoms of type A required to generate a scattering power equivalent to that of the current asymmetric unit at $\overline{s}$, and can be written as

$$\overline{N}_A = \sum_{i=1}^{N_{as}} \frac{f_i^2}{f_A^2} \qquad 6.3(a)$$

p is defined as $(N_r - N_p)$, where N_r is the number of reflections used in refinement and N_p is the number of parameters refined, and m, which takes a value of 8 for centrosymmetric structures and 4 for non-centrosymmetric structures.

126

Scattering factor data is not stored on the database. Thus within these studies we redefine equation 6.3 in terms of the atomic number:

$$N_A = \sum_{i=1}^{N_{ats}} \frac{Z_i^2}{Z_A^2} \qquad \text{6.3(b)}$$

which for carbon can be written as

$$N_C = \sum_{i=1}^{N_{ats}} \frac{Z_i^2}{Z_C^2} \qquad \text{6.3(c)}$$

In the CSD there is no estimate of $\bar{s}$ available so we have to assume that it does not vary significantly from structure to structure. This is probably a tenable assumption: the majority of structures are only collected to between 50 and 55° 2θ for molybdenum radiation and the majority of published structures will scatter to at least 40°. A few values of $\bar{s}$ are shown below in Table 6.2. One important point here is that some crystals may well scatter further than 55° but will not have been collected to the edge of their scattering because of constraints on diffractometer time. It is hard to predict whether this has a large effect on the e.s.d's.

Table 6.2. Ranges of values of $\bar{s}$ for a variety of values of θ. (Mo radiation)

θ	s_{max}	$\bar{s}$
15°	0.728	0.564
20°	0.962	0.745
25°	1.189	0.921
30°	1.407	1.090
35°	1.614	1.250

The other unavailable term is p, the difference between the number of reflections and the number of refined parameters. This quantity may well vary widely across all the crystallographic studies performed, due to its dependance on N_r, the number of reflections refined in the least squares process. This, in turn, is dependant on the size of the structure studied.

6.2 Methodology

Throughout the study the January 1992 release of the CSD was used. This was broken down into three subsets, which are described below. Data set 1 was extracted using the program QUEST (*Cambridge Structural Database Users Manual*, 1992). Data sets 2 and 3 were generated using locally written code, the details of which are given below.

6.2.1 Data sets Used

Data set 1.

This consisted of the 83516 entries for which atomic coordinates have been published since 1965. This data set was used to analyse the way in which CSD stored parameters were inter-related, e.g. the AS flag, R-factor, and Z_{max} (The maximum atomic number of any element within the asymmetric unit in a particular structure).

Data sets 2

In these data sets, the entries that had atomic e.s.d's recorded were extracted, subject to certain constraints:

a)   The structure was determined by X-ray (and not neutron) diffraction
b)   Intensity data were collected on a diffractometer
c)   No residual numerical errors remained after CSD checking and evaluating
      procedures
d)   No disorder or polymeric (catena) bonding was reported
e)   The R-factor was less than 10%

Workfiles 2 and 2A (fully defined below) were generated from entries satisfying these criteria and were used in the derivation of the predictive functions below. A total of 35747 CSD entries remained after file generation.

For a small number of entries in data set 2 the original literature was consulted and values of N_r (the number of reflections) and N_p (the number of refined parameters in least squares) were extracted for manual addition to the data set 3 workfile (see below). References were selected from three journals, *Acta Crystallographica Section C*, *Inorganic Chemistry*, and *Organometallics*. The reason for picking this selection of journals was to ensure that the chemical composition of this small data set was comparable to that of data set 2. (*Acta Crystallographica Section C* contains predominantly organic structures, so taken alone it is not representative of data set 2). Data set 3 was used to test the Cruickshank (1960) formulation more fully.

6.2.2 The local code.

Local code was written to extract data items from the CSD for inclusion in the workfile. Currently there is no way of abstracting the coordinate and bond length e.s.d's using the QUEST program, so the software was written to decode the database internal format. Each atomic coordinate and its attached e.s.d are stored on the database compactly as a single 32 bit integer. In decimal this can comprise a maximum of 9 digits, or as is the case in the CSD format, a variable length coordinate integer of up to 6 digits, e.s.d's of up to 2 digits and a final single integer that indicates the number of significant digits in the coordinate field. A program that performed the simple task of abstracting this data was written in FORTRAN. From this data and using the CSD connectivity field it was possible to calculate the e.s.d's for all C-C bonds present within the structure by using the formulation given by Muir and Mallinson (1993). Their formulation takes into account the possibility that positional coordinates may be correlated in oblique coordinate systems. Their derivation gave :

$$\sigma(X) = \sqrt{\begin{bmatrix} a^2\sigma^2(x) + b^2\cos^2\gamma\,\sigma^2(y) + c^2\cos^2\beta\,\sigma^2(z) + \\ 2ac\cos\beta\,\sigma(x)\sigma(y)\cos\beta^* + 2ab\cos\gamma\,\sigma(x)\sigma(y)\cos\gamma^* \\ +2bc\cos\beta\cos\gamma\,\sigma(y)\sigma(z)\cos\alpha^* \end{bmatrix}} \qquad 6.4(a)$$

$$\sigma(Y) = \sqrt{\begin{bmatrix} b^2\sin^2\gamma\,\sigma^2(y) + c^2\sin^2\beta\cos^2\alpha^*\sigma^2(z) \\ -2bc\sin\beta\sin\gamma\cos\alpha^*\sigma(y)\sigma(z) \end{bmatrix}} \qquad 6.4(b)$$

$$\sigma(Z) = c^2\sin^2\beta\sin^2\alpha^*\sigma^2(z) \qquad 6.4(c)$$

$$\sigma(E) = \sqrt{\frac{\sigma(X)^2 + \sigma(Y)^2 + \sigma(Z)^2}{3}} \qquad 6.4(d)$$

where X, Y and Z denote orthogonal coordinates, x, y and z denote fractional coordinates. a, b and c are the cell lengths, α, β and γ are the cell angles and α^*, β^* and γ^* are the reciprocal cell angles.

A mean of the individual $\sigma(C-C)$ values [ie $\overline{\sigma}(C-C)$] was calculated subject to a number of tests on their distribution across a structure. The tests applied were as follows:

1) The number of C-C bonds present in the structure (N_b) had to be ≥ 5.

2) The distribution of the individual $\sigma(C-C)$ values was analysed for each database entry. Firstly, the sample standard deviation (S) was calculated as:

$$S = \left\{ \frac{\sum_{i=1}^{N_b}\left[\sigma(C-C)_i - \overline{\sigma}(C-C)\right]^2}{N_b - 1} \right\}^{\frac{1}{2}} \qquad 6.5$$

where N_b is is the number of C-C bonds. Any $\sigma(C-C)$ for which $\left|\overline{\sigma}(C-C) - \sigma(C-C)\right| \geq 4S$ was disregarded. If the $N_b \geq 5$ condition was still satisfied then the mean was recalculated and used.

3) CSD entries that had very skewed distributions of $\sigma(C-C)$ were also eliminated: a term, M, as shown below was calculated using the maximum and the minimum of all $\sigma(C-C)$'s that passed the above test.

$$M = \frac{\sigma(C-C)_{max} + \sigma(C-C)_{min}}{2} \qquad 6.6$$

This approximation to the median was compared with the true calculated mean value, $\overline{\sigma}(C-C)$ by using:

$$D_{\sigma} = \frac{\left|\overline{\sigma}(C-C) - M\right|}{\overline{\sigma}(C-C)} \qquad 6.7$$

Entries having $D_{\sigma} > 0.25$ were rejected.

4) A few entries on the database have a systematic error in their coordinate e.s.d's, in that all values entered for a very few structures are incorrect by a power of ten. For example, 0.0021 may be recorded as 0.021 or 0.00021 due to a global CSD input error. In all these cases there is internal consistency and so they are not removed by the above tests. To try to eliminate these entries, any entry that had $\overline{\sigma}(C-C) > 0.04$ was rejected. Also, if the value of $\overline{\sigma}(C-C)$ did not agree with the AS flag recorded for the entry then the entry was rejected. A tolerance of 0.001Å was allowed for AS bands 1 and 2 and a higher tolerance of 0.002 for bands 3 and 4.[†]

[†] A tolerance was found to be necessary due to certain problems with the AS flags on the CSD. Some entries from around 1985 had had their AS flags manually calculated. The staff at the time had approximated the process somewhat and had not always used all of the C-C bond e.s.d's in AS flag assessment.

6.2.3 The information content and structure of the carbon-based workfiles 2 and 3

The local code was used to transform the structural information in data sets 2 and 3 from the binary CSD format into simple ASCII workfiles. These workfiles contained only those information fields that were required for the statistical and numerical analyses described in the following sections. These data items are summarized in Table 6.3 and were chosen as being related, directly or indirectly, to the precision of the coordinate set.

Table 6.3. Definition of data items included for each entry in the workfiles generated for data sets 2 and 3. Items superscripted with a "∗" are for data set 3 only.

Item	Definition
REFCODE	CSD reference code
R	Crystallographic R-factor
AS	CSD AS-flag
T	Temperature of data collection (K)
N_{nh}	Number of non-H atoms in asymmetric unit
N_h	Number of H atoms with coordinates reported
SPGN	Space group number (International Tables, 1960)
CENT	Non-centrosymmetric = 1, Centrosymmetric = 2
V	Unit-cell volume
Z_{max}	Atomic number of heaviest element in structure
Z_{rms}	$\left(\sum_{i=1}^{Na} Z_i^2 / 36 N_{nh} \right)^{\frac{1}{2}}$, Z_i are atomic numbers
$N_c^{\frac{1}{2}}$	$\left(\sum_{i=1}^{Na} Z_i^2 / 36 \right)^{\frac{1}{2}}$, Z_i are atomic numbers
RZ_{max}	The product RZ_{max}
RZ_{rms}	The product RZ_{rms}
$R\,N_c^{\frac{1}{2}}$	The product R $N_c^{\frac{1}{2}}$
$\overline{\sigma}(C\text{-}C)$	Mean calculated esd of C-C bond lengths
$\sigma(C\text{-}C)_{min}$	Minimum calculated esd of a C-C bond length
$\sigma(C\text{-}C)_{max}$	Maximum calculated esd of a C-C bond length
$\overline{\sigma}(C)$	Mean isotropic esd of C atoms
$\overline{\sigma}(A)$	Mean isotropic esd of heaviest element(s)
N_b	Number of C-C bonds contributing to $\overline{\sigma}(C\text{-}C)_c$
N_r∗	Number of independent reflections
N_p∗	Number of parameters refined by least squares
N_r/N_p∗	Ratio of N_r to N_p

The workfiles of data sets 2 and 3 consisted of a single formatted ASCII line for each CSD entry in the file. After all checks had been completed, data sets 2 and 3 contained 29,632 and 817 entries in their respective workfiles.

Workfiles 2 and 3 only contain e.s.d data for carbon atoms, for carbon-carbon bonds, and for the heaviest atom in each structure. These workfiles were used in the detailed analysis of carbon atom precision presented in section 6.4.

6.2.4 The information content and structure of the general atom workfile.

Later sections of this chapter (sections 6.6 onwards) present an analysis that considers e.s.d. values of all non-hydrogen element types, denoted as $\sigma(E)$, and derived from the original CSD data set 2. The mean isotropic $\sigma(E)$ values were calculated using equation 6.4(d) given above. In cases where a particular element occured more than once in a given structure the mean of the the values of $\sigma(E)$ for that particular element, $\overline{\sigma}(E)$, were calculated. Local code was written to apply equations 6.5 and 6.7 to $\sigma(E)$ values (rather than $\sigma(C-C)$ values) in order to prevent erroneous or internally disparate values from entering the workfile.

Table 6.4. Principal information fields in workfile 2A.

CSD entry information record (1 per entry).	
REFCODE	CSD reference code
R	Crystallographic R-factor
N	No. of non-hydrogen atoms
SPGN	Space-group number
CENT	Non-centrosymmetric = 1, Centrosymmetric = 2
V	Unit-cell volume
Z_{max}	Atomic number of heaviest element
$N_c^{\frac{1}{2}}$	As defined in Equation 6.3(c)
ITYPE	No. of element type records that follow for this entry
Element type records (ITYPE per entry)	
A	Element symbol
Z	Atomic number
$\overline{\sigma}(E)_a$	Mean isotropic esd (Å) for this element type
N_σ	Number of occurrences of element A contributing to the mean esd value.

The general-atom workfile has the structure and information content shown in Table 6.4. The general-atom workfile is denoted as data set 2A throughout this chapter, to distinguish it from the carbon-based data set 2, described above in Table 6.3

6.2.5 Methods of data analysis

Data sets 1, 2, 2A and 3 were used to generate simple descriptive statistics to analyse the distributions of the information contained in or derived from the CSD. Data sets 2 and 3 were then used in correlation experiments to investigate the interdependence of the extracted data items. Finally linear regression techniques were used to generate expressions that approximate the bond length and coordinate e.s.d's. The linear regressions were of two general forms:

$$y = a + g\left(f_p\right) \qquad \text{Type 1} \qquad 6.8(a)$$

$$y = a + bL + cM + ... \qquad \text{Type 2} \qquad 6.8(b)$$

where f_p is a function analogous to Cruickshank's derivation (1960), and L,M *etc*. are various data items from the respective workfile. In this work, calculations utilised statistical routines provided by the CAMAL library (Taylor, 1986) or from the NAG library (Numerical Algorithms Group, 1990).

6.3 Descriptive statistics for the retrieved data.

6.3.1 Data set 1 (83516 CSD entries)

Figure 6.2(a) shows the distribution of R-factors across all entries in data set 1. Figures 6.2(b) and (c) give the R-factor distributions for light ($Z_{max} \leq 18$) and heavy atom ($Z_{max} > 18$) structures respectively. The means of the distributions (b) and (c) are slightly different, with the mean R-factor being lower for heavy atom structures at 0.055 as compared to the light atom structures which have a mean R of 0.060. Very few

Figure 6.2: Distributions of R-factor.

(a) All data
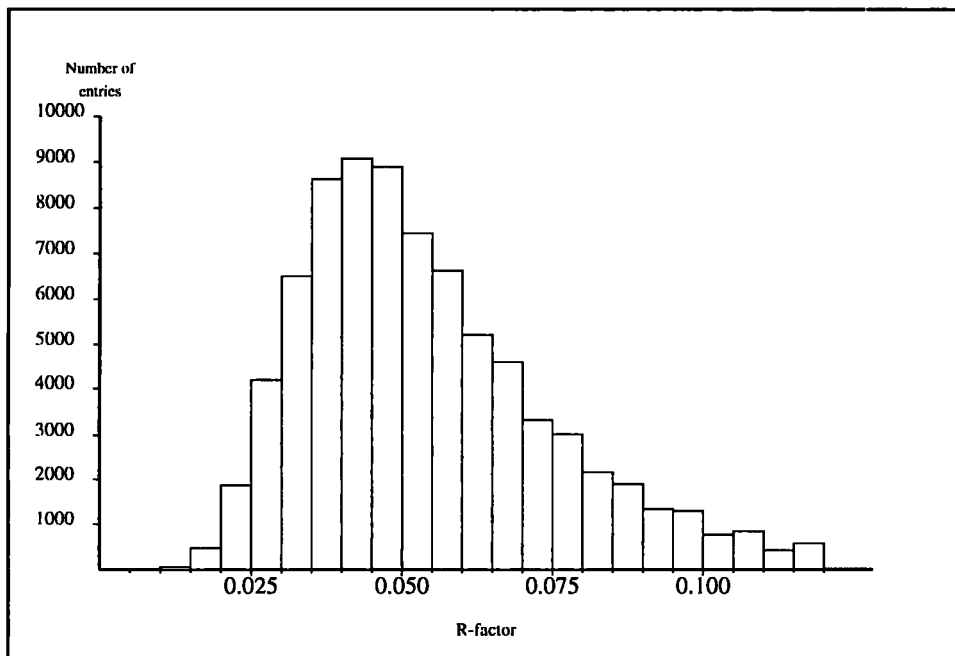
(b) Light atom ($Z_{max} \leq 18$) structures
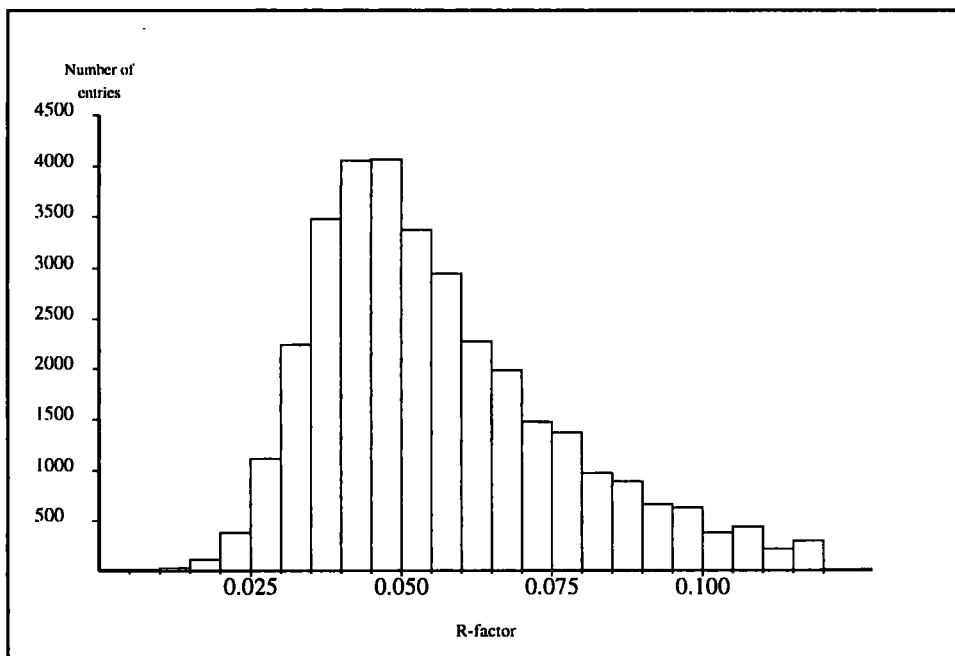
(c) Heavy ($Z_{max} > 18$) atom structures
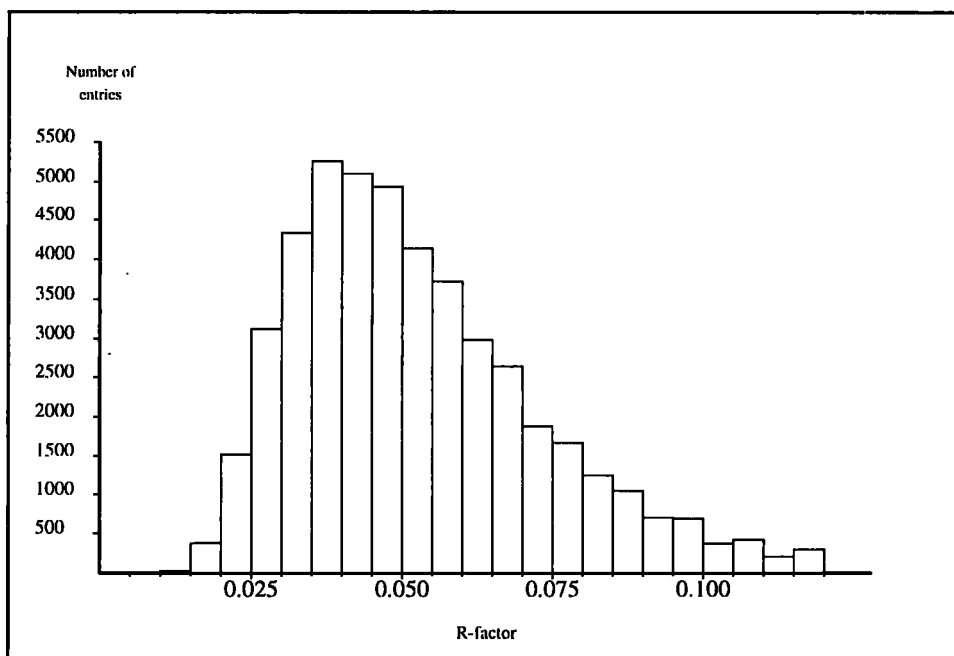
structures have R-factors less than 0.01. This result is in keeping with crystallographic theory and experience, namely that structures containing heavier atoms tend to scatter X-rays more strongly and are well located, which improves the R-factor. This is because the heavy elements have large scattering factors and so certain reflections are very intense. As a result of this certain reflections are more precisely measured and so the heavy atoms tend to be more precisely determined: the improvement in the R-factor reflects the high precision of the heavier elements within a particular structure.

Figure 6.3(a), (b) and (c) show similar overall, light atom and heavy atom distributions for the AS flag in the database. This gives an indication of the behaviour of $\overline{\sigma}(C-C)$ as a function of Z_{max}. The three figures show that the AS flag is considerably affected as a result of the presence of a heavy metal within the structure. This, again, is a result that is in accord with crystallographic theory and experience. The overall mean AS flag

Figure 6.3. Distributions of AS flag.

(a) All atoms
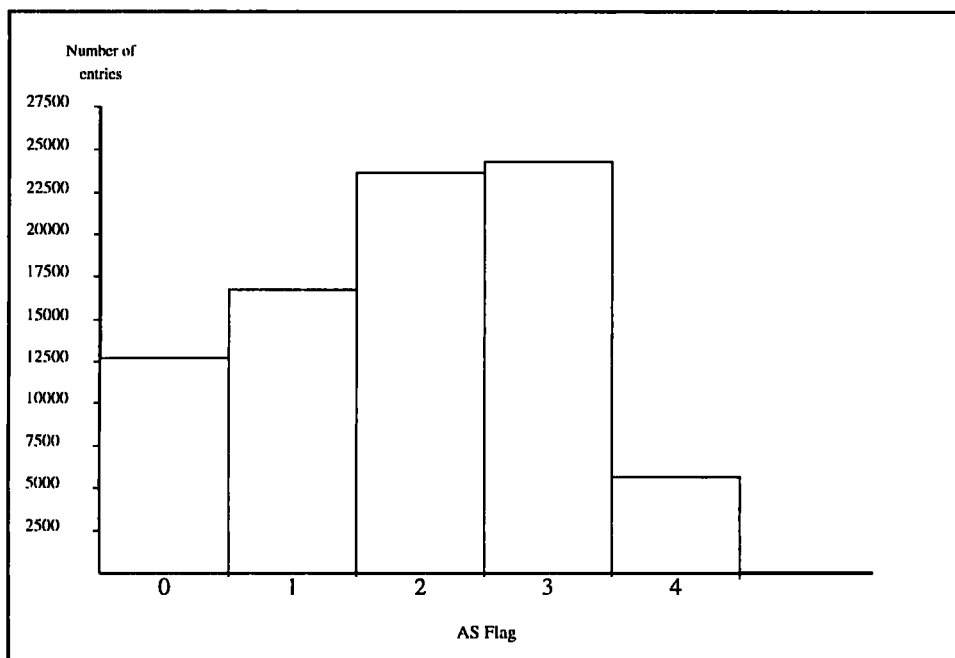
(b) Light ($Z_{max} \leq 18$) atom structures
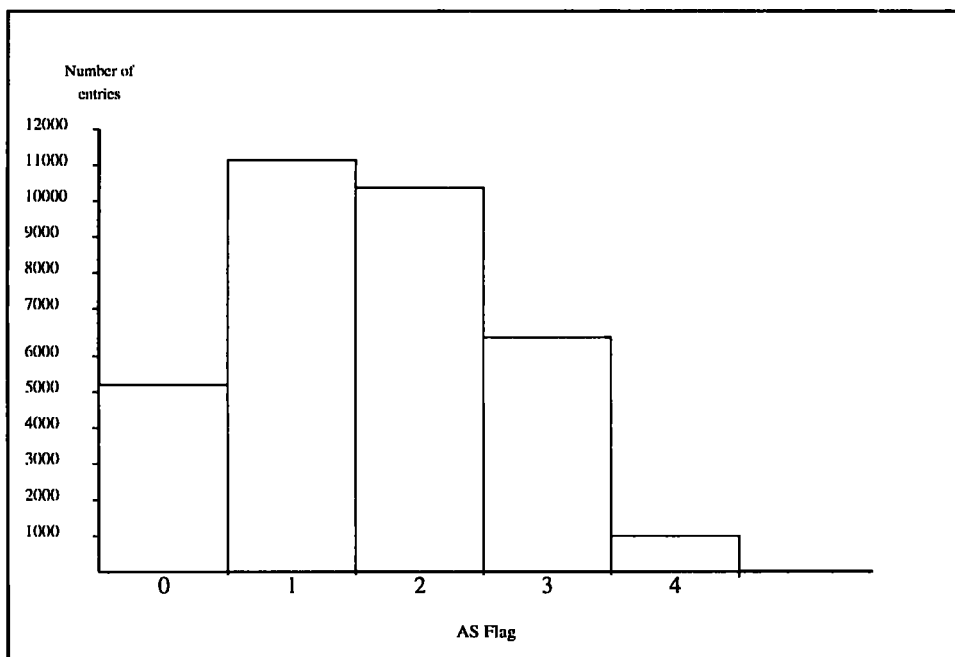
(c) Heavy ($Z_{max} > 18$) atom structures
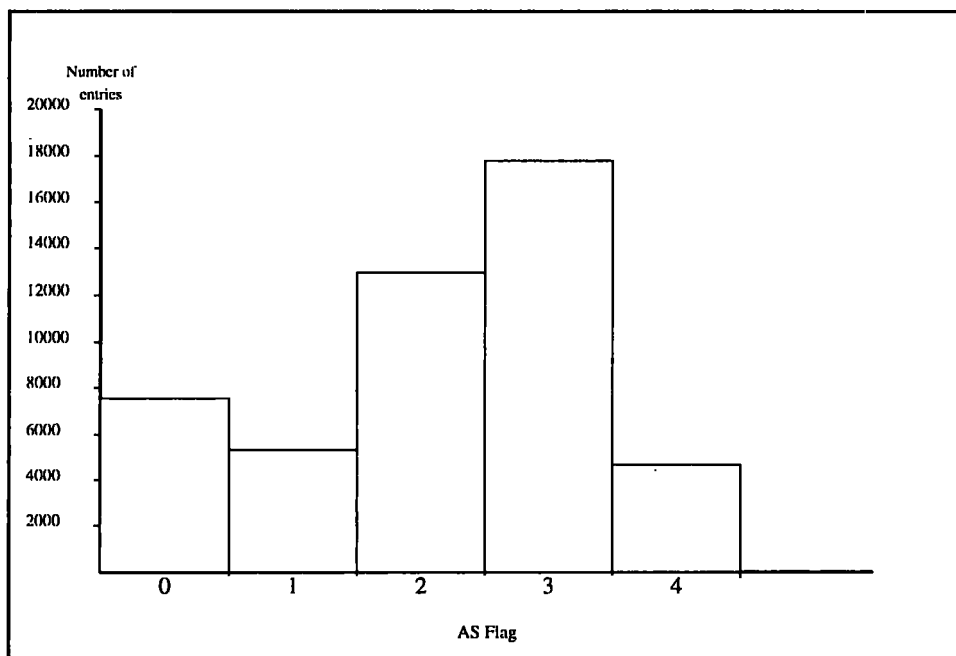
Table 6.5 The Distribution of the AS flag (data set 1) and of $\bar{\sigma}(C - C)$ (data set 2) for ranges of Z_{max}.

N_{ent} is the number of entries lying within the Z_{max} range. $<R>$, $<Z_{max}>$, $<AS>$ and $< \bar{\sigma}(C\text{-}C)>$ are the mean values over the particular Z_{max} range. The annotated values in parentheses for the AS flags (for data set 1) are the percentages of the total data set in the range.

	Z_{max} range				
	1	2	3	4	5
	6-10	11-18	19-36	37-57	58+
Data set 1					
N_{ent}	20137	28091	21777	8001	5510
$<R>$	0.059	0.056	0.060	0.053	0.051
$<Z_{max}>$	7.84	15.99	28.68	46.46	76.50
$<AS>$	1.71	2.28	2.47	2.61	3.06
AS=0	2697(3)	4608(6)	3664(4)	1221(1)	859(1)
AS=1	8212(10)	5243(6)	2563(3)	656(1)	104(-)
AS=2	6335(8)	8276(10)	6223(7)	2143(3)	708(1)
AS=3	2700(3)	8200(10)	7614(9)	3160(4)	2639(3)
AS=4	193(-)	1764(2)	1713(2)	821(1)	1200(1)
Data set 2					
N_{ent}	7831	4627	6005	4581	3485
$<R>$	0.052	0.051	0.050	0.044	0.042
$<Z_{max}>$	7.88	15.94	28.35	46.00	76.15
$< \bar{\sigma}(C\text{-}C)>$	0.0064	0.0076	0.0112	0.0127	0.0180

138

Table 6.6: Distributions of the AS-flag vs. R-factor for CSD entries having coordinates and published since 1965 (data set 1)

N_{ent} is the number of entries in a given R-factor range, <R> is the mean of the available (R >0.0) R-factors, <AS> is the mean (see text) of the available non-zero AS-flags. Integers (≥ 1) in parentheses are percentages of the total number of entries in each subdivision.

R-range	N_{ent}	<R>	AS=0	AS=1	AS=2	AS=3	AS=4	<AS>
(a) All entries (83516)								
not known	581(1)	-	261(-)	73(-)	95(-)	117(-)	35(-)	2.36
0.001-0.030	5832(7)	0.025	667(1)	2228(3)	2123(3)	773(1)	41(-)	1.73
0.031-0.040	14725(17)	0.035	1602(2)	5028(6)	4869(6)	3016(4)	210(-)	1.88
0.041-0.050	18268(22)	0.045	2249(3)	5191(6)	5806(7)	4475(5)	547(1)	2.02
0.051-0.060	14650(17)	0.054	2034(2)	2556(3)	4695(6)	4592(6)	773(1)	2.28
0.061-0.070	10267(12)	0.064	1599(2)	1023(1)	2937(4)	3866(5)	842(1)	2.52
0.071-0.080	6721(8)	0.074	1232(2)	392(-)	1572(2)	2720(3)	805(1)	2.71
0.081-0.100	7174(8)	0.088	1548(2)	213(-)	1208(1)	3028(4)	1177(1)	2.91
0.101-	5298(6)	0.125	1857(2)	74(-)	380(-)	1726(2)	1261(2)	3.21
Totals	82935(99)	0.057	13049(16)	16778(20)	23685(28)	24313(29)	5691(7)	2.27
(b) 'Organic' structures : CSD classes 1-59 (34992)								
not known	301(1)	-	119(-)	60(-)	59(-)	54(-)	9(-)	2.07
0.001-0.030	1334(4)	0.025	199(1)	751(2)	286(1)	90(-)	8(-)	1.43
0.031-0.040	5317(15)	0.035	571(2)	3167(9)	1259(4)	284(1)	36(-)	1.41
0.041-0.050	8110(23)	0.045	898(3)	3986(11)	2501(7)	670(2)	55(-)	1.55
0.051-0.060	6474(19)	0.054	795(2)	2081(6)	2549(7)	967(3)	82(-)	1.83
0.061-0.070	4484(13)	0.064	620(2)	832(2)	1814(5)	1129(3)	89(-)	2.12
0.071-0.080	2946(8)	0.074	502(1)	305(1)	1062(3)	986(3)	91(-)	2.35
0.081-0.100	3338(10)	0.088	682(2)	173(-)	869(3)	1389(4)	225(1)	2.63
0.101-	2688(8)	0.125	999(3)	50(-)	285(1)	946(3)	408(1)	3.01
Totals	34691(99)	0.060	5385(15)	11405(33)	10684(31)	6515(19)	1003(3)	1.90
(c) 'Metallo-organic' structures: CSD classes 60-86 (48524)								
not known	280(1)	-	142(-)	13(-)	36(-)	63(-)	26(-)	2.73
0.001-0.030	4498(9)	0.025	468(1)	1477(3)	1837(4)	683(1)	33(-)	1.82
0.031-0.040	9408(19)	0.035	1031(2)	1861(4)	3610(7)	2732(6)	174(-)	2.14
0.041-0.050	10158(21)	0.045	1351(3)	1205(3)	3305(7)	3805(8)	492(1)	2.40
0.051-0.060	8176(17)	0.054	1239(3)	475(1)	2146(4)	3625(7)	691(1)	2.65
0.061-0.070	5783(12)	0.064	979(2)	191(-)	1123(2)	2737(6)	753(2)	2.84
0.071-0.080	3775(8)	0.074	730(2)	87(-)	510(1)	1734(4)	714(2)	3.00
0.081-0.100	3836(8)	0.088	866(2)	40(-)	339(1)	1639(3)	952(2)	3.18
0.101-	2610(5)	0.124	858(2)	24(-)	95(-)	780(2)	853(2)	3.40
Totals	48244(99)	0.055	7664(16)	5373(11)	13001(27)	17798(37)	4688(10)	2.53

139

is 2.27 for all structures in data set 1, but rises to 2.53 for heavy atom structures and decreases to 1.91 for light atom structures: when heavier atoms dominate the scattering, then the precision with which lighter atoms can be located is reduced, even though the R-factor decreases with Z_{max}.

The patterns observed above are further quantified in Table 6.5, where a fuller breakdown on the basis of Z_{max} is shown. The relationships described above are shown more clearly in this table for both data sets 1 and 2.

Table 6.6 shows the relationship between the AS flag and the R-factor. In this table structures have been broken down into "organic" and "metallo-organic" structures by use of the CSD chemical classification system. In Table 6.6(a) we can see that the R-factor and AS flag are correlated. The <AS> value increases from 1.73 to 3.21 as <R> increases from 0.025 to 0.125. Similar effects are seen for the organic and metallo-organic subsets, although, as is expected in view of Table 6.5, the distributions of R-factor and AS flag are somewhat different in each subset: the <AS> values are lower for the lighter atom "organic" structures of Table 6.6(b) than for the "metallo-organics" of Table 6.6(c). It is hoped that this table (Allen, Cole and Howard,1995,1,2) may be used by researchers who wish to obtain some idea of the number of structures that will be rejected when the AS flag or the R-factor are used as secondary search criteria.

6.3.2 Data set 2 (29362 CSD entries)

Figure 6.4(a) shows the distribution of the calculated $\overline{\sigma}(C-C)$ values across the data set. The distribution for light atom structures (Zmax ≤ 18) is shown in Figure 6.4(b), while Figure 6.4(c) shows the same distribution for the heavy atom compounds (Zmax > 18). As expected, the distributions follow the same pattern as is observed for the AS flag; The means of the $\overline{\sigma}(C-C)$ distributions are:

$$\langle \overline{\sigma}(C - C) \rangle = 0.0103 \text{Å} \qquad \text{(All structures)} \qquad 6.9(a)$$

$$\langle \overline{\sigma}(C - C) \rangle = 0.0068 \text{Å} \qquad \text{(Light atom structures)} \qquad 6.9(b)$$

$$\langle \overline{\sigma}(C - C) \rangle = 0.0134 \text{Å} \qquad \text{(Heavy atom structures)} \qquad 6.9(c)$$

Equations 6.9(a)-(c) again show that the precision of the lighter atoms is strongly affected by the presence of heavier elements. Referral to Table 6.5 and 6.6 reinforces this point. Figure 6.4(a) also indicates how ill suited the current AS bandings are as precision indicators, particularly for light atom structures. In Figure 6.4(b), $\overline{\sigma}(C - C)$ are rarely above 0.02Å and the vast majority of structures have $\overline{\sigma}(C - C) < 0.01$Å.

Figures 6.4(a), (b) and (c): Distributions of $\overline{\sigma}(C - C)$ in data set 2

(a) All structures.
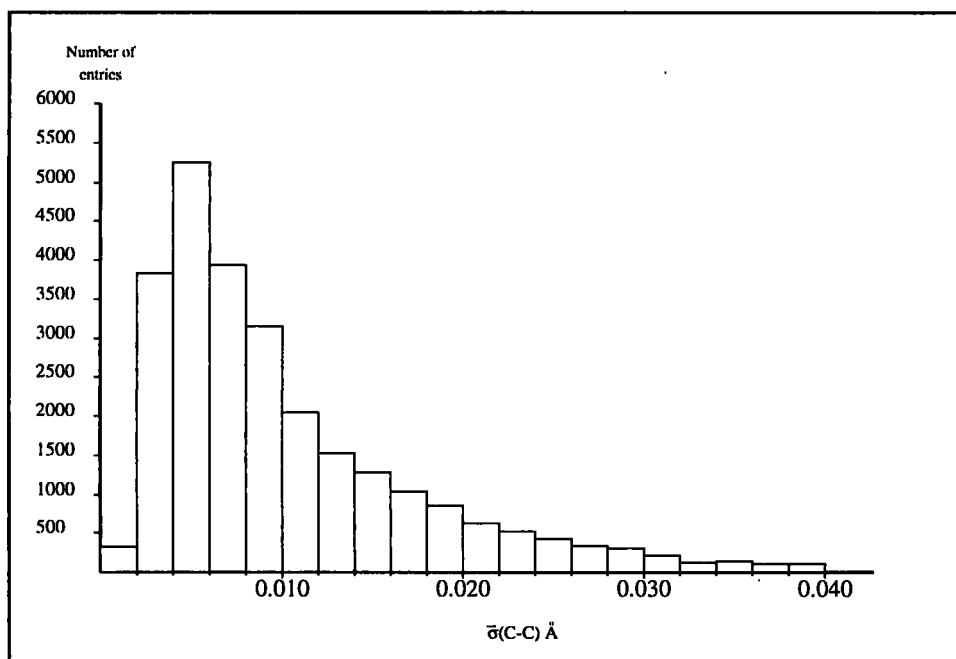
(b) Structures where $Z_{max} \leq 18$
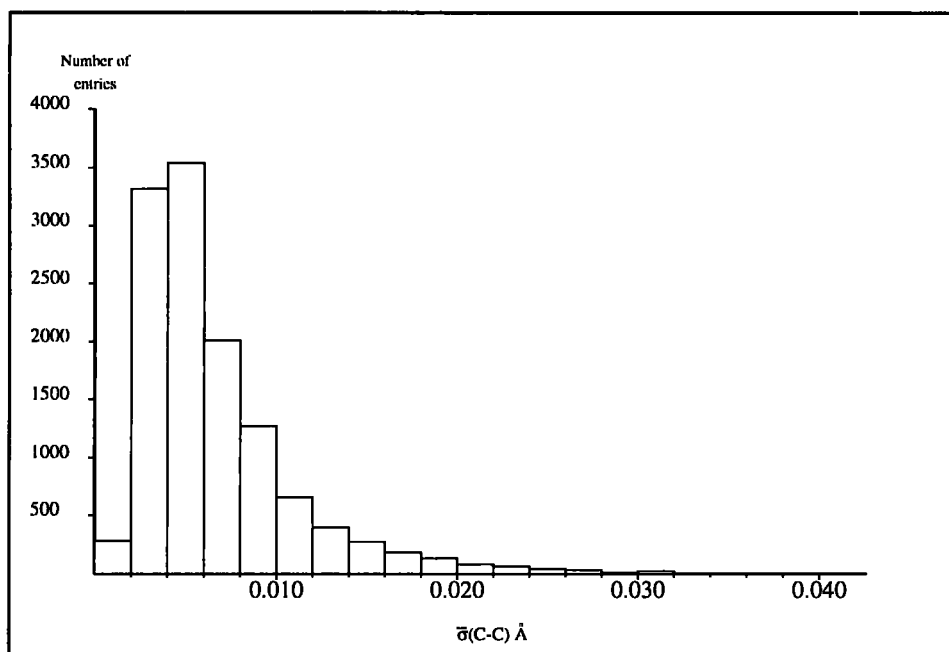
(c) Structures with $Z_{max} > 18$

142
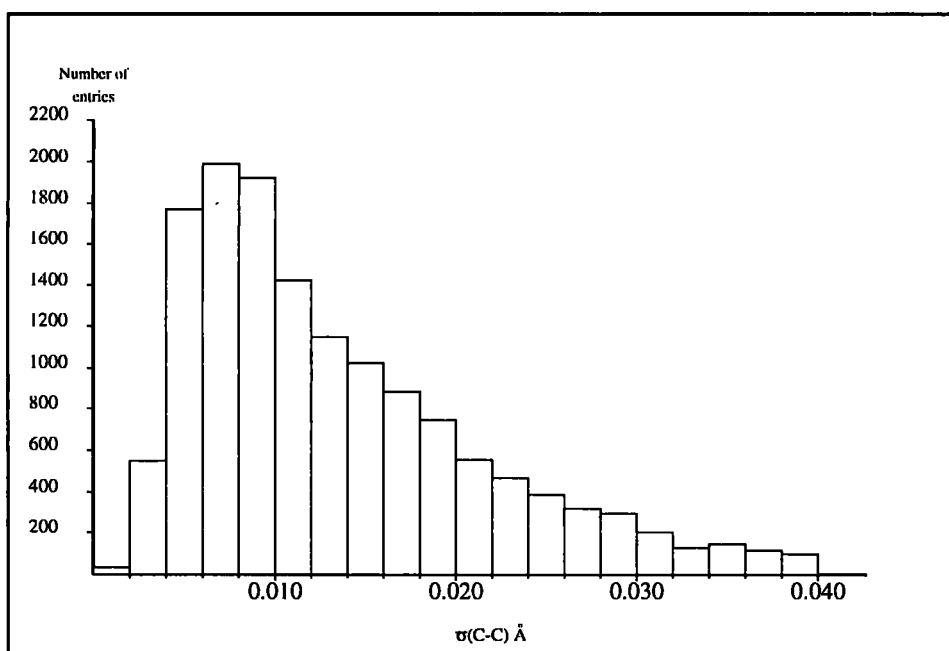
6.3.3 Data set 2A (124905 $\overline{\sigma}(E)$ values)

80 element types are represented within the workfile, ranging from lithium (with $Z = 3$) to neptunium (with $Z = 93$) and with the numbers of occurences ranging from 1 (for xenon) to 32715 occurences (for carbon). 24 elements occur in more than 500 structures within the workfile. It is inappropriate, due to the high number of elements represented to plot individual $\overline{\sigma}(E)$ distributions and a single plot of all $\overline{\sigma}(E)$ values for any element would also be inappropriate due to the dependancy of $\overline{\sigma}(E)$ on Z_E. To represent the variation in the $\overline{\sigma}(E)$ across the full element range in the workfile a "carbon-equivalent" value, $\overline{\sigma}(C)_e$, has been derived for each $\overline{\sigma}(E)$ within the workfiles. $\overline{\sigma}(C)_e$ can be derived as follows:

In any single structure the quantities R, $\overline{s}$ and p in Cruickshank's (1960) original expression (Eqn 6.2) are constants. This expression may be rewritten as:

$$\sigma(E) = CN_E^{\frac{1}{2}} \qquad\qquad 6.10$$

where C is a constant. If two different elements, A and B, are present in a particular structure then we can express the ratio between $\sigma(A)$ and $\sigma(B)$ using equation 6.10:

$$\frac{\sigma(A)}{\sigma(B)} = \frac{N_A^{\frac{1}{2}}}{N_B^{\frac{1}{2}}} \qquad\qquad 6.11$$

Using equation 6.3, we can also write

$$\sqrt{\frac{N_A}{N_B}} = \sqrt{\frac{\sum_i^{N_{ats}} f_i^2 \big/ f_A^2}{\sum_i^{N_{ats}} f_i^2 \big/ f_B^2}} = \frac{f_B}{f_A} \qquad\qquad 6.12$$

By approximating the scattering factors to the atomic number, (i.e. to the scattering factor at $\sin\theta/\lambda = 0$) we can write:

$$\sqrt{\frac{N_A}{N_B}} = \frac{Z_B}{Z_A} \qquad\qquad 6.13$$

And so

143

$$\frac{\sigma(A)}{\sigma(B)} = \frac{Z_B}{Z_A} \qquad\qquad 6.14$$

Using equation 6.14 it is possible to derive a normalized value of $\overline{\sigma}(E)$. For an element, E, we can write:

$$\overline{\sigma}(C)_e = \frac{\overline{\sigma}(E)_o . Z_E}{6} \qquad\qquad 6.15$$

Figure 6.5 shows the distribution of $\overline{\sigma}(C)_e$ for 124905 instances. The shape of the distribution across the workfile is similar to that observed for $\overline{\sigma}(C-C)$. The median of the distribution is circa 0.006Å, and, as in the $\overline{\sigma}(C-C)$ case, is highly skewed towards lower values. In the plot, some 2190 $\overline{\sigma}(E)_o$ values convert to $\overline{\sigma}(C)_e$ with values higher than 0.030Å. Figure 6.5 leads to the conclusion that the approximations involved in deriving equation 6.15 are reasonable.

Figure 6.5 The distribution of $\overline{\sigma}(C)_e$ across data set 2A
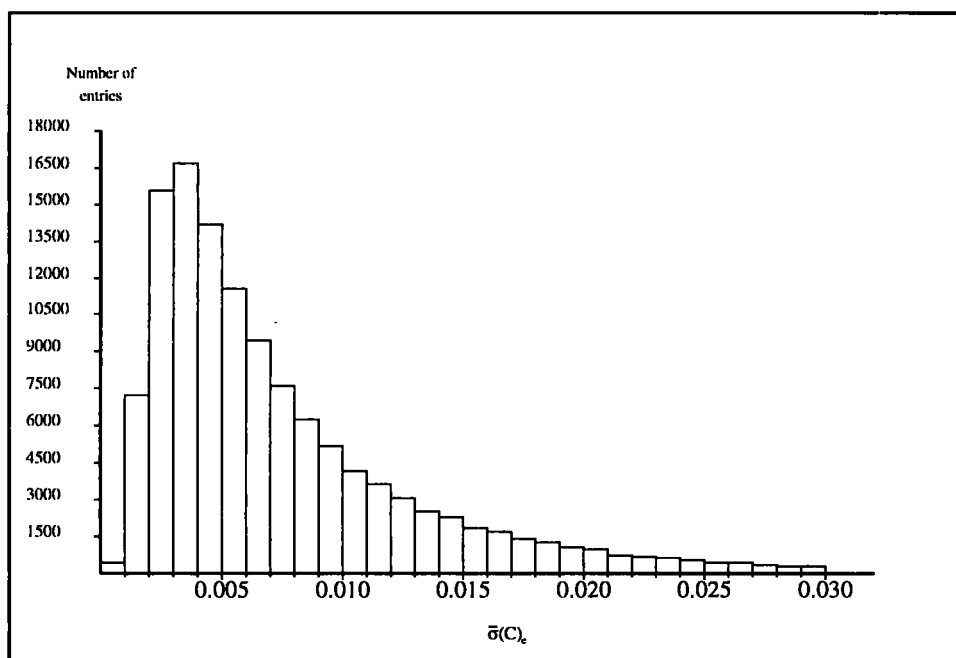
More quantitative evidence for this is given in Table 6.7, where values of $\langle\overline{\sigma}(C)_e\rangle$ are shown for the 24 most common elements in the database. The values of $\langle\overline{\sigma}(C)_e\rangle$ span the range from 0.0055-0.0116Å, with 17 values in the narrow range between 0.0055-0.0081Å. The remaining 7 elements comprise 3 halogens, (fluorine, bromine

and iodine) and 4 of the heaviest elements, (tungsten, rhenium, osmium and platinum). The high mean values for halogen elements are probably due to their monovalent nature. Halogen atoms are usually on the periphery of a molecule and hence are more susceptible to thermal motion. A consequence of this is that their positions are often less well determined.

Table 6.7. Values of $<\overline{\sigma}(E)_o>$ and $<\overline{\sigma}(C)_e>$ (see text) in Å for the 24 elements with ≥ 500 occurrences in the full workfile.

Values are given for (a) all data and (b) after removal of the upper and lower decile of the $<\overline{\sigma}(C)_e>$ distribution shown in Fig.6.5. El is the element symbol, Z is the atomic number. N_{occ} is the number of occurrences of each element type.

El	Z	(a) all data			(b) after decile removal		
		N_{occ}	$<\overline{\sigma}(E)_o>$	$<\overline{\sigma}(C)_e>$	N_{occ}	$<\overline{\sigma}(E)_o>$	$<\overline{\sigma}(C)_e>$
B	5	1583	0.00852	0.0071	1295	0.00828	0.0069
C	6	32715	0.00792	0.0079	26760	0.00678	0.0068
N	7	18564	0.00591	0.0069	15017	0.00531	0.0062
O	8	24574	0.00585	0.0078	19722	0.00488	0.0065
F	9	2192	0.00733	0.0110	1631	0.00520	0.0078
Si	14	1615	0.00249	0.0058	1316	0.00236	0.0055
P	15	6927	0.00276	0.0069	5693	0.00252	0.0063
S	16	6895	0.00236	0.0063	5518	0.00221	0.0059
Cl	17	6552	0.00286	0.0081	5374	0.00233	0.0066
Cr	24	582	0.00138	0.0055	471	0.00138	0.0055
Fe	26	1828	0.00136	0.0059	1493	0.00141	0.0061
Co	27	1100	0.00138	0.0062	872	0.00140	0.0063
Ni	28	759	0.00139	0.0065	626	0.00135	0.0063
Cu	29	1265	0.00136	0.0066	1034 ·	0.00132	0.0064
Br	35	1600	0.00158	0.0092	1374	0.00130	0.0076
Mo	42	1299	0.00104	0.0073	1042	0.00106	0.0074
Ru	44	1041	0.00094	0.0069	837	0.00094	0.0069
Rh	45	811	0.00097	0.0073	644	0.00100	0.0075
Sn	50	532	0.00095	0.0079	433	0.00084	0.0070
I	53	946	0.00120	0.0106	743	0.00097	0.0086
W	74	1042	0.00084	0.0103	741	0.00059	0.0073
Re	75	564	0.00078	0.0098	398	0.00055	0.0069
Os	76	587	0.00092	0.0116	427	0.00065	0.0082
Pt	78	941	0.00084	0.0109	639	0.00055	0.0071

Table 6.8. Values of $<\bar{\sigma}(E)_O>$ and $<\bar{\sigma}(C)_e>$ (see text) in Å for all 80 element types represented in the full workfile averaged over ranges of atomic number Z.

Values are given for (a) all data and (b) after removal of upper and lower deciles of the $<\bar{\sigma}(C)_e>$ distribution of Fig.1. N_{OCC} is the number of element occurrences in each range.

		(a) all data				(b) after decile removal			
Z_{min}	Z_{max}	N_{OCC}	$<Z>$	$<\bar{\sigma}(E)_O>$	$<\bar{\sigma}(C)_e>$	N_{OCC}	$<Z>$	$<\bar{\sigma}(E)_O>$	$<\bar{\sigma}(C)_e>$
3	10	79960	6.90	0.00670	0.0077	64707	6.88	0.00567	0.0065
11	18	22562	15.74	0.00267	0.0070	18354	15.74	0.00236	0.0062
19	36	9981	28.48	0.00145	0.0069	8214	28.57	0.00137	0.0065
37	57	6703	46.27	0.00102	0.0079	5330	46.21	0.00096	0.0074
58	92	5144	76.27	0.00083	0.0106	3639	76.21	0.00059	0.0075

Table 6.8 shows that heavier elements are relatively more imprecise as compared to light elements. The distribution is dramatically flattened if the upper and lower deciles of the $\bar{\sigma}(C)_e$ distribution are removed, as is also shown in Table 6.7. The range that lies within the upper and lower deciles of this distribution has $0.0023 \leq \bar{\sigma}(C)_e \leq 0.0160$Å. These deciles are denoted by σ+ and σ-

The above conclusions are further enhanced by Table 6.8, in which values of $\left\langle \bar{\sigma}(C)_e \right\rangle$ have been calculated for ranges of Z

6.3.4. Inter-relationships between $\bar{\sigma}(C - C)$, the R-factor and Z_{max}

Binning procedures were used to analyse how the R-factor is related to $\bar{\sigma}(C - C)$ over ranges of Z_{max}. Data set 2 was divided into 5 ranges of Z_{max}. Structures in these Z_{max} ranges were then binned into 5 different R-factor ranges, namely 0.001-0.035, 0.036-0.045, 0.046-0.055, 0.056-0.070 and 0.071-0.100. For each of these Z_{max} and R ranges values of $\left\langle \bar{\sigma}(C - C) \right\rangle$ and $\left\langle R \right\rangle$ were determined. Figure 6.6 shows a plot of $\left\langle R \right\rangle$ against $\left\langle \bar{\sigma}(C - C) \right\rangle$ for each range of Z_{max}. The size of each of the 25 bins described above ranged from 433 to 2251 entries, except for the highest R-factor bins in the two highest Z_{max} ranges, (4 and 5 in Figure 6.6), which contained 300 and 141 structures respectively. This is hardly surprising, considering the relationships already noted

Figure 6.6. $\langle R \rangle$ vs $\langle \overline{\sigma}(C-C) \rangle$ for 5 ranges of Z_{max}:

Ranges shown: (a) $3 \leq Z_{max} \leq 10$, (b) $11 \leq Z_{max} \leq 18$, (c) $19 \leq Z_{max} \leq 36$, (d) $37 \leq Z_{max} \leq 57$, (e) $58 \leq Z_{max} \leq 92$
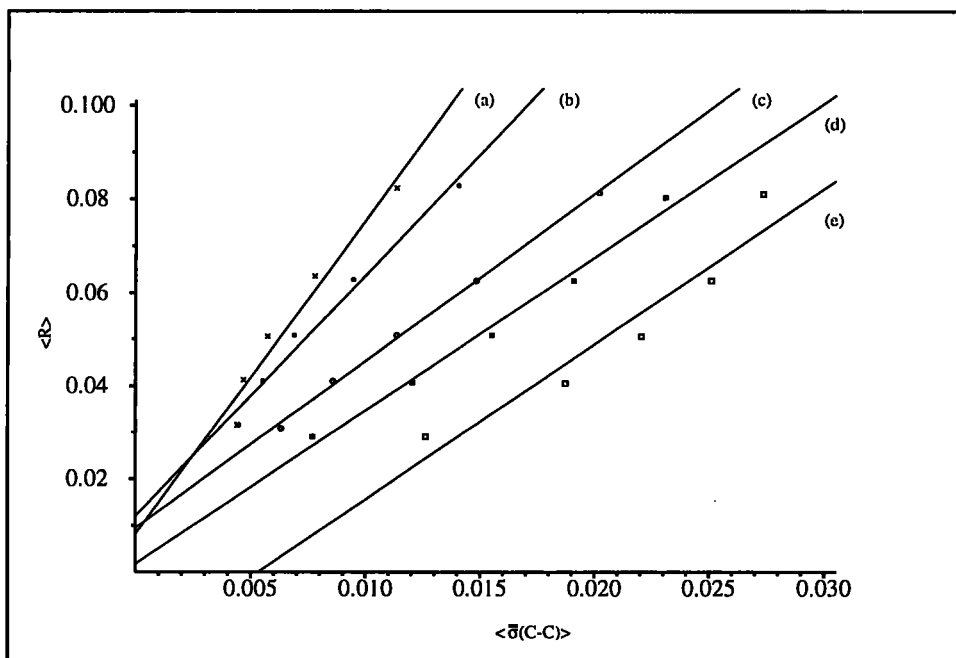
between R-factor and Z_{max} earlier in this chapter. Importantly in this exercise, in any Z_{max} range, the value of $\langle Z_{max} \rangle$ is roughly constant across all five R-factor bins.

Figure 6.6 shows that there is a linear relationship between $\langle \overline{\sigma}(C-C) \rangle$ and $\langle R \rangle$.in each Z_{max} sub-division. It also shows graphically how $\langle \overline{\sigma}(C-C) \rangle$ changes with Z_{max}. Even with the variation in bin sizes these results are in agreement with Cruickshank's (1960) derivation.

6.3.5 Data set 2: Tests of modified versions of the Cruickshank formula.

It is also possible, using binning methods, to test modified versions of the Cruickshank (1960) formulation (Equation 6.2). To do this, values of $\langle Rf(Z) \rangle$ where f(z) is some function of the atomic numbers of elements in each structure and R is the R-factor, were calculated and the data set was binned over various ranges of these values (see below).

Figure 6.7. Binned plots of $\langle \overline{\sigma}(C-C) \rangle$ vs (a) $\langle RN_c^{1/2} \rangle$, (b) $\langle RZ_{rms} \rangle$ and (c) $\langle RZ_{max} \rangle$
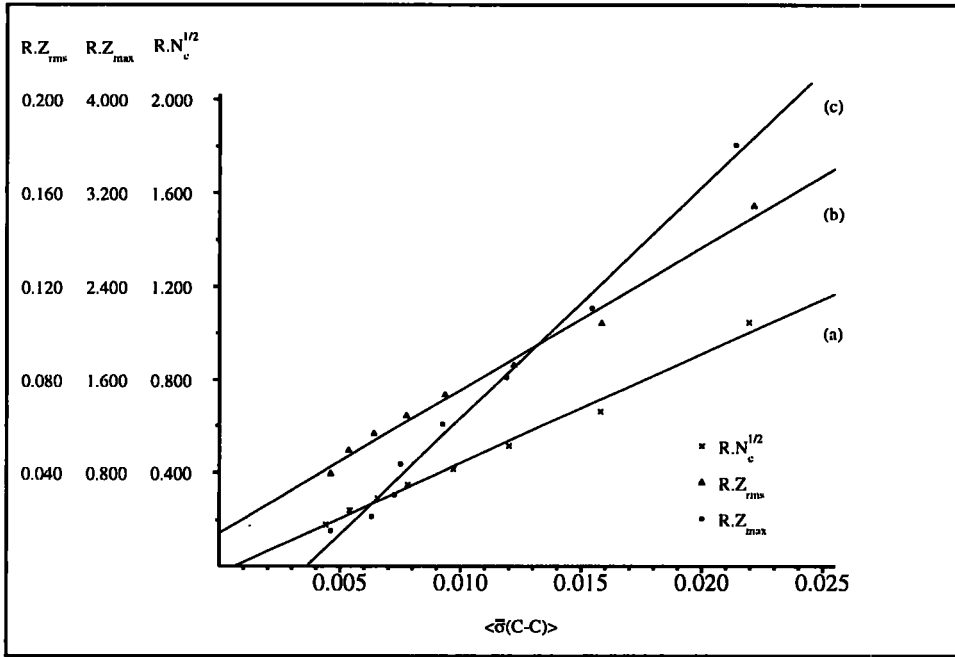
Within each bin a mean value of $\langle \overline{\sigma}(C-C) \rangle$ is calculated and the data points are plotted. In Figure 6.7, three different functions, $\langle RZ_{max} \rangle$, $\langle RZ_{rms} \rangle$, $\langle RN_c^{1/2} \rangle$ are shown.

The bin composition was generated from sorted lists so that each bin contained exactly 3248 entries. A linear relationship is shown to exist between all three functions and $\langle \overline{\sigma}(C-C) \rangle$, which is promising. Two of these functional forms, $\langle RZ_{rms} \rangle$, $\langle RN_c^{1/2} \rangle$, are notably better since they both pass close to the origin.

6.3.6 Data set 2: Correlation analyses.

A more exact test on the quality of each of these functions is given by correlation analysis. A full symmetric correlation matrix was generated for all the numerical data items (Table 6.3) that were extracted from the database. The larger correlation coefficients are shown in Table 6.9 below, for two subsets, denoted as S140 and S220. The formulation of these subsets is defined in Table 6.9(a), the coefficients are shown in Table 6.9(b).

Table 6.9. Correlation analysis of data set 2.

Two subsets of entries, denoted as S140 and S220, were chosen for analysis on the basis of $\overline{\sigma}$(C-C) and R-factor limits. The C_{ij} are correlation coefficients selected from the complete matrix for each subset.

(a) Subset definition

	Subset name	S140	S220
	$\overline{\sigma}$(C-C) minimum	0.001Å	0.002Å
	$\overline{\sigma}$(C-C) maximum	0.040Å	0.020Å
	R minimum	0.001	0.001
	R maximum	0.100	0.070
	No. of entries	25984	20334
	$<\overline{\sigma}$(C-C)$/\overline{\sigma}$(C)$>$	1.377	1.377

(b) Correlation coefficients

	Item (i)	Item (j)	C_{ij}	C_{ij}
	R	AS	0.359	0.214
	R	T	0.051	0.050
	R	N_{nh}	0.166	0.092
	R	N_h	0.022	0.038
	R	V	0.132	0.063
	R	Z_{max}	-0.227	-0.337
	R	Z_{rms}	-0.229	-0.343
	R	$N_c^{\frac{1}{2}}$	-0.114	-0.232
	R	$\overline{\sigma}$(A)	0.113	0.361
	R	$\overline{\sigma}$(C)	0.369	0.221
	$\overline{\sigma}$(C-C)	R	0.363	0.214
	$\overline{\sigma}$(C-C)	AS	0.851	0.888
	$\overline{\sigma}$(C-C)	T	0.044	0.045
	$\overline{\sigma}$(C-C)	N_{nh}	0.338	0.303
	$\overline{\sigma}$(C-C)	N_h	-0.030	0.026
	$\overline{\sigma}$(C-C)	V	0.351	0.303
	$\overline{\sigma}$(C-C)	Z_{max}	0.525	0.520
	$\overline{\sigma}$(C-C)	Z_{rms}	0.529	0.505
	$\overline{\sigma}$(C-C)	$N_c^{\frac{1}{2}}$	0.584	0.558
	$\overline{\sigma}$(C-C)	RZ_{max}	0.700	0.646
	$\overline{\sigma}$(C-C)	RZ_{rms}	0.737	0.657
	$\overline{\sigma}$(C-C)	$R\,N_c^{\frac{1}{2}}$	0.729	0.657
	$\overline{\sigma}$(C-C)	$\overline{\sigma}$(A)	0.031	0.035
	$\overline{\sigma}$(C-C)	$\overline{\sigma}$(C)	0.985	0.978

Many of these correlation coefficients confirm and quantify observations that are already apparent in the descriptive statistics of sections 6.3.1-6.3.5. R is positively correlated to the AS flag, and the functions of σ described in Table 6.9. R is negatively correlated to the functions of the unit cell constitution. Temperature does not correlate strongly with R-factor or with functions of σ. This merely reflects the dominance of room-temperature structural work in the CSD. The number of non-hydrogen atoms in the asymmetric unit, N_{nh}, and the unit cell volume, V, are both positively correlated with $\overline{\sigma}(C-C)$, but the R-factor also correlates positively with these quantities. (The functions of unit cell constitution also correlate slightly with V and N_{nh}. This is probably due to the fact that larger structures tend to have a higher concentration of light atoms). The most significant point from this analysis, however, is the strong correlation between the Cruickshank type functions (RZ_{max}, RZ_{rms}, $RN_c^{1/2}$) and $\overline{\sigma}(C-C)$, indicating the suitability of such quantities in the derivation of precision estimation equations.

6.3.7. Data set 2A: The relationship between $\langle \overline{\sigma}(C)_e \rangle$ and $\langle RN_c^{1/2} \rangle$

Above, in section 6.3.3, it was shown that the $\overline{\sigma}(C)_e$ is relatively constant across the range of Z by using a binning analysis. A similar analysis, as described above, has been used to study the relationship between $\langle \overline{\sigma}(C)_e \rangle$ and $\langle RN_c^{1/2} \rangle$ using data set 2A. In Table 6.10 values are shown of $\langle \overline{\sigma}(C)_e \rangle$, for ranges of $RN_c^{1/2}$ along with a coefficient, k which is defined below:

$$k = \frac{\langle \overline{\sigma}(C)_e \rangle}{\langle RN_c^{1/2} \rangle} \qquad 6.16$$

The coefficient is roughly constant across the range of bins, although a definite trend exists from low to high values of $\langle RN_c^{1/2} \rangle$. The k values suggest that a linear relationship between $\langle \overline{\sigma}(C)_e \rangle$ and $\langle RN_c^{1/2} \rangle$ would *under*-estimate lower values of $\overline{\sigma}(E)_o$ and *over*-estimate the very highest values as compared to those observed in

Table 6.10. Values of $\langle RN_c^{\frac{1}{2}} \rangle$, $\langle \overline{\sigma}(C)_e \rangle$ and the ratio k for ranges of $RN_c^{1/2}$

N_{occ} is the number of element-type occurrences in each range. Binning procedures are described in the text.

$R N_c^{\frac{1}{2}}$ range	N_{occ}	$\langle R N_c^{\frac{1}{2}} \rangle$	$\langle \overline{\sigma}(C)_e \rangle$	k
< 0.1	390	0.0818	0.0034	0.0412
0.1-0.2	10620	0.1669	0.0032	0.0190
0.2-0.3	26015	0.2507	0.0041	0.0164
0.3-0.4	23647	0.3476	0.0056	0.0160
0.4-0.5	17437	0.4466	0.0071	0.0159
0.5-0.6	12721	0.5462	0.0086	0.0157
0.6-0.7	9447	0.6473	0.0102	0.0157
0.7-0.8	6961	0.7469	0.0116	0.0155
0.8-0.9	4824	0.8466	0.0129	0.0153
0.9-1.0	3542	0.9468	0.0140	0.0148
1.0-1.1	2472	1.0484	0.0153	0.0146
1.1-1.2	1725	1.1439	0.0164	0.0144
1.2-1.3	1445	1.2462	0.0172	0.0138
1.3-1.5	1366	1.3875	0.0187	0.0135
1.5-2.0	1162	1.6621	0.0203	0.0122
> 2.0	576	2.2802	0.0233	0.0102

Figure 6.8. A plot of $\langle \overline{\sigma}(C)_e \rangle$ against $\langle RN_c^{\frac{1}{2}} \rangle$

151
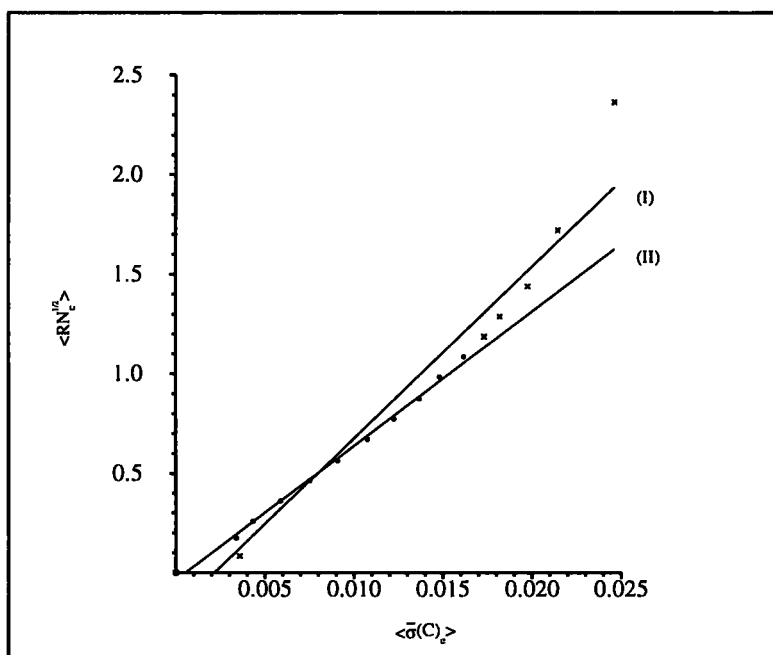
crystal structures. Such a function is more useful as a precision indicator. It is unlikely that a user would require a *maximum* precision, but if a *minimum* precision was specified the effect of using such a function would be that some data that is precise enough in reality would be omitted, rather than imprecise data being included. A scatter plot of the data given in Table 6.10 is shown in Figure 6.8.

The figure shows two lines. The first line, (denoted as (I)) is the fit of all points shown. The second line, (denoted as (II)), is the fit of all points within the range of $\left\langle RN_c^{1/2} \right\rangle$ of 0.1669-1.0484, (i.e. the upper and lower deciles were omitted: the line is denoted as (II) in Figure 6.8: points fitted in (II) are shown as circles).

The lines clearly show that by omitting the upper and lower ranges of the data, a more linear fit is obtained: (II) passes closer to the origin than (I). The quality of the fit is visibly better: nearly all the points in (II) intersect with the fitted line, at the scale shown.

6.4. Regression analyses of Data set 2

The results of the binning analyses above give a good qualitative idea of the suitability of possible models. In order to establish the best fitting models, multiple linear regression can be used to fit a series of independent variables as a linear expression to a set of observations, testing a linear function hypothesis, namely that the observations can be described as a linear combination of a set of functions of the independent variables, i.e.

$$Q = p_1 f^{(1)}(x_1) + p_2 f^{(2)}(x_2) + \ldots + p_p f^{(n)}(x_p)$$ 6.17(a)

or

$$Q = p_1 f^{(1)}(x_1, x_2 \ldots x_m) + p_2 f^{(2)}(x_{m+1}) + \ldots + p_p f^{(n)}(x_p)$$ 6.17(b)

where p_1 to p_p are coefficients, $f^{(1)}$ to $f^{(n)}$ etc. are functions and x_1 to x_p are independent variables. Q is an observation. Multiple linear regression gives estimates

of p_1 to p_p by minimising the difference between the observed values of Q and the values calculated by the linear hypothesis. Least squares is an example of a linear regression method.

A number of linear regressions were carried out on data set 2. The forms of these regressions are detailed in Table 6.11. Every regression was carried out on the two subsets of data set 2 that are denoted as S140 and S220. Type 1 regressions have the functional form defined by Cruickshank (1960), with the denominator assumed to be constant across all structures. Type 2 are similar but are not constrained such that the fit passes through the origin. Finally type 3 regressions use a multiple linear regression over individual data items. In the remaining discussion "observed" values are those which result from the stored σ's on the CSD, and are denoted by a subscripted "o". For example, an "observed" value of $\overline{\sigma}(C-C)$, will be denoted as "$\overline{\sigma}(C-C)_o$" and will have been calculated from the stored values of $\sigma(C)$ on the CSD. Similarly any σ-value that is derived from a regression is described as a "predicted" value and these will be subscripted with a "p".

Any analysis using linear regression needs some form of assessment criteria which describes how good a particular fit is as compared to another. The criteria are shown above in Table 6.12. The assessment criteria for regressions designed to estimate $\overline{\sigma}(C-C)$ also include criteria that show the success of prediction of the AS flag for each stored CSD entry.

Table 6.11. Types of regression used.

Enumeration and definition of the simple linear and multiple linear regressions performed on data set 2.

Regression No. (RN)	Functional Form
1. Simple linear, no constant	
1.1	$\overline{\sigma}(C-C)_p = kR$
1.2	$\overline{\sigma}(C-C)_p = kRZ_{max}$
1.3	$\overline{\sigma}(C-C)_p = kRZ_{rms}$
1.4	$\overline{\sigma}(C-C)_p = kRN_c^{1/2}$
2. Simple linear with constant	
2.1	$\overline{\sigma}(C-C)_p = a + kR$
2.2	$\overline{\sigma}(C-C)_p = a + kRZ_{max}$
2.3	$\overline{\sigma}(C-C)_p = a + kRZ_{rms}$
2.4	$\overline{\sigma}(C-C)_p = a + kRN_c^{1/2}$
3. Multiple linear	
3.2	$\overline{\sigma}(C-C)_p = a + bR + cZ_{max}$
3.3	$\overline{\sigma}(C-C)_p = a + bR + cZ_{rms}$
3.4	$\overline{\sigma}(C-C)_p = a + bR + cN_c^{1/2}$

Table 6.12 Parameters used to assess regression results

(For breivity, $\overline{\sigma}(C-C)_o$ and $\overline{\sigma}(C-C)_p$ are referred to as $\overline{\sigma}_0$ and $\overline{\sigma}_p$ respectively.

Item	Description
R_σ	A pseudo R-factor measuring the discrepancy between the $\overline{\sigma}_0$ and $\overline{\sigma}_p$ distributions where $R_\sigma = \Sigma|\overline{\sigma}_o - \overline{\sigma}_p|/\Sigma\overline{\sigma}_o$
rms(σ)	The root mean square error i.e. rms $(\sigma) = \left[\Sigma(\overline{\sigma}_o - \overline{\sigma}_p)^2 / n\right]^{\frac{1}{2}}$ for n observations in the subset
N_{50}	Percentage of entries with $|\overline{\sigma}_0 - \overline{\sigma}_p| \leq 0.0050$Å
N_{25}	Percentage of entries with $|\overline{\sigma}_0 - \overline{\sigma}_p| \leq 0.0025$Å
N_{10}	Percentage of entries with $|\overline{\sigma}_0 - \overline{\sigma}_p| \leq 0.0010$Å
n_{50}	Percentage of entries for which $\overline{\sigma}_p$ is within,
n_{25}	respectively, 50%,
n_{10}	25% and 10% of $\overline{\sigma}_0$
AS_0	Percentage of AS flags that are predicted exactly by the regression equation
AS_2	Percentage of AS flags that are predicted exactly or within 0.002Å of the relevant AS flag limits of Table 2

154

6.4.1. Prediction of $\overline{\sigma}(C - C)$ and $\overline{\sigma}(C)$: Preliminary regression analyses on Data sets 2 and 3

Regression results of all three types for the full subsets, S140 and S220, are shown below in Tables 6.13 and 6.14. Regressions 1.1 and 2.1, which are based on R alone, are notably worse than those based on regressions that include some function of the unit cell contents. Including Z_{max} as in regression type 1.2 improves the quality of the predictions, but not as much as using functions which invoke information about all of the unit cell contents. Very little difference in the assessment criteria exists between using Z_{rms} and $N_c^{1/2}$. The type 1 regressions seem better on average than the other regressions. In the S140 data set the best regression is 1.4 . The assessment of regressions for subset S220 is less clear cut in deciding between 1.3 and 1.4, but most of the statistics would suggest that 1.4 is slightly better.

Table 6.13. Overall results for the 25984 entries of subset S140 defined in Table 6.9(a)

RN	a	b or k	c	R_σ (%)	rms (σ) Å	N_{50} (%)	N_{25} (%)	N_{10} (%)	n_{50} (%)	n_{25} (%)	n_{10} (%)	AS_0 (%)	AS_2 (%)
1.1		0.2070(8)		50.8	0.0069	59.1	30.2	12.1	52.4	26.7	10.5	39.6	59.7
1.2		0.00651(2)		39.1	0.0058	73.2	47.5	21.4	67.0	35.3	14.0	59.0	76.9
1.3		0.1350(3)		35.5	0.0051	78.8	45.4	18.0	67.2	39.2	16.4	54.2	77.4
1.4		0.02157(6)		33.5	0.0051	79.3	54.8	24.4	75.1	44.0	18.4	61.1	83.2
2.1	0.0020(1)	0.171(3)		51.1	0.0069	58.0	28.7	11.3	51.4	26.1	10.4	39.5	59.0
2.2	0.00376(5)	0.00481(3)		36.9	0.0053	76.9	45.5	17.6	65.6	37.1	15.0	54.1	77.2
2.3	-0.00125(6)	0.1482(8)		34.2	0.0050	79.3	50.0	21.1	70.8	41.6	17.6	57.7	80.4
2.4	0.00138(6)	0.0194(1)		34.4	0.0051	79.6	50.0	20.2	70.0	40.9	17.4	57.0	80.7
3.2	-0.00755	0.23906	0.00021	35.9	0.0051	75.6	47.4	20.5	68.9	39.2	16.4	58.6	77.0
3.3	-0.01284	0.24010	0.00692	35.4	0.0051	76.4	48.7	21.2	70.1	40.3	16.9	58.5	78.4
3.4	-0.00855	0.20464	0.00092	35.1	0.0051	77.0	49.0	21.5	70.6	40.5	17.3	59.1	78.7

Table 6.14. Overall results for the 20334 entries of subset S220 defined in Table 6.9(a)

RN	a	b or k	c	R_σ (%)	rms(σ) Å	N_{50} (%)	N_{25} (%)	N_{10} (%)	n_{50} (%)	n_{25} (%)	n_{10} (%)	AS_0 (%)	AS_2 (%)
1.1		0.1755(7)		42.4	0.0044	77.2	43.9	17.9	62.6	32.0	13.1	40.1	68.9
1.2		0.00564(2)		38.5	0.0042	80.6	52.3	21.8	66.4	34.2	14.1	58.2	77.3
1.3		0.1169(3)		29.9	0.0032	89.1	63.2	27.6	78.1	47.6	20.3	57.1	85.6
1.4		0.01928(5)		30.2	0.0034	87.3	64.7	31.7	82.5	49.4	20.5	60.8	86.6
2.1	0.0047(1)	0.0777(2)		41.7	0.0042	79.9	42.2	16.3	61.6	32.3	13.1	40.3	61.2
2.2	0.00437(4)	0.00322(2)		31.1	0.0033	89.3	59.6	23.8	74.4	43.6	18.1	53.9	84.8
2.3	0.00079(6)	0.1068(9)		30.3	0.0032	89.4	61.7	26.0	76.2	46.3	19.6	55.1	84.9
2.4	0.00262(5)	0.0140(1)		30.4	0.0032	89.3	61.8	25.1	75.8	45.2	19.0	54.6	85.3
3.2	-0.00276	0.15959	0.00013	30.3	0.0032	89.4	60.7	26.1	76.4	45.9	19.3	56.8	84.2
3.3	-0.00627	0.15933	0.00457	30.7	0.0032	88.7	60.5	26.1	76.2	45.7	19.3	55.9	84.1
3.4	-0.00336	0.13170	0.00062	30.3	0.0032	89.0	61.0	26.3	76.7	46.0	19.5	56.4	84.7

For statistical reasons, if two models are not significantly different from one another then the better model is the one with fewer fitted parameters. Because of this, regressions of type 1 are preferable to the others. It is also noted that no good statistical reason exists for using Z_{rms} in preference to $N_c^{1/2}$.

6.4.2. Regressions that incorporate space-group information (Data set 2)

In the original Cruickshank formulation (1960), shown in equation 6.2, the denominator differs from centrosymmetric structures to non-centrosymmetric by a factor of $2^{1/2}$. In the next regressions subsets S140 and S220 are broken down further into four smaller subdivisions, S140c, S140nc, S220c and S220nc which, as the names suggest, divides the subsets into groups of centrosymmetric and non-centrosymmetric structures. Results of regressions of type 1 are shown below in Table 6.15 The inclusion of a factor of $2^{1/2}$ causes small but consistent improvement in the overall assessment criteria. Overall regressions type 1.3 and 1.4 appear better than the others. Of these two, type 1.4 again appears slightly better than type 1.3 on the basis of the assessment criteria.

Table 6.15(a). Results for the 15170 centrosymmetric structures of subset S220c

RN	a	b or k	c	R_σ (%)	rms (σ) Å	N_{50} (%)	N_{25} (%)	N_{10} (%)	n_{50} (%)	n_{25} (%)	n_{10} (%)	AS_0 (%)	AS_2 (%)
1.1		0.1725(7)		44.0	0.0044	76.1	42.0	17.6	60.9	30.5	12.2	39.9	67.9
1.2		0.00528(2)		34.2	0.0037	84.9	58.9	25.7	73.8	39.6	16.6	59.1	82.4
1.3		0.1109(3)		29.0	0.0031	90.2	66.0	28.1	78.6	48.0	20.4	61.2	87.5
1.4		0.01814(5)		28.4	0.0032	89.0	67.9	34.0	84.8	52.4	21.8	62.3	88.7

Table 6.15(b) Results for the 5164 non-centrosymmetric structures of subset S220nc

RN	a	b or k	c	R_σ (%)	rms (σ) Å	N_{50} (%)	N_{25} (%)	N_{10} (%)	n_{50} (%)	n_{25} (%)	n_{10} (%)	AS_0 (%)	AS_2 (%)
1.1		0.184(1)		38.1	0.0041	81.6	46.8	18.4	66.6	35.3	14.0	41.2	71.7
1.2		0.00828(6)		40.7	0.0045	77.0	47.3	18.1	65.5	29.7	11.5	41.8	77.4
1.3		0.1449(7)		28.3	0.0031	90.1	62.9	25.9	80.0	48.7	20.5	57.4	85.8
1.4		0.0254(1)		29.0	0.0033	88.2	64.0	29.9	83.4	50.9	20.4	60.1	86.2

Another good indication of the quality of regression 1.4 is the ratio of the derived coefficients from the two regressions. The Cruickshank equation would suggest that the proportionality constant between the centric and non-centric regressions should be $2^{\frac{1}{2}}$, i.e.

$$k_{nc} = 2^{\frac{1}{2}} k_c \qquad\qquad 6.19$$

In the regressions of type 1.4 (Table 6.14) this ratio, k_{nc}/k_c, is 1.399. This is very close to the expected value of 1.414, which gives confidence in the results presented. For the other regressions of type 1.1, 1.2 and 1.3 the ratio's are 1.071, 1.598 and 1.307. All of these are considerably different from the expected value which gives even greater confidence in the Cruickshank like regression that employs $N_c^{1/2}$.

In light of the above observations the recommended predictive functions are as follows:

For centrosymmetric structures:

$$\overline{\sigma}(C-C)_p = 0.01814 RN_c^{1/2} \qquad 6.20(a)$$

whereas, for non-centrosymmetric structures:

$$\overline{\sigma}(C-C) = 0.02537 RN_c^{1/2} \qquad 6.20(b)$$

One useful visual method for assessing these regressions are histograms showing their predictive power. In Figure 6.9 the composite numerical error distribution, $\left[\overline{\sigma}(C-C)_o - \overline{\sigma}(C-C)_p\right]$, is shown. Figure 6.10 shows the composite percentage error distribution which is given by

$$\frac{\overline{\sigma}(C-C)_o - \overline{\sigma}(C-C)_p}{\overline{\sigma}(C-C)_o} \qquad 6.21$$

Figures 6.9, 6.10 are drawn for data subset S220 and are annotated by a value of $\left\langle\overline{\sigma}(C-C)_o\right\rangle$ for the structures that fall into each bar of the histogram. Figure 6.9 shows that on average the higher values of $\overline{\sigma}(C-C)_o$ are underestimated, as would be expected. The distribution is slightly skewed towards positive values, where the predicted values are less than the observed values. Generally the lower values are predicted most closely on an absolute scale. Perhaps more revealing is Figure 6.10. This shows that the predictive power of these equations tends to be worse for the low values of $\overline{\sigma}(C-C)_o$ in percentage terms. Importantly the predictions tend to be overestimated rather than underestimated. Work by Kennard and Taylor (1986) has shown that observed coordinate e.s.d's tend to be underestimated by a factor of 1.4-1.5. A similar observation was made during an IUCr study where 17 independent data sets were collected and compared for tartaric acid (Abrahams, Hamilton and Mathieson, 1970; Hamilton and Abrahams, 1970; Mackenzie, 1974). The IUCr study suggested that e.s.d's were underestimated by as much as a factor of 2. Underestimation of the e.s.d's

Figure 6.9. The composite error distribution for regression 6.20, based on subset S220.

A value of the mean of $\overline{\sigma}(C - C)_o$ for all entries within a histogram bar is annotated above each histogram bar.
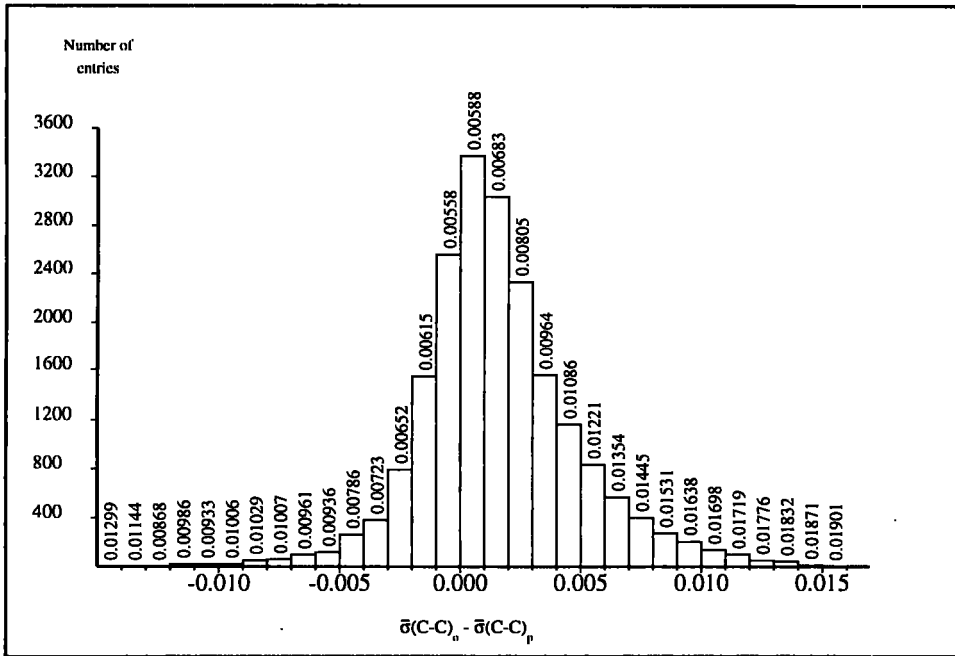
Figure 6.10. The percentage error distribution for regression 6.20, based on subset S220.
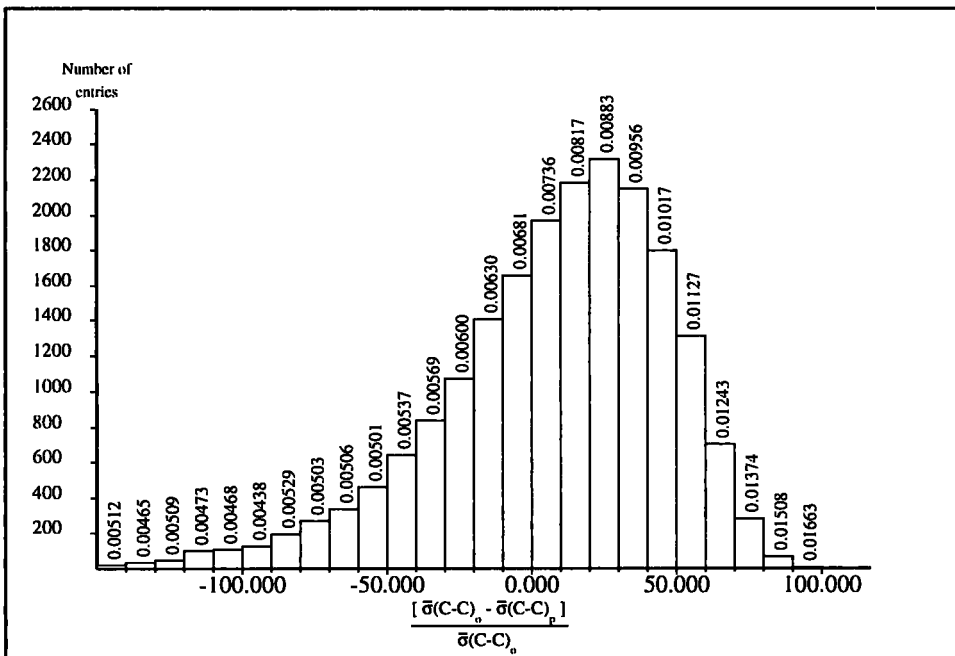
will only affect some of the entries in the study presented here. These entries will tend to be the smaller observed e.s.d's, so an overestimate of the smallest predicted e.s.d's is, perhaps, beneficial. Also smaller predicted values do not have to be too far away from their observed values for percentage deviations to be much larger than zero. This effect reduces as the general level of e.s.d's increases. Very few values are underestimated by more than 75%. This effect tends to occur for the highest values of $\overline{\sigma}(C-C)_o$.

The most promising feature of the equations is their ability to predict the AS flag within 0.002Å in nearly 90% of cases. This may well be useful for the ~16% of entries currently on the CSD that do not contain any indication of geometrical precision. It is also hoped that the results can be utilised in tandem with the AS flags to provide more detailed precision criteria.

6.4.3. Prediction of mean isotropic e.s.d's for carbon atoms ($\overline{\sigma}(C)_p$)

The regression analysis can also be used to predict values of $\overline{\sigma}(C)$, the mean isotropic e.s.d of a carbon atom for a particular structure, since an approximate relationship exists between the bond e.s.d. and the isotropic atom e.s.d. If correlations are ignored then:

$$\sigma(A-B) = \sqrt{\sigma^2(A) + \sigma^2(B)} \qquad 6.22$$

If A and B are the same type of element, and their atomic e.s.d's are both isotropic, then:

$$\sigma(A-A) \approx \sqrt{2\sigma^2(A)} = \sigma(A)\sqrt{2} \qquad 6.23$$

Thus, we would expect the mean atomic e.s.d and the mean bond e.s.d. to differ approximately by a factor of $\sqrt{2}$, so equations 6.20(a) and (b) can be re-written as:

$$\overline{\sigma}(C)_p = \frac{0.01814RN_c^{1/2}}{\sqrt{2}} = 0.01283RN_c^{1/2} \qquad 6.24(a)$$

for centrosymmetric structures

or

$$\bar{\sigma}(C)_p = \frac{0.02537 RN_c^{1/2}}{\sqrt{2}} = 0.01794 RN_c^{1/2} \qquad 6.24(b)$$

for non-centrosymmetric structures

The empirical nature of this study, however, suggests that it may be more appropriate to calculate a value for the denominator of this expression from data set 2. This empirical value. i.e.:

$$\left\langle \frac{\bar{\sigma}(C - C)}{\bar{\sigma}(C)} \right\rangle \qquad 6.25$$

can be calculated for the subsets, S140 and S220, as 1.377, remarkably close to the expected value of 1.414. This indicates that standard deviations in carbon atom positions are approximately isotropic over the very large sample of structures (29,367) used here. This has also been noted by Taylor and Kennard (1986) for a much smaller sample of 200 structures. Taking this into account we may re-write equations 6.24(a) and (b) as

$$\bar{\sigma}(C)_p = \frac{0.01814 RN_c^{1/2}}{1.377} = 0.01317 RN_c^{1/2} \qquad 6.26(a)$$

For centrosymmetric structures

$$\bar{\sigma}(C)_p = \frac{0.02537 RN_c^{1/2}}{1.377} = 0.01842 RN_c^{1/2} \qquad 6.26(b)$$

For non-centrosymmetric structures

The validity of the above two equations, 6.26(a)and (b), were tested by a regression of the form $\bar{\sigma}(C) = kRN_c^{1/2}$ for the sub-divisions of data set S220; S220c and S220nc. The results are shown below in Table 6.16. The composite error distribution, (Figure 6.11), is similar to that for C-C e.s.d's (Figure 6.9). Figure 6.11 does appear to be more normal than Figure 6.9 indicating that these equations tend to under-estimate and

Figure 6.11. Composite error distribution for the $\overline{\sigma}(C)$ regressions.
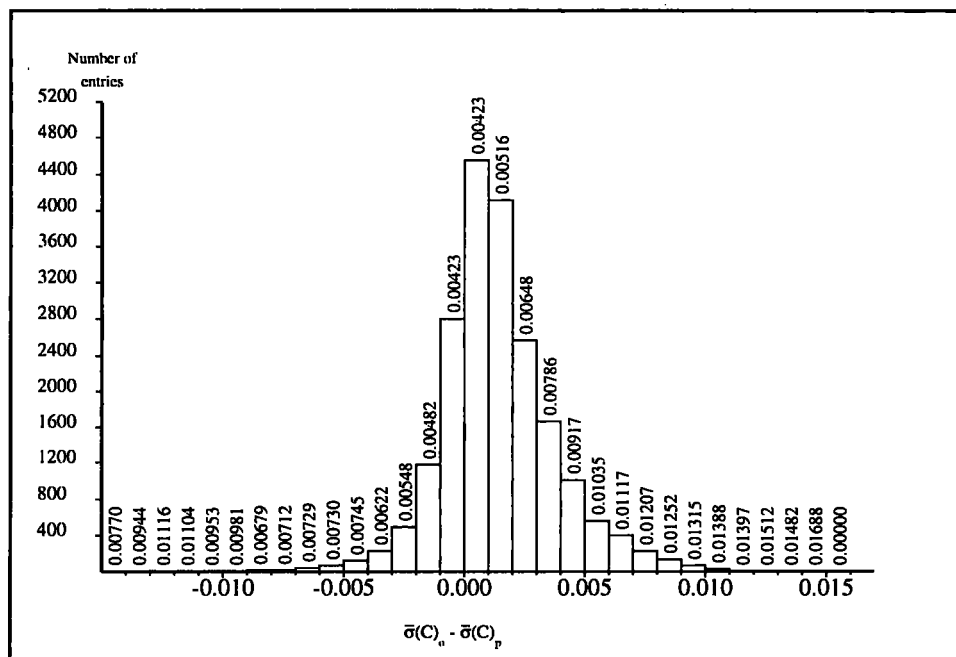
Table 6.16 Regression results for $\overline{\sigma}(C)$ based regressions.

Regression	k	R_σ (%)	rms (σ) Å	N_{50} (%)	N_{25} (%)	N_{10} (%)
Centrosymmetric	0.0132(4)	27.8	0.0023	95.4	79.9	45.8
Non-centrosymmetric	0.0187(1)	28.7	0.0026	94.6	76.5	40.0

over-estimate a similar number of e.s.d's. The ratio of the regression co-efficients from the centrosymmetric and non-centrosymmetric data sets is 1.417 which is exceptionally close to the expected value of 1.414 (Cruickshank, 1960). This is a very promising result, as it indicates that a real difference exists in the e.s.d's between centro-symmetric and non-centrosymmetric structres.

6.4.4. Estimated standard deviations of non-carbon atoms using data set 2.

As shown above in equation 6.14, we can write

$$\frac{\overline{\sigma}(A)}{\overline{\sigma}(B)} = \frac{Z_B}{Z_A}$$

and, for A as carbon and B as any element E, this can be rearranged to give

162

$$\sigma(E) = \frac{6\sigma(C)}{Z_E} \qquad 6.27$$

Thus, for predicting the mean value of $\sigma(E)$ we can write

$$\overline{\sigma}(E) = \frac{6\overline{\sigma}(C)}{Z_E} \qquad 6.28$$

This relationship can be tested using data set 2, since the mean coordinate e.s.d. is stored in this workfile for the heaviest element type as well as for carbon. To test this, a binning analysis of subsets S220 and S140 was performed. The mean "observed" heavy element e.s.d's, $\overline{\sigma}(E)_o$, were binned for fixed ranges of Z_E and the mean value of the atomic number within each bin, $\langle Z_E \rangle$, was calculated. Values of $\langle \overline{\sigma}(C)_o \rangle$ and $\langle \overline{\sigma}(E)_o \rangle$ were also calculated for each bin. We can obtain a value of $\langle \overline{\sigma}(E)_p \rangle$, the predicted mean for the heavy atom e.s.d's within a particular bin. Across a narrow range of Z_E it is possible to write

$$\langle \overline{\sigma}(E) \rangle_p = \frac{6 \langle \overline{\sigma}(C) \rangle_o}{\langle Z_E \rangle} \qquad 6.29$$

If this equation holds then a graph of $\langle \overline{\sigma}(E)_p \rangle$ versus $\langle \overline{\sigma}(E)_o \rangle$ should be linear, pass through the origin, and have a gradient of unity. Values for the binning analysis are shown in Table 6.17 and the graph of $\langle \overline{\sigma}(E)_p \rangle$ versus $\langle \overline{\sigma}(E)_o \rangle$ is shown in Figure 6.12 for both subsets S140 and S220.

It is noticable that the value of $K\left[= \langle \overline{\sigma}(E) \rangle_o \big/ \langle \overline{\sigma}(E) \rangle_p \right]$ falls across the Z_E range, although it is fairly constant from Z=20 onwards. One possible reason for this has been suggested (Cruickshank, 1993). If the information for each individual structure had been available concerning the mean reciprocal radius of the observed reflections, $\overline{s}$, then the correct estimating equation would have utilised the ratio of the scattering

Table 6.17. Mean values of $< \overline{\sigma}(E)_o >$ and $< \overline{\sigma}(C)>_o$ across ranges of Z_E

Values of $< \overline{\sigma}(E)>_o$ and $< \overline{\sigma}(C)>_o$, the observed isotropic esd's of a non-carbon atom (E) and a carbon atom (C) in Å, averaged over ranges of Z_E, the atomic number of E. N_{ent} is the number of entries in each range and $<Z_E>$ is the mean value for the range. The quantity $\overline{\sigma}(E)_p$ is calculated using Eq.16 with the appropriate $<Z_E>$ as denominator. K is the ratio $< \overline{\sigma}(E)>_o/ \overline{\sigma}(E)_p$. Calculations were carried out for subsets S140 and S220 of Table 6.

Z_E range	N_{ent}	$< \overline{\sigma}(E)_o >$	$< \overline{\sigma}(C)>_o$	$<Z_E>$	$< \overline{\sigma}(E)_p >$	K
(a) Subset S140						
7-9	7692	0.00332	0.00464	7.88	0.00353	0.941
10-19	4586	0.00190	0.00542	16.00	0.00203	0.936
20-29	3795	0.00117	0.00785	26.34	0.00179	0.654
30-39	1622	0.00126	0.00948	34.04	0.00167	0.755
40-49	2958	0.00088	0.00897	43.67	0.00123	0.715
50-69	1350	0.00088	0.01050	53.49	0.00118	0.746
70-	2979	0.00077	0.01338	77.04	0.00104	0.740
(b) Subset S220						
7-9	6051	0.00306	0.00427	7.89	0.00325	0.942
10-19	3786	0.00143	0.00472	16.02	0.00177	0.808
20-29	3088	0.00098	0.00623	26.33	0.00142	0.690
30-39	1243	0.00099	0.00727	34.01	0.00128	0.773
40-49	2425	0.00076	0.00710	43.63	0.00098	0.776
50-69	1019	0.00072	0.00787	53.20	0.00089	0.809
70-	1856	0.00063	0.00946	76.79	0.00074	0.851

Figure 6.12: $\left\langle \overline{\sigma}(E)_p \right\rangle$ versus $\left\langle \overline{\sigma}(E)_o \right\rangle$
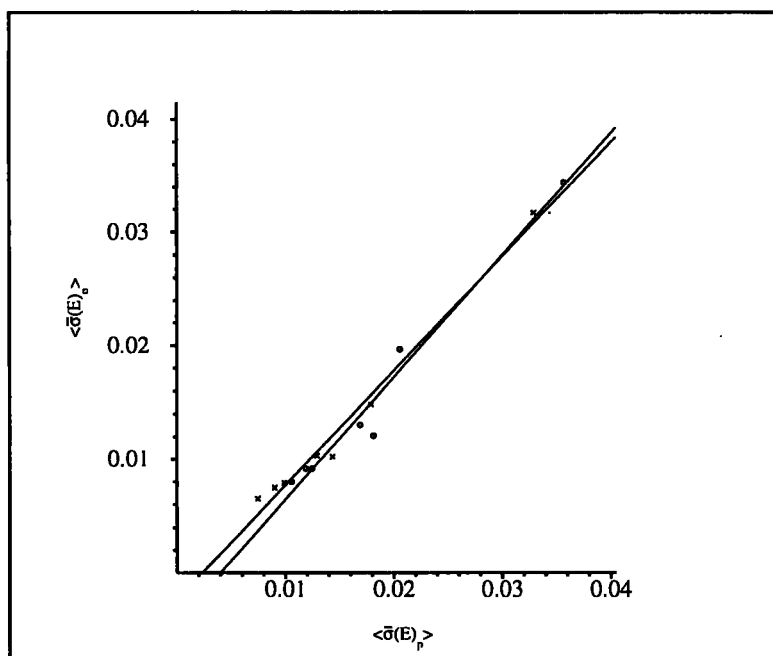
factors, $\overline{f_c}/\overline{f_e}$, at $\bar{s}$, rather than the more crude ratio of atomic numbers (that must be used here) in deriving values of $\langle\overline{\sigma(E)}_p\rangle$. Since scattering factors for the heavier elements decrease more slowly with θ than for carbon (and heavy atoms have smaller displacement parameters than C atoms in the same structure), the ratio $\overline{f_c}/\overline{f_e}$ will be smaller than Z_C/Z_E. Because of this difference it can be concluded that if the estimates for $\langle\overline{\sigma(E)}_p\rangle$ had been based on $\overline{f_c}/\overline{f_e}$ rather than Z_C/Z_E then higher values of K would have been observed for the higher Z_E ranges.

The lines shown in Figure 6.12 are almost co-linear. They also pass close to the origin. In the light of these results further regressions of type 1 were carried out using subset S220 with $\overline{\sigma(E)}_a$ as the dependent variable. Two regressions were performed, with $\overline{\sigma(C)}/Z_E$ and $RN_c^{1/2}/Z_E$ as the independent variables in regressions (i) and (ii) respectively. Naturally for (i) no discrimination between centrosymmetric and noncentrosymmetric structures is required. Because of the properties already shown to exist for this data set in Table 6.17 we would expect the constant in this simple regression to be somewhat less than the 'ideal' value of 6.0. In regression (ii) it is necessary to partition the data set into the two subdivisions, S220c and S220nc prior to regression analysis. The results for the two regressions, with an imposed intercept, were

(i)
$$\overline{\sigma(E)}_p = \frac{5.203\overline{\sigma}(C)}{Z_E}$$
6.30

(ii)
$$\overline{\sigma(E)}_p = \frac{0.0678RN_c^{\frac{1}{2}}}{Z_E}$$
6.31(a)

For centrosymmetric structures

$$\overline{\sigma(E)}_p = \frac{0.1006RN_c^{\frac{1}{2}}}{Z_E}$$
6.31(b)

For Non-centrosymmetric structures

The assessment criteria for these regressions are tabulated below in Table 6.18

Table 6.18. Regression assessment criteria for regressions (i) and (ii)

Regression	R_σ	r.m.s σ
(i)	28.7%	0.00088Å
(ii)Cent.	40.8%	0.00100Å
(ii)N.Cent.	34.9%	0.00140Å

The regression results are a confirmation of the validity of equation 6.28 above. The R_σ-values cited are, on average, a little higher than those observed for the earlier regressions. In principle, equation 6.28 could be used to estimate non-carbon e.s.d's from carbon e.s.d's. This is potentially very useful for database analyses where the AS flag could be used to estimate $\overline{\sigma}(C)$ and so some estimate of non-carbon precision could be derived. This short study of precision for the heaviest element in the asymmetric unit led to a more general study of all non-hydrogen atoms in each structure using data set 2A. These results are presented in Section 6.6 below.

6.5 Analysis of Data set 3

6.5.1 Descriptive statistics for Data set 3

A frequent "rule of thumb" used by crystallographers to assess the viability of a particular refinement is the ratio of the number of reflections used in refinement (N_r) and the number of parameters refined (N_p). Generally the stability of a refinement can be gauged from this ratio, since it indicates the number of reflections that contribute to the determination of each parameter. In Chapter 2, the expression for the least-squares parameter e.s.d. was given. To re-iterate, the e.s.d of the jth parameter p_j, can be expressed as:

$$\sigma_{p_j} = \left(\left(A^{-1} \right)_{jj} \frac{\sum_{i=1}^{N_r} w_i \left(F_{o,i}^2 - F_{c,i}^2 \right)^2}{N_r - N_p} \right)^{\frac{1}{2}}$$

where A^{-1} is the variance-covariance matrix that is calculated during least squares, w_i is an appropriately derived weight, $F_{o,i}^2$ is the square of the observed structure factor for the ith reflection and $F_{c,i}^2$ is the calculated value.

The expression shows that a parameter e.s.d. can be regarded as a scale factor multiplied by the appropriate diagonal term of the variance-covariance matrix. The scale factor's denominator shows that a parameter e.s.d. has an inverse dependency on $\left(N_r - N_p\right)^{1/2}$. The original Cruickshank formulation (Cruickshank, 1960) also incorporates this term in its denominator, thus the inclusion of such a term in a regression study should improve the predictive power of the function derived.

Because of this data set 3 was created by hand reference to the original literature. 817 entries had values of N_r and N_p appended, as described in Section 6.2.1

It is useful to begin by analysing the relationships between N_r, N_p and the precision indicators that are stored in the CSD. This is not just of benefit in this analysis but also of general interest to crystallographers. In Table 6.19, data set 3 has been divided into 8 unitary bins based on N_r/N_p, together with two other bins covering structures where this ratio is < 4 and ≥ 12 respectively. In the table, values of $\left\langle N_r/N_p \right\rangle$, $\left\langle \overline{\sigma}(C-C) \right\rangle$, $\left\langle R \right\rangle$, $\left\langle Z_{max} \right\rangle$, and $\left\langle RN_c^{1/2} \right\rangle$ are shown for each bin.

Table 6.19(a) shows that $\left\langle \overline{\sigma}(C-C) \right\rangle$ decreases rapidly over the first five bins as N_r/N_p increases, even though $\left\langle Z_{max} \right\rangle$ (and, of course, $\left\langle RN_c^{1/2} \right\rangle$) increase with $\left\langle N_r/N_p \right\rangle$, the value of $\left\langle R \right\rangle$ remains constant across the range. This is observed because structures that contain heavier elements tend to scatter more strongly at higher angles and so the number of observed reflections tends to increase with Z_{max}. This is reflected in the evaluated correlation coefficient between $\left\langle N_r/N_p \right\rangle$ and Z_{max} of 0.346 for subset S140 and 0.374 for S220. Despite this we still observe a negative correlation of $\left\langle N_r/N_p \right\rangle$

Table 6.19. Analysis of structural precision for ranges of N_r/N_p.

For each range, mean values are cited for $\langle N_r/N_p \rangle$, $\langle \bar{\sigma}(C\text{-}C) \rangle$, $\langle R \rangle$, $\langle Z_{max} \rangle$ and $\langle R\,N_c^{\frac{1}{2}} \rangle$. N_{ent} is the number of CSD entries in each range and the normalized quantity $\langle \bar{\sigma}(C\text{-}C) \rangle_n$ is described in the text.

N_r/N_p range	$\langle N_r/N_p \rangle$	$\langle \bar{\sigma}(C\text{-}C) \rangle$	$\langle R \rangle$	$\langle Z_{max} \rangle$	$\langle RN^{\frac{1}{2}}_c \rangle$	N_{ent}	$\langle \bar{\sigma}(C\text{-}C) \rangle_n$
(a) For all entries in data set 3							
0-4	3.16	0.01076	0.0488	10.29	0.319	31	0.01076
4-5	4.58	0.00943	0.0497	14.15	0.331	77	0.00910
5-6	5.52	0.00893	0.0487	16.13	0.345	105	0.00827
6-7	6.50	0.00854	0.0480	20.30	0.373	103	0.00730
7-8	7.53	0.00836	0.0488	22.50	0.399	103	0.00669
8-9	8.56	0.00866	0.0477	28.48	0.438	98	0.00631
9-10	9.47	0.00811	0.0464	28.72	0.414	76	0.00625
10-11	10.52	0.00784	0.0430	32.71	0.397	63	0.00630
11-12	11.34	0.00925	0.0440	36.30	0.459	53	0.00644
12-	15.28	0.00989	0.0438	45.68	0.541	108	0.00584
(b) For entries with $Z_{max} \leq 18$							
0-4	3.15	0.01066	0.0487	8.89	0.294	29	0.01155
4-5	4.56	0.00829	0.0499	10.12	0.291	64	0.00910
5-6	5.49	0.00691	0.0489	9.94	0.287	79	0.00768
6-7	6.48	0.00602	0.0492	11.01	0.300	69	0.00641
7-8	7.56	0.00516	0.0510	10.68	0.295	66	0.00557
8-9	8.59	0.00479	0.0488	11.94	0.291	50	0.00526
9-10	9.42	0.00440	0.0481	12.57	0.284	35	0.00495
10-11	10.48	0.00407	0.0463	13.15	0.267	26	0.00486
11-12	11.37	0.00435	0.0484	11.17	0.277	17	0.00501
12-	14.22	0.00501	0.0509	12.64	0.300	25	0.00533
(c) For entries with $Z_{max} > 18$							
0-4	3.32	0.01216	0.0500	30.50	0.675	2	0.00575
4-5	4.68	0.01504	0.0486	34.00	0.528	13	0.00910
5-6	5.61	0.01507	0.0481	34.92	0.520	26	0.00925
6-7	6.54	0.01364	0.0457	39.14	0.522	34	0.00834
7-8	7.49	0.01408	0.0448	43.59	0.583	37	0.00771
8-9	8.53	0.01268	0.0465	45.72	0.592	48	0.00684
9-10	9.51	0.01129	0.0450	42.51	0.525	41	0.00686
10-11	10.55	0.01048	0.0406	46.45	0.488	37	0.00686
11-12	11.33	0.01157	0.0419	48.16	0.545	36	0.00678
12-	15.60	0.01135	0.0417	55.63	0.613	83	0.00591

168

with $\langle \overline{\sigma}(C-C) \rangle$ of -0.146. Earlier studies have already proved that Z_{max} is positively correlated with $\langle \overline{\sigma}(C-C) \rangle$ which shows that N_r/N_p has an independent influence on the value of $\overline{\sigma}(C-C)$.

In order to assess the true effect of the parameterization within this system the values of $\langle \overline{\sigma}(C-C) \rangle$ in each bin have been normalised to the value of $\langle RN_c^{1/2} \rangle$ from the first bin by using the proportionality between $\overline{\sigma}(C-C)$ and $RN_c^{1/2}$, so

$$\langle \overline{\sigma}(C-C) \rangle_n = \frac{\langle \overline{\sigma}(C-C) \rangle \langle RN_c^{\frac{1}{2}} \rangle_n}{\langle RN_c^{\frac{1}{2}} \rangle_1} \qquad 6.32$$

where $\langle \overline{\sigma}(C-C) \rangle_n$ is the normalised value associated with a specific $\langle RN_c^{1/2} \rangle$. The denominator here has a value of 0.31955. The normalised binning results are shown in Table 6.19(a). The relationship between N_r/N_p and $\overline{\sigma}(C-C)$ now becomes more clear for larger values of N_r/N_p. In Figure 6.13 a plot of the value of $\langle N_r/N_p \rangle$ against $\langle \overline{\sigma}(C-C) \rangle_n$ is shown. The effect on precision is clearly seen to be non-linear in this function; The shape is approximately hyperbolic, suggesting an approximate relationship shown below:

$$\frac{1}{\left(N_r/N_p\right)^a} \propto \overline{\sigma}(C-C) \qquad 6.33$$

This is expected, considering the functional form of the e. s. d., which has an inverse relation to $\left(N_r-N_p\right)^{1/2}$. This function varies in approximately the same way as $\left(N_r/N_p\right)$ for the range of N_r and N_p values that are observed within data set 3. This is easily shown by evaluating the proportionality factor between the two functions as follows. Firstly we can write

$$b.\left(N_r-N_p\right)^{\frac{1}{2}} = \left(\frac{N_r}{N_p}\right)^{\frac{1}{2}} \qquad 6.34$$

which gives a value of b as:

$$b = \left(\frac{N_r/N_p}{N_r - N_p} \right)^{\frac{1}{2}} = \frac{1}{\sqrt{N_p}} \left(\frac{N_r}{N_r - N_p} \right)^{\frac{1}{2}} \qquad 6.35$$

If $N_r \gg N_p$ then we can write

$$b \approx \frac{1}{\sqrt{N_p}} \qquad 6.36$$

This is a reasonable approximation for most structures in data set 3. The ratio N_r/N_p is rarely less than 4 for the structures in this data set. Only 31 entries lie in this range out of 817.

Figure 6.13: $\langle N_r/N_p \rangle$ against (a) $\langle \overline{\sigma}(C-C) \rangle$ and (b) $\langle \overline{\sigma}(C-C) \rangle_n$.
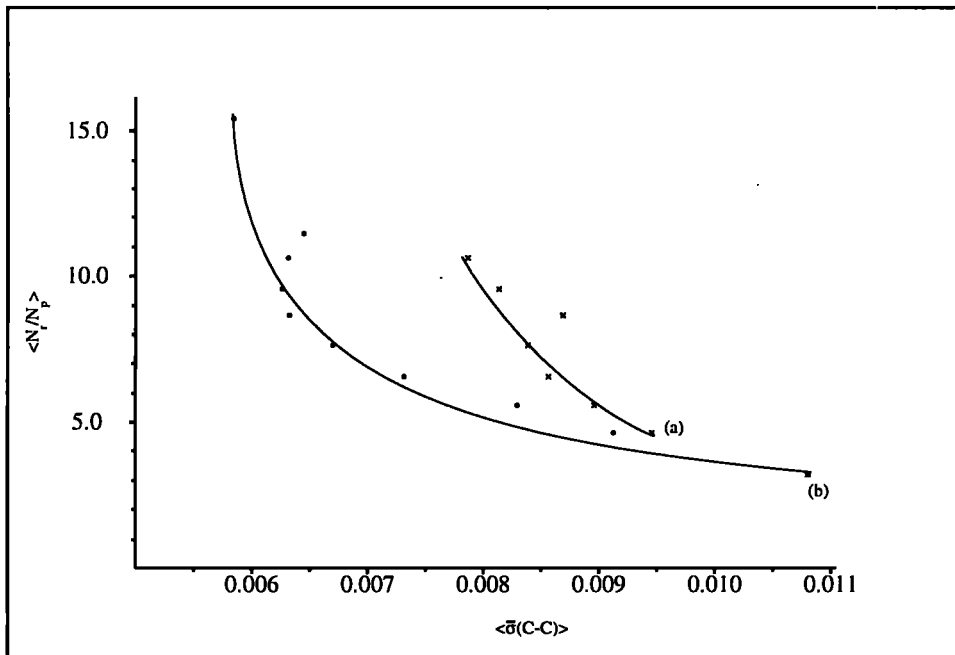
Structures on the data base are very varied in size, from small molecules with only 4 or 5 non-hydrogen atoms in the asymmetric unit to large structures with as many as 150 non-hydrogen atoms, but the majority of compounds have between 25 and 75 non-hydrogen atoms within the asymmetric unit. The number of refined parameters in most structures is approximately 8 times the number of atoms. This means that in most compounds N_p will be between 250 and 750, which means that b will be between 0.063 and 0.037 for most structures. This is a relatively small range and so b can be regarded as approximately constant. Because of this $\left(N_r/N_p\right)^{1/2}$ can be regarded as being approximately proportional to $\left(N_r - N_p\right)^{1/2}$ for the majority of compounds in data set 3.

In Tables 6.19(b) and 6.19(c) data set 3 has been split into organic ($Z_{max} \leq 18$) and metallo-organic structures ($Z_{max} > 18$). The values of $\langle \overline{\sigma}(C-C)\rangle$ for organic structures show similar trends to the normalised values in the full table, but those for the 'metallo-organic structures' are misleading, in that they appear roughly constant across the range of N_r/N_p. The true behaviour of $\langle \overline{\sigma}(C-C)\rangle$ is only seen in the normalised values. The dependency of N_r/N_p on Z_{max} is again reinforced by the distribution of the data. Only 26.7% of the entries shown in Table 6.19 which have $N_r/N_p \leq 7$ are metallo-organic. 62.8% of entries with $N_r/N_p > 7$ in Table 6.19 are metallo-organic even though less than half (43.7%) of the whole data set are in this class. The effect of the scattering power of heavier elements is clear from this data: the heavier elements cause more reflections to be observed.

6.5.2 Regression experiments for Data set 3

Simple linear regressions were used to test the merit of including some function of the N_r and N_p within a regression experiment. $\left(N_r/N_p\right)^{1/2}$, $\left(N_r - N_p\right)^{1/2}$ and $N_r^{1/2}$, were sequentially incorporated as in a series of regressions which took the form shown below in Table 6.20

Table 6.20 Types of regression used.

Type 1.4:	$\overline{\sigma}(C-C) = kRN_c^{1/2}$	6.37(a)
Type 4.4:	$\overline{\sigma}(C-C) = \dfrac{kRN_c^{1/2}}{\left(N_r/N_p\right)^{1/2}}$	6.37(b)
Type 5.4:	$\overline{\sigma}(C-C) = \dfrac{kRN_c^{1/2}}{\left(N_r\right)^{1/2}}$	6.37(c)
Type 6.4:	$\overline{\sigma}(C-C) = \dfrac{kRN_c^{1/2}}{\left(N_r - N_p\right)^{1/2}}$	6.37(d)
Type 7.4:	$\overline{\sigma}(C-C) = \dfrac{kRN_c^{1/2}}{\overline{s}}$	6.37(e)
Type 8.4:	$\overline{\sigma}(C-C) = \dfrac{kRN_c^{1/2}}{\overline{s}\left(N_r\right)^{1/2}}$	6.37(f)
Type 9.4:	$\overline{\sigma}(C-C) = \dfrac{kRN_c^{1/2}}{\overline{s}\left(N_r - N_p\right)^{1/2}}$	6.37(g)

In types 7.4 to 9.4 an added term, $\overline{s}$, is also included. This is an estimate of the r.m.s reciprocal radius that can be derived from N_r together with other parameters that are available on the database . This derivation (Cruickshank, 1993) is shown below:

Let us suppose that there are no unobserved reflections within the limiting sphere, and we ignore the effects of principal zones, where the reflection multiplicity may be less than that of a general reflection (m^*). If this is the case then we can estimate s_{max} $\left[= 2\sin(\theta_{max})\right]$. We can write an expression for the number of observed reflections, N_r as:

$$N_r = \frac{V_s^*}{m^* V_c^*} = \frac{4\pi s_{max}^3}{3m^* V_c^*} = \frac{4\pi s_{max}^3 V_c}{3m^*} \qquad 6.38$$

In data set 3 a value of N_r is available, thus rearrangement of equation 6.38 can give an estimate of s_{max}:

$$s_{max} = \left(\frac{3m^* N_r}{4\pi V_c}\right)^{\frac{1}{3}} \qquad 6.39$$

The root mean square radius $\overline{s}^2$ for a solid sphere can be calculated by

$$\overline{s}^2 = \frac{\int_0^{s_{max}} s^2 4\pi s^2 ds}{\int_0^{s_{max}} 4\pi s^2 ds} \qquad\qquad 6.40$$

Which can be evaluated as

$$\overline{s}^2 = \frac{\frac{1}{5}\left(4\pi s_{max}^5\right)}{\frac{1}{3}\left(4\pi s_{max}^3\right)} = \frac{3}{5}s_{max}^2 \qquad\qquad 6.41$$

and so the r.m.s reciprocal radius, $\overline{s}$ is given by

$$\overline{s} = \left\{\tfrac{3}{5}\right\}^{\frac{1}{2}} s_{max} \qquad\qquad 6.42$$

By substituting for S_{max}, this can be written as

$$\overline{s} = \left\{\frac{3}{5}\right\}^{\frac{1}{2}} \left\{\frac{3m^* N_r}{4\pi V_c}\right\}^{\frac{1}{3}} \qquad\qquad 6.43$$

The first regression in this series serves as a "benchmark" for this study, i.e. data set 3 has been regressed with the same criteria as were used in the best regressions using data set 2. This indicates whether incorporating functions of Nr and Np has a truly beneficial influence on the quality of the results.

The benchmark regressions results, shown below in Table 6.21 are similar to those of the same form on data set 2, although the assessment criteria in the data set 3 regressions are on average slightly lower than those for data set 2. This is expected due to the lower amount of data utilised in data set 3. The values of k are fairly close to one

Table 6.21(a) Results for 371 centrosymmetric structures from subset S220 [Table 6(a)]) for which N_r and N_p values are available in data set 3.

RN	a	b or k	c	R_σ (%)	rms (σ) Å	N_{50} (%)	N_{25} (%)	N_{10} (%)	n_{50} (%)	n_{25} (%)	n_{10} (%)	AS_0 (%)	AS_2 (%)
1.4		0.0194(4)		27.4	0.0028	92.2	74.4	38.5	84.9	55.0	20.0	69.3	90.6
4.4		0.0539(9)		25.0	0.0025	94.1	75.2	41.0	89.0	59.6	24.0	70.3	93.0
5.4		0.923(16)		27.2	0.0027	92.5	73.9	35.6	85.4	54.7	20.2	69.8	92.7
6.4		0.854(15)		26.9	0.0027	92.5	75.2	36.1	86.0	55.5	18.9	70.1	92.7
7.4		0.0138(3)		26.0	0.0028	92.5	74.1	46.1	90.8	58.0	27.5	70.1	92.7
8.4		0.616(13)		27.5	0.0032	90.6	75.2	43.9	89.8	58.2	26.2	70.2	92.5
9.4		0.567(13)		27.6	0.0032	90.0	75.5	43.7	89.0	57.7	28.3	69.8	92.7

Table 6.21(b). Results for 184 non-centrosymmetric structures from subset S220 [Table 6.9(a)] for which N_r and N_p values are available in data set 3.

RN	a	b or k	c	R_σ (%)	rms (σ) Å	N_{50} (%)	N_{25} (%)	N_{10} (%)	n_{50} (%)	n_{25} (%)	n_{10} (%)	AS_0 (%)	AS_2 (%)
1.4		0.0253(7)		28.3	0.0030	92.4	64.7	31.0	83.7	48.9	20.1	67.5	91.9
4.4		0.059(2)		27.8	0.0030	92.4	70.1	35.3	87.0	55.4	25.0	69.7	92.4
5.4		1.08(2)		20.8	0.0021	96.2	79.9	43.5	89.1	66.9	26.6	72.1	96.2
6.4		0.82(3)		24.9	0.0036	94.6	76.1	46.2	94.0	64.7	34.2	71.7	94.0
7.4		0.0220(6)		26.0	0.0028	94.0	71.2	37.0	88.0	53.8	23.9	69.9	92.9
8.4		0.90(2)		21.2	0.0024	95.6	80.4	47.3	92.4	65.8	32.6	70.8	94.5
9.4		0.69(2)		26.6	0.0036	91.9	73.9	45.1	94.6	59.2	27.2	70.5	94.0

another for each respective regression. The ratio, k_{nc}/k_c, appears a little low, at 1.30, although the e.s.d. on the predicted regression constants are appreciably higher than in data set 2, due to the lower number of data points. Simple propagation of errors suggests that the approximate e.s.d. in the above ratio is ~0.06, so the ratio is less than 2σ away from its predicted value of 1.414.

Regressions 1.4 through to 9.4 give somewhat inconsistent results for the centrosymmetric and non-centrosymmetric subsets. In general they suggest that inclusion of a term involving N_r appears to improve the results marginally. It would

appear that inclusion of $\bar{s}$ has no real effect on the quality of the results obtained. The best result for centrosymmetric data sets appears to be the regression of type 4.4 which gives an R_σ of 25.0%. For the non-centric data the result is different. Type 5.4 gives an R_σ of 20.8% which is a little puzzling. The ratio, k_{nc}/k_c, varies widely across the regressions but the precision is again low. The low number of data in these regressions is probably partially to blame for this inconsistency; it is not possible to observe the small effect of incorporating some function of N_r and N_p with the small number of data points available. This anomaly can be investigated further by examining the distributions and inter-relationships between the independent variables used in the Cruickshank formulation. The $\bar{s}$ distribution is almost normal. Values within this distribution range from 0.48-1.37Å^{-2}. Of these values, 88%, however, lie within a band between 0.6 and 1.0Å^{-2}; 58% lie within 0.7 and 0.9Å^{-2}. As is apparent from this data, the majority of $\bar{s}$ values are very close to one another. The range of $N_r - N_p$ is much broader, being between 241 and 9896. The corresponding range of $\left(N_r - N_p\right)^{1/2}$ is 15.5-99.5, although the majority of entries lie between 26.0 and 75.0.

Another useful picture is given by correlation analysis. Below, in Table 6.22, correlation results are given. The table shows that the R-factor is essentially independent of $\bar{s}$ and $f\left(N_r, N_p\right)$ since the correlation coefficients, C_{ij} where i = R are all low. The correlations between $N_c^{1/2}$ and $f\left(N_r, N_p\right)$ are appreciable ranging from 0.27 to 0.65. It is interesting here that the highest correlation coefficient is between $\left(N_r - N_p\right)^{1/2}$ and $N_c^{1/2}$. This reflects the increase of observed data that occurs as a result of having more strongly scattering elements within the unit cell. The ratio of data to parameters is less strongly correlated to $N_c^{1/2}$. This is logical because of the definition of $N_c^{1/2}$. The value of $N_c^{1/2}$ increases with cell volume. This is not the case for N_r/N_p because an increase in the cell volume causes an increase in both the denominator and

Table 6.22.: Correlation coefficients for subset S220.

Correlation coefficients C_{ij} derived from the 555 entries of subset S220 for which N_r,N_p values were available in data set 3. Names of data items are defined in the text or in Table 3.

	Item i		
Item j	$\overline{\sigma}(C-C)$	R	$N_c^{\frac{1}{2}}$
N_r/N_p	-0.141	-0.059	0.273
$\left(N_r/N_p\right)^{1/2}$	-0.147	-0.055	0.270
N_r	0.102	-0.031	0.364
$N_r^{1/2}$	0.108	-0.054	0.650
$\left(N_r - N_p\right)^{1/2}$	0.086	-0.060	0.644
$\overline{s}$	-0.334	-0.114	-0.219
$\overline{s}\left(N_r - N_p\right)^{1/2}$	-0.079	-0.098	0.437

the numerator of this function, so it remains approximately constant. $N_r - N_p$ however will be correlated with volume because an increase in volume causes a greater absolute increase in N_r than in N_p. This dependency on cell volume is possibly the cause of some of the correlation.

The expected inverse correlations between $\overline{\sigma}(C-C)$, $\overline{s}$ and $f\left(N_r,N_p\right)$ are not seen. $\overline{s}$ is correlated to $\overline{\sigma}(C-C)$ but not as strongly as might be expected. Also $\overline{s}$ is correlated to $N_c^{1/2}$, by nearly as much as to $\overline{\sigma}(C-C)$. The functions of N_r and N_p show a highest inverse correlation of -0.147. A possible reason for the low correlations is the relatively narrow ranges of the variables within the data set. It would seem that the combination of effects through the data set tend to cancel one another out and so no benefit is seen in the prediction of the e.s.d's.

6.6 Regressions using data set 2A: Prediction of e.s.d's in the general atom case

As was shown above in equation 6.28, the mean e.s.d. of any element type can be approximately derived from the mean e.s.d. of the carbon atoms within any particular structure *via*

$$\overline{\sigma}(E) = \frac{Z_C \overline{\sigma}(C)}{Z_E}$$

We can also express this function by substituting for $\overline{\sigma}(C)$ to form an expression that gives $\overline{\sigma}(E)$ as a function of R and $N_c^{1/2}$:

$$\overline{\sigma}(E) = \frac{Z_C k R N_c^{1/2}}{Z_E} = \frac{K R N_c^{1/2}}{Z_E} \qquad 6.44$$

In this expression, k should be approximately equivalent to the regression constants shown in equations 6.26(a) and (b); the $\overline{\sigma}(C)$-based equations of Section 6.4.3

In data set 2A, atomic e.s.d's were extracted for all element types except hydrogen in each structure and values of $\overline{\sigma}(E)$ were calculated for each type. Equation 6.44 was tested using regressions where $\overline{\sigma}(E)_o$ was the dependant variable and $R N_c^{1/2}/Z_E$ was the independent variable in estimating the quantity K which should be equal to $6k$ where k is the regression constant in equations 6.26(a) and (b). The regression experiments were assessed in a similar way to the regressions for carbon atoms. The same criteria were used, but N_{50}, N_{25} and N_{10} are probably not as suitable for assessing these regressions due to the lower overall numerical values of $\overline{\sigma}(E)$ which are lower for a general element E than for the carbon atoms which were treated earlier. Again in this study, the data set was broken down into two smaller subsets of centrosymmetric compounds and noncentrosymmetric compounds prior to regression analysis. A series of regressions were performed with a variety of data selection criteria on these two

Table 6.23(a). Data selection criteria for regrssions based on equation 6.44

Summary of data subsets used in regression analyses based on Eq.6.44. RN is the regression number; Z indicates the range of element types covered by each analysis; R_{min}, R_{max} define R-factor limits; σ_{min}, σ_{max} are $\overline{\sigma}(C)_e$ limits (see text); C/NC indicates centrosymmetric or non-centrosymmetric structures included; N_{ent} is the number of CSD entries used; N_{tot} is the total number of $\overline{\sigma}(E)_o$ values available; N_- and N_+ are numbers of atoms with $\overline{\sigma}(C)_e$ below and above the σ-limits; N_{at} is the total number of $\overline{\sigma}(E)_o$ values finally used in each regression.

RN	44.1		44.2		44.3		44.4		44.5	
Z	all		all		all		$Z \geq 10$		$3 \leq Z < 10$	
R_{min}	0.01		0.01		0.02		0.01		0.01	
R_{max}	0.10		0.10		0.07		0.10		0.10	
σ_{min}	0.0001		0.0023		0.0022		0.0023		0.0023	
σ_{max}	0.0400		0.0160		0.0140		0.0160		0.0160	
C/NC	C	NC	C	NC	C	NC	C	NC	C	NC
N_{ent}	25685	8066	25685	8066	22757	7087	25685	8066	25685	8066
N_{tot}	97874	27031	97874	27031	87068	23919	37345	7577	60529	19454
N_-	1	0	10465	1419	8979	1127	4260	468	6205	951
N_+	375	179	9470	3307	9159	3251	3540	1117	5930	2190
N_{at}	97480	26852	77939	22305	68930	19541	29545	5992	48394	16313

Table 6.23(b). Results of regression analyses based on Eq.6.44.

The parameters used as assessment criteria are defined in Table 6.12. RN is the regression number and C/NC indicates that only centrosymmetric or non-centrosymmetric structures were included.

RN	44.1		44.2		44.3		44.4		44.5	
C/NC	C	NC	C	NC	C	NC	C	NC	C	NC
K	0.09286	0.12542	0.08143	0.10789	0.08093	0.10868	0.06160	0.08581	0.08354	0.10948
$\sigma(K)$	0.00014	0.00037	0.00011	0.00030	0.00012	0.00030	0.00014	0.00046	0.00015	0.00035
R_σ	36.5	36.1	30.9	31.0	30.0	30.1	31.8	31.2	28.8	30.2
rms(σ)	0.0029	0.0034	0.0019	0.0022	0.0017	0.0019	0.0008	0.0009	0.0024	0.0026
pc(σ)	53.1	48.0	37.5	35.0	37.2	34.8	34.4	34.0	29.4	30.9
N_{50}	93.1	90.1	96.5	95.1	97.8	97.1	99.9	99.9	94.7	93.5
N_{25}	80.5	73.8	85.3	81.1	88.0	84.8	98.4	97.4	78.2	75.5
N_{10}	49.3	41.4	57.8	51.3	61.2	54.5	85.6	81.4	45.3	42.2
n_{50}	63.6	67.7	76.0	78.8	76.2	79.0	77.3	78.9	84.3	83.3
n_{25}	37.0	39.4	45.6	46.6	45.9	47.2	44.5	46.4	52.0	49.7
n_{10}	15.6	16.5	19.1	19.6	19.4	19.9	18.4	19.4	22.1	21.0

subsets, including breaking the data into "light" atom (Zmax $\leq$ 10) and heavy atoms (Zmax > 10) and applying more or less stringent criteria of R-factor and $\overline{\sigma}(C)_e$ criteria. Details of these regressions, along with the data criteria for each one, are given in Tables 6.23(a) and (b).

In the first regression (Table 6.23(b), 44.1) nearly all $\overline{\sigma}(E)_o$ values were used but. 554 entries for which $\overline{\sigma}(C)_e$ exceeded 0.040Å were omitted. The R_σ values for the centrosymmetric and noncentrosymmetric subsets are relatively high at 36.5 and 36.1% respectively compared with their counterparts for $\overline{\sigma}(C)$ and $\overline{\sigma}(C-C)$ regressions shown above in Table 6.13,6.14, and 6.16 respectively. The values of n50, n25 and n10 for these regressions are also lower than in the carbon case. Another important indicator is the distribution of $\overline{\sigma}(E)_o - \overline{\sigma}(E)_p$. In both the centosymmetric and non-centrosymmetric cases the distributions are skewed towards positive values (indicating that the predicted values are smaller than the observed values). The distribution also shows that the larger values are more commonly under-estimated than the smaller values. The absolute errors in values of $\overline{\sigma}(E)_o$ is above +0.005Å in 6.9% of cases. In contrast to this, the absolute percentage errors, given by

$$b = \frac{100\left(\overline{\sigma}(E)_o - \overline{\sigma}(E)_p\right)}{\overline{\sigma}(E)_o} \qquad 6.45$$

is skewed strongly towards negative values of b, for low values of $\overline{\sigma}(E)_o$. The explanation for this is that a given absolute error, $\overline{\sigma}(E)_o - \overline{\sigma}(E)_p$, will give a larger *percentage* error for error for a smaller $\overline{\sigma}(E)_o$ than for a higher value. Unfortunately some 13.6% of entries have negative percentage errors of greater than 100%. This is much larger than was observed in earlier regressions and seems somewhat unsatisfactory. In Figure 6.8 above we see that the upper and lower bands of $\left\langle RN_c^{1/2} \right\rangle$ deviate from linearity.

Because of this deviation, more stringent criteria were imposed on the data included in further regressions, the results of which are also presented in Table 6.23(b). (regression

44.2) In these regressions the selection criteria were derived from the full distribution of $\overline{\sigma}(C)_e$ for both centric and non-centric structures together, values below the lower decile and above the upper decile being removed. The criteria applied truncated the data with lower and upper limits of $\overline{\sigma}(C)_e$ of 0.0023 and 0.0160Å respectively. These criteria reduced the number of data in the regressions by 24661 entries. (some 19.7%). Of these 11884 values were less than 0.0023Å and 12777 entries were greater than 0.0160Å.

The number of data truncated from either side of the distribution was markedly different in the non-centrosymmetric subset. In centric structures the numbers truncated from the upper and lower regions of the distribution (defined as N+ and N- respectively) were approximately the same, but in non-centric systems it was found that N-<<N+. This reflects the domination of centric structures in the crystallographic database. We expect centric structures to have lower e.s.d's than their non-centric equivalents; there is a factor of $\sqrt{2}$ between these two subsets as can be seen in the original Cruickshank (1960) derivation. The domination of centric structures means that the median of the distribution is lower than the median in non-centric structures and so a larger number of non-centric structures with larger e.s.d's are removed.

The results of the regressions are significantly improved by the removal of these data points. The R_σ values are close to 30% and the value of pc(σ), the mean percentage error, defined as

$$pc(\sigma) = \frac{100\left(\sum_{i=1}^{N_{obs}}\left[\frac{\overline{\sigma}(E)_o - \overline{\sigma}(E)_p}{\overline{\sigma}(E)_o}\right]\right)}{N_{obs}} \qquad 6.46$$

also decreases. The predictive power of the equations, described by n50, n25, n10, N50, N25 and N10, increases. Some 82% of $\overline{\sigma}(E)_p$ values are within 0.0025Å of their observed values. 80% of values are within 50% of their observed values. These observations are taken over a very wide range of numerical values of $\overline{\sigma}(E)_o$.

The composite absolute and percentage error distribution (the centrosymmetric and non-centrosymmetric distributions taken together) for regression 44.2 in Table 6.23(b) are shown in Figures 6.14(a) and (b). The absolute error distribution is much less skewed; only 2.9% of atoms have an absolute error that exceeds 0.005Å. The percentage error is still strongly skewed toward negative values signifying that the lower values of $\overline{\sigma}(E)_p$ are over-estimated. This is due to reasons already noted, but in this analysis the number of seriously over-estimated e.s.d's is lower; only some 6.5% of values have a negative percentage error that exceeds 100%.

These results show the overall benefit of removing the upper and lower deciles from the distribution. Further regressions were carried out to try to establish the main contributing factor to the discrepancy in the highly mis-estimated data. The regressions of type 44.3 in Table 6.23(b) relate to a data set where additional acceptance critera have been applied on the value of the crystallographic R-factor. In these regressions the R-factors of entries used were restricted to values between 2 and 7%. The assessment criteria in the regression only show marginal improvement, despite a removal of some 11.6% entries from the regression . The constants in both sets of regressions, K_c and K_{nc} are practically equal. These points suggest that the deviant entries are not in the high R-factor regions.

The regression ratios, K_{nc}/K_c are 1.351, 1.325 and 1.343 for the three sets of regressions described above. All are in fair agreement with the expected value of 1.414, although the value is not as close as the values of 1.399 and 1.417 observed for the

Figures 6.14(a) : The composite absolute error distribution of $\overline{\sigma}(C)_e$ for regression 44.2
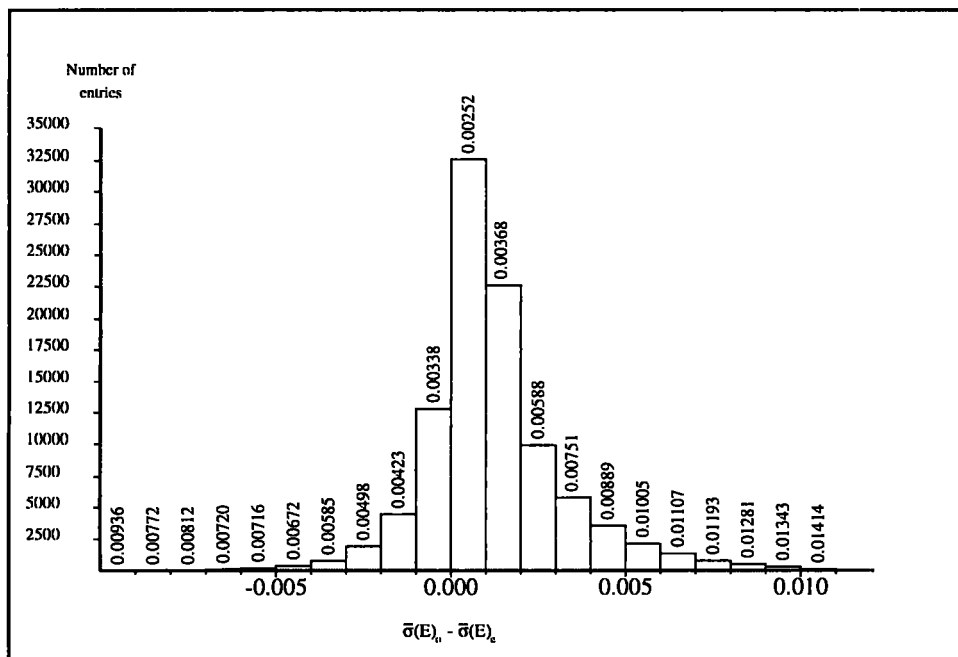
Figures 6.14(b): The percentage error distribution of $\overline{\sigma}(C)_e$ for regression 44.2
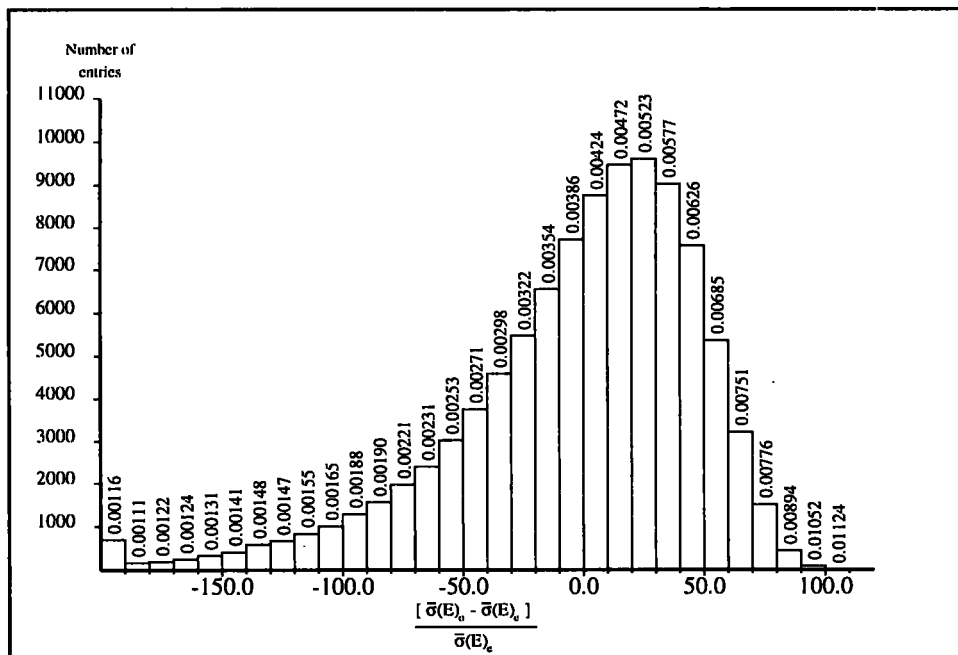
regressions based on carbon. This is probably due to the broad range of element types that are represented in the data set. Earlier in Table 6.8 and Figure 6.8, it was shown that the ratio, k, $\left[= \left\langle \overline{\sigma}(C)_e \right\rangle / \left\langle RN_c^{1/2} \right\rangle \right]$ steadily decreases with increasing atomic number.

Data set 2A is dominated by light atoms, C,N,O; some 61% of the observations in regressions 44.1 and 44.2 are e.s.d's from light atoms. The evidence above suggests that heavier atoms should be treated separately from light atoms. The earlier regressions using the heaviest atom in the asymmetric unit (Table 6.15) also suggest a non-linear dependancy; the ratio between the regression constants here is 1.483, appreciably higher than the expected value of 1.414.

Because of the evidence above further regressions were carried out on smaller subsets of data set 2A. These subsets split data set 2A into first row elements and other elements. The results of these regressions (44.4 and 44.5) are also shown in Table 6.23(b). The heavier metal regression deals with values of $\overline{\sigma}(E)_a$ which are smaller. Because of this the absolute assesment criteria, N_{50}, N_{25} and N_{10} appear better. A more realistic assessment is given by n_{50}, n_{25} and n_{10} which are comparable to the earlier regressions for the heavier elements, and are notably better for the lighter atom regressions. This is consistent with a non-linear dependance of $\sigma(E)$ on Z; the heavy atom regressions still possess a wide range of values of Z, whereas the light atom regressions do not. One very promising aspect of these regressions is that the negative percentage error only exceeds 100% in 3.5% of cases. The constants in these regressions (centric and non-centric) are very close to those observed for data set 2 based on the heaviest element present in the asymetric unit. The ratio of regression constants between 44.4 and 44.5 is 1.393 which compares very favourably to the expected value of 1.414 (Cruickshank, 1960).

183

6.6.1 The effect of incorporating scattering factors

The constants observed in regression 44.4 are somewhat lower than those observed for carbon atoms, or compared to the regressions incorporating only light atoms, such as 44.5. As was explained in Section 6.4.4, this effect is probably due to the use of Z_C/Z_E as opposed to $\bar{f}_C/\bar{f}_E$.

Due to the effects described in Section 6.4.4, it was decided to analyse the effect of incorporating the scattering factor into a further three sets of regressions. A constant value of $\bar{s}$ was assumed. Scattering factors that corresponded to $\bar{s}$ were used to modify equation 6.44 thus:

$$\bar{\sigma}(E)_p = \frac{\bar{f}_C kRN_c^{\frac{1}{2}}}{\bar{f}_E} \qquad 6.47$$

The majority of structural studies are measured to 55° in θ for Cu Kα radiation. This gives an s_{max} $\left[= 2\sin(\theta_{max})\right]$ of 1.063Å-1 which corresponds to an $\bar{s}$ (by equation 6.42) of 0.823Å^{-1}. This gives a $\bar{\theta}$ $\left[= \bar{s}\lambda/2\right]$ of approximately 0.3Å^{-1}. Scattering factors (*International Tables of Crystallography*, 1992) at this value of $\sin\theta/\lambda$ were used in the following analysis.

To study the overall effects of the scattering factor required a re-calculation of the workfile for data set 2A to incorporate a value of $\overline{N_c}$, as an alternative to N_c. This is given in equation 6.48 below:

$$\overline{N_c} = \sum_{i=1}^{N_{nH}} \frac{\bar{f}_i^2}{\bar{f}_c^2} \qquad 6.48$$

The following three equations were tested. In each case the incorporation of the scattering factor data takes a different form; The first regression only analyses the effect of substituting $\overline{N_c}$ for N_c. In the second regression the more fundamental problem of

diferential fall off is analysed by using $\overline{f}_C/\overline{f}_E$ rather than the ratio of Z_C/Z_E. The third regression attempts to consider both possible effects.

$$\overline{\sigma}(E)_p = \frac{6kR\overline{N}_c^{\frac{1}{2}}}{Z_E} = \frac{KR\overline{N}_c^{\frac{1}{2}}}{Z_E} \qquad \text{6.49(a)}$$

$$\overline{\sigma}(E)_p = \frac{2.494kRN_c^{\frac{1}{2}}}{\overline{f}_E} = \frac{KRN_c^{\frac{1}{2}}}{\overline{f}_E} \qquad \text{6.49(b)}$$

$$\overline{\sigma}(E)_p = \frac{2.494kR\overline{N}_c^{\frac{1}{2}}}{\overline{f}_E} = \frac{KR\overline{N}_c^{\frac{1}{2}}}{\overline{f}_E} \qquad \text{6.49(c)}$$

Here, 2.494 is the value of the scattering factor for carbon at $\sin\theta/\lambda$ of $0.30\text{Å-}1$. Scattering factors for any element at this value of $\sin\theta/\lambda$ are denoted as $\overline{f}_E$

An acceptance criterion for earlier regressions on data set 2A restricted the allowed values of the observed "equivalent carbon" e.s.d's , $\overline{\sigma}(C)_e$, which are derived by normalising each $\overline{\sigma}(E)_o$ value to carbon using Z_C/Z_E This criterion is appropriate for regressions based on equation 6.49(a) but, in equations 6.49(b) and (c), the e.s.d's are normalised via the ratio of atomic scttering factors at $0.30\text{Å-}1$. Because of this it is inappropriate to use $\overline{\sigma}(C)_e$ to define the inclusion criteria. For regressions based on equations 6.49(b) and (c), a new distribution of $\overline{\sigma}(C)_f$ was derived where:

$$\overline{\sigma}(C)_f = \frac{\overline{f}_E}{\overline{f}_C}\overline{\sigma}(E)_o \qquad \text{6.50}$$

Regressions based on equations 6.49(b), (c) utilised this distribution. The upper and lower decile threshold values of $\overline{\sigma}(C)_f$ calculated were 0.0028 and 0.0200Å. These are slightly higher than the values of $\sigma+$ and $\sigma-$ observed for the $\overline{\sigma}(C)_e$ distribution (0.0023 and 0.0160 respectively). This is because $\overline{f}_e/\overline{f}_c > Z_E/Z_C$. The $\overline{\sigma}(C)_f$ distribution is essentially of the same form as for $\overline{\sigma}(C)_e$, but the magnitudes of the trends are generally enhanced.

185

Below, in Table 6.24, results are shown for the three regressions based on equation 6.49(a). These three regressions are equivalent to the three counterparts 44.2 and 44.4 and 44.5 in Table 6.23. Table 6.24 shows that the substitution of $\overline{N}_c^{1/2}$ for $N_c^{1/2}$ does not improve the assessment criteria. Further, the regression constants, K_c and K_{nc}, for the heavier atom regressions (Table 6.24, 49(a).5.) are lower than those for their light atom counter-parts (Table 6.24,49(a).4). We can conclude from this that using $\overline{N}_c^{1/2}$ instead of $N_c^{1/2}$ does not correct for the differences between the light and heavy atom regressions; this alone is not correcting for the effects that were noted by Cruickshank (1993).

Regressions using equation 6.49(b) are shown in Table 6.25. In these regressions the original expression for $N_c^{1/2}$ is used in conjunction with the ratio of the scattering factors. Again, the regressions correspond to their counterparts, 44.2, 44.4 and 44.5, in Table 6.23. The assessment criteria are improved as compared to those observed in Table 6.23. The percentage differences, particularly, are better. Most improvement is seen for the heavy atom regressions, as shown by comparing regression 44.5 with 49(b).5. The light atom results are not significantly improved. The K values also indicate the improvement of the regressions. The ratios K_{heavy}/K_{light} are 0.910 (centrosymmetric structures) and 0.965 (non-centrosymmetric structures) by comparison with 0.737 and 0.781 observed for the comparable regressions in Table 6.23(b). The expected value for these regressions is, of course 1.0, which would indicate that no independant effect is present due to the atomic constitution of a particular crystal.

Table 6.24 Results of regression analyses based on equation 6.49(a)

The parameters used as assessment criteria are defined in Table 6.12. RN is the regression number and C/NC indicates that only centrosymmetric or non-centrosymmetric structures were included. The selection of data for regressions 49(a).2,49(a).4,49(a).5 is exactly as given in Table 6.23(a) for regressions 44.2,44.4,44.5 respectively.

RN	49(a).2		49(a).4		49(a).5	
C/NC	C	NC	C	NC	C	NC
K	0.05526	0.08180	0.04022	0.05932	0.05700	0.08378
$\sigma(K)$	0.00008	0.00024	0.00010	0.00033	0.00010	0.00029
$R\sigma$	32.2	32.5	33.6	32.7	29.8	30.9
rms(σ)	0.0021	0.0024	0.0008	0.0009	0.0025	0.0027
pc(σ)	38.4	36.6	35.9	35.7	28.9	29.7
N_{50}	96.1	94.5	99.9	99.9	94.0	92.9
N_{25}	84.3	79.9	98.1	97.1	77.4	74.8
N_{10}	57.0	50.5	84.3	80.3	45.3	43.5
n_{50}	75.5	77.9	75.2	77.2	85.0	84.4
n_{25}	44.0	45.3	41.4	44.7	51.0	49.8
n_{10}	18.1	18.8	17.2	18.1	21.1	20.7

Table 6.25 Results of regression analyses based on Eq.6.49(b).

The parameters used as assessment criteria are defined in Table 6.12. RN is the regression number and C/NC indicates that only centrosymmetric or non-centrosymmetric structures were included. The selection criteria for regressions 49(b).2,49(b).4,49(b).5 are analogous to those used for regressions 44.2,44.4,44.5 of Table 6.23(b). Mnemonics are given in Table 6.12

RN	49(b).2		49(b).4		49(b).5	
Z	all		$Z \geq 10$		$Z \leq 10$	
R_{min}	0.01		0.01		0.01	
R_{max}	0.10		0.10		0.10	
σ_{min}	0.0028		0.0028		0.0028	
σ_{max}	0.0200		0.0200		0.0200	
C/NC	C	NC	C	NC	C	NC
N_{ent}	25685	8066	25685	8066	25685	8066
N_{tot}	97874	27031	37345	7577	60529	19454
N-	10877	1684	2535	280	8342	1404
N+	9781	3208	5461	1593	4320	1615
N_{at}	77216	22139	29349	5704	47867	16435
K	0.03814	0.05041	0.03487	0.04869	0.03831	0.05046
$\sigma(K)$	0.00006	0.00014	0.00008	0.00026	0.00007	0.00016
$R\sigma$	30.0	30.6	31.1	30.5	29.6	30.7
rms(σ)	0.0022	0.0025	0.0007	0.0008	0.0027	0.0029
pc(σ)	31.4	31.2	31.9	31.6	29.5	30.8
N_{50}	95.5	93.8	99.9	99.9	92.7	91.6
N_{25}	84.0	79.0	98.9	98.2	74.9	72.3
N_{10}	59.4	51.4	87.0	83.1	42.7	40.4
n_{50}	81.5	82.5	80.3	82.0	84.2	83.1
n_{25}	49.2	49.3	46.7	47.7	51.4	49.7
n_{10}	20.8	20.7	19.5	20.6	21.9	20.8

Table 6.26. Results of regression analyses based on Eq.6.49(c)

The parameters used as assessment criteria are defined in Table 6.12. RN is the regression number and C/NC indicates that only centrosymmetric or non-centrosymmetric structures were included. The selection criteria for regressions 49(c).2,49(c).4,49(c).5 are analogous to those used for regressions 2,4,5 of Table 6 and Table 7. Mnemonics are as given in Table 6.12.

RN	49(c).2		49(c).4		49(c).5	
Z	all		$Z\geq 10$		$Z\leq 10$	
R_{min}	0.01		0.01		0.01	
R_{max}	0.10		0.10		0.10	
σ_{min}	0.0028		0.0028		0.0028	
σ_{max}	0.0200		0.0200		0.0200	
C/NC	C	NC	C	NC	C	NC
N_{ent}	25685	8066	25685	8066	25685	8066
N_{tot}	97874	27031	37345	7577	60529	19454
N-	10877	1684	2535	280	8342	1404
N+	9781	3208	5461	1593	4320	1615
N_{at}	77216	22139	29349	5704	47867	16435
K	0.02556	0.03762	0.02289	0.03397	0.02570	0.03776
$\sigma(K)$	0.00004	0.00011	0.00005	0.00019	0.00005	0.00013
$R\sigma$	31.2	31.6	32.8	31.7	30.7	31.5
$rms(\sigma)$	0.0023	0.0026	0.0008	0.0009	0.0029	0.0030
$pc(\sigma)$	31.9	31.4	33.0	32.5	29.6	30.2
N_{50}	95.0	93.4	99.9	99.9	92.0	91.1
N_{25}	83.2	78.4	98.6	97.7	74.0	71.8
N_{10}	58.3	50.9	85.8	82.3	41.9	40.2
n_{50}	80.7	82.3	78.5	80.4	83.8	83.8
n_{25}	46.7	47.4	44.3	46.5	49.2	48.2
n_{10}	19.3	19.7	18.1	19.0	20.3	19.8

The final regressions utilised both $\overline{N_c}^{1/2}$ and the structure factor ratio as based on equation 6.49(c) shown above. Again, the counterparts to regressions 44.2,44.4 and 44.5 are shown. The overall assessment criteria for the three regressions (Table 6.26) are somewhat worse than for the regressions based on Eq. 6.49(b) (Table 6.25). The improvements seen are not maintained in these regressions. Also notable are the ratios, K_{heavy}/K_{light}, which are seen to be 0.890 and 0.899. These are better than for regressions based on the original formulation (44.2,44.4 and 44.5) in Table 6.23, but are worse than the equivalent regressions based on Eq 6.49(b). (Table 6.25).

The systematic effect observed throughout these regressions may be due to the difference in the relative scattering power of heavy atom structures compared to that for light atom structures. One would expect that $\bar{s}$ for heavy atom structures would be

higher than for light atom structures due to the stronger scattering of the heavier elements. This may explain why the regressions of Tables 6.24 and 6.26 are poor; using a different $\bar{s}$ would induce a K_{heavy}/K_{light} ratio that is closer to unity in regressions based on Equation 6.49(c). [In those shown in Table 6.25, based on Eq. 6.49(b), the ratio will be somewhat greater than unity; this would be compensated by using $\overline{N_c}^{1/2}$ in regressions based on Eq. 6.49(c)]. This effect is probably not linear across the range of atomic numbers because of limitations on data collection: In most structural studies crystallographers only collect "enough" data to solve and refine a particular compound to an acceptable level, rather than collect to the edge of the scattering sphere, so $\bar{s}$ values will tend towards a maximum value.

6.7 Conclusions

The results presented within this Chapter show that it is indeed possible to obtain an order of magnitude estimate of the mean e.s.d. for a particular element within any particular structure, and across a wide variety of element types, from data stored within the CSD. In carbon based regressions it is possible to predict values of $\bar{\sigma}(C-C)$ that are within 0.005Å of their true values in 86% of cases. The work using data set 3 has shown that including some additional information pertaining to the strength of the scattering power of a particular crystal as suggested by the Cruickshank formulation (1960) only improves the results marginally. This may be due, in part, to the approximations that have to be used in deriving values of $\bar{s}$ and also due to the relative dominance of data set 3 by a low range of values of N_r and N_p. The original Cruickshank expression is only an approximation to the true e.s.d. and thus to expect more accurate results might be considered optimistic.

It is hoped that these results will be used, either alone or in conjunction with the stored AS flag, to predict a reasonable estimate of the coordinate e.s.d's for the large number of data that do not have this information available in the CSD. Furthermore the equations described within this Chapter could allow us to extend the information

content of the AS flags stored within the database. It may be possible to describe an "atomic AS flag" which is based on the carbon AS flag, since data set 2A has confirmed that a relationship exists between the uncertainties of different elements within any particular structure.

Reliable estimates of precision for all atomic positions may well allow "semi-weighted mean" estimates of geometric parameters to be derived as suggested by Taylor & Kennard (1983), although the equations tend to over-estimate the e.s.d. This may not be unreasonable, however, since a study by Taylor & Kennard (1986) has shown that the published standard uncertainties of atomic coordinates are often underestimated by up to 50% in the more precise structural studies.

It is recommended that the following equations should be used in generation of e.s.d. estimates. For estimates of carbon-carbond bond length e.s.d's, equations 6.20(a) and (b) should be used for centrosymmetric and non-centrosymmetric structures respectively:

$$\overline{\sigma}(C-C)_p = 0.01814 RN_c^{1/2} \qquad 6.20(a)$$

$$\overline{\sigma}(C-C) = 0.02537 RN_c^{1/2} \qquad 6.20(b)$$

For general atomic e.s.d. estimates for an element, E, equations based on regression set 6.49(b) should be used. The appropriate equations are:

$$\overline{\sigma}(E)_p = \frac{0.0381 RN_o^{1/2}}{\overline{f}_E} \qquad 6.51(a)$$

and

$$\overline{\sigma}(E)_p = \frac{0.0504 RN_o^{1/2}}{\overline{f}_E} \qquad 6.51(b)$$

for centrosymmetric and non-centrosymmetric structures respectively.

.

6.8 References

Abrahams, S.C., Hamilton, W.C., Mathieson, A. McL.; *Acta Crystallographica*, **A26**, 1, 1970

Acta Crystallographica Section C, Inorganic Chemistry, and *Organometallics*.

Numerical Algorithms Group, Subroutine Library Numerical Algorithms Group Ltd, Wilkinson House, Jordan Hill Road, Oxford OX2 5DR, England, 1990

Allen, F. H.; Cole, J. C.; Howard, J. A. K.; *Acta Cryst.* ; **A51**, 95, 1995,1

Allen, F. H.; Cole, J. C.; Howard, J. A. K.; *Acta Cryst.* ; **A51**, 112, 1995,2

Allen, F. H.; Doyle, M. J.; *Acta Cryst.*; **A43**, C291, 1987

Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G.; *J. Chem. Inf. Comput. Sci.*; **31**, 187, 1991

Cruickshank, D. W. J., *Acta Cryst.* ; **13**, 774, 1960

Cruickshank, D. W. J., Private Communication, 1993

Hamilton, W. C., Abrahams, S.C.; *Acta Cryst.*; **A26**, 18, 1970

International Tables for Crystallography, Vol. A, Edited by Theo Hahn, Third, revised edition. Kluwer Academic Publishers, London, 1992

Mackenzie, J. K.; *Acta Cryst.*; **A30**, 607, 1974

Muir, K. W.; Mallinson, P. R., *J. Appl. Cryst.*; **26**, 142, 1993

Pauling, L.; *The Nature of the Chemical Bond*, Ithaca, New York, Cornell Univ. Press, 1940

Pimental, G. C.; McClellan, A. L.; *The Hydrogen Bond*, San Francisco, 1960

QUEST; *Cambridge Structural Database Users Manual*, 1992

Sutton, L. E.; Editor. *Tables of Interatomic distances and Configuration in Molecules and Ions*, Special Publication No. 11, London: The Chemical Society, 1958

Sutton, L. E.; Editor. *Tables of Interatomic distances and Configuration in Molecules and Ions (Supplement)*, Special Publication No. 18, London: The Chemical Society, 1965

Taylor, R.; The CAMAL library, *J. Appl. Cryst.*; **19**, 90, 1986

Taylor, R.; Kennard, O.; *Acta Cryst.*; **B39**, 517, 1983

Taylor, R. ; Kennard, O.; *Acta Cryst.*; **B42**, 112, 1986

Chapter 7.

The relationship between space group, crystallographic molecular point group, and chemical constitution

7.1 Introduction

A fundamental aspect of crystallography is the determination of the space group of a crystal. The selection of a particular space group reflects the solid state energy minimum of molecular packing and conformation in the crystalline state, and can give us direct information about the chemistry of the solid state. The space group also governs many macroscopic crystal properties of an optical nature, primarily because of the symmetry restrictions that it imposes on the outcome of interactions between crystals and light.

In recent years, the science of molecular self-assembly has provoked interest within the chemical community. The benefits of designing systems that spontaneously self-assemble are yet to be fully realised, but one can quite easily see how such an area has academic interest, in that it could be thought to parallel the processes of replication that occur in life. In forming a crystal, a large number of like molecules have to aggregate together in an ordered way and so the science of crystal growth should be of interest to all physical scientists, not just to crystallographers.

Because of these facts a large amount of effort has been dedicated to establishing why crystals form in the way that they do. One approach that has been employed on numerous occasions has been to compile statistics of space group frequency from literature sources.

The earliest studies of space group frequency came from Nowacki (1942,1943). These studies were compiled by hand when a mere 3500 structures of all types had been

elucidated. The statistics were later revised as more structures were solved. Even in this early work, the striking dominance of a few low symmetry space groups was apparent for molecular crystals. Some scepticism was voiced as to the significance of the distribution: Mackay (1967) published a statistical report based on the distribution that Nowacki had observed. The distribution offered a unique chance to test a solution to a problem that had been challenging statisticians for over 40 years at that time. This was: "Is it possible to predict the number of possible events that would occur in a data set of infinite size from a finite data set in which not all possible events had occurred". Space group theory dictates that in 3 dimensions there are only 219 space groups (when enatiomorphic space groups are counted only once) and so this distribution offers a unique example where the number of "possible" events in an infinite dataset is known *a priori*. Mackay found that the theory predicted that the number of possible types of event in an infinite sample would be between 218 and 222 from the published distribution. From this he concluded that the theory was correct and that the abscence of certain space groups in the distribution may well have arisen by chance.

More recent work on space group occurence has used the *NBS Crystal Identification File* for empirical analysis of space group frequency. Mighell and Himes (1983) have produced a space group distribution for organic molecules in the NBS File. Their distribution shows the predominance of the space groups $P2_1/c$, $P\bar{1}$, $P2_12_12_1$ and $C2/c$, as predicted by Kitiagorodski (1961) who described these space groups as close-packed, without any special spatial requirements, (as opposed to "limitingly" closed packed where close-packing is only deemed possible for certain specific molecular orientations). Mighell and Rodgers (1980) have also produced a distribution for inorganic structures from the same source. A similar study by Baur & Kassner (1992) on the Inorganic Crystal Structures Database (ICSD) showed a much wider spread of crystallographic space group. No one space group dominates their distribution. The most common space-group in this distribution is Pnma (8.3%), closely followed by $P2_1/c$ (8.1%). The ICSD has a far higher population of high symmetry space groups:

15.6% of all structures on the ICSD are in cubic space groups (as compared to 0.5% on the CSD)

The most recent general study for organic systems was carried out in 1987 using the CSD (Allen *et al*, 1991) by Padmaja *et al* (1990). The results coincided with those found previously. They also analysed the distribution for systems with a reported Z value greater than the space group multiplicity (i.e. structures with more than one independent molecule in the asymmetric unit). They noted that the distribution in this case was less accentuated toward the space group $P2_1/c$, although in their paper they attribute some part of this change to poor space group determination. Baur and Tillmans (1986) have estimated that at least 3% of structures may be reported in the wrong space group, a very common error being a missed inversion centre. Marsh (1986) recommends refinement in the centrosymmetric space group in cases where possible ambiguity exists between a centrosymmetric space group and its non-centrosymmetric counterpart, even if this means using a slightly disordered model. Baur and Kassner (1992) have also reviewed the problem of incorrect space group determination. They conclude that the most commonly mis-assigned space group is Cc.

These studies, however, do not break down the frequency to take account of possible molecular symmetry which may have considerable influence on space group usage. They also do not consider the influence of charged species, or analyse the effect of the presence of discrete and different molecular moities on the distribution of space-groups. There is an example of a study of some 13000 space group occurences of organic molecules by Belsky and Zorkii (1977) in which the frequencies are broken down by point group usage. A similar approach is used in this study, although a much larger number of data have been used and the study covers not only organic molecules but metallo-organic examples as well.

Another recent study of space group frequency (Brock and Dunitz, 1994) has focussed on analysing the change in the space group distribution as a function of Z', the number

of independent "molecules" in the asymmetric unit. They draw a number of conclusions using space groups that possess only one form of point symmetry of any one particular order (for example, Pbca, which only has point inversion symmetry). Their study correctly assumes that structures in such space groups that have Z' < 1, have to utilise the only point symmetry available to be consistent. They also analyse the distribution of space groups where Z' > 1.

Kitiagorodskii (1961) explains the predominance of a few space-groups in simple organic molecules by assuming that, for an arbitrary molecule, all long distance intermolecular contacts in the lattice are attractive. From this assumption the lowest energy is achieved at the point of closest packing which corresponds to the densest possible arrangement in the crystal. He considers layer packing initially, and then "packs layers" to find the likely space groups for arbitrarily shaped moities. This approach, considering the number of examples that Kitiagorodski had available, works remarkably well for the vast majority of structures. The model does not, however, take into account the effect of specific molecular shape, or strong "structure determining" interactions. (An example of such an interaction dominating packing has been observed in the structures of mono-alcohols by Brock and Duncan (1994), where n-fold screwaxes can be encouraged due to H...OH non-bonding interactions, depending on the shape of the alcohol.).

Figure 7.1. Bad packing of an arbitrary figure due to mirror planes in crystal structures
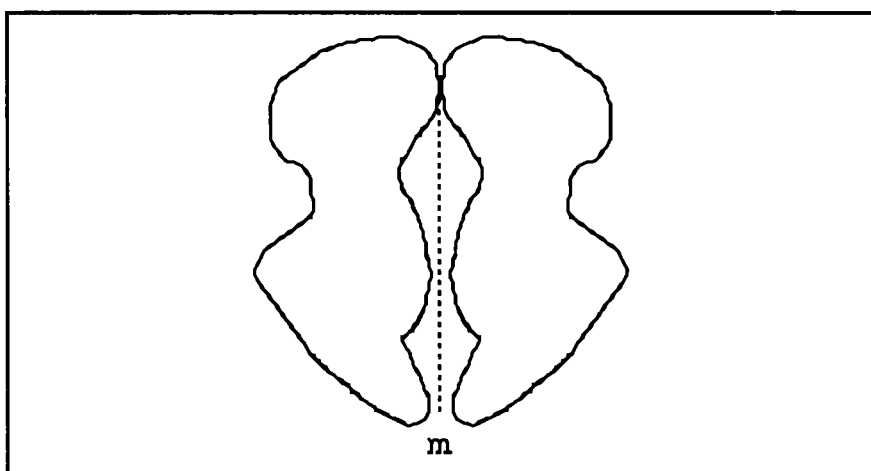
Figure 7.2. Favourable packing due to 2_1 axes.
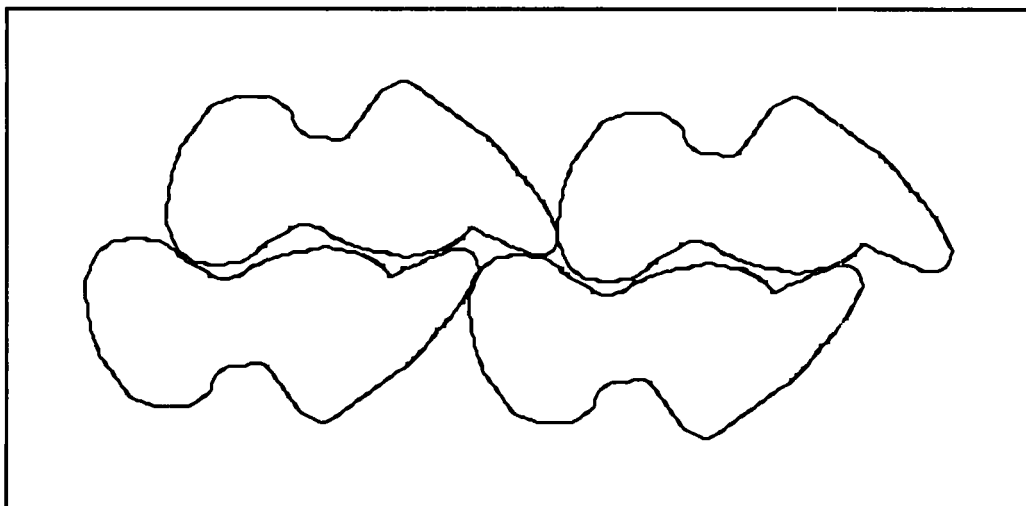
More recently, Wilson (1988, 1990, 1991, 1993) has developed a theoretical frequency model by assigning a positive or negative score to each symmetry element in the space group. The basis for this approach is that certain symmetry elements (such as mirror planes) would hinder packing by forcing unfavourable packing arrangements in the lattice (see Figure 7.1) whereas other elements (such as a 2_1 axis) assist in packing (see Figure 7.2). The most favourable types of symmetry elements involve translation. These can exist without specific requirements of molecular shape since they do not hinder six fold layer coordination. Wilson has used regression techniques to derive a model which relates the presence of specific symmetry elements within a space group to the observed space group frequency. In this analysis he has introduced the concept of the difference between "*encumbered*" and "*free*" symmetry elements. An encumbered axis is one that intersects with another symmetry axis, in effect one that does not operate without influence from another, and so Wilson weights this type of axis differently to its free counterparts in his model. An example of such a system occurs in $R\bar{3}$ where two different types of inversion centre exist. The inversion centre that is incorporated in the $\bar{3}$ axis would regarded as "encumbered" by the 3-fold axis.

Figure 7.3. Alignment of C=O dipoles.
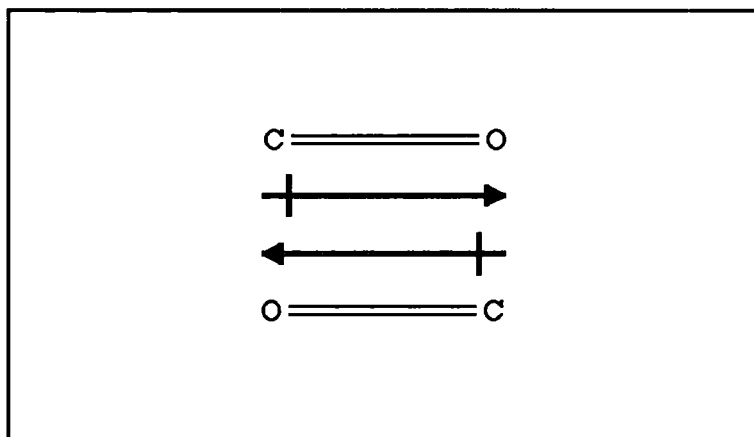
One explanation for the popularity of space groups P$\bar{1}$ and P2$_1$/c is that the overall molecular dipoles align in an opposite fashion to maximise favourable dipole-dipole interactions. A recent paper (Whitesell *et al*, 1994) discredits this point of view. They performed a statistical analysis of molecular dipole moments in three simple space groups, P1, P$\bar{1}$ and P2$_1$. The study revealed two points. Firstly, in space group P2$_1$, there was no correlation between the molecular dipole and the angle between symmetry related molecules. Secondly there was no statistical difference between the distribution of the magnitudes of molecular dipole moments in any of the three space groups; the dipole did not influence space group selection. Their conclusion is that a molecular dipole has little or no influence on centrosymmetric packing. This argument does not, however, refute the idea of localised bonding dipoles being implicated in determining molecular orientations (indeed, a recent study by Taylor *et al* (1990) on C=O bonds has suggested that C=O bonds always tend to align with their dipoles paired in opposing fashion, as shown below in Figure 7.3). Whitesell and co-worker's study also did not consider ionic species, particularly zwitterionic complexes. Their study was only focused on specific space groups.

In this chapter we consider the effect on space group frequency of a number of chemical and crystallographic factors. The influence of charge, and the number of discrete residues were both factors that were considered, along with the effect of occupation of cystallographic special positions. Full tables that cross crystallographic point group with

197

space group for a range different structures with varying numbers of residues (discrete bonded units: see Chapter 5) in the asymmetric unit are produced. There is a good reason for analysing space group distributions as a function of the number of residues (as opposed to a function of Z'): the number of residues is less susceptible to space group errors. Also the number of residues in the asymmetric unit is independent of the space group symmetry, so any effect on space group symmetry, as a result of change in the number of residues, can be regarded as an independent effect. A sub-group dataset has been generated using database bit-screens (see Chapter 5) to analyse similar statistics for charged systems, uncharged systems and "zwitterionic" compounds.

7.2 Methodology

7.2.1 Description of the data sets used

Data set 1

Data set 1 comprised all 120451 entries in the April 1994 release of the CSD. This data set was broken down into a number of subsets which were used to analyse differences in the distributions as a result of varying chemical and crystallographic conditions. Subset 1(a) contained only organic structural studies. Metallo-organic compounds comprised subset 1(b).A further subset, 1(c), was generated that comprised disordered structures only. For each structure, the space group number and symbol were retrieved along with the CSD refcode. A final subset, 1(d), was generated that contained all ordered "polymeric" (catena) structures on the CSD.

Data set 2

Data set 2 comprised 81832 entries from the April 1994 release of the CSD that fulfiled the criteria shown in Table 7.1

Table 7.1. Acceptance criteria for data set 2

No polymeric structures

No disorder reported in the publication

Entry error-free to the 0.002Å level

Coordinates were present.

Table 7.2. Information content of data set 2.

For each entry:

REFCODE	CSD reference code for the entry
R_f	Crstallographic R-factor
S	Space group
N_{at}	Number of non-hydrogen atoms stored for the entry
N_{sat}	Number of non-hydrogen "*satoms*" (see Chapter 5) stored for the entry
N_{hat}	Number of hydrogen atoms stored for the entry
N_{hsat}	Number of hydrogen "*satoms*" (see Chapter 5) stored for the entry
Z	The number of independent molecules within the unit cell
Z'	The number of independent molecules within the asymmetric unit
N_{res}	The number of crystallographic residues

For each residue within each entry:

P_r	Crystallographic point group of residue r
OP_r	Order of the point group P_r
$n_{at,r}$	Number of non-hydrogen atoms in residue r
$n_{sat,r}$	Number of non-hydrogen "*satoms*" in residue r
$n_{hat,r}$	Number of hydrogen atoms in residue r
$n_{hsat,r}$	Number of hydrogen "*satoms*" in residue r

A file containing the refcodes of the appropriate CSD entries was generated utilising the QUEST3D program (Allen et al, 1991). This was then used as input to an in-house program to generate data set 2 . For each refcode the in-house software decoded the atomic coordinates from standard CSD formats. These coordinates were then separated into individual CSD defined residues (see Chapter 5) and the site symmetries of each residue were determined using the method described in Section 7.2.2.

As for data set 1, data set 2 was split into a number of smaller subsets. Subsets 2(a) and 2(b) contained organic and metallo-organic structures respectively. Subset 2(c) contained only structures in which no residue had a net charge (i.e. non-ionic and zwitterionic compounds). Ionic structures were all included in subset 2(d). A fifth small data set (subset 2(e)) of structures that contained "zwitterionic" residues was also derived by using a CSD assigned "*bit-screen*" (see Chapter 5).

A number of other items were also retrieved for each entry. These are shown in Table 7.2.

7.2.2 Point group determination.

The symmetry atoms ("*satoms*", defined in Chapter 5) were used to identify the molecular point group of each residue. Each "satom" is stored as a single integer (see Chapter 5). Contained in this integer there is a reference to the symmetry operator, i, which was used to generate the fractional coordinates of the satom from its parent atom.

Table 7.3. Trace and determinant properties for the twelve non-translational crystallographic symmetry operations.

Operator	1	2	3	4	6	$\bar{1}$	$m(\bar{2})$	$\bar{3}$	$\bar{4}$	$\bar{6}$
Trace	3	-1	0	1	2	-3	1	0	-1	-2
Determinant	1	1	1	1	1	-1	-1	-1	-1	-1

Each symmetry operator consists of a 3 × 3 rotational matrix and a 3 × 1 translational vector.

The type of symmetry operator is easily identified from its 3 × 3 rotational matrix. The trace and determinant of this matrix uniquely identify the nature of the symmetry operator, if we consider Table 7.3 where the properties for all non-translational[§] operators are shown (Giacovazzo, 1993).

This process does not discriminate between translational and non-translational symmetry elements. Because of this, polymeric (catena) crystal structures were omitted from data set 2. (An isolated residue cannot have symmetry atoms generated by a translational symmetry element, such as a 2_1 screw axis: see Figure 7.4)

Any non-translational operator used can thus be easily identifed. Other stored operators also need testing, however, since it is possible for a molecule to lie so that it is entirely within a symmetry element: i.e. with all atoms lying on special positions. A very obvious example of such a case is a flat molecule, which can lie fully within a mirror plane (an example is shown in Figure 7.5). No satoms would be stored for this entry on the CSD. To detect such situations, all symmetry operators that were not independently used for symmetry related atom generation were tested on the full coordinate set.

Figure 7.4. Polymeric nature of a 2_1 axis

The figure shows that if a symmetry atom is generated via a translational symmetry axis (In this case a 2_1 axis), then a polymeric chain is implied by generating further symmetry atoms from the first one

---
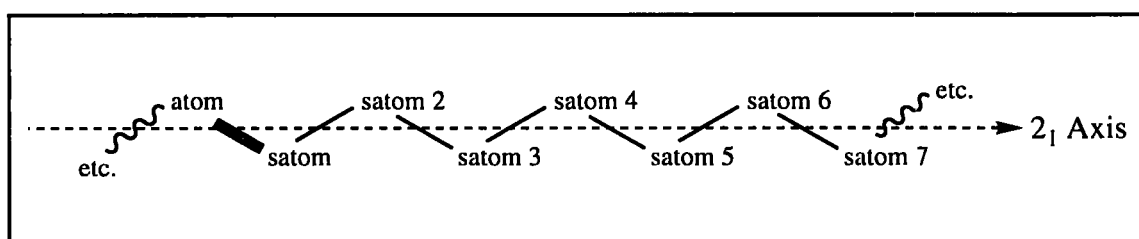
[§] i.e. operators that do not involve a component of translation in the axis direction, e.g. a 3_1 axis

Figure 7.5 An example of a molecule lying fully within a mirror plane.

The structure (Refcode BRTETZ) is in space group $P2_1/m$. As is apparent, the molecule lies fully in the mirror plane at $x, \frac{1}{4}, z$
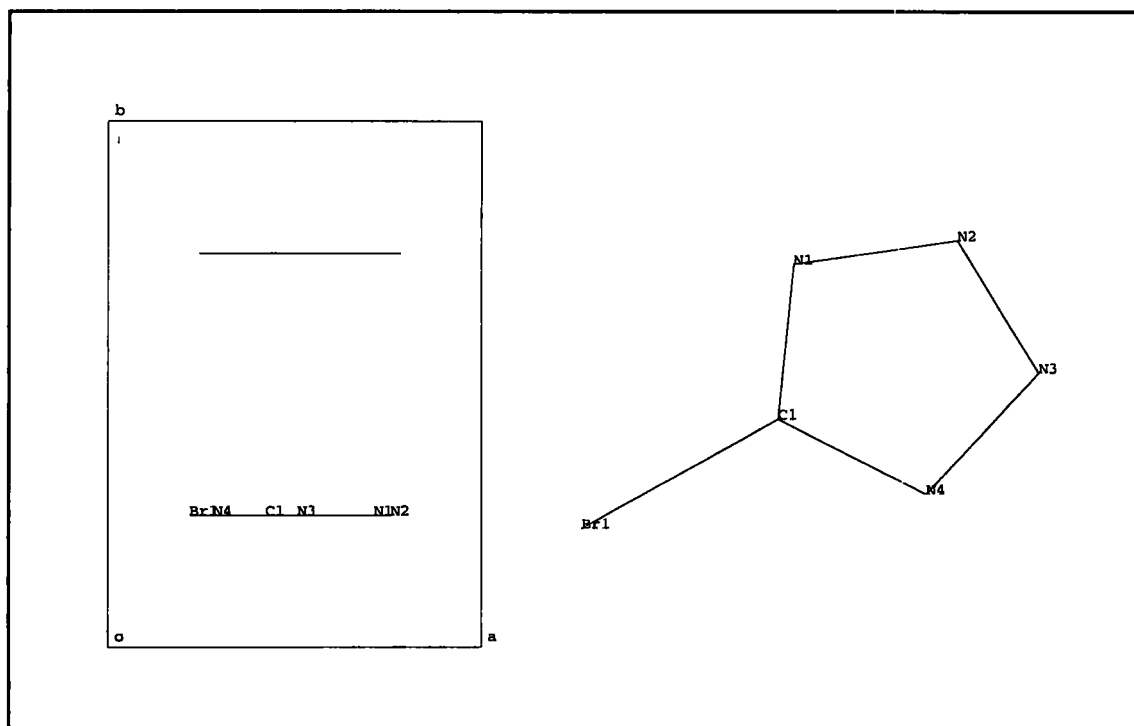
To test each operator the following process was used. Consider a fractional coordinate $X_{o,r}$ in a residue r which has the set of atomic coordinates and symmetry atom coordinates, $\{X_{s,r}\}$. By applying an operator, g, to this coordinate a new coordinate, $X_{g,r}$, is generated. If $X_{g,r}$ is a member of $\{X_{s,r}\}$ for all atomic coordinates, then the operator, g, is part of the residue's crystallographic point group.

In some cases, the original coordinates, $\{X_{s,r}\}$, can lie on a symmetry element that is related to a CSD stored element by a unit cell translation. To test for such situations, if the coordinate $X_{g,r}$ was not a member of $\{X_{s,r}\}$ then unit cell translations were applied to $X_{g,r}$ (i.e. if $X_{g,r} = x_{g,r}, y_{g,r}, z_{g,r}$ then the coordinates in the range $x_{g,r} \pm 2, y_{g,r} \pm 2, z_{g,r} \pm 2$ were generated). If any of these variants were members of the set $\{X_{s,r}\}$ then the operator, g, was regarded as a member of the residue's crystallographic point group.

Table 7.4. Point group table for the point group 6/mmm (D_{6h}).

6/mmm	1	2×6	2×3	2	3×2	3×2	$\bar{1}$	$2\times\bar{6}$	$2\times\bar{3}$	m	$3\times m$	$3\times m$
D_{6h}	E	$2C_6$	$2C_3$	C_2	$3C_2'$	$3C_2''$	i	$2S_3$	$2S_6$	σ_h	$3\sigma_d$	$3\sigma_d'$
A_{1g}	1	1	1	1	1	1	1	1	1	1	1	1
A_{2g}	1	1	1	1	-1	-1	1	1	1	1	-1	-1
B_{1g}	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1
B_{2g}	1	-1	1	-1	-1	1	1	-1	1	-1	-1	1
E_{1g}	2	1	-1	-2	0	0	2	1	-1	-2	0	0
E_{2g}	2	-1	-1	2	0	0	2	-1	-1	2	0	0
A_{1u}	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1
A_{2u}	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1
B_{1u}	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1
B_{2u}	1	-1	-1	-1	-1	1	-1	1	-1	1	1	-1
E_{1u}	2	1	-1	-2	0	0	-2	-1	1	2	0	0
E_{2u}	2	-1	-1	2	0	0	-2	1	1	-2	0	0

A crystallographic point group can be represented as a pair of six digit integers. The first integer is used to indicate the number of rotational symmetry elements within the point group. The second integer does likewise for improper rotational elements. The nth digit within each integer shows the number of nth order axes present within the point group. This is best seen with an example. The point group table for 6/mmm (D_{6h}) is shown in Table 7.4.

The table shows the number of symmetry operators of each type that are present in the point group. The two integers, I_1 and I_2, described above are defined mathematically in equations 7.1 and 7.2. For proper rotations:

$$I_1 = \sum_{i=1}^{6} N_{Pr_i} \cdot 10^{(i-1)} \qquad (7.1)$$

where N_{Pr_i} is the number of operators of order i in the point group. For the point group 6/mmm (D_{6h}), I_1 can be written as

$$I_1\left(\tfrac{6}{mmm}\right) = 2\times10^{(6-1)} + 2\times10^{(3-1)} + 7\times10^{(2-1)} + 1\times10^{(1-1)} = 200271$$

Equation 7.2 defines the second integer, for improper axes:

$$I_2 = \sum_{i=1}^{6} N_{\text{Impr}_i} \cdot 10^{(i-1)} \tag{7.2}$$

In this case N_{Impr_i} is the number of improper axes of order i in the point group again; this gives

$$I_2\left(\tfrac{6}{mmm}\right) = 2 \times 10^{(6-1)} + 2 \times 10^{(3-1)} + 7 \times 10^{(2-1)} + 1 \times 10^{(1-1)} = 200271$$

A full table of the derived integers for all 32 point groups is shown in Table 7.5. By using equations 7.1 and 7.2, two integers could be derived for each residue in each entry of data set 2. By comparing these integers to those in Table 7.5, each residue's crystallographic point group could be determined.

Point group assignments were checked computationally in two ways. Primarily, any impossible groupings of symmetry elements that were found to be used by the database were reported. Secondly a look-up table (written by the author) using the special position multiplicities for each space group (*International Tables for Crystallography*, 1992) was used to identify molecules that appeared to occupy special positions that did not occur in the reported space group. Each entry that manifested one or more of these problems was checked by hand and in every case an error had occured either in the original publication or on input into the database. These errors have been noted and will be investigated at the CCDC.

Table 7.5. Integer representations of point groups

Point Group	Proper Rotation	Improper Rotation	Point Group	Proper Rotation	Improper Rotation
1	1	0	$\bar{3}$	201	201
$\bar{1}$	1	1	32	231	0
2	11	0	3m	201	30
m	1	10	$\bar{3}$m	231	231
2/m	11	11	6	200211	0
222	31	0	$\bar{6}$	201	200010
mm2	11	20	6/m	200211	200211
mmm	31	31	622	200271	0
4	2011	0	6mm	200211	60
$\bar{4}$	11	2000	$\bar{6}$2m	231	200040
4/m	2011	2011	6/mmm	200271	200271
422	2051	0	23	831	0
4mm	2011	40	m$\bar{3}$	831	831
$\bar{4}$2m	31	2020	432	6891	0
4/mmm	2051	2051	$\bar{4}$3m	831	6060
3	201	0	m$\bar{3}$m	6891	6891

7.2.3. Analysis of data sets 1 and 2

Two standalone computer programs were written to compile tables of space group frequencies from the two data sets. One program compiled statistics of space group frequencies, and was used to analyse data set 1. The other analysed the relationship between space group and point group and allowed the user to produce a 230 × 32 matrix of populations for all possible combinations of space group and point group. Both programs accounted for duplicate refcodes within the CSD. These occur for a number of reasons: The CSD stores two entries when a structure of a particular compound is determined on more than one occasion. These entries are related through their respective refcodes. An example is oxalic acid dihydrate. Numerous studies of oxalic acid dihydrate have been carried out through the years, for a variety of reasons. Each has its own refcode. Some of these are:

As is apparent from these examples, the first six characters of refcodes of duplicate compound determinations are the same. Naturally, any frequency tables have to be corrected for the prescence of duplicate determinations to prevent certain compounds from being counted twice. Unfortunately, the "duplicate refcode" problem is somewhat deeper, since polymorphic structures are also assigned the same refcode (they are studies of the same chemical *compound*, and so they are classed together by this refcode assignment criterion).

To solve the polymorph problem, structures with duplicate refcodes but different space groups were counted individually. It was accepted that some polymorphic compounds would be missed as a result of this decision (since some polymorphic pairs have the same space group, just different unit cell dimensions) but the frequency tables should not be strongly influenced due to the relative rarity of "same space group" polymorphism within the CSD.

The two programs did not account for incorrect space group determinations (see Marsh, 1986, Baur *et al*, 1986, 1992 and references therein), i.e. cases where the space group reported is actually a sub-group of the true space group. The R-factor does not indicate such situations: usually refinement in a sub-group results in a similar value as in the true space group[†]. Such situations can be detected within the refinement process, since positional and thermal parameters in the low symmetry model will be strongly correlated. In principle, it would be possible to detect such situations through analysis of the stored coordinates, but to do this for every compound on the CSD would be too CPU intensive. To reiterate the extent of this problem, it should be recalled from

---

[†] Such situations arise due to missed molecular symmetry. This can mean that the lower symmetry model is over-paramaterized, if the true structural model utilises molecular symmetry. In such cases the R-factor in the lower symmetry space group can be slightly *lower* due to the higher number of degrees of freedom available in the lower symmetry model.

Section 7.1 that an estimated 3% of structures have had their space group mis-determined (Baur and Tillmanns, 1986).

A possible alternative approach would be to only count the first occurence of each refcode within a duplicate set. This method of analysis would be worse than the method employed, since incorrect space group assignments will usually be in the chronologically earliest study. These studies are usually the first studies on the CSD and so have the first refcode: a table based only on the earliest refcodes would contain the most mis-assigned space groups.

7.2.4. A note on data presentation.

The analysis of these two data sets involves the generation of numerous large *"sparse"*[†] result matrices. It is impratical to present all of these matrices in the primary text of this thesis, due to the large amount of space that would be required. Because of this, the full data matrices are included in a microfiche format at the end of this thesis. The results included in the primary text will focus on specific space groups which will be shown, or will refer the reader to the microfiched data. Tables within the microfiched data are denoted "7M" (e.g. the third microfiched table would be Table 7M.3)

7.3. Descriptive statistics for data sets 1 and 2

Data set 1

A breakdown of the data content of data set 1 is given below in Table 7.6. A number of general features about the CSD can be noted from this table. The CSD has an approximately equal representation of organic and organometallic structures. 15.1% of

---

[†] "sparse" is a term used to describe a matrix which has a large number of elements that have a value of zero.

Table 7.6. A summary of the data content of data set 1

Values given in parentheses refer to the percentage of the entries in the column subset. This table approximately follows subsets (a) to (d), defined in Section 7.2.1 (Subset (d) for polymeric structures does not contain disordered structures. In this table polymeric structures are shown with disordered structures to provide an indication of the numbers of polymeric structures that are disordered.

Subset	All	Organic	Organometallic	Polymeric	Disordered
All	120481 (-)	61551 (-)	58930 (-)	3531 (-)	18228 (-)
Organic	61551 (51.1)	-	-	637 (18.0)	7003 (38.4)
Organometallic	58930 (48.9)	-	-	2894 (82.0)	11225 (61.6)
Polymeric	3531 (2.9)	637 (1.0)	2894 (4.9)	-	664 (3.6)
Disordered	18228 (15.1)	7003 (11.4)	11225 (19.1)	664 (18.8)	-

structures show some form of disorder. To check the CSD for disordered solvated molecules, a search was used that asked the question "does the text string 'solv' occur in the disorder text field or do both the text strings 'solv' and 'diso' occur in the remarks text field" for disordered structures. This search yielded 3428 hits (18.8% of the disordered structures on the CSD). This gives an idea of the number of structures with disordered solvated molecules, although this estimate does not account for situations where both the primary molecular fragment and the solvent are disordered, and some disordered solvents will not be recorded in this way.

Table 7.6 also shows that disordered structures are predominantly organometallic. This may well be due to the nature of the materials studied, as a result of the types of ligands used which tend to be conformationally flexible. Another point that is apparent from Table 7.6 is that polymeric structures are more likely to be organometallic rather than organic.

Data set 2

A survey of the data content of data set 2 is given in Tables 7.7 and 7.8.It is apparent, from Table 7.7, that organometallic entries are more severly affected by disorder or higher R-factors (or more organometallic structures do not have coordinates) because a more drastic reduction occurs in the number of organometallic structures as a result of

Table 7.7. A break down of the numbers of entries of various categories

Number of entries (including duplicates):	81832
Number of entries (including duplicates) with R-factor ≤ 10%:	77553
Number of unique entries with R-factor ≤ 10%:	73765
Number of unique organic entries with R-factor ≤ 10%:	40228[†]
Number of unique organometallic structures with R-factor ≤ 10%:	33540[†]

Table 7.8. A break down of numbers of structures as a function of ionic nature and the number of residues

N_{res}	All	Ionic[†]	Non-ionic[†Δ]
1	47436 (+3 in Fd3m)	0	47436
2	16805	7276	9532
3	4859	3331	1529
4	2442	1847	596
5	881	-	-

imposing these precision criteria. Table 7.8 shows that the majority of compounds on the CSD are non-ionic, although the relative proportions of ionic systems to non-ionic systems increases with residue (as would be expected: discrete ionic systems[Δ] consist of at least one cation and at least one anion, and so by definition require more than one residue.

[†] The slight discrepancies between the sums of various groups and the quoted sums of these groups is due to a small number of erroneous entries on the CSD, where duplicate refcodes have been assigned to compounds which do not have the same molecular constitution.

[Δ] "Non-ionic" also encapsulates the small number of zwitterionic complexes on the CSD: I.e. the definition of a "non-ionic" system is "having no residue with a net charge"

7.4. Analysis of data set 1.

The full distribution of space group populations for all unduplicated refcodes in the April 1994 version of the CSD is given in Table 7M.1. The distribution observed is naturally similar to that observed by Mighell, Himes and Rodgers (1983) and further confirms that a relatively small number of space groups dominate organic and metallo-organic crystal structures.

The most common chiral space group is $P2_12_12_1$. This result seems somewhat counter intuitive (it is unusual that a higher symmetry space group should be more popular than a lower one). This is neatly explained by Brock & Dunitz in a recent publicatation (1994), by a wider consideration of the effects of screw axes and glide planes on space group frequency. Screw axes are seen to feature heavily in the more popular space groups, and they conclude from this that a screw axis is more favourable in crystal packing than translation thus the high density of screw axes in $P2_12_12_1$ is favourable. More evidence for this conclusion can be drawn from semi-empirical energy calculations. Work in this area has been carried out by Filippini and Gavezzotti (1991) They have used a crystal packing potential energy function of the form

$$PPE = \sum_{i=1}^{N_{ats}} \sum_{j=1}^{N_{ats}} \left[A\,e^{-Br_{ij}} - \frac{C}{r_{ij}^6} \right]$$

to calculate the interaction energy between near neighbours in the crystal field for a range of non-hydrogen bonded organic compounds. In their equation, r_{ij} is the inter-atomic distance between atoms i and j and A,B and C are parameters that vary depending on the type of contact. In their study they considered several different compounds in a number of different space groups. For each of these compounds, they used the pair potential given above to calculate the relative contribution to the cohesion energy caused by nearest neighbour interactions created as a result of specific symmetry elements within the lattice. In this way the relative importance of each symmetry element could be established for each compound. Generally, 2_1 axes made large

contributions to the lattice cohesion energy for compounds in the five space groups that were considered ($P\bar{1}$, $P2_1$, $P2_12_12_1$, $P2_1/c$, and Pbca), although the glide plane dominated in Pbca. (They argue that the prescence of three glide planes in Pbca inhibits the action of other symmetry operators, since the inherent screw axes do not make a large contribution to the cohesion energy.)

In the centric examples the inversion operator usually makes a very large contribution. In $P\bar{1}$ they conclude that the inversion centre is responsible for nearly all of the cohesion energy, very little is due to translation, although translation does account for up to ~50% of the cohesion energy in other space groups.

As previously noted some 38.5% of all organic and metallo-organic systems in this study occur in the space group $P2_1/c$. The high frequency of this space group has been noted in previous studies. This is easily explained by considering the overall packing of molecules. Kitiagorodski explains that the inversion centre operator applies no restriction on molecular orientation with respect to a specific crystallographic direction and so imposes no energy penalty within the lattice. Since 2_1 axes have been shown to be strong contributors to the lattice cohesion energy (Filippini and Gavezzotti, 1991), this observation is not surprising.

Data sets 1(a) and (b): Organic structures compared to metallo-organic structures

Tables 7M.2 and 7M.3 show distributions for the organic and metallo-organic complexes on the CSD. The organic distribution compares reasonably well with the space group frequency table generated for organic molecules by Mighell, Himes and Rodgers (1983). Some difference in order is apparent, but the percentage change is not vast, except, perhaps, for $P2_1/c$, which has decreased in popularity by ~3%. Non-centrosymmetric space groups seem to have increased slightly in popularity.

211

Table 7.9. The top ten organic and metallo-organic space groups.

Organic			Metallo-organic			Mighell et al : organics		
Number of organic entries: 56321			Number metallo-organic entries: 54299			Number of entries: 29059		
Space Group	Nent	% of total	Space Group	Nent	% of total	Space Group	Nent	% of total
P2₁/c	18511	32.9	P2₁/c	21182	39.0	P2₁/c	10450	36.0
P1̄	8840	15.7	P1̄	12240	22.5	P1̄	3986	13.7
P2₁2₁2₁	8084	14.4	C2/c	4641	8.5	P2₁2₁2₁	3359	11.6
P2₁	5080	9.0	P2₁2₁2₁	2475	4.6	P2₁	1957	6.7
C2/c	3253	5.8	Pbca	2148	4.0	C2/c	1930	6.6
Pbca	2245	4.0	P2₁	1393	2.6	Pbca	1261	4.3
Pna2₁	875	1.6	Pnma	1038	1.9	Pnma	548	1.9
Pnma	755	1.3	Pna2₁	866	1.6	Pna2₁	513	1.8
P1	749	1.3	Cc	620	1.1	Pbcn	341	1.2
C2	648	1.2	Pbcn	593	1.1	P1	305	1.0

The top ten space groups in data sets 1(a) and (b) are shown in Table 7.9, along with the top ten space groups from Mighell and co-worker's study. The distribution of organic compounds is different from that for metallo-organics: non-centrosymmetric space groups are more heavily populated for organic compounds.

This effect can be investigated further. A summary of the ten most populated non-centrosymmetric space groups is given in Table 7.10. The relative population of non-centrosymmetric space groups can again be seen to be far higher in organic structures than metallo-organics. This is probably because many more organic crystals are formed from resolved chiral enantiomers: one very influential source of such molecules is organic natural products. Further evidence for this is provided by Table 7.10. The chiral space groups are relatively more popular for organic molecules than for metallo-organic molecules, whereas the popularity of achiral non-centrosymmetric space groups (in Table 7.10 these constitute all space groups containing glide planes) is not enhanced. The relative popularity of chiral space groups is enhanced by a factor of approximately three times from metallo-organics to organics. These data suggests that

Table 7.10. The top ten non-centrosymmetric space groups for organic and metallo-organic compounds in the CSD.

Organic			Metallo-organic		
Total Number of entries in subset: 56321			Total Number of entries in subset: 54299		
Space group	N_{ent}	% of subset	Space group	N_{ent}	% of subset
$P2_12_12_1$	8084	14.4	$P2_12_12_1$	2475	4.6
$P2_1$	5080	9.0	$P2_1$	1393	2.6
$Pna2_1$	875	1.6	$Pna2_1$	866	1.6
P1	749	1.3	Cc	620	1.1
C2	648	1.2	$Pca2_1$	371	0.7
Cc	495	0.9	P1	350	0.6
$Pca2_1$	476	0.8	C2	287	0.5
$P2_12_12$	337	0.6	Pc	221	0.4
$P4_12_12 + P4_32_12$	278	0.5	$P4_12_12 + P4_32_12$	194	0.4
Pc	214	0.4	Fdd2	183	0.3
Total	17236	30.6	Total	6960	12.8

the chance of obtaining a non-centrosymmetric crystal structure from an arbitrary molecule is considerably lower (nearer the overall metallo-organic rate of observation of 12.8%) than one would estimate from the full CSD.

Another possible reason for a difference in these two distributions was considered. It is possible that more metallo-organic compounds are inherently centrosymmetric, and so occupy centres of symmetry. This is discussed more fully for data set 2 in Section 7.5.3. The relative lack of enhancement of achiral non-centrosymmetric space groups tends to lead to the former conclusion.

Data set 1(c): Disordered structures on the CSD

A full distribution of disordered space groups is given in Table 7M.4. A notable feature involves space groups that possess mirror symmetry. As noted in Table 7.6, an average of 15% of structures contain some type of disorder. The average rate of disorder for the six most common space groups which contain mirror planes (Pnma, $P2_1/m$, C2/m, $Cmc2_1$, Cmca and Cmcm) is considerably higher at 31%: Cmcm and C2/m show rates

of 39.1 and 38.4% respectively. The lowest rate is observed for Cmc2$_1$ at 21.7%. This data shows that the presence of a mirror plane in a space group is more likely to be associated with a disorder.

By comparing Table 7M.1 (all structures) with Table 7M.4 (disordered structures), it is apparent that a fair proportion of high symmetry space groups contain some form of disorder. This, again, is particularly true of mirror symmetric space groups. For example, 8 of the 11 structures in Pm$\bar{3}$m, 15 of the 26 structures in Fm$\bar{3}$m and 14 of the 18 in Im$\bar{3}$m are disordered.

Data set 1(d) "Polymeric" (Catena) structures

This small subset of data (Table 7M.5), slightly surprisingly, does not seem to show much change from the distribution observed for metallo-organic complexes (there is a slight change in the relative populations, probably reflecting that not all structures in the polymeric data set are metallo-organic). A summary of the top ten space groups is given below in Table 7.11.

Table 7.11. Top ten space groups for polymeric structures

Total Number of entries in subset: 2759		
P2$_1$/c	910	33.0
P$\bar{1}$	415	15.0
C2/c	264	9.6
P2$_1$2$_1$2$_1$	190	6.9
Pbca	110	4.0
P2$_1$	107	3.9
Pnma	88	3.2
Pna2$_1$	60	2.2
C2/m	44	1.6
Cc	41	1.5

7.5 Analysis of data set 2.

7.5.1. A note on data analysis of data set 2.

In the following sections, a number of conclusions are drawn from the data that have been extracted by the methods described. One problem with a study of this kind is that the vast quantity of data can, perhaps, obscure the more subtle trends which may be seen if smaller data subsets were taken. The method described in earlier sections has been primarily designed as a *search* technique: it is hoped that it can be integrated into the CSD software, so that a user is able to create a CSD query where a specific crystallographic point group symmetry is a requirement of the search. As a result of this, it is probable that not all possible aspects of the derived data set have been analysed, because the appropriate questions have not been asked of it.

7.5.2. Preliminary features of data set 2.

The full point group - space group matrix for all systems, regardless of the number of residues present is shown in Table 7M.6. For structures with $N_{res} > 1$, only the residue with the "highest" point group is counted in the point group table. (For example, if a structure contained two residues in $P\bar{1}$, with one residue sitting on a $\bar{1}$ site, then only the point group $\bar{1}$ for the space group $P\bar{1}$ would be incremented). The "highest" point group is based primarily on the highest order symmetry element that exists in the point group, rather than the order of the point group. (i.e. a 4-fold is "higher" than mmm, even though mmm has an order of 8). Some ambiguity exists as a result of this (for example, P4/n has both a 4 fold site and a $\bar{4}$ site. In these tables the $\bar{4}$ site is the "higher" order axis, even though its actual order is the same as a 4 fold axis). By breaking this table in this way, a number of features can be noted. (a) Mirror symmetric space groups do not occur without occupation of the mirror planes. (b) Two fold axes do not impose such strict criteria on packing: a 2-fold axis is occupied in around 50% of cases where such an axis is available, although this varies from space group to space group. (c) The "rate of axis occupation" increases for higher order axes: 3-folds are occupied in around 80%

of examples of 3-fold symmetric space groups. (d) Space groups with higher order axes occur very rarely. In most cases where they do, a high symmetry special position is occupied. These points are discussed more fully for structures where N_{res} = 1. In such cases no imposed ambiguity exists as a result of only showing the "highest" point group. A separate analysis has been used for a small subset of space groups with N_{res} = 2, where combinations of point group occupations have been analysed.

7.5.3 Differences between Organic and Metallo-organic compounds.

Tables 7M.7 and 7M.8 show full point group - space group matrices for organic and metallo-organic compounds respectively. Naturally, the differences in space group frequency observed for data set 1 again occur in these tables. The benefit of breaking these structures down by point group is that it gives an estimate of the frequency of occurence of specific point groups as well as space groups. The frequencies of individual point groups are summarised below in Table 7.12. As is apparent from this table, special positions are occupied more frequently in metallo-organic complexes. In section 7.4, it was noted that chiral space groups occur more frequently for organic compounds. This was ascribed to a "sociological" effect: chemists wish to design chiral organic compounds, and so such crystals are more common in the organic subset. Table 7.12 probably shows this effect too. The organic subset has a larger percentage of chiral space groups with molecules in general positions, and so the relative percentage of special position occupation is reduced. An alternative rationale is the reverse of this argument: more organometallic molecules have inherent symmetry, and so more occupy special positions in achiral space groups. The former argument is favoured, because of the point noted earlier, namely that achiral non-centrosymmetric space groups are no more popular than centrosymmetric ones.

Table 7.12. Point group frequencies for organic and metallo-organic structures

	Organic		Organometallic			Organic		Organometallic	
	Nent in subset: 40228		Nent in subset 33540			Nent in subset: 40228		Nent in subset 33540	
Point group	Nent	% of subset	Nent	% of subset	Point group	Nent	% of subset	Nent	% of subset
1	33712	83.8	23467	70.0	422	4	<0.1	3	<0.1
-1	3271	8.1	5583	16.6	4mm	1	<0.1	3	<0.1
2	1757	4.4	2586	7.7	-42m	16	<0.1	16	<0.1
m	843	2.1	940	2.8	4/mmm	2	<0.1	1	<0.1
2/m	112	0.3	173	0.5	6	0	0.0	0	0.0
222	28	<0.1	49	0.1	-6	46	0.1	18	<0.1
mm2	61	0.2	74	0.2	6/m	0	0.0	0	0.0
mmm	10	<0.1	7	<0.1	622	0	0.0	0	0.0
3	119	0.3	271	0.8	6mm	0	0.0	0	0.0
-3	52	0.1	93	0.3	-62m	0	0.0	0	0.0
32	16	<0.1	50	0.1	6/mmm	0	0.0	0	0.0
3m	22	<0.1	17	<0.1	23	7	<0.1	8	<0.1
-3m	5	<0.1	0	0.0	m-3	1	<0.1	0	0.0
4	9	<0.1	15	<0.1	432	0	0.0	0	0.0
-4	119	0.3	153	0.5	-43m	4	<0.1	5	<0.1
4/m	5	<0.1	8	<0.1	m-3m	6	<0.1	0	0.0

7.5.4. The effect of charge on space group popularity.

Data set 2(d): Charged systems

Table 7M.9 shows the full point group - space group matrix for subset 2(d). Charged residues are present in some 23% of database entries, a fairly sizeable subset. It is apparent, from this table, that P2₁/c is the chosen space group in 36% of cases. This compares to 38% for all structures and 39% for non-ionic systems. Charge seems to have little effect on the overall selection of space group, although little conclusion can be drawn from analysing charge alone. In Section 7.5.5, the influence of numbers of

residues in the asymmetric unit is explored. In charged systems $N_{res} \geq 2$, and this may influence the overall results.

Data set 2(e): "Zwitterionic" systems.

A similar analysis to that for charged systems was carried out using a dataset 2(e) (Zwitterionic complexes). The results are shown in Table 7M.10. One problem in this subset that has to be considered prior to data analysis is the definition of a zwitterionic system. In a literal sense this means a molecule that possesses both positive and negative charges. The database, in using this considers any molecule that has a formally ionic bond within the molecule to lie within this category, and thus any statistics in this sub-group will contain molecules of that type (For example, a formally charged transition metal coordinating to a pendant arm macrocycle with a formally negatively charged ligating oxygen would be in the data set). The real interest in zwitterionic molecules lies in molecules with a large charge separation which creates a potentially large dipole moment. Unfortunatly the dataset does not only represent these examples.

The most interesting feature of this data set is that the percentage of molecules in the space group $P2_1/c$ increases to 46%, as compared to 40% in the equivalent search for the full database. The percentage of space group $P\bar{1}$ is 13% as compared to 16% for the full set.

7.5.5. Effects on space group frequencies as a function of N_{res}

All structures in data set 2

Table 7.13 shows the variation of space group frequency as a function of the number of residues in the asymmetric unit for the ten most popular space groups in each case. We can see from these that the most popular space groups are commonly utilised in all

Table 7.13. The effect of numbers of residues on space group frequency

Nres = 1			Nres = 2			Nres = 3			Nres = 4		
S.G.	%	N_{ent}	S.G.	%	N_{ent}	S.G	%	N_{ent}	S.G	%	N_{ent}
$P2_1/c$	41.0	19423	$P2_1/c$	35.5	5972	$P2_1/c$	33.2	1614	$P2_1/c$	29.7	724
$P\bar{1}$	16.2	7672	$P\bar{1}$	23.0	3871	$P\bar{1}$	24.7	1200	$P\bar{1}$	25.6	626
$P2_12_12_1$	11.8	5618	$P2_12_12_1$	8.5	1430	$P2_12_12_1$	8.3	401	$P2_1$	8.0	195
$C2/c$	6.9	3258	$P2_1$	7.4	1245	$C2/c$	6.5	317	$P2_12_12_1$	7.3	179
$P2_1$	5.8	2742	$C2/c$	6.3	1050	$P2_1$	6.1	296	$C2/c$	5.6	137
Pbca	5.1	2396	Pbca	3.4	569	Pbca	3.6	174	P1	3.1	75
$Pna2_1$	1.6	772	$Pna2_1$	1.6	273	$Pna2_1$	1.8	87	Pbca	2.9	71
Pnma	1.4	668	P1	1.3	221	Pnma	1.7	81	$Pna2_1$	1.9	46
Pbcn	1.0	461	$Pca2_1$	1.2	199	Cc	1.4	69	Cc	1.7	42
Cc	0.9	412	Cc	1.1	184	P1	1.2	57	C2	1.4	34
Total	91.7		Total	89.3		Total	88.5		Total	87.2	

systems, regardless of the number of residues, although the ranking of space groups does vary slightly.

Several features should be noted. There is a rise in the popularity of $P\bar{1}$. $P2_1/c$ remains the most popular space group followed by $P\bar{1}$ in all cases, but the dominance of $P2_1/c$ is significantly reduced. The sum of the percentages of $P2_1/c$ and $P\bar{1}$ remains approximately constant across the N_{res} range at a value between 55.3% and 58.5% which suggests that the decrease in the popularity of $P2_1/c$ is accounted for by the increase in the popularity of $P\bar{1}$. It is quite possible that these structures in $P\bar{1}$, with $N_{res} > 1$, are pseudo-$P2_1/c$. This point of view is supported in a paper by Scaringe (1991).

In order to establish the direct effect of having more than one independent molecule in the asymmetric unit , a study was carried out on a sample where the number of residues equalled Z', the number of independent molecules in the asymmetric unit. This, of course, means that all the residues in the asymmetric unit are alike in a chemical sense (although not necessarily structurally identical). The results are displayed in Table 7.14.

As is quite apparent from this that the effect described above (i.e. $P2_1/c$ decreasing in frequency as $P\bar{1}$ increases) appears to be more accentuated as Z' increases. Z' is not frequently above two. This was noted in a recent publication by Brock and Dunitz (1994). They explained the relative infrequency of Z' > 2 as an artefact of the difficulty of solution and refinement of structures that show this property. They note that there is no entropy penalty from having Z' > 1 because the entropy of perfectly ordered crystals is zero at absolute zero (Pauling and Tolman, 1925). Brock and Dunitz consider the vibrational entropy of a crystal to be the same in systems with any Z'. It is, however, possible that unknown kinetic effects influence the process of crystallisation. The kinetics of crystal formation is an area that is still very poorly understood.

Table 7.14. The effect of Independant molecules on Space group frequency.

Nres = 1 Z' = 1			Nres = 2 Z' = 2			Nres = 3 Z'=3		
S. G.	%	Nent	S. G.	%	Nent	S. G.	%	Nent
$P2_1/c$	43.9	16531	$P2_1/c$	31.5	1276	$P2_1/c$	28.6	52
$P\bar{1}$	16.4	6190	$P\bar{1}$	28.7	1163	$P\bar{1}$	27.5	50
$P2_12_12_1$	14.7	5529	$P2_1$	14.6	589	$P2_1$	9.3	17
$P2_1$	7.1	2679	$P2_12_12_1$	7.9	318	$P2_12_12_1$	8.8	16
Pbca	5.5	2078	P1	3.7	148	C2/c	6.0	11
$Pna2_1$	2.0	760	$Pca2_1$	2.8	112	P1	5.5	10
Cc	1.1	404	$Pna2_1$	2.4	97	Cc	5.0	9
$Pca2_1$	0.9	323	Pbca	2.4	97	Pbca	2.2	4
C2	0.6	210	C2/c	1.5	59	$P3_1$	2.2	4
P1	0.4	157	Cc	1.0	42	$Pna2_1$	1.7	3
Total	92.6		Total	96.5		Total	96.8	

The comparable results for "non-ionic" structures are shown in Table 7.15. The decrease in popularity of $P2_1/c$ is very apparent in this subset of data. Of all the groups of data studied this is the only example in which $P\bar{1}$ becomes more popular than $P2_1/c$, although the trend seems to be paralleled in the case of varying Z', but to a lesser extent. This suggests that there is a definite correlation between the number of residues in the asymmetric unit and space group selection.

Table 7.15. The variation of space group frequencies with N_{res} for "Non-ionic" Systems

$N_{res} =$	1		$N_{res} =$	2		$N_{res} =$	3		$N_{res} =$	4	
S. G.	%	N_{ent}	S. G.	%	N_{ent}	S. G.	%	N_{ent}	S. G.	%	N_{ent}
$P2_1/c$	41.0	19423	$P2_1/c$	32.4	3091	$P\bar{1}$	28.9	442	$P\bar{1}$	29.4	175
$P\bar{1}$	16.2	7672	$P\bar{1}$	27.6	2633	$P2_1/c$	27.9	427	$P2_1/c$	23.3	139
$P2_12_12_1$	11.8	5618	$P2_1$	9.6	910	$P2_12_12_1$	10.1	155	$P2_1$	12.6	75
$C2/c$	6.9	3258	$P2_12_12_1$	8.6	815	$P2_1$	9.7	149	$P2_12_12_1$	10.6	63
$P2_1$	5.8	2742	$C2/c$	5.3	504	$C2/c$	5.7	87	P1	6.4	38
Pbca	5.1	2396	Pbca	2.4	227	P1	2.4	36	$C2/c$	4.2	25
$Pna2_1$	1.6	772	P1	2.0	191	Pbca	2.2	33	C2	2.5	15
Pnma	1.4	668	$Pna2_1$	1.5	138	Cc	1.5	23	Cc	2.0	12
Pbcn	1.0	461	$Pca2_1$	1.4	131	C2	1.4	22	Pbca	1.0	6
Cc	0.9	412	C2	1.2	117	$Pna2_1$	1.2	19	$Pna2_1$	1.0	6
Total	91.7		Total	92.0		Total	91.0		Total	93.0	

Data set 2(d). Charged systems

The full table of how charged system frequencies behave with residue is shown in Table 7.16 below. It must be noted that in cases where $N_{res} > 2$, not all residues necessarily have to be charged, since the bit screen 55 in the CSD only implies that a charged moiety is present within the structure. The distribution follows a similar pattern as the full database with the $P2_1/c:P\bar{1}$ ratio decreasing as N_{res} increases; The case of $N_{res} = 2$ for ionic systems compares well with the case of $N_{res} = 1$ for non-ionic

systems. The drop off in favourability of $P2_1/c$ is less marked than in non-ionic systems. These two features can be explained by considering an ionic system to represent a single residue non-ionic system (i.e. if the charges hold the two residues in relatively close proximity to one another, one might regard the ion pair as a single residue of balanced charge). The reason for the less marked fall in favourability could be due to examples where all residues possess charge, which may compare to the non-ionic $N_{res} = 1$ case, although this argument is tentative and unproven. It is probably only true for some examples which create the apparent trend.

Zwitterionic structures cannot be compared directly with charged systems since individual charged residues cannot exist in a crystal without a counter-ion. The closest comparison between zwitterions and charged complexes is between all zwitterions (the size of data set 2(e) is too small too be broken down by residue) and charged systems with $N_{res} = 2$. Zwitterions favour $P2_1/c$ more strongly than other types of structural residue.

Table 7.16. Effect of residue on the frequency of systems with charged species.

Nres =2			Nres = 3			Nres = 4		
S. G.	%	N_{ent}	S. G.	%	N_{ent}	S. G.	%	N_{ent}
$P2_1/c$	39.6	2881	$P2_1/c$	35.7	1188	$P2_1/c$	31.7	585
$P\bar{1}$	17.0	1240	$P\bar{1}$	22.8	758	$P\bar{1}$	24.4	451
$P2_12_12_1$	8.5	615	$P2_12_12_1$	7.4	246	$P2_1$	6.6	121
C2/c	7.5	546	C2/c	6.9	230	$P2_12_12_1$	6.3	116
Pbca	4.7	343	$P2_1$	4.4	147	C2/c	6.1	112
$P2_1$	4.6	335	Pbca	4.2	141	Pbca	3.5	65
$Pna2_1$	1.9	135	Pnma	2.2	73	$Pna2_1$	2.2	40
Pnma	1.7	127	$Pna2_1$	2.0	68	P1	2.0	37
Cc	1.5	110	Cc	1.4	46	Cc	1.6	30
$Pca2_1$	0.9	68	Pbcn	1.2	40	Pnma	1.5	28
Total	87.9		Total	88.2		Total	85.9	

7.5.6. The effect on the numbers of centrosymmetric and non-centrosymmetric space groups by varying numbers of residues.

Table 7.17 shows the ratio of centrosymmetric to non-centrosymmetric space groups for a range of values of N_{res}, for all of data set 2. This ratio is relatively constant across the range of values. The correlation coefficient between N_{res} and the ratio was calculated as -0.76.

A similar analysis is shown for the uncharged subset 2(c) in Table 7.18. The decrease in the centrosymmetric:non-centrosymmetric ratio for uncharged systems is more marked than in the case where all structures are considered. In this case, the correlation coefficient between N_{res} and the ratio was calculated as -0.96. Caution has to be taken in interpreting this result. In the higher values of N_{res} the statistics may well be biased by enatiomerically resolved compounds which, by necessity, have to crystallise in chiral space groups. Molecules of this type are frequently solvated in the crystalline form[◊] and so can have high values of N_{res} on average as compared to the bulk of compounds. The trend suggested, however, is that high numbers of discrete residues within the lattice may encourage non-centrosymmetric space groups.

Table 7.17. Popularity of centrosymmetric versus non-centrosymmetric space groups for all structures as a function of N_{res}

N_{res}	Centrosymmetric	Non-centrosymmetric	Ratio
1	35343	12093	2.9:1
2	12479	4326	2.9:1
3	3655	1204	3.0:1
4	1717	725	2.4:1
5	627	254	2.5:1

[◊] This is particularly true of proteins

Table 7.18. Popularity of centrosymmetric space groups versus non-centrosymmetric space groups for non-ionic and zwitterionic systems as a function of N_{res}

N_{res}	Centrosymmetric	Non-centrosymmetric	Ratio
1	35343	12093	2.9:1
2	6827	2705	2.5:1
3	1048	481	2.2:1
4	360	236	1.5:1
5	94	58	1.6:1

7.6. The effect of symmetry on space group occupation

The full point group - space group matrix for structures with $N_{res} = 1$ is shown in Table 7M.11. It is easier to draw conclusions from this table, rather than the full table (Table 7M.1) because the symmetry shown is not effected by the hierarchical order defined for displaying these systems. Later sections deal with cases where $N_{res} = 2$.

7.6.1 The effect of mirror planes.

The overall statistics for space group frequencies in Table 7M.11 are changed somewhat when one considers only crystals in which no special position is occupied. In this sub-set, space groups possessing mirror planes do not occur. This has been previously noted by a number of authors (Kitiagorodski, 1961, Wilson, 1988, Brock and Dunitz, 1994). A summary of the more common mirror symmetric space groups is given in Table 7.19 to emphasize this point. This rule is also true for the higher symmetry point groups mm2 and mmm where space groups only occur if a molecule occupies the symmetry site. This applies more generally for all structures. In all structures that exist in mirror symmetric space groups at least one residue appears to occupy a mirror symmetric site. (Table 7M.1)

Table 7.19. Point group versus space group for the more common mirror symmetric space groups. All structures have $N_{res} = 1$.

Space group	N_{ent}	Point group					
		1	$\bar{1}$	2	m	2/m	mm2
Pnma	668	0	0	-	668	-	-
P2$_1$/m	249	0	0	-	249	-	-
C2/m	117	0	0	0	31	86	-
Cmc2$_1$	73	0	-	-	73	-	-
Cmca	62	0	0	0	23	39	-
Cmcm	45	0	0	0	0	0	45
Pbcm	36	0	0	0	36	-	-

In the full study of molecules (Table 7M.1) of all site symmetries only three examples were originally observed where the published space group contained an unoccupied mirror plane. In two of these cases, refcodes VERNOV (Wang *et al*, 1989) and KULZOG (Gomez *et al*, 1992) there were clear errors in the ascribed space groups. In VERNOV the author had later amended the symmetry operations that were stored on the database, but the space group stored had not been corrected. In KULZOG an obvious error could be detected by analysing the closest contacts in the system. The closest "contact" stored was 0.139Å between two symmetry related H atoms, which is quite clearly incorrect. These errors may be in the primary literature. More likely is that the structures have been incorrectly entered in the CSD. The third example is DACTYL (Koski *et al*, 1975). (the structure of deutero-acetylene at low temperature, space group Acam). This appeared to have an intermolecular D-D contact of 0.99Å. Investigation showed that the stored symmetry elements for this structure were incorrect.

Table 7.19 shows that the more popular space groups that possess both mirror planes and 2-fold axes are C-centred. The more common space groups do not have 2-fold axes parallel to the mirror plane. A final point is that mirror symmetric space groups tend to

have *independent* inversion symmetry (i.e. an inversion centre that does not lie in the mirror plane). The only exception amongst the popular space groups is $Cmc2_1$.

7.6.2. The effect of 2-fold axes on space group frequency.

Table 7.20 shows the relative popularity of space groups with 2-fold axes for molecules with one residue. This suggests that 2-fold axes are not as influential as Kitiagorodskii originally believed. He postulated that the space group C2/c was unsuitable for molecules that lacked a centre of symmetry. Inspection of a number of the examples in this space group that possess molecules in general positions soon shows that this is not the case. It is possible that some relationship exists between the packing of this set of structures. One wonders whether these structures undergo a form of strong dimeric packing around the inversion centre. Scaringe (1991) analysed the frequency of space groups with molecules in general positions only. He concluded that both mirror planes and 2-fold axes are inherently unfavourable for close-packing from this analysis, although he does admit that this is only true for some 94% of the data surveyed. He did not analyse the data for all space groups. This, for the purposes that Scaringe required, was an adequate conclusion[†] although considerable interest should exist as to why the 6% can pack in 2-fold symmetric space-groups: in any distribution, it is frequently the points at the edge of the distribution and the outliers that purvey the largest amount of information.

Table 7.20 shows that the symmetry sites are not always occupied within these space groups. Generally 2-fold axes are occupied in 45.6% of occurences of 2-fold symmetric space groups. Several reasons have been suggested for structures which pack around 2-fold axes. The occurence of *chiral* 2-fold symmetric space groups is more common in the protein data bank than in the CSD (Padmaja *et al*, 1990). Brock (1994) has

---

[†] The purpose of the study in this case was to reduce the complexity of the problem of establishing, by energetic minimisation, the space group that a particular molecule occupies. If certain plane group layers can be disregarded, the size of the potential hypersurface is considerably reduced. (The number of dimensions is curtailed).

anecdotally attributed this to the packing of homochiral "bean-shaped" molecules; she notes that a large number of molecules on the protein data bank (Bernstein et al, 1977) follow this trend. Another possibility is as described in the introduction of this Chapter, that certain inter-molecular forces dominate within the lattice, forcing the alignment of molecules about a 2-fold axis. Inter-molecular non-bonding contacts have been shown to influence the orientation of molecules in the crystalline phase an example of such an interaction is the hydrogen bond.(Etter, 1990, has reviewed this subject). Unusual space group distributions have been noted for mono-alcohols, as a result of the favourability of packing around particular symmetry elements to maximise H...OH interactions (Brock and Duncan, 1994). Wilson (1988) has postulated that centrosymmetric structures can be described by considering such structures as pairs of molecules formed into inversion dimers. Perhaps a similar idea can be used to explain packing around a two fold axis, in cases where local electrostatic forces dominate.

Table 7.20. The more common two fold symmetric space groups with $N_{res} = 1$

Space Group	N_{ent}	Point group				
		1	$\bar{1}$	2	m	2/m
C2	282	215	-	67	-	-
C2/m	117	0	0	0	31	86
P2/c	109	45	5	59	-	-
C2/c	3258	1528	323	1407	-	-
P21212	151	89	-	62	-	-
Fdd2	203	89	-	114	-	-
Pccn	171	94	17	60	-	-
Pbcn	461	149	24	288	-	-
$P4_12_12 + P4_32_12$	209	91	-	118	-	-

7.6.3 Higher order proper rotation axes.

Table 7M.11 shows that space groups with molecules in general positions become very infrequent as the highest order rotation axis increases. Trigonal space groups do occur with molecules in general positions*, although they are more frequent if the special positions are occupied. The occurence of tetragonal space groups with 4-fold axes is considerably less common, particularly with molecules in general positions. One effective example occurs in I4/m, with the molecule occupying the mirror plane, seven examples in I4 and one in P4/n. Closer inspection of these refcodes was carried out and it was found that one example was clearly incorrect: the refcode PHXTCN (Noerenberg *et al*, 1977) is reported to be in I4. Inspection using the XP program on a SHELXL format file indicated that the compound had intermolecular nitrile-nitrile "contacts" of between 1.4 and 1.6Å. The crystals may have been twinned. A second, slightly dubious, structure was detected for FORPOR (Shinomoto *et al*, 1987). This compound was reported with a residual of 3%. In normal circumstances this would be relatively good, but the structure contains a uranium atom which accounts for nearly half the electrons in the asymmetric unit. The database also commented in the "remarks" record on a short H-H contact. As a result of this one atom had been deleted by the database staff. In a third refcode, PHMTPT (Bassett *et al*, 1980), the authors, in original paper, reported slightly different numerical values for the a and b cell parameters, despite the obvious requirement that they should be identical, so some ambiguity exists with regard to the space group in this case. In three examples (ALAALA (Fletterick *et al*, 1971), BABFUF (Norrestam *et al*, 1981), KEXNUW (Konda *et al*, 1990) extensive non-bonded networks were observed using XP with very short O...H contacts (<2Å) which appeared to hold the lattice together. One compound in this group (ALAALA) is also Zwitterionic. The remaining three examples (EMPCRU (Wartchow and Berthold, 1977),

---

* A high number of crystals occur in the space group $R\bar{3}$ with molecules lieing on general positions. It is quite possible that these are examples where a 3-fold twinning (A drilling, from the German) has been overlooked during the refinement process. This argument is given added weight when one considers that the sub-group that results from the removal of the 3-fold axis is P-1, the second most common space group.

Table 7.21. Frequencies of special position occupation for rotation axes

2-folds[a]			
Space group	Nent	Nocc	% occupancy
C2/c	3258	1407	43.2
C2/m	117	86	73.5
$P2_12_12$	151	62	41.1
$C222_1$	44	27	61.3
Fdd2	203	114	56.2
Pccn	171	60	35.1
Pbcn	461	288	62.4
$P4_12_112 + P4_32_112$	209	118	56.4
$P3_121 + P3_221$	51	28	54.9
Total	4665	2190	46.9

3-folds[b]			
Space group	Nent	Nocc	% occupancy
R3	69	44	63.8
$P\bar{3}$	21	18	85.7
$R\bar{3}$	172	97	56.3
R3c	59	46	78.0
$P6_3$	21	17	81.0
$P6_3/m$	30	23	76.7
Pa3	21	21	100.0
Total	393	266	67.7

4-folds[c]			
Space group	Nent	Nocc	% occupancy
P4/n	7	5	71.4
P4nc	3	3	100.0
I4/m	12	11	91.7
P4/nnc	6	4	66.7
Total	28	23	82.1

6-folds
No examples were found

[a] 2/m sites occupation was included as 2-fold occupation

[b] $\bar{3}$ sites, $\bar{3}2$ sites and $\bar{6}$ (3/m)sites were all included as 3-fold occupation since they all contain inherent 3-fold axes

[c] 4/m sites were included as 4-folds since they contain an inherent 4 fold axis.

MBYZMN (Sheldrick *et al*, 1978), SIKFOH (Pramod *et al*, 1990)) have no obvious inter-molecular contacts within the lattice and no obvious errors. Two points of

similarity occur in these cases; They are all reported in I4 and they are all examples of chiral molecules. This result is a little puzzling. Kitiagorodskii argues that it is impossible to close-pack efficiently in a tetragonal space group that possesses a 4-fold axis. What these examples must represent are systems where either the energetic minimum does not coincide with the point of closest packing, or where the molecular shape "happens" to cause the closest packed arrangement to be around a 4-fold axis.

No examples of space groups with 6-fold axes are observed in the tables even with the axis occupied. This may be due to the rarity of molecules with high order point groups possessing 6-fold axes in the crystallographic database, which does not deal with purely inorganic molecules. The lack of examples of molecules in general positions in this space group is probably due to the difficulty of packing molecules efficiently about such an axis.

A similar result is observed for systems where the number of residues in the asymmetric unit is greater than 1. The frequency of space group P3 increases from one incidence in the case of $N_{res} = 1$ to 10 occurences for $N_{res} = 3$.This may be an artefact of missed symmetry, possible super-cells being R3, or $R\bar{3}$, with an added molecular symmetry element, both of which are more common. All these 10 examples have Z' equal to 1. Closer inspection revealed that in all cases there were 3 molecules in the asymmetric unit, and all three lay on 3-fold axes. The structures all appeared to be at least pseudo R-centred (in some cases there may be missed symmetry).

4-fold axes are rarely unoccupied when N_{res} is greater than one. There are only four examples where this is the case. Analysing them individually revealed that in one case there is an extensive hydrogen bonding network. A second system appeared to be a layer structure consisting of molecular layers and solvent layers. A third had a solvent molecule on the special position that was not included on the database. The fourth was mirror symmetric ionic system but no apparent reason lent itself to explaining its packing around a four-fold axis.

7.6.4. Higher order improper rotation axes.

Care has to be taken when comparing improper rotation axes. The Hermann-Mauguin notation belies the fact that a 3-fold axis has a lower multiplicity than a $\bar{3}$ axis. Similarily a $\bar{6}$ axis has an equal multiplicity to a $\bar{3}$ axis. Another important difference is that a $\bar{3}$ axis is inherently centrosymmetric compared to a $\bar{4}$ or $\bar{6}$. Table 7M.11 shows that, in the single residue case, space groups containing $\bar{3}$, $\bar{4}$ and $\bar{6}$ axes do occur, and that the occupation of these axes is not necessarily a pre-requisite. In $\bar{3}$ symmetric space groups it is common for the 3-fold axis to be occupied. Earlier evidence with respect to P$\bar{1}$ and P2$_1$/c indicates that it is favourable for molecules to pack around inversion centres, thus the $\bar{3}$ axis is chosen in preference to a normal 3-fold in an alternative space group that lacks the added inversion symmetry. $\bar{3}$ axis space groups are about as common as $\bar{4}$ space groups. 133 examples of I4$_1$/a occur with 57 examples where the molecules occupy general positions in the single residue case. The few examples of hexagonal space groups include some 29 with $\bar{6}$ axes in the single residue case, all in space group P6$_3$/m. These examples always have at least m symmetry (7) and usually display the full $\bar{6}$ symmetry. This has to be the case due to the inherent m symmetry that a $\bar{6}$ axis contains. Without occupation of at least the mirror plane it would be impossible to satisfy necessary spatial or electrostatic arguments.

7.6.5 Higher order space groups.

The occurrence of cubic space groups is relatively rare in the CSD. In all examples, regardless of the number of residues (Table 7M.1), at least one residue occupies a special position. The distribution of cubic space group frequencies is highly skewed towards two space groups, Pa$\bar{3}$ and P2$_1$3, the latter being the inversion sub-group of the former. These two space groups are the lowest symmetry cubic systems, the highest symmetry being $\bar{3}$ in Pa$\bar{3}$ and 3 in P2$_1$3. One surprising element of analysing the data *via* the number of residues is that some twenty examples occur in Pa$\bar{3}$ with $N_{res} = 2$, a comparable number to those with $N_{res} = 1$. Also, the incidence of P2$_1$3 appears to be

more common with $N_{res} = 2$. Inspecting these refcodes individually revealed that they were all charged systems and several were iso-structural with one another. Most of the systems have both residues on the 3-fold axis, although this is not exclusively the case. One residue always lies on the 3-folds as the tables show. It is interesting that these charged systems occupy this space group. The comparable subset in $Pa\bar{3}$ (i.e. $N_{res} = 2$) shows a somewhat different pattern. Not all of the systems in this space group were charged, and those that were charged usually had a centrosymmetric ion on the inversion centre. Only 2 examples out of 20 did not coincide with this trend. One of them, BAICTC (Trop *et al*, 1980), had a centric ion in the lattice, it just did not lie on the centre of symmetry. The other example, MAMCOB (Calderazzo *et al*, 1981), was more puzzling, since it did not seem to follow the pattern that was observed in the other 18 examples. It would be interesting to look in depth at the charge distributions in these compounds to try to rationalise the differences. The data certainly suggests that high symmetry non-centrosymmetric ionic pairs are much more likely to crystallise in $P2_13$ as opposed to $Pa\bar{3}$, although it has to be noted that very few data are available. These conclusions were also noted by Brock and Dunitz in their analysis based on Z'.

7.7. The frequency of occupation of more than one symmetry site.

In certain space groups it is possible for more than one symmetry site to be occupied. An example of such a space group is $P\bar{1}$ where there are eight independent sets of inversion centres[†]. In this study it is impossible to discern which operator is occupied but it is possible to analyse how often more than one residue occupies a symmetry site within structures. A program has been written to analyse the workfile in this way and some generalisations for the more common space groups will be presented here. A few statistics for rarer groups will also be highlighted.

7.7.1. Inversion centres.

[†] there are 3 pairs in the faces, 3 sets of four on the edges 1 set of eight for the corners and a further single inversion centre located at the centre of the cell. This information is described in *International Tables for Crystallography* (1992) by the Wyckoff positions.

Inversion centres occur in most of the more common space groups, particularly P2₁/c, P1̄, C2/c and Pbca, four of the six most popular space groups. Of these four, three (P2₁/c, P1̄ and Pbca) can have special positions with inversion only, C2/c also has 2-fold symmetry, and so only the former three space groups can be considered together. Below, in Table 7.22, the relative popularity of usage of more than one inversion centre is shown. These data are compared to cases where only one or no element of symmetry is used for structures with Nres = 2 for these 3 space groups.

The most important feature of Table 7.22 is that the space group Pbca does not show any examples where both inversion centres are occupied, whereas P1̄ shows this situation in a large number of examples. The trend follows the number of Wyckoff positions of inversion symmetry available in the lattice. The most probable reason for Pbca never having all of its special positions occupied is due to their distribution in the lattice. In Pbca the orthorhombic symmetry constrains each Wyckoff position's location within the lattice relative to one another due to the angular restraint of orthorhombic symmetry. If two different residues both occupy symmetry sites then all of the Wyckoff positions are occupied in Pbca and the packing is unfavourable; the special positions force 4-fold layer coordination. Kitiagorodski (1961) has stressed the unfavourable nature of this as compared to the optimal 6-fold coordination. This situation does not occur as much in P2₁/c due to the flexibility of the monoclinic β angle and because not all inversion centres have to be occupied. The β angle can change to allow repositioning of certain symmetry sites relative to others. In the triclinic system there is complete angular freedom.

Table 7.22. The relative popularity of point group occupation for space groups with only inversion symmetric special positions. N_{pos} refers to the number of independent Wycoff positions of symmetry $\bar{1}$

Space group	N_{ent}	N_{pos}	1,1	$\bar{1},1$	$\bar{1},\bar{1}$
P$\bar{1}$	3871	8	2733	730	408
P2$_1$/c	5972	4	4710	1038	224
Pbca	569	2	504	65	0

7.7.2. The usage of 2-fold axes in structures with $N_{res} > 1$

The most common 2-fold symmetric space group is C2/c. This space group also has inversion symmetry and all combinations of position occupation occur. Table 7.23 shows that inversion symmetry is only frequently used by structures in C2/c and Pccn. Use of two available inversion centres in 2-fold symmetric space groups is not very common. This result would be expected considering the previous points highlighted for structures having only inversion symmetry. This pattern is also true for all the examples in data set 2 with $N_{res} > 1$, and again seems to be correlated to the number of independent inversion centres within the lattice. Only structures where more than two independent inversion centres exist, utilise more than one in any individual structure. The frequencies do not seem to show any particular correlations to the number of independent 2-fold axes, but this is to be expected. It is quite possible that more than one molecule can occupy the same 2-fold axis which adds an extra dimension to this problem. Of course, these statistics are skewed by the fact that the two residues do not necessarily have to have 2-fold symmetry. We can draw the conclusion, however, that if two residues are present in the unit cell in a 2-fold symmetric space group they do not both have to occupy Wyckoff positions. This is the case except for the space group Pnna. In this space group both residues tend to be on a 2-fold axis. Inspection of this small set of entries suggested that the majority of compounds had both independent axes occupied, although entries with single atom counter-ions tended to utilise only one axis.

Table 7.23. The more common 2-fold symmetric space groups with $N_{res} = 2$.

Space Group	Nres	$N_{pos}(\bar{1})$	$N_{pos}(2)$	1,1	$\bar{1}$,1	2,1	$\bar{1}$,$\bar{1}$	2,$\bar{1}$	2,2
C2/c	1050	4	1	313	79	373	18	161	106
C2	151	-	2	84	-	38	-	-	29
Pbcn	111	2	1	27	0	54	0	6	24
P2/c	103	4	2	22	0	13	2	5	61
P2₁2₁2	71	-	2	27	-	29	-	-	15
Pccn	46	2	2	9	4	12	0	2	20
Fdd2	36	-	1	15	-	17	-	-	4
C222₁	29	-	2	11	-	5	-	-	13
Pnna	21	2	2	1	0	0	0	0	20

7.8. Discussion

The popularity of space groups for molecular crystals is an area which has been well studied by Kitiagorodski (1966). He uses close-packing principles to explain why certain symmetry elements are unfavourable in crystal packing.

A simple electrostatic argument explains why mirror planes are exclusive in packing; in order to pack around a mirror plane a molecule has to be aligned opposing its mirror image. In this hypothetical situation any localised bonding dipoles in the molecules will be aligned against one another creating an unfavourable interaction relative to some other molecular orientation.

The electrostatic argument can also be used to explain why lattice centering appears to be a good structural feature in combination with mirror planes. Consider the simple space group Pm (a very unpopular space group). Two independent mirror planes are present in this space group. One lies in the ac plane, the other parallel to this plane at $b = 1/2$ As a result of this configuration, a molecule in the ac plane has to be parallel with another molecule in the opposite face of the unit cell. (See Figure 7.6(a)). This forces alignment of both the molecular and bonding dipoles entirely and so is not

favourable. In the space group Cm, the C-centering induces occupation of the central mirror plane (i.e. it is not independent any more - occupation of one forces occupation of the other) with a set of like molecules that are translated by a 1/2 along the b axis. (See Figure 7.6(b)), providing a partitioning layer which could help to negate the electrostatic influence present in Pm.

Similar arguments can be used to explain why Pnma and P2₁/m are popular, and Pmna is not. Pnma and P2₁/m both have inversion centres sandwiched between the "would be independent" mirror planes, forcing them both to be occupied with molecules that have their dipoles in an antiparallel arrangement, whereas the inversion centre in Pmna lies in the mirror plane, which does not force the central mirror plane to be occupied.

Figure 7.6(a). Bad alignment in Pm.
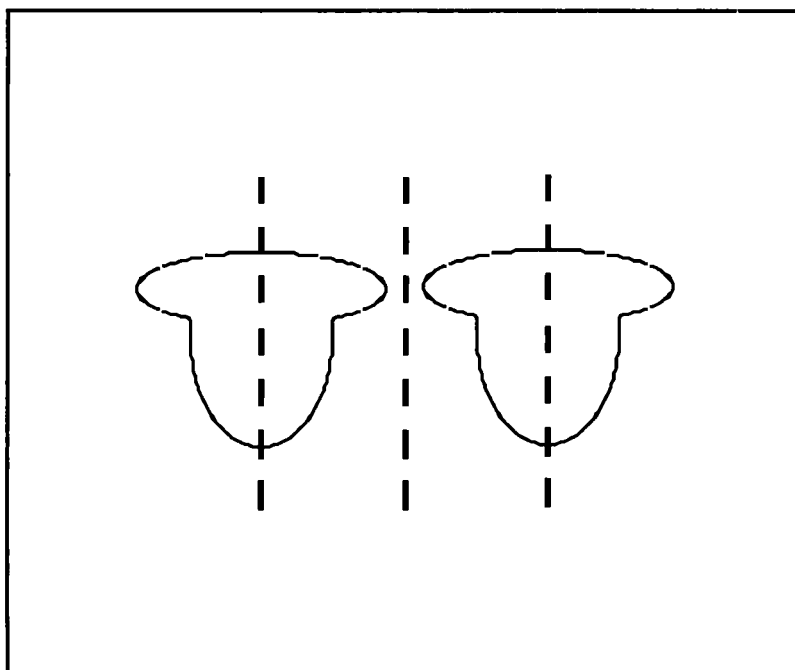
Figure 7.6(b). Improvement in Cm.
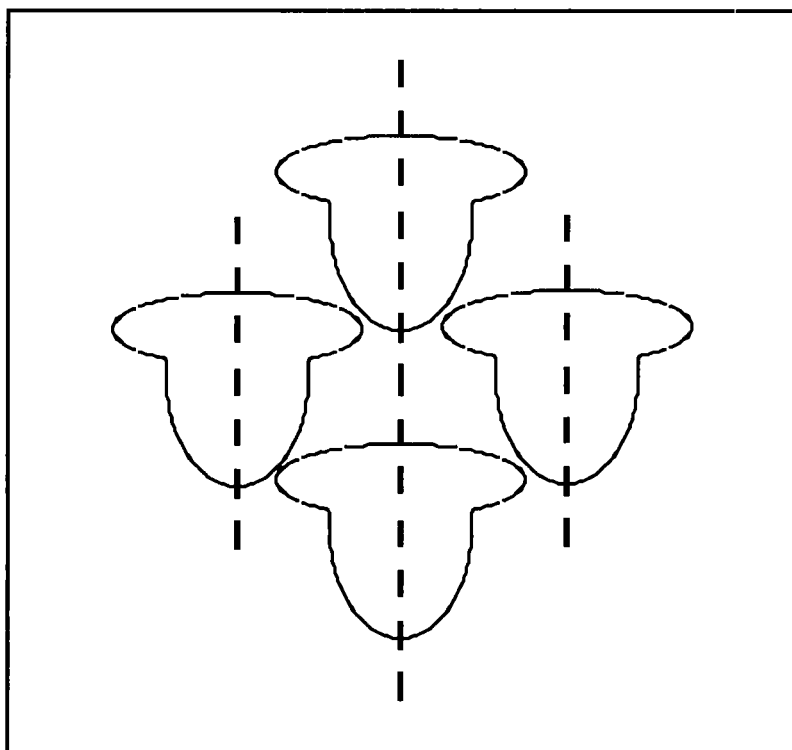
An alternative argument to the electrostatic model involves the consideration of free volume within the crystal lattice. This was the idea on which Kitiagorodskii (1961) based his theory of close-packing. The premise states "maximise density; minimise free volume". In effect, if over long ranges we can consider a molecule to be inherently attracted to another, then within a crystal the lowest energy system will pertain to a relatively high packing coefficient. This argument does not, of course, consider the relative strength of any interaction or consider the repulsions due to ionic behaviour. These, necessarily, will cause molecules to arrange themselves such that the crystal achieves a global energy minimum, but notwithstanding this, the close packed arrangement will approach a minimum of energy. In this context, Brock and Dunitz (1994) have neatly explained the bad effect that mirror planes have on packing, by describing the free volume that special positions create. If a mirror plane is unoccupied then it forces creation of a region of free volume that is equal to the volume of an infinite sheet of thickness R where R is the closest contact between moieties that are symmetry related about the m-plane. One can consider a comparable volume for other symmetry elements; A two fold axis creates an infinite "diad cylinder" of radius

R/2, an inversion centre creates a volume that is equivalent to a sphere of radius R/2. Clearly the relative volumes would indicate that mirror planes would be the most unfavourable followed by an axis followed the inversion centre. The fact that free volume is not the only important influence in crystal packing is thus indicated by the lack of popularity of P1 as a space group in preference for P$\bar{1}$. The popularity of inversion centres is also explained by the electrostatic consideration, in that they cause localised alignment of bond dipoles in a favourable way, as well as favourable alignment of the overall molecular dipole, although this latter contribution has been refuted as a dominating factor (Whitesell *et al*, 1994).

The free volume argument also neatly explains why higher order axes are unpopular. If we consider the closest contact distance between atoms to be R (so that a 2-fold axis creates a free volume equivalent to an infinite diad cylinder), then a 3-fold axis creates a volume that is equal to at least an infinite trigonal prism of edge length R, a 4-fold axis forms an infinite rectangular prism of edge length R and a 6-fold forms an infinite hexagonal prism of edge length R. Brock and Dunitz (1994) argue that the free volume is effectively a cylinder of the same radius for all rotation axes. However, the trends noted here for axis occupation would suggest that a higher penalty has to be paid for higher order axes. Indeed, the volumes of the solids described above can be easily derived and are shown in Table 7.24.

Table 7.24. Free volumes created by rotation axes

L is the length of the axis within one unit cell, R is the inter-molecular separation between symmetry related molecules.

Axis	2	3	4	6
Free Volume	$L.\left(\pi\dfrac{R^2}{4}\right)$	$L.\left(\pi\dfrac{R^2}{3}\right)$	$L.\left(\pi\dfrac{R^2}{2}\right)$	$L.\left(\pi R^2\right)$

The free volume clearly increases with order. A 6-fold axis would create four times the free volume of a 2-fold axis. In their paper Brock and Dunitz (1994) also argue that the free volume depends on the direction in which the axis lies. They note that $P3_121$ and $P3_221$ occur a number of times, whereas $P3_112$ and $P3_212$ do not. The directions of the 2-fold axes differ in the two space groups, causing the free volume created to be 75% larger for $P3_112$ than for $P3_121$. The dramatic influence this factor has on the space group frequency is apparent. Since a much larger percentage change in free volume (3-folds occupy 33% more space than 2-folds, 4-folds occupy 100% more volume than 2-folds, etc.) occurs by increasing axis order, one has to consider higher order axes as being more unfavourable than lower ones.

In conjunction with free volume, it can be noted that there seem to be "rough" sub-group super-group relationships between popular space groups. This was first postulated by Wilson (1988). For example, by removing the mirror in Pnma, the remaining possible sub-groups are $P2_12_12_1$, $P2_1/c$ and $Pna2_1$. All three are relatively popular "general symmetry" space groups. $P2_1/m$ reduces to $P2_1$ and $P\bar{1}$. C2/c can be reduced to $P2_1/c$ by removing the C-centering. What this suggests is that certain arrangements of symmetry elements are extrememly popular for a large number of molecules. Within this framework additional elements are sometimes present: in mirror symmetric space groups this is due to molecular symmetry.

Structures probably form in such a way that minimises both the free volume and the electrostatic forces. Clearly a crystal can be regarded as an energetic minimum, and minimising the free volume corresponds to maximising the contacts between atoms, which are usually attractive over a long range. This was essentially Kitagorodskii's (1961) rationâle.

A final, somewhat anecdotal, conclusion can be drawn by looking at a few examples of structures on the CSD that are reported in rare space groups. Some of these structures appear correct, and in themselves are very interesting. It is likely that each rare space

group occurs for very good specific reasons which are yet to be discovered. Unfortunately, just by looking at a few examples on the CSD, one can tell that the issue is clouded by space group mis-determinations. A search for the space group number eight (space group Cm) revealed such evidence. Several examples were noted where an inversion centre appeared to have been missed. Marsh (1996) has recently reviewed this subject, and he rightly concludes that such errors are unacceptable, particularly when an inversion centre is missed.

7.9 Conclusions

The primary aim of the work in this chapter was to provide a means for deriving the crystallographic point group for each independent residue within the asymmetric unit. This aim has been fulfilled.

The work in this chapter shows that molecular symmetry, a structural composition has influence on space group frequency. It has been confirmed that mirror planes are unsuitable for molecular packing. Structures with higher numbers of crystallographic residues have differing space group distributions: the lower symmetry space group $P\bar{1}$ becomes more favourable, the space group $P2_1/c$ less favourable, although $P2_1/c$ is generally still the most popular space group.

A similar study to the work in this chapter has very recently come to light in a publication by Zorki and co-workers (1995). They have analysed 13000 organic structures by hand to come to similar conclusions. Zorki states that there is no way of deriving this information from the CSD. This is currently true of the CSD software, but hopefully inclusion of the searching routines described in this chapter will alleviate this problem for future work.

7.10 List of microfiched tables

Table 7M.1. Space group frequencies for all compounds in Data set 1

Table 7M.2. Space group frequencies for organic compounds in Data set 1.

Table 7M.3. Space group frequencies for metallo-organic compounds in Data set 1

Table 7M.4. Space group frequencies for disordered compounds in Data set 1.

Table 7M.5. Space group frequencies for polymeric compounds in Data set 1.

Table 7M.6. Space group frequencies versus crystallographic point group for all compounds in Data set 2.

Table 7M.7. Space group frequencies versus crystallographic point group for organic compounds in Data set 2.

Table 7M.8. Space group frequencies versus crystallographic point group for metallo-organic compounds in Data set 2.

Table 7M.9. Space group frequencies versus crystallographic point group for structures containing charged residues in Data set 2.

Table 7M.10. Space group frequencies versus crystallographic point group for zwitterionic compounds in Data set 2.

Table 7M.11. Space group frequencies versus crystallographic point group for compounds with $N_{res} = 1$ in Data set 2.

7.11. References

Allen, F.H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M., Watson, D. G.; *J. Chem. Inf. Comput. Sci.* **31**, 187, 1991

Baur, W.H.; Kassner, D. *Acta cryst.* ; **B48**, 356, 1992

Baur, W. H.; Tillmanns, E.; *Acta. Cryst.,*; **B42**, 95, 1986

Belsky, V. K.; Zorkii, P. M.; *Acta. Cryst.*; **A33**, 1004, 1977

Bernstein, F.C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.,Jr.;Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. J.; *Mol. Biol.* ; **112**, 535,1977

Brock, C.P; Dunitz, J.D. *Chem. Mater.,* **6**, 1118, 1994

Brock, C. P.; Duncan, L. L.; *Chem. Mater.,* **6**, 1307, 1994

Etter, M. C.; *Acc. Chem. Res.,* **23**, 120, 1990

Filippini, G.; Gavezzotti, A., *Acta. Cryst..*; **B48**, 230, 1991

Giacovazzo, C.; Monaco, H. L.; Viterbo, D.; Scordari, F.; Gilli, G.; Zanotti, G.; Catti, M.; *Fundamentals of Crystallography,* Published by the IUCr and Oxford University Press, 1992

International Tables for Crystallography, Vol. A, Edited by Theo Hahn, Third, revised edition. Kluwer Academic Publishers, London, 1992

Kettle, S.F.A. *Symmetry and Structure*; Wiley & Sons, 1985

Kitiagorodskii, A.I. *Organic Chemical Crystallography*; Consultant's Bureau: New York, 1961. (English translation of the Russian original published by Press of the Academy of Sciences of the USSR, Moscow, 1955).

Mackay, A.L., *Acta. Cryst.*; **22**, 329, 1967

Marsh, R. E.; *Acta. Cryst.*; **B42**, 193, 1986

Marsh, R.E.; Paper currently in press, 1996

Mighell, A.D.; Himes, V.L,. *Acta. Cryst.*; **A36**, 321, 1980

Mighell, A.D.; Himes, V.L.; Rogers, J.D. *Acta. Cryst.*; **A39**, 737, 1983

Nowacki, W. *Helv. Chim. Acta.*; **25**, 863, 1942

Nowacki, W. *Helv. Chim. Acta.*; **26**, 459, 1943

Nowacki, W. *Helv. Chim. Acta.*; **34**, 1957, 1951

Nowacki, W. *Crystal Data*; J.D.H.Donnay and W.Nowacki, Ed. Geol. Soc. Amer.; 85, 1954

Padmaja, N.; Ramakumar, S.;Viswamitra, M.A. *Acta Cryst.*; **A46**, 725, 1990

Pauling, L.; Tolman, R. C.; *J. Am. Chem. Soc.*; **47**, 2148, 1925

Scaringe, R.P. In *Electron crystallography of Organic Molecules*; Fryer, J.R. and Dorset, D.L., Ed.; Kluwer: Dordrecht; **Vol.328**; 85, 1991

Taylor, R.; Mullaley, A.; Mullier, G. W.; *Pesticide Science*; **29**, 197, 1990

Whitesell, J.K.; Davis, R.E.; Wong, M-S.; Chang, N-L. *J. Am. Chem. Soc.*; **116**, 523, 1994

Wilson, A.J.C. *Acta Cryst.*; **A44**, 715, 1988

Wilson, A.J.C. *Acta Cryst.*; **A46**, 742, 1990

Wilson, A.J.C. *Z. Krystallogr.*; **197**, 85, 1991

Wilson, A.J.C. *Acta Cryst.*; **A46**, 795, 1993

CSD refcodes referred to in this Chapter:

ALAALA: Fletterick, R. J.; Tsai, C.; Hughes, R. E.; *J. Phys. Chem.*; 75, 918, 1971

BABFUF: Norrestam, R.; Bock, K.; Thogersen, H.; *Acta Cryst.*; **B37**, 1689, 1981

BAICTC: Trop, H. S.; Davison, A.; Jones, A. G.; Davis, M. A.; Szalda, D. J.; Lippard, S. J.; *Inorg. Chem.*; **19**, 1105, 1980

DACTYL: Koski, H. K.; Sandor, E.; *Acta. Cryst.*; **B31**, 350, 1975

EMPCRU: Wartchow, R.; Berthold, H. J.; *Z. Kristallog.*; **145**, 240, 1977

FORPOR: Shinomoto, R.; Zalkin, A.; Edelstein, N. M.; Zhang, D.; *Inorg. Chem.*; **26**, 2868, 1987

KEXNUW: Konda, Y.; Iguchi, M.; Harigaya, Y.; Takayanagi, H.; Ogura, H.; Xian Li; Hong Xiang Lou; Onda, M.; *Tetrahedron Letters*; **31**, 5315, 1990

KULZOG: Gomez, F. A.; Johnson, S. E.; Knobler, C. B.; Hawthorne, M. F.; *Inorg. Chem.*; **31**, 3558, 1992

MAMCOB: Calderazzo, F.; Fachinetti, G.; Marchetti, F.; Zanazzi, P. F.; *J. Chem. Soc., Chem. Comm.*; **181**, 1981

MBYZMN: Sheldrick, G. M.; Davison, B. E.; Trotter, J.; *Acta Cryst.*; **B34**, 1387, 1978

PHMTPT: J. -M. Bassett, Mandl, J. R. ; Schmidbaur, H.; *Chem. Ber.,* **110**, 1284, 1977

PHXTCN: Noernberg, H.; Kratzin, H.; Boldt, P.; Sheldrick, W. S.; *Chem. Ber.*; **110**, 1284, 1977

SIKFOH: Pramod, K.; Eaton, P. E.; Gilardi, R.; Flippen-Anderson, J. L.; *J. Org. Chem.,* **55**, 6105, 1990

Oxalic acid references;

Brill, R.; Hermann, C.; Peters, C. *Ann. Phys. (Leipzig)*; **42**, 357, 1942

Dunitz, J.D.; Robertson, J.M. *J. Chem. Soc.*; 142, 1947

Sabine, T.M.; Cox, G.W.; Craven, B.M. *Acta Cryst.*; **B25**, 2437, 1969

Feld, R.H.; Brown, P.J.; Lehmann, M.S. *Acta Cryst.*; **A34**, S28, 1978

Chapter 8

Summary of the work in this thesis and suggestions for future studies

8.1. Summary of crystallographic studies.

Two crystallographic studies of novel sets of compounds have been presented in this thesis. The first of these studies (Chapter 3) analysed a number of reactants, products and precursors of potential and existing dyotropic molecules. The results for the dyotropic pairs add to the studies that have already been carried out on similar systems. This project still requires some attention, but the structural results have added to the knowledge base currently available. The structures of the precursors have shown that certain proposed dyotropic systems would be viable.

The second set of compounds (Chapter 4) have potential use in non-linear optics. The studies fulfilled the rôle of identifying the synthesised compounds. In one case this confirmed the chemically proposed structure. In the other two samples, the structures were unknown. As such, the crystal structures provided a means of compound identification. The structures have provided the necessary coordinate data so that the hyperpolarisabilities of the respective molecules could be calculated.

8.2. Summary of studies utilising the Cambridge Structural Database (CSD)

8.2.1. Estimating the e.s.d's from CSD stored precision indicators

In chapter 6, an analytical estimate of the coordinate e.s.d's (Cruickshank, 1960) was approximated with available precision indicators in the CSD to attempt to produce a more reliable precision indicator for ~44000 CSD entries that do not store the coordinate e.s.d's. Correlation and regression techniques were used to derive empirical

functions that provided an estimate of the e.s.d. for a particular atom type in a structure. The best estimates were given by a single parameter regression using the atomic numbers of the constituent elements of the asymmetric unit and the crystallographic R-factor. Subdivision of the data set into centrosymmetric and non-centrosymmetric structures, as indicated by Cruickshank's (1960) analysis led to improved e.s.d. estimates.

A somewhat anomalous observation was that the inclusion in the regressions of a function of the least squares over-determination of the crystallographic refinements caused little improvement in the results. The estimates obtained in the best regressions were felt to be of high enough quality to be used as precision criteria, particularly in cases where the CSD "AS" flag was not stored. This information should improve the selectivity of retrieved data sets from the CSD.

8.2.2. The effect of symmetry, the number of crystallographic residues and a number of chemical features on space group frequency.

Chapter 7 used new software to separate each suitable CSD entry into individual residues and to derive the crystallographic point group for each residue.

Little difference was observed in organic and metallo-organic space group frequencies. The only major difference observed was in the population of chiral space groups. This was attributed to a sociological effect: organic chemists try to synthesise more chiral molecules, and also chiral natural products are, almost exclusively, members of an organic class.

It was suggested that pairs of ionic residues behaved like a single non-ionic residue. The number of crystallographic residues in the asymmetric appeared to influence the space group distribution. The lower symmetry space group, $P\bar{1}$ becomes more favourable at the expense of the higher symmetry space group $P2_1/c$. Non-centrosymmetric space groups are more common for structures with $N_{res} > 1$.

245

Mirror planes were conclusively shown to be structure determining: mirror symmetric space groups only occur when at least one crystallographic mirror plane is occupied. Proper rotation axes become more unfavourable with increasing order, and are also more frequently occupied with increasing order. Other improper rotation axes have less influence on space group frequency primarily because it is possible to pack molecules efficiently around improper axes, so axis occupation is not a pre-requisite.

Ionic structures with residues having three-fold axes prefer to crystallise in the space group $P2_13$ rather than its centrosymmetric counterpart, $Pa\bar{3}$.

8.3. Suggestions for future work: Crystallography.

Clearly the work in Chapter 3 can be extended by solving and refining structures of more dyotropic systems. It is possible that with enough structures the structure-reactivity correlation principle may be applied to the compounds to try to identify those structural features that have the strongest influence on reactivity. Such details may well provide enlightenment on the reaction mechanism.

Structures shown in Chapter 4 were primarily solved as a service to the synthetic chemists. It is assured that more similar samples will be provided for crystallographic study, since the chemists need atomic coordinates for estimating the molecular hyperpolarisabilty. Such details are important for directing further work: such compounds can be incorporated into suitable media (such as Langmuir-Blodgett films: see Ashwell *et al*, 1990).

8.4. Suggestions for future work: CSD studies.

8.4.1. Future work with the estimates of e.s.d's.

The work on estimating the e.s.d's from CSD precision indicators is almost complete. One possible further piece of work would be to attempt constrained regressions. It is possible that "better" results would be obtained by constraining the estimates to lie within the ranges given by the CSD "AS" flag for the entries where both the AS flag and the e.s.d's were present. Equations from such regressions may give better estimates of the e.s.d's in cases where the AS flag only is stored.

Application of the equations to the full data base is also needed. Following such work an assesment can be made as to the precision of actual sets of data that would be retrieved by applying a precision constraint based on the empirical equations. If they prove suitable, it may then be possible to introduce a single precision criterion for all entries on the CSD for future use.

8.4.2. Future work on space group frequencies and symmetry.

A number of additional pieces of work could be carried out utilising the data provided in Chapter 7. Firstly, searching on residue symmetry could be implemented on the CSD, to facilitate easier extraction of this data in the future. A clear question arises from the "rare" space groups that are present on the CSD, particularly those that have no inherent molecular symmetry, as to why these structures form in the particular space group. Some examples of such space groups have been individually inspected in Chapter 7, but there are more examples that have not (particularly the rarer orthorhombic structures). Brock and co-workers (1994) have shown that mono-alcohols utilise rare space groups due to structure determining H...OH interactions. It may be possible to rationalise "classes" of molecules that crystallise in specific space groups, perhaps due to specific orientations of functional groups at the molecular periphery.

One purpose of analysing space group frequencies is to try to maximise the chance of crystallising a sample in a particular space group, or class of space groups. An example is in non-linear optical design where a non-centrosymmetric crystal structure is desirable as a medium for second order harmonic generation. Because of this, it would be useful to find out if any symmetry has an influence on crystalline packing. In Chapter 7 it was possible to analyse the occupation of special positions. This data alone, however, cannot be used to answer questions such as "Do molecules with three fold axes tend to crystallise in non-centrosymmetric structures". Brock and Dunitz (1994) have suggested that three fold axes do cause more usage of non-centrosymmetric space groups by studying the value of Z'. However this seems a little naive, since they have no idea of the number of structures that possess three fold axes, but do not utilise them in the crystal packing. Examples of molecules with molecular mirror planes that do not utilise them in crystal packing are given in Chapter 3. Some work has been started in this area, although no final results have yet been derived (Cole *et al*, 1995).

To deal with this problem, a study needs to be carried out that determines the *molecular* symmetry of each crystalline residue, so that the rate of axis *occupation* can be estimated. Brock and Dunitz (1994) have pointed out that a higher than average number of molecules seem to occupy centres of inversion, as compared to the number of molecules that have ever been synthesised. Some analysis would be worthwhile to establish how many examples of symmetry site occupation are due to solvent molecules or uniatomic ions. It is hoped that the derived workfile can be used to produce these results, although this form of analysis has not yet been carried out.

8.5. Conclusions

The research in this thesis has shown that it is still useful to derive chemical information via crystallography, since it provides chemical researchers with a detailed analysis of the accurate *experimental* chemical structure. Some of the research has also emphasised

that it is possible to analyse large numbers of crystallographic studies to derive crystallographic trends, by utilising crystallographic databases. The information content of such databases is, however, limited by the entries within the database. Little can be concluded about the packing of molecules with six-fold symmetry, since few examples of such compounds exist. In interpreting any result, care has to be taken to accomodate the possibility that a specific effect of database content is not manifesting itself. An example of such an effect can be seen in Table 7M.1 in Chapter 7, where it appears that chiral space groups with "N_1"(where N = 3,4 or 6) axes tend to occur more frequently than their counterpart space groups with "N_2", "N_3" or "N_5" (i.e.3_2, 4_3 or 6_5) axes. This is probably due to two factors. Firstly, the "N_1" space groups are first in International Tables, (1992) (and are usually selected first by automatic space group determination programs): in some studies it is impratical to determine the enantiomorph due to a lack of heavier atoms, and so the first is selected. Secondly, natural products which occur in such space groups may bias the population of one set of enantiomorphic space groups, since the structure of the opposite enatiomorph will not be derived. This effect could be erroneously ascribed some chemical significance.

Limitations such as these are only remedied by careful data analysis, and a greater quantity of structural data. Subtle chemical effects may well still be hidden in the CSD. These will be far more apparent when much more data is available. More data also allows much finer sub-divisions to be used in studies (for example, a study of C=O groups can currently be broken down into aldehydes, ketones carboxylic acids, esters etc. etc.: more data may allow each of these sub-divisions to be further divided: e.g. the nature of the atom(s) α to C=O groups could be used to divide the data set further). More data also allows more stringent secondary criteria to be applied, increasing the internal precision of the data set. The conclusion is thus that crystallography still has a crucial rôle in the future of chemical analysis.

8.6. References

Ashwell, G. J.; Bryce, M. R.; Dawnay, Grainger, A. M.; Hasan, M.; E. J. C.; Kuczynski, A. P.; Szablewski, M.; Sandy, I. M.; *J. Chem. Soc. Faraday Trans.*; **86**, 1117, 1990

Brock, C. P.; Duncan, L. L.; *Chem. Mater.*; **6**, 1307, 1994

Brock, C.P; Dunitz, J.D. *Chem. Mater.*; **6**, 1118, 1994

Cole, J. C.; Allen, F.H.; Howard, J. A. K.; *Unpublished work,* 1995

International Tables for Crystallography, Vol. A, Edited by Theo Hahn, Third, revised edition. Kluwer Academic Publishers, London, 1992

Appendix 3A

Tables of structural data for structures in Chapter 3

The labels of chemically equivalent atoms in each independant molecule in structures (1) and (2) can be derived by adding 5 to the first numeral in the label, for example Cl1 in molecule (1) of structure 1(1) is chemically equivalent to Cl51 in molecule (2) of structure (1), C10 in molecule (1) is chemically equivalent to C60 in molecule (2).

Table 3A.1.1. Atomic coordinates and equivalent isotropic displacement parameters for (1)

Molecule 1

Atom	x	y	z	U(eq)
C(1)	0.7883(2)	0.2633(2)	0.5103(3)	0.041(1)
C(2)	0.6922(2)	0.2318(2)	0.6273(3)	0.042(1)
Cl(1)	0.8095(1)	0.3835(1)	0.4331(1)	0.058(1)
C(3)	0.7188(2)	0.2816(2)	0.7872(3)	0.051(1)
C(4)	0.8087(3)	0.2501(2)	0.8748(3)	0.059(1)
C(9)	0.7835(3)	0.1503(2)	0.8823(3)	0.063(1)
C(10)	0.6695(3)	0.0863(2)	0.8019(3)	0.057(1)
C(11)	0.6634(2)	0.1142(2)	0.6350(3)	0.045(1)
C(12)	0.7478(2)	0.0948(2)	0.5217(3)	0.048(1)
Cl(12)	0.7100(1)	-0.0346(1)	0.4599(1)	0.074(1)
C(13)	0.8751(2)	0.1472(2)	0.5764(3)	0.049(1)
Cl(13)	0.9629(1)	0.0856(1)	0.6493(1)	0.080(1)
C(14)	0.8992(2)	0.2475(2)	0.5703(3)	0.046(1)
Cl(14)	1.0247(1)	0.3432(1)	0.6348(1)	0.068(1)
C(15)	0.7448(2)	0.1701(2)	0.3974(3)	0.046(1)
Cl(15)	0.6043(1)	0.1459(1)	0.3053(1)	0.061(1)
C(16)	0.6021(3)	0.2336(2)	0.8676(4)	0.066(1)
C(17)	0.5725(3)	0.1172(2)	0.8740(3)	0.068(1)

Molecule 2

Atom	x	y	z	U(eq)
C(51)	0.7406(2)	-0.3092(2)	1.0330(2)	0.042(1)
Cl(51)	0.6762(1)	-0.3257(1)	1.2083(1)	0.061(1)
C(52)	0.6546(2)	-0.3073(2)	0.8965(2)	0.039(1)
C(53)	0.5620(2)	-0.4077(2)	0.8312(3)	0.045(1)
C(54)	0.6185(2)	-0.4748(2)	0.7575(3)	0.055(1)
C(59)	0.6905(3)	-0.4312(2)	0.6522(3)	0.060(1)
C(60)	0.7043(2)	-0.3221(2)	0.6260(3)	0.053(1)
C(61)	0.7409(2)	-0.2563(2)	0.7736(3)	0.042(1)
C(62)	0.8636(2)	-0.2378(2)	0.8562(3)	0.044(1)
Cl(62)	0.9833(1)	-0.1486(1)	0.7696(1)	0.068(1)
C(63)	0.8789(2)	-0.3383(2)	0.8918(3)	0.044(1)
Cl(63)	0.9642(1)	-0.3900(1)	0.7965(1)	0.068(1)
C(64)	0.8059(2)	-0.3806(2)	0.9981(3)	0.043(1)
Cl(64)	0.7773(1)	-0.4983(1)	1.0708(1)	0.061(1)
C(65)	0.8432(2)	-0.2075(2)	1.0165(3)	0.047(1)
Cl(65)	0.8085(1)	-0.0951(1)	1.0361(1)	0.072(1)
C(66)	0.4938(2)	-0.3753(2)	0.7020(3)	0.051(1)
C(67)	0.5782(2)	-0.3225(2)	0.5813(3)	0.058(1)

Table 3A.1.2. Bond lengths [Å] for (1).

Molecule 1

C(1)-C(14)	1.512(3)	C(1)-C(15)	1.541(3)
C(1)-C(2)	1.549(3)	C(1)-Cl(1)	1.761(2)
C(2)-C(3)	1.529(3)	C(2)-C(11)	1.555(3)
C(3)-C(4)	1.500(4)	C(3)-C(16)	1.554(4)
C(4)-C(9)	1.319(4)	C(9)-C(10)	1.495(4)
C(10)-C(11)	1.537(3)	C(10)-C(17)	1.547(5)
C(11)-C(12)	1.551(4)	C(12)-C(13)	1.511(4)
C(12)-C(15)	1.542(3)	C(12)-Cl(12)	1.766(2)
C(13)-C(14)	1.327(3)	C(13)-Cl(13)	1.698(2)
C(14)-Cl(14)	1.701(3)	C(15)-Cl(15)	1.772(3)
C(16)-C(17)	1.536(4)	C(51)-C(64)	1.508(3)

Molecule 2

C(51)-C(65)	1.542(3)	C(51)-C(52)	1.563(3)
C(51)-Cl(51)	1.751(2)	C(52)-C(53)	1.537(3)
C(52)-C(61)	1.550(3)	C(53)-C(54)	1.495(4)
C(53)-C(66)	1.543(3)	C(54)-C(59)	1.315(4)
C(59)-C(60)	1.495(4)	C(60)-C(61)	1.532(3)
C(60)-C(67)	1.561(4)	C(61)-C(62)	1.562(3)
C(62)-C(63)	1.512(3)	C(62)-C(65)	1.525(3)
C(62)-Cl(62)	1.769(2)	C(63)-C(64)	1.324(3)
C(63)-Cl(63)	1.697(3)	C(64)-Cl(64)	1.700(2)
C(65)-Cl(65)	1.766(3)	C(66)-C(67)	1.529(4)

Table 3A.1.3. Bond Angles [°] for (1).

Molecule 1

C(14)-C(1)-C(15)	97.3(2)	C(14)-C(1)-C(2)	111.0(2)
C(15)-C(1)-C(2)	101.2(2)	C(14)-C(1)-Cl(1)	115.2(2)
C(15)-C(1)-Cl(1)	116.1(2)	C(2)-C(1)-Cl(1)	114.0(2)
C(3)-C(2)-C(1)	119.6(2)	C(3)-C(2)-C(11)	110.1(2)
C(1)-C(2)-C(11)	102.8(2)	C(4)-C(3)-C(2)	111.5(2)
C(4)-C(3)-C(16)	105.8(2)	C(2)-C(3)-C(16)	104.5(2)
C(9)-C(4)-C(3)	115.1(3)	C(4)-C(9)-C(10)	114.7(3)
C(9)-C(10)-C(11)	111.6(2)	C(9)-C(10)-C(17)	106.5(2)
C(11)-C(10)-C(17)	104.3(2)	C(10)-C(11)-C(12)	120.7(2)
C(10)-C(11)-C(2)	109.0(2)	C(12)-C(11)-C(2)	102.3(2)
C(13)-C(12)-C(15)	97.1(2)	C(13)-C(12)-C(11)	111.5(2)
C(15)-C(12)-C(11)	101.2(2)	C(13)-C(12)-Cl(12)	114.9(2)
C(15)-C(12)-Cl(12)	115.8(2)	C(11)-C(12)-Cl(12)	114.3(2)
C(14)-C(13)-C(12)	107.5(2)	C(14)-C(13)-Cl(13)	127.8(2)
C(12)-C(13)-Cl(13)	124.4(2)	C(13)-C(14)-C(1)	107.3(2)
C(13)-C(14)-Cl(14)	127.7(2)	C(1)-C(14)-Cl(14)	124.7(2)
C(1)-C(15)-C(12)	92.7(2)	C(1)-C(15)-Cl(15)	115.7(2)
C(12)-C(15)-Cl(15)	114.6(2)	C(17)-C(16)-C(3)	109.3(3)
C(16)-C(17)-C(10)	110.1(2)	C(64)-C(51)-C(65)	97.5(2)

Molecule 2

C(64)-C(51)-C(52)	111.0(2)	C(65)-C(51)-C(52)	101.6(2)
C(64)-C(51)-Cl(51)	115.1(2)	C(65)-C(51)-Cl(51)	116.0(2)
C(52)-C(51)-Cl(51)	113.8(2)	C(53)-C(52)-C(61)	109.9(2)
C(53)-C(52)-C(51)	120.1(2)	C(61)-C(52)-C(51)	102.0(2)
C(54)-C(53)-C(52)	111.2(2)	C(54)-C(53)-C(66)	105.8(2)
C(52)-C(53)-C(66)	105.1(2)	C(59)-C(54)-C(53)	114.9(3)
C(54)-C(59)-C(60)	115.1(3)	C(59)-C(60)-C(61)	111.4(2)
C(59)-C(60)-C(67)	105.8(2)	C(61)-C(60)-C(67)	104.4(2)
C(60)-C(61)-C(52)	109.3(2)	C(60)-C(61)-C(62)	120.2(2)
C(52)-C(61)-C(62)	102.6(2)	C(63)-C(62)-C(65)	97.8(2)
C(63)-C(62)-C(61)	111.0(2)	C(65)-C(62)-C(61)	101.3(2)
C(63)-C(62)-Cl(62)	114.9(2)	C(65)-C(62)-Cl(62)	115.5(2)
C(61)-C(62)-Cl(62)	114.4(2)	C(64)-C(63)-C(62)	107.3(2)
C(64)-C(63)-Cl(63)	127.7(2)	C(62)-C(63)-Cl(63)	124.7(2)
C(63)-C(64)-C(51)	107.1(2)	C(63)-C(64)-Cl(64)	127.8(2)
C(51)-C(64)-Cl(64)	124.8(2)	C(62)-C(65)-C(51)	92.6(2)
C(62)-C(65)-Cl(65)	116.3(2)	C(51)-C(65)-Cl(65)	115.7(2)
C(67)-C(66)-C(53)	109.6(2)	C(66)-C(67)-C(60)	109.9(2)

Table 3A.1.4. Anisotropic displacement parameters ($\mathring{A}^2$) for (1).

Molecule 1

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C(1)	0.045(1)	0.038(1)	0.044(1)	0.006(1)	-0.002(1)	0.019(1)
C(2)	0.042(1)	0.041(1)	0.042(1)	0.002(1)	-0.003(1)	0.016(1)
Cl(1)	0.068(1)	0.046(1)	0.067(1)	0.019(1)	0.011(1)	0.026(1)
C(3)	0.059(2)	0.043(1)	0.048(1)	-0.004(1)	0.001(1)	0.016(1)
C(4)	0.062(2)	0.063(2)	0.043(1)	-0.005(1)	-0.010(1)	0.013(1)
C(9)	0.079(2)	0.070(2)	0.042(1)	0.004(1)	-0.014(1)	0.030(2)
C(10)	0.078(2)	0.042(1)	0.044(1)	0.007(1)	-0.005(1)	0.013(1)
C(11)	0.052(1)	0.039(1)	0.041(1)	0.001(1)	-0.007(1)	0.011(1)
C(12)	0.066(2)	0.037(1)	0.043(1)	-0.004(1)	-0.011(1)	0.022(1)
Cl(12)	0.109(1)	0.043(1)	0.071(1)	-0.013(1)	-0.014(1)	0.031(1)
C(13)	0.056(2)	0.054(1)	0.046(1)	-0.002(1)	-0.008(1)	0.033(1)
Cl(13)	0.101(1)	0.087(1)	0.078(1)	-0.008(1)	-0.028(1)	0.070(1)
C(14)	0.043(1)	0.049(1)	0.049(1)	0.001(1)	-0.005(1)	0.019(1)
Cl(14)	0.046(1)	0.069(1)	0.080(1)	-0.001(1)	-0.010(1)	0.009(1)
C(15)	0.052(1)	0.050(1)	0.041(1)	0.001(1)	-0.007(1)	0.023(1)
Cl(15)	0.064(1)	0.070(1)	0.047(1)	-0.004(1)	-0.019(1)	0.027(1)
C(16)	0.068(2)	0.069(2)	0.058(2)	-0.002(1)	0.013(1)	0.020(2)
C(17)	0.076(2)	0.066(2)	0.046(2)	0.006(1)	0.007(1)	0.005(2)

Molecule 2

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C(51)	0.044(1)	0.046(1)	0.034(1)	0.002(1)	-0.001(1)	0.016(1)
Cl(51)	0.063(1)	0.085(1)	0.036(1)	0.005(1)	0.005(1)	0.026(1)
C(52)	0.042(1)	0.042(1)	0.038(1)	0.002(1)	-0.003(1)	0.020(1)
C(53)	0.040(1)	0.045(1)	0.048(1)	0.009(1)	-0.006(1)	0.013(1)
C(54)	0.056(2)	0.044(1)	0.065(2)	-0.010(1)	-0.024(1)	0.021(1)
C(59)	0.056(2)	0.070(2)	0.058(2)	-0.025(1)	-0.019(1)	0.033(2)
C(60)	0.050(2)	0.074(2)	0.034(1)	0.001(1)	-0.002(1)	0.022(1)
C(61)	0.043(1)	0.042(1)	0.042(1)	0.006(1)	-0.002(1)	0.016(1)
C(62)	0.040(1)	0.044(1)	0.044(1)	0.006(1)	0.000(1)	0.011(1)
Cl(62)	0.049(1)	0.070(1)	0.075(1)	0.024(1)	0.008(1)	0.007(1)
C(63)	0.040(1)	0.051(1)	0.044(1)	-0.002(1)	-0.006(1)	0.022(1)
Cl(63)	0.060(1)	0.093(1)	0.067(1)	-0.008(1)	-0.001(1)	0.047(1)
C(64)	0.046(1)	0.041(1)	0.041(1)	0.003(1)	-0.011(1)	0.017(1)
Cl(64)	0.073(1)	0.047(1)	0.064(1)	0.012(1)	-0.014(1)	0.022(1)
C(65)	0.051(1)	0.042(1)	0.048(1)	-0.004(1)	-0.007(1)	0.015(1)
Cl(65)	0.091(1)	0.045(1)	0.078(1)	-0.014(1)	-0.011(1)	0.026(1)
C(66)	0.042(1)	0.059(2)	0.053(1)	0.006(1)	-0.010(1)	0.021(1)
C(67)	0.054(2)	0.073(2)	0.044(1)	0.010(1)	-0.012(1)	0.021(1)

Table 3A.1.5. Hydrogen coordinates and isotropic displacement parameters [Å^2] for (1).

Molecule 1

Atom	x	y	z	U(eq)
H(2)	0.6255(20)	0.2425(4)	0.5852(13)	0.050
H(3)	0.7426(8)	0.3574(23)	0.7848(3)	0.061
H(4)	0.8786(24)	0.2985(17)	0.9214(16)	0.071
H(9)	0.8315(19)	0.1228(11)	0.9323(19)	0.076
H(10)	0.6556(5)	0.0159(24)	0.8107(4)	0.068
H(11)	0.5828(24)	0.0798(10)	0.5934(13)	0.055
H(15)	0.8039(18)	0.1740(2)	0.3231(22)	0.055
H(16A)	0.6105(3)	0.2636(7)	0.9700(24)	0.079
H(16B)	0.5386(15)	0.2479(4)	0.8120(13)	0.079
H(17A)	0.4980(18)	0.0821(9)	0.8203(13)	0.081
H(17B)	0.5664(4)	0.0974(6)	0.9779(25)	0.081

Molecule 2

Atom	x	y	z	U(eq)
H(52)	0.6146(12)	-0.2616(13)	0.9242(8)	0.047
H(53)	0.5082(15)	-0.4448(11)	0.9109(23)	0.054
H(54)	0.6039(6)	-0.5423(23)	0.7842(10)	0.066
H(59)	0.7306(15)	-0.4664(13)	0.5970(20)	0.072
H(60)	0.7572(18)	-0.2944(9)	0.5490(25)	0.064
H(61)	0.7347(3)	-0.1915(19)	0.7550(6)	0.050
H(65)	0.9089(21)	-0.2038(2)	1.0848(21)	0.057
H(66A)	0.4304(13)	-0.4365(12)	0.6559(9)	0.061
H(66B)	0.4571(8)	-0.3277(10)	0.7433(9)	0.061
H(67A)	0.5803(2)	-0.2524(15)	0.5723(4)	0.069
H(67B)	0.5507(6)	-0.3585(8)	0.4836(22)	0.069

Table 3A.2.1. Atomic coordinates and equivalent isotropic displacement parameters [Å^2] for (2).

Molecule 1

Atom	x	y	z	U(eq)
C(1)	0.2013(3)	0.9039(2)	0.2621(2)	0.022(1)
Cl(1)	0.1523(1)	0.8061(1)	0.3016(1)	0.035(1)
C(2)	0.0830(3)	0.9433(2)	0.1887(2)	0.018(1)
C(3)	0.0004(3)	1.0355(2)	0.2316(2)	0.018(1)
C(4)	0.0983(3)	1.1350(2)	0.3224(2)	0.021(1)
C(9)	0.1581(3)	1.1746(2)	0.2720(3)	0.024(1)
C(10)	0.1033(3)	1.1003(2)	0.1506(2)	0.024(1)
C(11)	0.1558(3)	0.9917(2)	0.1322(2)	0.021(1)
C(12)	0.3057(3)	0.9768(2)	0.1820(2)	0.023(1)
Cl(12)	0.4052(1)	0.9894(1)	0.1072(1)	0.037(1)
C(13)	0.3680(3)	1.0401(2)	0.3078(2)	0.022(1)
Cl(13)	0.4874(1)	1.1508(1)	0.3721(1)	0.033(1)
C(14)	0.3065(3)	0.9968(2)	0.3548(2)	0.022(1)
Cl(14)	0.3283(1)	1.0408(1)	0.4915(1)	0.033(1)
C(15)	0.2817(3)	0.8647(2)	0.1821(2)	0.025(1)
Cl(1A)	0.1860(1)	0.7630(1)	0.0502(1)	0.036(1)
Cl(1B)	0.4347(1)	0.8187(1)	0.2394(1)	0.034(1)
O(16)	-0.0341(2)	1.0646(2)	0.1354(2)	0.022(1)
C(30)	-0.1280(3)	1.0073(2)	0.2485(2)	0.021(1)
C(31)	-0.2201(3)	0.9089(2)	0.1481(3)	0.028(1)
C(32)	-0.2069(3)	1.1015(2)	0.2603(3)	0.034(1)
C(33)	-0.0890(3)	0.9874(2)	0.3560(3)	0.029(1)

Molecule 2

Atom	x	y	z	U(eq)
Cl(51)	0.3524(1)	0.3478(1)	-0.4432(1)	0.038(1)
Cl(62)	0.0850(1)	0.4814(1)	-0.1502(1)	0.035(1)
Cl(63)	0.0711(1)	0.6507(1)	-0.2695(1)	0.038(1)
Cl(64)	0.2428(1)	0.5728(1)	-0.4468(1)	0.046(1)
C(51)	0.3109(3)	0.4347(2)	-0.3325(2)	0.023(1)
C(52)	0.4292(3)	0.4796(2)	-0.2159(2)	0.019(1)
C(53)	0.5413(3)	0.5804(2)	-0.1665(2)	0.022(1)
C(54)	0.4761(3)	0.6752(2)	-0.1799(3)	0.027(1)
C(59)	0.4108(3)	0.7047(2)	-0.1106(3)	0.028(1)
C(60)	0.4323(3)	0.6285(2)	-0.0532(2)	0.025(1)
C(61)	0.3522(3)	0.5148(2)	-0.1335(2)	0.020(1)
C(62)	0.2005(3)	0.4875(2)	-0.2136(2)	0.021(1)
C(63)	0.1687(3)	0.5529(2)	-0.2845(2)	0.025(1)
C(64)	0.2339(3)	0.5217(2)	-0.3542(2)	0.027(1)
C(65)	0.2048(3)	0.3807(2)	-0.3050(3)	0.025(1)
Cl(6A)	0.2613(1)	0.2812(1)	-0.2553(1)	0.037(1)
Cl(6B)	0.0466(1)	0.3237(1)	-0.4184(1)	0.039(1)
O(66)	0.5663(2)	0.6058(2)	-0.0493(2)	0.024(1)
C(80)	0.6742(3)	0.5644(2)	-0.1874(3)	0.031(1)
C(81)	0.7257(3)	0.4651(3)	-0.1716(4)	0.045(1)
C(82)	0.7815(3)	0.6622(3)	-0.1029(3)	0.042(1)
C(83)	0.6515(4)	0.5540(3)	-0.3060(3)	0.049(1)

Table 3A.2.2. Bond lengths [Å] for (2).

Molecule 1

C(1)-C(14)	1.521(4)	C(1)-C(2)	1.565(4)
C(1)-C(15)	1.573(4)	C(1)-Cl(1)	1.761(3)
C(2)-C(11)	1.565(4)	C(2)-C(3)	1.597(4)
C(2)-H(2)	0.96(3)	C(3)-O(16)	1.461(3)
C(3)-C(30)	1.530(3)	C(3)-C(4)	1.531(4)
C(4)-C(9)	1.329(4)	C(4)-H(4)	0.95(3)
C(9)-C(10)	1.507(4)	C(9)-H(9)	0.86(3)
C(10)-O(16)	1.454(3)	C(10)-C(11)	1.559(4)
C(10)-H(10)	0.95(3)	C(11)-C(12)	1.553(4)
C(11)-H(11)	0.90(3)	C(12)-C(13)	1.525(4)
C(12)-C(15)	1.557(4)	C(12)-Cl(12)	1.762(3)
C(13)-C(14)	1.333(4)	C(13)-Cl(13)	1.705(3)
C(14)-Cl(14)	1.696(3)	C(15)-Cl(1A)	1.776(3)
C(15)-Cl(1B)	1.782(3)	C(30)-C(33)	1.534(4)
C(30)-C(32)	1.537(4)	C(30)-C(31)	1.542(4)
C(31)-H(31A)	0.97(2)	C(31)-H(31B)	0.97(2)
C(31)-H(31C)	0.97(2)	C(32)-H(32A)	0.98(2)
C(32)-H(32B)	0.98(2)	C(32)-H(32C)	0.98(2)
C(33)-H(33A)	0.98(2)	C(33)-H(33B)	0.98(2)
C(33)-H(33C)	0.98(2)		

Molecule 2

Cl(51)-C(51)	1.763(3)	Cl(62)-C(62)	1.756(3)
Cl(63)-C(63)	1.701(3)	Cl(64)-C(64)	1.703(3)
C(51)-C(64)	1.516(4)	C(51)-C(52)	1.561(4)
C(51)-C(65)	1.566(4)	C(52)-C(61)	1.566(4)
C(52)-C(53)	1.586(4)	C(52)-H(52)	0.97(3)
C(53)-O(66)	1.458(3)	C(53)-C(54)	1.521(4)
C(53)-C(80)	1.537(4)	C(54)-C(59)	1.323(4)
C(54)-H(54)	0.91(3)	C(59)-C(60)	1.509(4)
C(59)-H(59)	0.90(3)	C(60)-O(66)	1.447(3)
C(60)-C(61)	1.565(4)	C(60)-H(60)	0.92(3)
C(61)-C(62)	1.557(4)	C(61)-H(61)	0.92(3)
C(62)-C(63)	1.522(4)	C(62)-C(65)	1.560(4)
C(63)-C(64)	1.326(4)	C(65)-Cl(6A)	1.769(3)
C(65)-Cl(6B)	1.781(3)	C(80)-C(83)	1.531(5)
C(80)-C(81)	1.536(5)	C(80)-C(82)	1.541(4)
C(81)-H(81A)	0.93(2)	C(81)-H(81B)	0.93(2)
C(81)-H(81C)	0.93(2)	C(82)-H(82A)	0.98(2)
C(82)-H(82B)	0.98(2)	C(82)-H(82C)	0.98(2)
C(83)-H(83A)	0.99(2)	C(83)-H(83B)	0.99(2)
C(83)-H(83C)	0.99(2)		

Table 3A.2.3(a). Angles [°] for 2: *Molecule 1*

C(14)-C(1)-C(2)	110.9(2)	C(14)-C(1)-C(15)	98.2(2)
C(2)-C(1)-C(15)	99.5(2)	C(14)-C(1)-Cl(1)	115.9(2)
C(2)-C(1)-Cl(1)	115.9(2)	C(15)-C(1)-Cl(1)	113.9(2)
C(1)-C(2)-C(11)	102.5(2)	C(1)-C(2)-C(3)	127.0(2)
C(11)-C(2)-C(3)	101.2(2)	C(1)-C(2)-H(2)	108.20(14)
C(11)-C(2)-H(2)	108.20(14)	C(3)-C(2)-H(2)	108.20(13)
O(16)-C(3)-C(30)	110.6(2)	O(16)-C(3)-C(4)	100.2(2)
C(30)-C(3)-C(4)	117.0(2)	O(16)-C(3)-C(2)	96.1(2)
C(30)-C(3)-C(2)	118.9(2)	C(4)-C(3)-C(2)	110.2(2)
C(9)-C(4)-C(3)	106.6(2)	C(9)-C(4)-H(4)	126.7(2)
C(3)-C(4)-H(4)	126.7(2)	C(4)-C(9)-C(10)	106.4(3)
C(4)-C(9)-H(9)	126.8(2)	C(10)-C(9)-H(9)	126.8(2)
O(16)-C(10)-C(9)	101.0(2)	O(16)-C(10)-C(11)	97.5(2)
C(9)-C(10)-C(11)	111.9(2)	O(16)-C(10)-H(10)	114.78(13)
C(9)-C(10)-H(10)	114.8(2)	C(11)-C(10)-H(10)	114.78(14)
C(12)-C(11)-C(10)	125.6(2)	C(12)-C(11)-C(2)	103.3(2)
C(10)-C(11)-C(2)	101.8(2)	C(12)-C(11)-H(11)	108.3(2)
C(10)-C(11)-H(11)	108.3(2)	C(2)-C(11)-H(11)	108.28(14)
C(13)-C(12)-C(11)	111.6(2)	C(13)-C(12)-C(15)	98.5(2)
C(11)-C(12)-C(15)	99.2(2)	C(13)-C(12)-Cl(12)	115.4(2)
C(11)-C(12)-Cl(12)	114.2(2)	C(15)-C(12)-Cl(12)	115.9(2)
C(14)-C(13)-C(12)	107.7(2)	C(14)-C(13)-Cl(13)	127.6(2)
C(12)-C(13)-Cl(13)	124.6(2)	C(13)-C(14)-C(1)	107.8(2)
C(13)-C(14)-Cl(14)	127.6(2)	C(1)-C(14)-Cl(14)	124.4(2)
C(12)-C(15)-C(1)	92.5(2)	C(12)-C(15)-Cl(1A)	115.6(2)
C(1)-C(15)-Cl(1A)	114.1(2)	C(12)-C(15)-Cl(1B)	113.4(2)
C(1)-C(15)-Cl(1B)	114.1(2)	Cl(1A)-C(15)-Cl(1B)	107.0(2)
C(10)-O(16)-C(3)	97.6(2)	C(3)-C(30)-C(33)	109.7(2)
C(3)-C(30)-C(32)	108.6(2)	C(33)-C(30)-C(32)	108.6(2)
C(3)-C(30)-C(31)	111.2(2)	C(33)-C(30)-C(31)	110.1(2)
C(32)-C(30)-C(31)	108.6(2)	C(30)-C(31)-H(31A)	109.5(2)
C(30)-C(31)-H(31B)	109.5(2)	H(31A)-C(31)-H(31B)	109.5
C(30)-C(31)-H(31C)	109.5(2)	H(31A)-C(31)-H(31C)	109.5
H(31B)-C(31)-H(31C)	109.5	C(30)-C(32)-H(32A)	109.5(2)
C(30)-C(32)-H(32B)	109.5(2)	H(32A)-C(32)-H(32B)	109.5
C(30)-C(32)-H(32C)	109.5(2)	H(32A)-C(32)-H(32C)	109.5
H(32B)-C(32)-H(32C)	109.5	C(30)-C(33)-H(33A)	109.5(2)
C(30)-C(33)-H(33B)	109.5(2)	H(33A)-C(33)-H(33B)	109.5
C(30)-C(33)-H(33C)	109.5(2)	H(33A)-C(33)-H(33C)	109.5
H(33B)-C(33)-H(33C)	109.5		

Table 3A.2.3(b) Bond Angles [°] for 2: *Molecule 2*

C(64)-C(51)-C(52)	111.0(2)	C(64)-C(51)-C(65)	98.0(2)
C(52)-C(51)-C(65)	99.9(2)	C(64)-C(51)-Cl(51)	116.2(2)
C(52)-C(51)-Cl(51)	115.5(2)	C(65)-C(51)-Cl(51)	113.7(2)
C(51)-C(52)-C(61)	102.6(2)	C(51)-C(52)-C(53)	127.2(2)
C(61)-C(52)-C(53)	101.5(2)	C(51)-C(52)-H(52)	107.97(14)
C(61)-C(52)-H(52)	107.97(14)	C(53)-C(52)-H(52)	107.97(14)
O(66)-C(53)-C(54)	100.0(2)	O(66)-C(53)-C(80)	110.1(2)
C(54)-C(53)-C(80)	117.6(2)	O(66)-C(53)-C(52)	96.5(2)
C(54)-C(53)-C(52)	110.6(2)	C(80)-C(53)-C(52)	118.2(2)
C(59)-C(54)-C(53)	107.1(3)	C(59)-C(54)-H(54)	126.4(2)
C(53)-C(54)-H(54)	126.4(2)	C(54)-C(59)-C(60)	105.8(3)
C(54)-C(59)-H(59)	127.1(2)	C(60)-C(59)-H(59)	127.1(2)
O(66)-C(60)-C(59)	101.3(2)	O(66)-C(60)-C(61)	97.7(2)
C(59)-C(60)-C(61)	112.2(2)	O(66)-C(60)-H(60)	114.56(14)
C(59)-C(60)-H(60)	114.6(2)	C(61)-C(60)-H(60)	114.6(2)
C(62)-C(61)-C(60)	126.2(2)	C(62)-C(61)-C(52)	102.9(2)
C(60)-C(61)-C(52)	101.0(2)	C(62)-C(61)-H(61)	108.44(14)
C(60)-C(61)-H(61)	108.4(2)	C(52)-C(61)-H(61)	108.44(14)
C(63)-C(62)-C(61)	111.6(2)	C(63)-C(62)-C(65)	98.0(2)
C(61)-C(62)-C(65)	99.7(2)	C(63)-C(62)-Cl(62)	115.5(2)
C(61)-C(62)-Cl(62)	114.4(2)	C(65)-C(62)-Cl(62)	115.5(2)
C(64)-C(63)-C(62)	107.7(2)	C(64)-C(63)-Cl(63)	128.1(2)
C(62)-C(63)-Cl(63)	124.2(2)	C(63)-C(64)-C(51)	108.0(3)
C(63)-C(64)-Cl(64)	127.9(2)	C(51)-C(64)-Cl(64)	124.0(2)
C(62)-C(65)-C(51)	92.4(2)	C(62)-C(65)-Cl(6A)	114.6(2)
C(51)-C(65)-Cl(6A)	114.8(2)	C(62)-C(65)-Cl(6B)	113.9(2)
C(51)-C(65)-Cl(6B)	114.3(2)	Cl(6A)-C(65)-Cl(6B)	106.8(2)
C(60)-O(66)-C(53)	97.3(2)	C(83)-C(80)-C(81)	109.6(3)
C(83)-C(80)-C(53)	109.7(3)	C(81)-C(80)-C(53)	111.0(2)
C(83)-C(80)-C(82)	109.3(3)	C(81)-C(80)-C(82)	108.9(3)
C(53)-C(80)-C(82)	108.3(2)	C(80)-C(81)-H(81A)	109.5(2)
C(80)-C(81)-H(81B)	109.5(2)	H(81A)-C(81)-H(81B)	109.5
C(80)-C(81)-H(81C)	109.5(2)	H(81A)-C(81)-H(81C)	109.5
H(81B)-C(81)-H(81C)	109.5	C(80)-C(82)-H(82A)	109.5(2)
C(80)-C(82)-H(82B)	109.5(2)	H(82A)-C(82)-H(82B)	109.5
C(80)-C(82)-H(82C)	109.5(2)	H(82A)-C(82)-H(82C)	109.5
H(82B)-C(82)-H(82C)	109.5	C(80)-C(83)-H(83A)	109.5(2)
C(80)-C(83)-H(83B)	109.5(2)	H(83A)-C(83)-H(83B)	109.5
C(80)-C(83)-H(83C)	109.5(2)	H(83A)-C(83)-H(83C)	109.5
H(83B)-C(83)-H(83C)	109.5		

Table 3A.2.4. Anisotropic displacement parameters [Å^2] for (2).

Molecule 1

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C(1)	0.024(1)	0.020(1)	0.024(1)	0.008(1)	0.011(1)	0.007(1)
Cl(1)	0.041(1)	0.028(1)	0.050(1)	0.024(1)	0.021(1)	0.012(1)
C(2)	0.017(1)	0.019(1)	0.018(1)	0.005(1)	0.006(1)	0.002(1)
C(3)	0.019(1)	0.020(1)	0.017(1)	0.010(1)	0.007(1)	0.004(1)
C(4)	0.021(1)	0.018(1)	0.022(1)	0.006(1)	0.007(1)	0.005(1)
C(9)	0.020(1)	0.020(1)	0.034(2)	0.013(1)	0.008(1)	0.003(1)
C(10)	0.021(1)	0.031(2)	0.026(2)	0.018(1)	0.009(1)	0.001(1)
C(11)	0.018(1)	0.029(1)	0.015(1)	0.007(1)	0.007(1)	0.000(1)
C(12)	0.020(1)	0.032(2)	0.019(1)	0.008(1)	0.011(1)	0.004(1)
Cl(12)	0.024(1)	0.060(1)	0.028(1)	0.014(1)	0.015(1)	0.003(1)
C(13)	0.016(1)	0.024(1)	0.020(1)	0.004(1)	0.003(1)	0.004(1)
Cl(13)	0.021(1)	0.032(1)	0.033(1)	0.004(1)	0.005(1)	-0.004(1)
C(14)	0.021(1)	0.028(1)	0.015(1)	0.007(1)	0.006(1)	0.010(1)
Cl(14)	0.036(1)	0.045(1)	0.016(1)	0.010(1)	0.009(1)	0.012(1)
C(15)	0.018(1)	0.025(1)	0.021(1)	-0.001(1)	0.006(1)	0.005(1)
Cl(1A)	0.025(1)	0.033(1)	0.029(1)	-0.009(1)	0.008(1)	0.004(1)
Cl(1B)	0.025(1)	0.031(1)	0.036(1)	0.004(1)	0.007(1)	0.013(1)
O(16)	0.018(1)	0.031(1)	0.022(1)	0.017(1)	0.006(1)	0.003(1)
C(30)	0.019(1)	0.020(1)	0.025(1)	0.007(1)	0.011(1)	0.002(1)
C(31)	0.022(1)	0.029(2)	0.031(2)	0.009(1)	0.011(1)	-0.003(1)
C(32)	0.026(2)	0.029(2)	0.051(2)	0.015(2)	0.021(2)	0.009(1)
C(33)	0.035(2)	0.033(2)	0.028(2)	0.014(1)	0.019(1)	0.003(1)

Molecule 2

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
Cl(51)	0.034(1)	0.034(1)	0.030(1)	-0.006(1)	0.016(1)	0.002(1)
Cl(62)	0.024(1)	0.043(1)	0.045(1)	0.018(1)	0.022(1)	0.006(1)
Cl(63)	0.034(1)	0.033(1)	0.050(1)	0.020(1)	0.014(1)	0.019(1)
Cl(64)	0.057(1)	0.060(1)	0.034(1)	0.033(1)	0.017(1)	0.017(1)
C(51)	0.024(1)	0.020(1)	0.022(1)	0.004(1)	0.011(1)	0.005(1)
C(52)	0.019(1)	0.019(1)	0.023(1)	0.009(1)	0.010(1)	0.005(1)
C(53)	0.024(1)	0.020(1)	0.018(1)	0.004(1)	0.009(1)	0.000(1)
C(54)	0.034(2)	0.018(1)	0.028(2)	0.008(1)	0.011(1)	0.000(1)
C(59)	0.030(2)	0.019(1)	0.031(2)	0.004(1)	0.010(1)	0.005(1)
C(60)	0.023(1)	0.030(2)	0.020(1)	0.006(1)	0.010(1)	0.004(1)
C(61)	0.020(1)	0.023(1)	0.024(1)	0.013(1)	0.011(1)	0.006(1)
C(62)	0.019(1)	0.023(1)	0.026(1)	0.012(1)	0.012(1)	0.007(1)
C(63)	0.023(1)	0.022(1)	0.026(2)	0.009(1)	0.005(1)	0.007(1)
C(64)	0.030(2)	0.028(2)	0.024(2)	0.013(1)	0.007(1)	0.006(1)
C(65)	0.018(1)	0.019(1)	0.034(2)	0.009(1)	0.007(1)	0.003(1)
Cl(6A)	0.030(1)	0.022(1)	0.066(1)	0.022(1)	0.019(1)	0.005(1)
Cl(6B)	0.021(1)	0.035(1)	0.041(1)	0.000(1)	0.005(1)	-0.002(1)
O(66)	0.020(1)	0.029(1)	0.022(1)	0.009(1)	0.008(1)	0.003(1)
C(80)	0.028(2)	0.027(2)	0.036(2)	0.004(1)	0.019(1)	-0.004(1)
C(81)	0.026(2)	0.038(2)	0.076(3)	0.017(2)	0.031(2)	0.010(1)
C(82)	0.032(2)	0.036(2)	0.051(2)	0.003(2)	0.022(2)	-0.009(1)
C(83)	0.052(2)	0.054(2)	0.042(2)	0.006(2)	0.034(2)	-0.011(2)

Table 3A.2.5. Hydrogen coordinates and isotropic displacement parameters [Å^2] for (2).

Molecule 1

Atom	x	y	z	U(eq)
H(2)	0.0192(21)	0.8827(20)	0.1318(19)	0.022
H(4)	0.1135(6)	1.1631(10)	0.4000(26)	0.025
H(9)	0.2166(24)	1.2315(23)	0.3026(13)	0.029
H(10)	0.1091(3)	1.1310(11)	0.1010(18)	0.029
H(11)	0.1229(12)	0.9514(15)	0.0582(27)	0.025
H(31A)	-0.2381(18)	0.9201(7)	0.0795(14)	0.042
H(31B)	-0.3054(19)	0.8976(10)	0.1570(9)	0.042
H(31C)	-0.1749(13)	0.8469(13)	0.1444(11)	0.042
H(32A)	-0.2888(21)	1.0849(8)	0.2723(20)	0.051
H(32B)	-0.2326(21)	1.1151(12)	0.1921(16)	0.051
H(32C)	-0.1497(14)	1.1646(14)	0.3241(18)	0.051
H(33A)	-0.0369(22)	1.0526(14)	0.4204(13)	0.044
H(33B)	-0.0332(21)	0.9302(17)	0.3516(8)	0.044
H(33C)	-0.1718(16)	0.9666(17)	0.3651(10)	0.044

Molecule 2

Atom	x	y	z	U(eq)
H(52)	0.4747(15)	0.4214(19)	-0.2056(4)	0.023
H(54)	0.4812(4)	0.7058(12)	-0.2268(19)	0.033
H(59)	0.3632(19)	0.7592(22)	-0.0999(5)	0.034
H(60)	0.4238(4)	0.6541(10)	0.0150(25)	0.030
H(61)	0.3669(6)	0.4726(15)	-0.0939(14)	0.024
H(81A)	0.6707(21)	0.4048(15)	-0.2311(18)	0.067
H(81B)	0.7232(26)	0.4652(11)	-0.1052(20)	0.067
H(81C)	0.8146(23)	0.4652(11)	-0.1688(23)	0.067
H(82A)	0.7452(13)	0.7264(15)	-0.1061(15)	0.063
H(82B)	0.8627(21)	0.6581(10)	-0.1212(14)	0.063
H(82C)	0.8047(20)	0.6640(11)	-0.0279(17)	0.063
H(83A)	0.5821(27)	0.4915(20)	-0.3605(13)	0.074
H(83B)	0.7377(22)	0.5452(22)	-0.3185(9)	0.074
H(83C)	0.6199(28)	0.6187(18)	-0.3156(8)	0.074

Table 3A.3.1. Atomic coordinates and equivalent isotropic displacement parameters [$Å^2$] for (3).

Atom	x	y	z	U(eq)
C(1)	0.6365(2)	0.2178(3)	0.9354(1)	0.037(1)
Cl(1)	0.5801(1)	0.1727(1)	1.0194(1)	0.057(1)
C(2)	0.6392(2)	0.3988(3)	0.9148(1)	0.036(1)
C(3)	0.7007(2)	0.5302(3)	0.9522(1)	0.037(1)
C(4)	0.7979(2)	0.4739(2)	0.9624(1)	0.033(1)
C(5)	0.8545(2)	0.5791(3)	1.0164(1)	0.040(1)
Cl(5)	0.8381(1)	0.5574(1)	1.1154(1)	0.065(1)
C(6)	0.9168(2)	0.6751(3)	0.9925(1)	0.040(1)
Cl(6)	0.9790(1)	0.7902(1)	1.0586(1)	0.066(1)
C(7)	0.9411(2)	0.6731(3)	0.9107(2)	0.042(1)
Cl(7)	1.0272(1)	0.7921(1)	0.8858(1)	0.069(1)
C(8)	0.9031(2)	0.5728(3)	0.8590(1)	0.043(1)
Cl(8)	0.9457(1)	0.5409(1)	0.7686(1)	0.076(1)
C(9)	0.8250(2)	0.4703(3)	0.8748(1)	0.035(1)
C(10)	0.7384(2)	0.5270(3)	0.8292(1)	0.039(1)
C(11)	0.6648(2)	0.3989(3)	0.8279(1)	0.037(1)
C(12)	0.6718(2)	0.2181(3)	0.8084(1)	0.040(1)
Cl(12)	0.6657(1)	0.1760(1)	0.7071(1)	0.070(1)
C(13)	0.7458(2)	0.1369(3)	0.8563(2)	0.042(1)
Cl(13)	0.8355(1)	0.0544(1)	0.8177(1)	0.067(1)
C(14)	0.7252(2)	0.1367(3)	0.9309(1)	0.040(1)
Cl(14)	0.7822(1)	0.0511(1)	1.0084(1)	0.062(1)
C(15)	0.5942(2)	0.1523(3)	0.8555(1)	0.043(1)
Cl(1A)	0.4905(1)	0.2338(1)	0.8272(1)	0.061(1)
Cl(1B)	0.5842(1)	-0.0583(1)	0.8525(1)	0.060(1)
C(31)	0.6600(2)	0.6187(4)	1.0185(2)	0.052(1)
C(101)	0.7445(2)	0.6117(4)	0.7511(2)	0.058(1)
O(16)	0.7046(1)	0.6360(2)	0.8856(1)	0.043(1)

Table 3A.3.2. Bond lengths [Å] for (3)

C(1)-C(14)	1.523(3)	C(1)-C(2)	1.555(3)
C(1)-C(15)	1.569(3)	C(1)-Cl(1)	1.753(2)
C(2)-C(11)	1.551(3)	C(2)-C(3)	1.559(3)
C(2)-H(2)	0.97(2)	C(3)-O(16)	1.442(3)
C(3)-C(31)	1.513(3)	C(3)-C(4)	1.559(3)
C(4)-C(5)	1.509(3)	C(4)-C(9)	1.571(3)
C(4)-H(4)	0.95(2)	C(5)-C(6)	1.328(3)
C(5)-Cl(5)	1.727(2)	C(6)-C(7)	1.462(3)
C(6)-Cl(6)	1.725(2)	C(7)-C(8)	1.325(3)
C(7)-Cl(7)	1.723(2)	C(8)-C(9)	1.508(3)
C(8)-Cl(8)	1.726(2)	C(9)-C(10)	1.568(3)
C(9)-H(9)	0.96(2)	C(10)-O(16)	1.443(3)
C(10)-C(101)	1.513(3)	C(10)-C(11)	1.554(3)
C(11)-C(12)	1.553(3)	C(11)-H(11)	0.99(3)
C(12)-C(13)	1.517(3)	C(12)-C(15)	1.570(3)
C(12)-Cl(12)	1.755(2)	C(13)-C(14)	1.326(3)
C(13)-Cl(13)	1.701(2)	C(14)-Cl(14)	1.697(2)
C(15)-Cl(1A)	1.768(3)	C(15)-Cl(1B)	1.769(3)
C(31)-H(313)	0.93(3)	C(31)-H(312)	0.95(3)
C(31)-H(311)	0.92(3)	C(101)-H(10C)	0.96(3)
C(101)-H(10B)	0.91(4)	C(101)-H(10A)	1.00(3)

Table 3A.3.3. Bond angles [°] for (3)

C(14)-C(1)-C(2)	112.7(2)	C(14)-C(1)-C(15)	97.1(2)
C(2)-C(1)-C(15)	99.2(2)	C(14)-C(1)-Cl(1)	115.3(2)
C(2)-C(1)-Cl(1)	114.5(2)	C(15)-C(1)-Cl(1)	115.7(2)
C(11)-C(2)-C(1)	103.1(2)	C(11)-C(2)-C(3)	102.1(2)
C(1)-C(2)-C(3)	128.1(2)	C(11)-C(2)-H(2)	110.0(14)
C(1)-C(2)-H(2)	106(2)	C(3)-C(2)-H(2)	107(2)
O(16)-C(3)-C(31)	108.8(2)	O(16)-C(3)-C(4)	101.0(2)
C(31)-C(3)-C(4)	119.8(2)	O(16)-C(3)-C(2)	99.3(2)
C(31)-C(3)-C(2)	112.6(2)	C(4)-C(3)-C(2)	112.4(2)
C(5)-C(4)-C(3)	113.6(2)	C(5)-C(4)-C(9)	114.6(2)
C(3)-C(4)-C(9)	101.7(2)	C(5)-C(4)-H(4)	105.0(14)
C(3)-C(4)-H(4)	111.0(14)	C(9)-C(4)-H(4)	111.2(14)
C(6)-C(5)-C(4)	124.2(2)	C(6)-C(5)-Cl(5)	120.7(2)
C(4)-C(5)-Cl(5)	115.0(2)	C(5)-C(6)-C(7)	120.8(2)
C(5)-C(6)-Cl(6)	121.0(2)	C(7)-C(6)-Cl(6)	117.8(2)
C(8)-C(7)-C(6)	120.9(2)	C(8)-C(7)-Cl(7)	121.0(2)
C(6)-C(7)-Cl(7)	117.9(2)	C(7)-C(8)-C(9)	124.3(2)
C(7)-C(8)-Cl(8)	121.0(2)	C(9)-C(8)-Cl(8)	114.6(2)
C(8)-C(9)-C(10)	113.3(2)	C(8)-C(9)-C(4)	114.5(2)
C(10)-C(9)-C(4)	101.7(2)	C(8)-C(9)-H(9)	106.1(14)
C(10)-C(9)-H(9)	109.9(14)	C(4)-C(9)-H(9)	111.5(14)
O(16)-C(10)-C(101)	109.4(2)	O(16)-C(10)-C(11)	99.0(2)
C(101)-C(10)-C(11)	113.1(2)	O(16)-C(10)-C(9)	100.8(2)
C(101)-C(10)-C(9)	118.9(2)	C(11)-C(10)-C(9)	112.9(2)
C(2)-C(11)-C(12)	103.1(2)	C(2)-C(11)-C(10)	102.1(2)
C(12)-C(11)-C(10)	128.0(2)	C(2)-C(11)-H(11)	112(2)
C(12)-C(11)-H(11)	107(2)	C(10)-C(11)-H(11)	105(2)
C(13)-C(12)-C(11)	112.3(2)	C(13)-C(12)-C(15)	97.6(2)
C(11)-C(12)-C(15)	99.7(2)	C(13)-C(12)-Cl(12)	115.6(2)
C(11)-C(12)-Cl(12)	113.9(2)	C(15)-C(12)-Cl(12)	115.6(2)
C(14)-C(13)-C(12)	107.6(2)	C(14)-C(13)-Cl(13)	127.7(2)
C(12)-C(13)-Cl(13)	124.6(2)	C(13)-C(14)-C(1)	108.1(2)
C(13)-C(14)-Cl(14)	127.2(2)	C(1)-C(14)-Cl(14)	124.6(2)
C(1)-C(15)-C(12)	92.0(2)	C(1)-C(15)-Cl(1A)	114.8(2)
C(12)-C(15)-Cl(1A)	114.8(2)	C(1)-C(15)-Cl(1B)	113.8(2)
C(12)-C(15)-Cl(1B)	113.7(2)	Cl(1A)-C(15)-Cl(1B)	107.50(13)
C(3)-C(31)-H(313)	112(2)	C(3)-C(31)-H(312)	108(2)
H(313)-C(31)-H(312)	109(3)	C(3)-C(31)-H(311)	112(2)
H(313)-C(31)-H(311)	110(3)	H(312)-C(31)-H(311)	106(3)
C(10)-C(101)-H(10C)	113(2)	C(10)-C(101)-H(10B)	109(2)
H(10C)-C(101)-H(10B)	108(3)	C(10)-C(101)-H(10A)	111(2)
H(10C)-C(101)-H(10A)	110(3)	H(10B)-C(101)-H(10A)	106(3)
C(3)-O(16)-C(10)	99.7(2)		

Table 3A.3.4. Anisotropic displacement parameters [Å^2] for (3).

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C(1)	0.036(1)	0.039(1)	0.035(1)	0.000(1)	0.003(1)	-0.005(1)
Cl(1)	0.061(1)	0.065(1)	0.047(1)	0.002(1)	0.013(1)	-0.020(1)
C(2)	0.029(1)	0.037(1)	0.042(1)	0.000(1)	0.004(1)	0.003(1)
C(3)	0.041(1)	0.030(1)	0.041(1)	-0.001(1)	0.007(1)	0.001(1)
C(4)	0.038(1)	0.028(1)	0.034(1)	0.000(1)	0.001(1)	-0.002(1)
C(5)	0.046(1)	0.038(1)	0.034(1)	-0.003(1)	0.000(1)	0.000(1)
Cl(5)	0.079(1)	0.081(1)	0.035(1)	-0.005(1)	0.001(1)	-0.013(1)
C(6)	0.043(1)	0.030(1)	0.046(1)	-0.005(1)	-0.003(1)	-0.001(1)
Cl(6)	0.070(1)	0.056(1)	0.071(1)	-0.017(1)	-0.012(1)	-0.019(1)
C(7)	0.038(1)	0.036(1)	0.054(1)	0.006(1)	0.002(1)	-0.005(1)
Cl(7)	0.056(1)	0.064(1)	0.088(1)	0.008(1)	0.009(1)	-0.026(1)
C(8)	0.038(1)	0.051(1)	0.040(1)	0.003(1)	0.008(1)	-0.003(1)
Cl(8)	0.055(1)	0.126(1)	0.049(1)	-0.012(1)	0.021(1)	-0.020(1)
C(9)	0.036(1)	0.033(1)	0.035(1)	-0.003(1)	0.003(1)	-0.001(1)
C(10)	0.040(1)	0.040(1)	0.038(1)	0.006(1)	0.004(1)	0.001(1)
C(11)	0.035(1)	0.041(1)	0.035(1)	0.004(1)	-0.002(1)	0.001(1)
C(12)	0.042(1)	0.045(1)	0.034(1)	-0.006(1)	-0.001(1)	0.000(1)
Cl(12)	0.082(1)	0.089(1)	0.038(1)	-0.018(1)	0.002(1)	-0.009(1)
C(13)	0.036(1)	0.036(1)	0.054(1)	-0.009(1)	0.002(1)	0.001(1)
Cl(13)	0.047(1)	0.059(1)	0.096(1)	-0.031(1)	0.010(1)	0.009(1)
C(14)	0.039(1)	0.031(1)	0.048(1)	0.000(1)	-0.007(1)	-0.002(1)
Cl(14)	0.067(1)	0.044(1)	0.073(1)	0.016(1)	-0.029(1)	-0.005(1)
C(15)	0.038(1)	0.044(1)	0.044(1)	-0.005(1)	-0.005(1)	-0.005(1)
Cl(1A)	0.035(1)	0.076(1)	0.070(1)	-0.001(1)	-0.012(1)	-0.003(1)
Cl(1B)	0.064(1)	0.046(1)	0.068(1)	-0.011(1)	-0.003(1)	-0.017(1)
C(31)	0.055(2)	0.044(1)	0.058(2)	-0.012(1)	0.016(1)	0.003(1)
C(101)	0.060(2)	0.065(2)	0.048(2)	0.021(1)	-0.002(1)	-0.008(2)
O(16)	0.047(1)	0.031(1)	0.051(1)	0.005(1)	0.004(1)	0.006(1)

Table 3A.3.5. Hydrogen coordinates and isotropic displacement parameters [Å^2] for (3).

Atom	x	y	z	$U(eq)$
H(2)	0.5796(16)	0.4379(29)	0.9180(14)	0.035(6)
H(4)	0.8018(15)	0.3707(30)	0.9855(14)	0.033(6)
H(9)	0.8394(15)	0.3642(30)	0.8586(14)	0.035(6)
H(11)	0.6176(18)	0.4438(34)	0.7922(17)	0.054(8)
H(313)	0.6516(20)	0.5520(39)	1.0612(19)	0.063(9)
H(312)	0.6048(23)	0.6585(40)	0.9986(19)	0.068(10)
H(311)	0.6922(18)	0.7065(36)	1.0347(16)	0.048(8)
H(10C)	0.7831(21)	0.7026(40)	0.7545(18)	0.062(9)
H(10B)	0.6902(24)	0.6469(45)	0.7346(22)	0.080(11)
H(10A)	0.7633(21)	0.5371(39)	0.7095(20)	0.062(9)

Table 3A.4.1. Atomic coordinates and equivalent isotropic displacement parameters [Å^2] for (4)

Atom	x	y	z	U(eq)
C(1)	0.4016(2)	0.2595(3)	0.4347(2)	0.018(1)
Cl(1)	0.4716(1)	0.1885(1)	0.3864(1)	0.026(1)
C(2)	0.3579(2)	0.1981(2)	0.4939(2)	0.018(1)
C(3)	0.3845(2)	0.1685(2)	0.5749(2)	0.017(1)
C(31)	0.4296(3)	0.0735(3)	0.5805(2)	0.023(1)
C(4)	0.4094(2)	0.2585(2)	0.6213(2)	0.014(1)
C(5)	0.4799(2)	0.2889(2)	0.6508(2)	0.017(1)
Cl(5)	0.5654(1)	0.2303(1)	0.6266(1)	0.024(1)
C(6)	0.4817(2)	0.3708(2)	0.6986(2)	0.017(1)
Cl(6)	0.5674(1)	0.4045(1)	0.7405(1)	0.024(1)
C(7)	0.4139(2)	0.4256(2)	0.7116(2)	0.018(1)
Cl(7)	0.4165(1)	0.5285(1)	0.7685(1)	0.027(1)
C(8)	0.3435(2)	0.3974(2)	0.6783(2)	0.017(1)
Cl(8)	0.2625(1)	0.4713(1)	0.6889(1)	0.030(1)
C(9)	0.3417(2)	0.3123(2)	0.6349(2)	0.016(1)
C(10)	0.2766(2)	0.2552(2)	0.5958(2)	0.018(1)
C(101)	0.1949(2)	0.2602(3)	0.6262(2)	0.028(1)
C(11)	0.2834(2)	0.2584(3)	0.5085(2)	0.015(1)
C(12)	0.2926(2)	0.3493(2)	0.4560(2)	0.017(1)
Cl(12)	0.2038(1)	0.4102(1)	0.4374(1)	0.026(1)
C(13)	0.3581(2)	0.4189(2)	0.4835(2)	0.017(1)
Cl(13)	0.3558(1)	0.5404(1)	0.4438(1)	0.025(1)
C(14)	0.4336(2)	0.3577(3)	0.4678(2)	0.017(1)
Cl(14)	0.5050(1)	0.4175(1)	0.4120(1)	0.025(1)
C(15)	0.3306(2)	0.2960(2)	0.3878(2)	0.019(1)
Cl(1A)	0.2739(1)	0.1971(1)	0.3502(1)	0.024(1)
Cl(1B)	0.3537(1)	0.3732(1)	0.3114(1)	0.026(1)
O(16)	0.3073(1)	0.1551(2)	0.6058(1)	0.020(1)

Table 3A.4.2. Bond lengths (Å) for (4).

C(1)-C(2)	1.542(5)	C(1)-C(14)	1.550(5)
C(1)-C(15)	1.564(5)	C(1)-Cl(1)	1.766(3)
C(2)-C(11)	1.541(5)	C(2)-C(3)	1.575(5)
C(2)-H(2)	0.94(3)	C(3)-O(16)	1.453(4)
C(3)-C(31)	1.498(5)	C(3)-C(4)	1.530(5)
C(31)-H(311)	0.97(3)	C(31)-H(312)	0.95(4)
C(31)-H(313)	0.93(4)	C(4)-C(5)	1.388(5)
C(4)-C(9)	1.393(5)	C(5)-C(6)	1.397(4)
C(5)-Cl(5)	1.726(3)	C(6)-C(7)	1.399(5)
C(6)-Cl(6)	1.718(3)	C(7)-C(8)	1.404(5)
C(7)-Cl(7)	1.722(3)	C(8)-C(9)	1.385(4)
C(8)-Cl(8)	1.725(3)	C(9)-C(10)	1.531(5)
C(10)-O(16)	1.459(4)	C(10)-C(101)	1.510(5)
C(10)-C(11)	1.571(5)	C(101)-H(10A)	1.00(4)
C(101)-H(10B)	0.97(4)	C(101)-H(10C)	0.95(4)
C(11)-C(12)	1.552(5)	C(11)-H(11)	0.90(3)
C(12)-C(13)	1.547(5)	C(12)-C(15)	1.563(5)
C(12)-Cl(12)	1.767(3)	C(13)-C(14)	1.565(5)
C(13)-Cl(13)	1.784(3)	C(13)-H(13)	0.93(3)
C(14)-Cl(14)	1.778(3)	C(14)-H(14)	0.97(3)
C(15)-Cl(1B)	1.766(3)	C(15)-Cl(1A)	1.784(3)

Table 3A.4.3. Bond angles [°] for (4)

C(2)-C(1)-C(14)	111.5(3)	C(2)-C(1)-C(15)	99.1(3)
C(14)-C(1)-C(15)	102.5(3)	C(2)-C(1)-Cl(1)	112.4(2)
C(14)-C(1)-Cl(1)	114.0(2)	C(15)-C(1)-Cl(1)	116.1(2)
C(11)-C(2)-C(1)	103.9(3)	C(11)-C(2)-C(3)	102.6(3)
C(1)-C(2)-C(3)	129.0(3)	C(11)-C(2)-H(2)	110(2)
C(1)-C(2)-H(2)	108(2)	C(3)-C(2)-H(2)	103(2)
O(16)-C(3)-C(31)	110.0(3)	O(16)-C(3)-C(4)	98.4(2)
C(31)-C(3)-C(4)	119.7(3)	O(16)-C(3)-C(2)	96.8(3)
C(31)-C(3)-C(2)	115.3(3)	C(4)-C(3)-C(2)	112.5(3)
C(3)-C(31)-H(311)	107(2)	C(3)-C(31)-H(312)	112(2)
H(311)-C(31)-H(312)	110(3)	C(3)-C(31)-H(313)	111(2)
H(311)-C(31)-H(313)	107(3)	H(312)-C(31)-H(313)	109(3)
C(5)-C(4)-C(9)	120.7(3)	C(5)-C(4)-C(3)	133.3(3)
C(9)-C(4)-C(3)	105.9(3)	C(4)-C(5)-C(6)	119.1(3)
C(4)-C(5)-Cl(5)	121.0(3)	C(6)-C(5)-Cl(5)	119.8(3)
C(5)-C(6)-C(7)	120.0(3)	C(5)-C(6)-Cl(6)	119.7(3)
C(7)-C(6)-Cl(6)	120.3(3)	C(6)-C(7)-C(8)	120.4(3)
C(6)-C(7)-Cl(7)	120.1(3)	C(8)-C(7)-Cl(7)	119.5(3)
C(9)-C(8)-C(7)	118.8(3)	C(9)-C(8)-Cl(8)	121.4(3)
C(7)-C(8)-Cl(8)	119.7(3)	C(8)-C(9)-C(4)	120.7(3)
C(8)-C(9)-C(10)	133.6(3)	C(4)-C(9)-C(10)	105.7(3)
O(16)-C(10)-C(101)	109.5(3)	O(16)-C(10)-C(9)	98.2(2)
C(101)-C(10)-C(9)	119.6(3)	O(16)-C(10)-C(11)	97.0(2)
C(101)-C(10)-C(11)	115.4(3)	C(9)-C(10)-C(11)	112.9(3)
C(10)-C(101)-H(10A)	110(2)	C(10)-C(101)-H(10B)	107(2)
H(10A)-C(101)-H(10B)	107(3)	C(10)-C(101)-H(10C)	110(2)
H(10A)-C(101)-H(10C)	110(3)	H(10B)-C(101)-H(10C)	112(3)
C(2)-C(11)-C(12)	103.2(3)	C(2)-C(11)-C(10)	102.5(3)
C(12)-C(11)-C(10)	129.4(3)	C(2)-C(11)-H(11)	111(2)
C(12)-C(11)-H(11)	106(2)	C(10)-C(11)-H(11)	105(2)
C(13)-C(12)-C(11)	111.0(3)	C(13)-C(12)-C(15)	102.9(3)
C(11)-C(12)-C(15)	99.0(3)	C(13)-C(12)-Cl(12)	114.2(2)
C(11)-C(12)-Cl(12)	113.1(2)	C(15)-C(12)-Cl(12)	115.3(2)
C(12)-C(13)-C(14)	103.2(3)	C(12)-C(13)-Cl(13)	114.4(2)
C(14)-C(13)-Cl(13)	115.4(2)	C(12)-C(13)-H(13)	112(2)
C(14)-C(13)-H(13)	109(2)	Cl(13)-C(13)-H(13)	103(2)
C(1)-C(14)-C(13)	102.9(3)	C(1)-C(14)-Cl(14)	114.6(2)
C(13)-C(14)-Cl(14)	115.9(2)	C(1)-C(14)-H(14)	112(2)
C(13)-C(14)-H(14)	110(2)	Cl(14)-C(14)-H(14)	101(2)
C(12)-C(15)-C(1)	92.8(2)	C(12)-C(15)-Cl(1B)	115.7(2)
C(1)-C(15)-Cl(1B)	115.3(2)	C(12)-C(15)-Cl(1A)	114.2(2)
C(1)-C(15)-Cl(1A)	113.3(2)	Cl(1B)-C(15)-Cl(1A)	105.6(2)
C(3)-O(16)-C(10)	99.8(2)		

Table 3A.4.4. Anisotropic displacement parameters [Å^2] for (4).

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C(1)	0.020(2)	0.019(2)	0.013(2)	-0.003(2)	0.001(2)	0.003(2)
Cl(1)	0.027(1)	0.031(1)	0.021(1)	-0.005(1)	0.005(1)	0.007(1)
C(2)	0.022(2)	0.012(2)	0.019(2)	-0.002(2)	-0.001(2)	-0.002(2)
C(3)	0.020(2)	0.017(2)	0.015(2)	0.003(2)	0.002(2)	0.000(2)
C(31)	0.030(3)	0.016(2)	0.023(2)	0.000(2)	-0.002(2)	-0.001(2)
C(4)	0.017(2)	0.015(2)	0.011(2)	0.002(1)	0.001(2)	0.000(2)
C(5)	0.020(2)	0.019(2)	0.013(2)	0.007(2)	0.001(2)	0.002(2)
Cl(5)	0.019(1)	0.023(1)	0.030(1)	0.003(1)	0.002(1)	0.004(1)
C(6)	0.017(2)	0.021(2)	0.013(2)	0.004(2)	-0.001(2)	-0.004(2)
Cl(6)	0.023(1)	0.030(1)	0.021(1)	0.002(1)	-0.004(1)	-0.008(1)
C(7)	0.026(2)	0.020(2)	0.007(2)	0.001(2)	-0.003(2)	-0.004(2)
Cl(7)	0.036(1)	0.023(1)	0.022(1)	-0.007(1)	-0.001(1)	-0.002(1)
C(8)	0.020(2)	0.020(2)	0.011(2)	0.004(2)	0.006(2)	0.004(2)
Cl(8)	0.027(1)	0.029(1)	0.032(1)	-0.006(1)	-0.001(1)	0.012(1)
C(9)	0.017(2)	0.019(2)	0.011(2)	0.006(2)	0.002(2)	-0.001(2)
C(10)	0.020(2)	0.016(2)	0.018(2)	0.001(2)	0.000(2)	0.000(2)
C(101)	0.022(2)	0.034(2)	0.028(2)	0.001(2)	0.004(2)	-0.004(2)
C(11)	0.012(2)	0.017(2)	0.017(2)	-0.003(2)	-0.001(2)	-0.006(2)
C(12)	0.013(2)	0.019(2)	0.019(2)	-0.002(2)	-0.002(2)	0.002(2)
Cl(12)	0.020(1)	0.027(1)	0.031(1)	-0.001(1)	-0.007(1)	0.005(1)
C(13)	0.023(2)	0.016(2)	0.012(2)	0.004(2)	0.000(2)	-0.001(2)
Cl(13)	0.033(1)	0.016(1)	0.026(1)	0.004(1)	-0.006(1)	-0.002(1)
C(14)	0.018(2)	0.020(2)	0.014(2)	0.001(2)	0.000(2)	-0.004(2)
Cl(14)	0.022(1)	0.032(1)	0.023(1)	0.006(1)	0.002(1)	-0.008(1)
C(15)	0.022(2)	0.019(2)	0.015(2)	0.000(2)	0.000(2)	-0.003(2)
Cl(1A)	0.028(1)	0.026(1)	0.020(1)	-0.006(1)	-0.004(1)	-0.007(1)
Cl(1B)	0.034(1)	0.029(1)	0.015(1)	0.005(1)	-0.003(1)	-0.004(1)
O(16)	0.021(2)	0.019(1)	0.021(1)	0.002(1)	0.004(1)	-0.006(1)

Table 3A.4.5. Hydrogen coordinates and isotropic displacement parameters for 4.

Atom	x	y	z	$U(eq)$
H(2)	0.3455(19)	0.1360(25)	0.4725(17)	0.021
H(311)	0.3965(20)	0.0216(26)	0.5606(19)	0.028
H(312)	0.4441(19)	0.0583(24)	0.6305(21)	0.028
H(313)	0.4738(20)	0.0762(25)	0.5508(20)	0.028
H(10A)	0.1958(19)	0.2526(25)	0.6815(22)	0.034
H(10B)	0.1665(21)	0.2037(27)	0.6060(21)	0.034
H(10C)	0.1712(21)	0.3214(27)	0.6128(21)	0.034
H(11)	0.2411(19)	0.2257(23)	0.4921(17)	0.018
H(13)	0.3540(18)	0.4313(23)	0.5346(19)	0.021
H(14)	0.4620(18)	0.3471(23)	0.5137(18)	0.021

Table 3A.5.1 Atomic coordinates for (5)

Atom	x	y	z	U(eq)
C(1)	0.2626(3)	0.3424(3)	1.1023(3)	0.040(1)
Cl(1)	0.3846(1)	0.4158(1)	1.1495(1)	0.063(1)
C(2)	0.2463(3)	0.2673(3)	1.1831(3)	0.044(1)
C(3)	0.1883(4)	0.2767(3)	1.2816(3)	0.047(1)
C(4)	0.0632(3)	0.3170(3)	1.2571(3)	0.039(1)
C(5)	0.0185(4)	0.3959(3)	1.2942(3)	0.045(1)
Cl(5)	0.1137(1)	0.4680(1)	1.3767(1)	0.074(1)
C(6)	-0.1046(4)	0.4143(3)	1.2657(3)	0.048(1)
Cl(6)	-0.1627(1)	0.5099(1)	1.3161(1)	0.075(1)
C(7)	-0.1785(4)	0.3558(3)	1.1971(3)	0.051(1)
Cl(7)	-0.3287(1)	0.3811(1)	1.1573(1)	0.079(1)
C(8)	-0.1324(4)	0.2760(3)	1.1604(3)	0.046(1)
Cl(8)	-0.2273(1)	0.2058(1)	1.0732(1)	0.073(1)
C(9)	-0.0105(3)	0.2550(3)	1.1927(3)	0.040(1)
C(10)	0.0662(4)	0.1715(3)	1.1781(3)	0.048(1)
C(11)	0.1659(4)	0.1955(3)	1.1138(3)	0.046(1)
C(12)	0.1495(4)	0.2357(3)	1.0013(3)	0.045(1)
Cl(12)	0.1089(2)	0.1515(1)	0.9032(1)	0.083(1)
C(13)	0.0658(3)	0.3195(3)	0.9898(3)	0.040(1)
Cl(13)	0.0074(1)	0.3541(1)	0.8570(1)	0.062(1)
C(14)	0.1434(3)	0.3934(2)	1.0602(3)	0.034(1)
Cl(14)	0.1612(1)	0.4997(1)	0.9969(1)	0.052(1)
C(15)	0.2768(4)	0.2798(3)	1.0073(3)	0.047(1)
Cl(1A)	0.2970(1)	0.3386(1)	0.8926(1)	0.060(1)
Cl(1B)	0.3981(1)	0.2005(1)	1.0358(1)	0.078(1)
C(16)	0.1509(4)	0.1744(3)	1.2886(3)	0.057(1)
C(17)	0.0041(5)	0.0793(3)	1.1472(4)	0.074(2)

Table 3A.5.2. Bond lengths [Å] for (5).

C(1)-C(14)	1.543(5)	C(1)-C(2)	1.547(5)
C(1)-C(15)	1.562(5)	C(1)-Cl(1)	1.759(4)
C(2)-C(3)	1.546(5)	C(2)-C(11)	1.551(6)
C(2)-H(2)	0.98	C(3)-C(4)	1.510(5)
C(3)-C(16)	1.555(6)	C(3)-H(3)	0.98
C(4)-C(5)	1.377(5)	C(4)-C(9)	1.391(5)
C(5)-C(6)	1.397(6)	C(5)-Cl(5)	1.719(4)
C(6)-C(7)	1.387(6)	C(6)-Cl(6)	1.719(4)
C(7)-C(8)	1.393(6)	C(7)-Cl(7)	1.718(4)
C(8)-C(9)	1.396(5)	C(8)-Cl(8)	1.732(4)
C(9)-C(10)	1.527(5)	C(10)-C(17)	1.532(6)
C(10)-C(16)	1.559(6)	C(10)-C(11)	1.564(6)
C(11)-C(12)	1.546(5)	C(11)-H(11)	0.98
C(12)-C(13)	1.534(5)	C(12)-C(15)	1.565(6)
C(12)-Cl(12)	1.760(4)	C(13)-C(14)	1.564(5)
C(13)-Cl(13)	1.789(3)	C(13)-H(13)	0.98
C(14)-Cl(14)	1.779(3)	C(14)-H(14)	0.98
C(15)-Cl(1A)	1.765(4)	C(15)-Cl(1B)	1.777(4)
C(16)-H(16A)	0.97	C(16)-H(16B)	0.97
C(17)-H(17A)	0.96	C(17)-H(17B)	0.96
C(17)-H(17C)	0.96		

Table 3A.5.3 Bond Angles [°] for (5)

C(14)-C(1)-C(2)	111.6(3)	C(14)-C(1)-C(15)	102.7(3)
C(2)-C(1)-C(15)	99.3(3)	C(14)-C(1)-Cl(1)	113.9(3)
C(2)-C(1)-Cl(1)	112.4(3)	C(15)-C(1)-Cl(1)	115.7(3)
C(3)-C(2)-C(1)	128.2(3)	C(3)-C(2)-C(11)	103.9(3)
C(1)-C(2)-C(11)	102.7(3)	C(3)-C(2)-H(2)	106.8(2)
C(1)-C(2)-H(2)	106.8(2)	C(11)-C(2)-H(2)	106.8(2)
C(4)-C(3)-C(2)	113.1(3)	C(4)-C(3)-C(16)	97.5(3)
C(2)-C(3)-C(16)	97.3(3)	C(4)-C(3)-H(3)	115.4(2)
C(2)-C(3)-H(3)	115.4(2)	C(16)-C(3)-H(3)	115.4(2)
C(5)-C(4)-C(9)	122.1(4)	C(5)-C(4)-C(3)	130.1(4)
C(9)-C(4)-C(3)	107.6(3)	C(4)-C(5)-C(6)	118.8(4)
C(4)-C(5)-Cl(5)	119.8(3)	C(6)-C(5)-Cl(5)	121.4(3)
C(7)-C(6)-C(5)	119.9(4)	C(7)-C(6)-Cl(6)	120.8(3)
C(5)-C(6)-Cl(6)	119.3(3)	C(6)-C(7)-C(8)	120.7(4)
C(6)-C(7)-Cl(7)	120.0(3)	C(8)-C(7)-Cl(7)	119.3(4)
C(7)-C(8)-C(9)	119.6(4)	C(7)-C(8)-Cl(8)	119.2(3)
C(9)-C(8)-Cl(8)	121.2(3)	C(4)-C(9)-C(8)	118.7(4)
C(4)-C(9)-C(10)	107.5(3)	C(8)-C(9)-C(10)	133.7(4)
C(9)-C(10)-C(17)	119.1(4)	C(9)-C(10)-C(16)	97.5(3)
C(17)-C(10)-C(16)	115.6(4)	C(9)-C(10)-C(11)	111.9(3)
C(17)-C(10)-C(11)	113.5(4)	C(16)-C(10)-C(11)	95.8(3)
C(12)-C(11)-C(2)	103.8(3)	C(12)-C(11)-C(10)	128.0(4)
C(2)-C(11)-C(10)	104.3(3)	C(12)-C(11)-H(11)	106.3(2)
C(2)-C(11)-H(11)	106.3(2)	C(10)-C(11)-H(11)	106.3(2)
C(13)-C(12)-C(11)	110.6(3)	C(13)-C(12)-C(15)	103.0(3)
C(11)-C(12)-C(15)	99.5(3)	C(13)-C(12)-Cl(12)	114.4(3)
C(11)-C(12)-Cl(12)	112.4(3)	C(15)-C(12)-Cl(12)	115.6(3)
C(12)-C(13)-C(14)	102.9(3)	C(12)-C(13)-Cl(13)	115.0(3)
C(14)-C(13)-Cl(13)	115.3(2)	C(12)-C(13)-H(13)	107.7(2)
C(14)-C(13)-H(13)	107.7(2)	Cl(13)-C(13)-H(13)	107.72(12)
C(1)-C(14)-C(13)	103.2(3)	C(1)-C(14)-Cl(14)	114.3(2)
C(13)-C(14)-Cl(14)	115.5(2)	C(1)-C(14)-H(14)	107.8(2)
C(13)-C(14)-H(14)	107.8(2)	Cl(14)-C(14)-H(14)	107.81(11)
C(1)-C(15)-C(12)	92.6(3)	C(1)-C(15)-Cl(1A)	115.2(3)
C(12)-C(15)-Cl(1A)	115.1(3)	C(1)-C(15)-Cl(1B)	114.2(3)
C(12)-C(15)-Cl(1B)	114.3(3)	Cl(1A)-C(15)-Cl(1B)	105.6(2)
C(3)-C(16)-C(10)	95.6(3)	C(3)-C(16)-H(16A)	112.6(2)
C(10)-C(16)-H(16A)	112.6(2)	C(3)-C(16)-H(16B)	112.6(2)
C(10)-C(16)-H(16B)	112.6(2)	H(16A)-C(16)-H(16B)	110.1
C(10)-C(17)-H(17A)	109.5(3)	C(10)-C(17)-H(17B)	109.5(3)
H(17A)-C(17)-H(17B)	109.5	C(10)-C(17)-H(17C)	109.5(3)
H(17A)-C(17)-H(17C)	109.5	H(17B)-C(17)-H(17C)	109.5

Table 3A.5.4. Anisotropic displacement parameters [Å^2] for (5)

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C(1)	0.031(2)	0.043(2)	0.045(2)	-0.004(2)	0.005(2)	0.002(2)
Cl(1)	0.036(1)	0.074(1)	0.076(1)	-0.009(1)	0.001(1)	-0.010(1)
C(2)	0.039(2)	0.046(2)	0.045(2)	0.007(2)	0.003(2)	0.013(2)
C(3)	0.050(2)	0.054(2)	0.035(2)	0.006(2)	0.001(2)	0.006(2)
C(4)	0.048(2)	0.041(2)	0.029(2)	0.006(2)	0.008(2)	0.001(2)
C(5)	0.057(2)	0.046(2)	0.037(2)	-0.001(2)	0.018(2)	-0.008(2)
Cl(5)	0.089(1)	0.064(1)	0.066(1)	-0.023(1)	0.011(1)	-0.009(1)
C(6)	0.058(2)	0.048(2)	0.046(2)	0.013(2)	0.028(2)	0.014(2)
Cl(6)	0.096(1)	0.063(1)	0.080(1)	0.011(1)	0.050(1)	0.027(1)
C(7)	0.041(2)	0.070(3)	0.045(2)	0.021(2)	0.018(2)	0.006(2)
Cl(7)	0.043(1)	0.119(1)	0.079(1)	0.032(1)	0.023(1)	0.014(1)
C(8)	0.048(2)	0.057(2)	0.035(2)	0.010(2)	0.010(2)	-0.012(2)
Cl(8)	0.063(1)	0.094(1)	0.058(1)	0.000(1)	0.002(1)	-0.030(1)
C(9)	0.048(2)	0.041(2)	0.031(2)	0.007(2)	0.012(2)	-0.003(2)
C(10)	0.063(3)	0.036(2)	0.047(2)	0.006(2)	0.014(2)	0.002(2)
C(11)	0.061(2)	0.030(2)	0.051(2)	0.001(2)	0.017(2)	0.011(2)
C(12)	0.056(2)	0.039(2)	0.043(2)	-0.012(2)	0.015(2)	-0.004(2)
Cl(12)	0.123(1)	0.065(1)	0.067(1)	-0.035(1)	0.036(1)	-0.024(1)
C(13)	0.038(2)	0.051(2)	0.030(2)	0.003(2)	0.006(1)	-0.003(2)
Cl(13)	0.053(1)	0.093(1)	0.034(1)	0.008(1)	0.000(1)	-0.007(1)
C(14)	0.033(2)	0.035(2)	0.035(2)	0.005(2)	0.008(1)	0.002(2)
Cl(14)	0.053(1)	0.041(1)	0.066(1)	0.015(1)	0.022(1)	0.005(1)
C(15)	0.049(2)	0.043(2)	0.052(2)	-0.001(2)	0.021(2)	0.012(2)
Cl(1A)	0.062(1)	0.068(1)	0.057(1)	0.001(1)	0.032(1)	0.003(1)
Cl(1B)	0.072(1)	0.069(1)	0.102(1)	0.007(1)	0.040(1)	0.036(1)
C(16)	0.068(3)	0.051(2)	0.052(2)	0.020(2)	0.010(2)	0.016(2)
C(17)	0.104(4)	0.042(3)	0.080(3)	0.004(2)	0.026(3)	-0.009(3)

Table 3A.5.5. Hydrogen coordinates and isotropic displacement parameters [Å^2] for (5).

Atom	x	y	z	$U(eq)$
H(2)	0.3253(3)	0.2386(3)	1.2065(3)	0.053
H(3)	0.2411(4)	0.3012(3)	1.3446(3)	0.057
H(11)	0.2152(4)	0.1401(3)	1.1142(3)	0.056
H(13)	-0.0027(3)	0.3044(3)	1.0229(3)	0.048
H(14)	0.1047(3)	0.4062(2)	1.1203(3)	0.041
H(16A)	0.1083(4)	0.1633(3)	1.3457(3)	0.069
H(16B)	0.2185(4)	0.1326(3)	1.2938(3)	0.069
H(17A)	0.0638(5)	0.0319(5)	1.1510(27)	0.103(12)
H(17B)	-0.0414(26)	0.0831(8)	1.0765(11)	0.103(12)
H(17C)	-0.0489(24)	0.0651(12)	1.1947(18)	0.103(12)

Table 3A.6.1. Atomic coordinates and equivalent isotropic displacement parameters [Å^2] for (6)

Atom	x	y	z	U(eq)
C(1)	-0.0935(3)	1.3845(2)	0.8921(2)	0.025(1)
C(2)	-0.0556(2)	1.2377(2)	0.8289(2)	0.023(1)
C(3)	0.1253(2)	1.1192(2)	0.8317(2)	0.018(1)
C(4)	0.2424(2)	1.0616(2)	0.7120(2)	0.018(1)
C(5)	0.3531(2)	0.9012(2)	0.7233(2)	0.018(1)
Cl(5)	0.2337(1)	0.7683(1)	0.7828(1)	0.028(1)
C(6)	0.5277(2)	0.8561(2)	0.6860(2)	0.020(1)
Cl(6)	0.6418(1)	0.6665(1)	0.6931(1)	0.030(1)
C(7)	0.6284(2)	0.9669(2)	0.6287(2)	0.021(1)
Cl(7)	0.8552(1)	0.9033(1)	0.5893(1)	0.033(1)
C(8)	0.5464(2)	1.1094(2)	0.6047(2)	0.019(1)
Cl(8)	0.6665(1)	1.2317(1)	0.5215(1)	0.027(1)
C(9)	0.3495(2)	1.1774(2)	0.6422(2)	0.018(1)
C(10)	0.3091(2)	1.3247(2)	0.7083(2)	0.020(1)
C(11)	0.1274(3)	1.4421(2)	0.7050(2)	0.024(1)
C(12)	0.0105(3)	1.4978(2)	0.8237(2)	0.024(1)
C(13)	0.0197(3)	1.3810(2)	0.6507(2)	0.029(1)
C(14)	-0.0692(3)	1.2840(2)	0.7102(2)	0.028(1)
C(15)	0.2353(2)	1.1738(2)	0.8882(2)	0.018(1)
C(16)	0.2502(3)	1.1231(2)	0.9959(2)	0.024(1)
C(17)	0.3566(3)	1.1747(3)	1.0424(2)	0.030(1)
C(18)	0.4489(3)	1.2768(3)	0.9812(2)	0.030(1)
C(19)	0.4360(2)	1.3277(2)	0.8725(2)	0.025(1)

Table 3A.6.2. Bond lengths [Å] for (6)

C(1)-C(2)	1.540(3)	C(1)-C(12)	1.540(3)
C(1)-H(1B)	0.99(2)	C(1)-H(1A)	0.98(2)
C(2)-C(14)	1.507(3)	C(2)-C(3)	1.589(3)
C(2)-H(2)	0.95(2)	C(3)-C(15)	1.513(3)
C(3)-C(4)	1.562(3)	C(3)-H(3)	0.98(2)
C(4)-C(5)	1.510(3)	C(4)-C(9)	1.579(3)
C(4)-H(4)	1.00(2)	C(5)-C(6)	1.333(3)
C(5)-Cl(5)	1.734(2)	C(6)-C(7)	1.471(3)
C(6)-Cl(6)	1.729(2)	C(7)-C(8)	1.332(3)
C(7)-Cl(7)	1.733(2)	C(8)-C(9)	1.516(3)
C(8)-Cl(8)	1.739(2)	C(9)-C(10)	1.560(3)
C(9)-H(9)	0.99(2)	C(10)-C(20)	1.509(3)
C(10)-C(11)	1.591(3)	C(10)-H(10)	0.95(2)
C(11)-C(13)	1.513(3)	C(11)-C(12)	1.547(3)
C(11)-H(11)	0.97(2)	C(12)-H(12B)	0.96(2)
C(12)-H(12A)	0.97(2)	C(13)-C(14)	1.323(3)
C(13)-H(13)	0.97(2)	C(14)-H(14)	0.94(2)
C(15)-C(16)	1.386(3)	C(15)-C(20)	1.397(3)
C(16)-C(17)	1.390(3)	C(16)-H(16)	0.92(2)
C(17)-C(18)	1.380(3)	C(17)-H(17)	0.93(3)
C(18)-C(19)	1.394(3)	C(18)-H(18)	0.94(2)
C(19)-C(20)	1.395(3)	C(19)-H(19)	0.97(2)

Table 3A.6.3 Bond angles [°] for (6)

C(2)-C(1)-C(12)	113.1(2)	C(2)-C(1)-H(1B)	108.9(13)
C(12)-C(1)-H(1B)	111.5(13)	C(2)-C(1)-H(1A)	108.8(14)
C(12)-C(1)-H(1A)	108.0(14)	H(1B)-C(1)-H(1A)	106(2)
C(14)-C(2)-C(1)	107.8(2)	C(14)-C(2)-C(3)	113.1(2)
C(1)-C(2)-C(3)	113.6(2)	C(14)-C(2)-H(2)	108.8(13)
C(1)-C(2)-H(2)	106.6(13)	C(3)-C(2)-H(2)	106.7(13)
C(15)-C(3)-C(4)	108.50(14)	C(15)-C(3)-C(2)	115.1(2)
C(4)-C(3)-C(2)	113.3(2)	C(15)-C(3)-H(3)	105.0(11)
C(4)-C(3)-H(3)	108.4(11)	C(2)-C(3)-H(3)	106.0(11)
C(5)-C(4)-C(3)	109.8(2)	C(5)-C(4)-C(9)	114.3(2)
C(3)-C(4)-C(9)	112.8(2)	C(5)-C(4)-H(4)	104.7(11)
C(3)-C(4)-H(4)	108.1(11)	C(9)-C(4)-H(4)	106.5(11)
C(6)-C(5)-C(4)	125.5(2)	C(6)-C(5)-Cl(5)	120.0(2)
C(4)-C(5)-Cl(5)	114.38(13)	C(5)-C(6)-C(7)	120.0(2)
C(5)-C(6)-Cl(6)	122.0(2)	C(7)-C(6)-Cl(6)	117.76(14)
C(8)-C(7)-C(6)	120.6(2)	C(8)-C(7)-Cl(7)	121.7(2)
C(6)-C(7)-Cl(7)	117.73(14)	C(7)-C(8)-C(9)	125.2(2)
C(7)-C(8)-Cl(8)	119.9(2)	C(9)-C(8)-Cl(8)	114.87(14)
C(8)-C(9)-C(10)	109.0(2)	C(8)-C(9)-C(4)	114.0(2)
C(10)-C(9)-C(4)	112.0(2)	C(8)-C(9)-H(9)	106.3(11)
C(10)-C(9)-H(9)	107.5(11)	C(4)-C(9)-H(9)	107.6(11)
C(20)-C(10)-C(9)	108.2(2)	C(20)-C(10)-C(11)	114.8(2)
C(9)-C(10)-C(11)	113.1(2)	C(20)-C(10)-H(10)	107.0(12)
C(9)-C(10)-H(10)	108.7(12)	C(11)-C(10)-H(10)	104.7(12)
C(13)-C(11)-C(12)	108.0(2)	C(13)-C(11)-C(10)	113.8(2)
C(12)-C(11)-C(10)	113.1(2)	C(13)-C(11)-H(11)	106.6(13)
C(12)-C(11)-H(11)	107.7(13)	C(10)-C(11)-H(11)	107.3(12)
C(1)-C(12)-C(11)	113.4(2)	C(1)-C(12)-H(12B)	110.0(13)
C(11)-C(12)-H(12B)	110.8(13)	C(1)-C(12)-H(12A)	108.9(14)
C(11)-C(12)-H(12A)	108.1(14)	H(12B)-C(12)-H(12A)	105(2)
C(14)-C(13)-C(11)	118.8(2)	C(14)-C(13)-H(13)	122.2(14)
C(11)-C(13)-H(13)	119.0(14)	C(13)-C(14)-C(2)	118.0(2)
C(13)-C(14)-H(14)	122.6(14)	C(2)-C(14)-H(14)	119.3(14)
C(16)-C(15)-C(20)	119.7(2)	C(16)-C(15)-C(3)	123.1(2)
C(20)-C(15)-C(3)	117.1(2)	C(15)-C(16)-C(17)	120.6(2)
C(15)-C(16)-H(16)	118.6(14)	C(17)-C(16)-H(16)	120.8(13)
C(18)-C(17)-C(16)	120.0(2)	C(18)-C(17)-H(17)	123(2)
C(16)-C(17)-H(17)	117(2)	C(17)-C(18)-C(19)	120.0(2)
C(17)-C(18)-H(18)	118.8(14)	C(19)-C(18)-H(18)	121.2(14)
C(18)-C(19)-C(20)	120.2(2)	C(18)-C(19)-H(19)	122.9(13)
C(20)-C(19)-H(19)	116.9(13)	C(19)-C(20)-C(15)	119.5(2)
C(19)-C(20)-C(10)	123.5(2)	C(15)-C(20)-C(10)	116.9(2)

Table 3A.6.4. Anisotropic displacement parameters [Å^2] for (6).

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C(1)	0.019(1)	0.022(1)	0.029(1)	-0.004(1)	-0.004(1)	-0.001(1)
C(2)	0.016(1)	0.022(1)	0.032(1)	-0.002(1)	-0.005(1)	-0.006(1)
C(3)	0.018(1)	0.015(1)	0.021(1)	0.001(1)	-0.004(1)	-0.006(1)
C(4)	0.017(1)	0.019(1)	0.019(1)	0.000(1)	-0.007(1)	-0.006(1)
C(5)	0.024(1)	0.014(1)	0.018(1)	-0.001(1)	-0.005(1)	-0.008(1)
Cl(5)	0.031(1)	0.018(1)	0.034(1)	0.000(1)	-0.002(1)	-0.011(1)
C(6)	0.024(1)	0.017(1)	0.020(1)	-0.003(1)	-0.009(1)	-0.002(1)
Cl(6)	0.030(1)	0.018(1)	0.036(1)	-0.001(1)	-0.010(1)	0.002(1)
C(7)	0.017(1)	0.027(1)	0.018(1)	-0.006(1)	-0.005(1)	-0.005(1)
Cl(7)	0.017(1)	0.035(1)	0.044(1)	-0.006(1)	-0.005(1)	-0.003(1)
C(8)	0.021(1)	0.024(1)	0.016(1)	-0.002(1)	-0.004(1)	-0.012(1)
Cl(8)	0.027(1)	0.031(1)	0.024(1)	0.002(1)	-0.002(1)	-0.016(1)
C(9)	0.021(1)	0.019(1)	0.015(1)	0.001(1)	-0.006(1)	-0.007(1)
C(10)	0.022(1)	0.017(1)	0.020(1)	0.001(1)	-0.003(1)	-0.009(1)
C(11)	0.030(1)	0.014(1)	0.024(1)	0.003(1)	-0.008(1)	-0.004(1)
C(12)	0.026(1)	0.017(1)	0.030(1)	-0.003(1)	-0.009(1)	-0.003(1)
C(13)	0.033(1)	0.023(1)	0.028(1)	-0.002(1)	-0.017(1)	0.006(1)
C(14)	0.024(1)	0.024(1)	0.041(1)	-0.009(1)	-0.020(1)	0.003(1)
C(15)	0.016(1)	0.016(1)	0.020(1)	-0.004(1)	-0.003(1)	-0.001(1)
C(16)	0.026(1)	0.022(1)	0.019(1)	-0.001(1)	-0.003(1)	-0.001(1)
C(17)	0.030(1)	0.034(1)	0.022(1)	-0.006(1)	-0.013(1)	0.005(1)
C(18)	0.023(1)	0.036(1)	0.032(1)	-0.015(1)	-0.013(1)	0.000(1)
C(19)	0.020(1)	0.025(1)	0.030(1)	-0.009(1)	-0.005(1)	-0.006(1)
C(20)	0.017(1)	0.018(1)	0.019(1)	-0.005(1)	-0.004(1)	-0.002(1)

Table 3A.6.5. Hydrogen coordinates and isotropic displacement parameters [Å^2] for 1.

Atom	x	y	z	$U(eq)$
H(1B)	-0.0706(29)	1.3563(26)	0.9667(20)	0.030(6)
H(1A)	-0.2188(32)	1.4359(27)	0.9078(20)	0.035(6)
H(2)	-0.1455(28)	1.1891(25)	0.8674(18)	0.025(6)
H(3)	0.0946(24)	1.0313(23)	0.8792(17)	0.014(5)
H(4)	0.1633(25)	1.0519(22)	0.6672(16)	0.016(5)
H(9)	0.3131(24)	1.2088(22)	0.5719(17)	0.015(5)
H(10)	0.3949(27)	1.3785(23)	0.6720(17)	0.019(5)
H(11)	0.1544(27)	1.5309(26)	0.6587(19)	0.026(6)
H(12B)	0.0790(29)	1.5207(25)	0.8667(19)	0.027(6)
H(12A)	-0.0718(31)	1.5944(28)	0.8135(19)	0.033(6)
H(13)	0.0205(31)	1.4120(27)	0.5730(22)	0.038(6)
H(14)	-0.1373(30)	1.2413(26)	0.6806(20)	0.032(6)
H(16)	0.1911(28)	1.0524(25)	1.0348(18)	0.023(5)
H(17)	0.3614(32)	1.1366(28)	1.1152(23)	0.043(7)
H(18)	0.5204(31)	1.3106(27)	1.0141(20)	0.035(6)
H(19)	0.4983(28)	1.3991(25)	0.8265(19)	0.025(5)

Table 3A.7.1. Atomic coordinates and equivalent isotropic displacement parameters [Å^2] for (7)

Atom	x	y	z	$U(eq)$
C(1)	0.7900(5)	0.4027(2)	0.6206(1)	0.033(1)
C(2)	0.9501(6)	0.4552(2)	0.5873(1)	0.041(1)
C(3)	0.9423(6)	0.5372(2)	0.5982(1)	0.046(1)
C(4)	0.7702(6)	0.5681(2)	0.6425(1)	0.045(1)
C(5)	0.6119(6)	0.5167(2)	0.6763(1)	0.042(1)
C(6)	0.6212(5)	0.4336(2)	0.6664(1)	0.032(1)
C(7)	0.4575(5)	0.3786(2)	0.7064(1)	0.034(1)
C(8)	0.5035(5)	0.3772(2)	0.7833(1)	0.036(1)
C(9)	0.7614(5)	0.4085(2)	0.8087(1)	0.041(1)
C(10	0.8374(6)	0.3680(2)	0.8741(1)	0.050(1)
C(11)	0.8289(6)	0.2782(2)	0.8677(1)	0.048(1)
C(12)	0.6809(6)	0.2408(2)	0.8243(1)	0.046(1)
C(13)	0.4918(5)	0.2839(2)	0.7808(1)	0.037(1)
C(14)	0.5114(5)	0.2866(2)	0.7031(1)	0.035(1)
C(15)	0.7660(5)	0.2644(2)	0.6721(1)	0.032(1)
C(16)	0.7814(6)	0.1756(2)	0.6554(1)	0.041(1)
Cl(16)	0.9515(2)	0.1137(1)	0.7089(1)	0.062(1)
C(17)	0.6649(6)	0.1476(2)	0.6009(1)	0.047(1)
Cl(17)	0.6749(2)	0.0473(1)	0.5786(1)	0.081(1)
C(18)	0.5391(6)	0.2048(2)	0.5553(1)	0.045(1)
Cl(18)	0.3047(2)	0.1672(1)	0.5019(1)	0.069(1)
C(19)	0.6055(5)	0.2820(2)	0.5565(1)	0.038(1)
Cl(19)	0.4809(2)	0.3492(1)	0.4996(1)	0.053(1)
C(20)	0.8012(5)	0.3125(2)	0.6078(1)	0.033(1)

Table 3A.7.2 Bond lengths [Å] for (7)

C(1)-C(2)	1.392(4)	C(1)-C(6)	1.393(4)
C(1)-C(20)	1.519(3)	C(2)-C(3)	1.378(4)
C(3)-C(4)	1.383(4)	C(4)-C(5)	1.380(4)
C(5)-C(6)	1.393(4)	C(6)-C(7)	1.500(4)
C(7)-C(14)	1.553(4)	C(7)-C(8)	1.569(3)
C(8)-C(9)	1.519(4)	C(8)-C(13)	1.549(4)
C(9)-C(10)	1.529(4)	C(10)-C(11)	1.495(4)
C(11)-C(12)	1.312(4)	C(12)-C(13)	1.488(4)
C(13)-C(14)	1.580(3)	C(14)-C(15)	1.528(4)
C(15)-C(16)	1.512(4)	C(15)-C(20)	1.542(3)
C(16)-C(17)	1.331(4)	C(16)-Cl(16)	1.722(3)
C(17)-C(18)	1.467(4)	C(17)-Cl(17)	1.724(3)
C(18)-C(19)	1.327(4)	C(18)-Cl(18)	1.729(3)
C(19)-C(20)	1.524(4)	C(19)-Cl(19)	1.719(3)

Table 3A.7.3. Bond angles [°] for (7)

C(2)-C(1)-C(6)	119.3(2)	C(2)-C(1)-C(20)	120.5(2)
C(6)-C(1)-C(20)	120.2(2)	C(3)-C(2)-C(1)	121.2(3)
C(2)-C(3)-C(4)	119.5(3)	C(5)-C(4)-C(3)	119.9(3)
C(4)-C(5)-C(6)	121.1(3)	C(1)-C(6)-C(5)	118.9(2)
C(1)-C(6)-C(7)	121.0(2)	C(5)-C(6)-C(7)	120.1(2)
C(6)-C(7)-C(14)	117.9(2)	C(6)-C(7)-C(8)	118.0(2)
C(14)-C(7)-C(8)	90.3(2)	C(9)-C(8)-C(13)	112.7(2)
C(9)-C(8)-C(7)	116.3(2)	C(13)-C(8)-C(7)	88.7(2)
C(8)-C(9)-C(10)	110.5(2)	C(11)-C(10)-C(9)	110.9(2)
C(12)-C(11)-C(10)	122.9(3)	C(11)-C(12)-C(13)	122.7(3)
C(12)-C(13)-C(8)	115.8(2)	C(12)-C(13)-C(14)	123.0(2)
C(8)-C(13)-C(14)	90.0(2)	C(15)-C(14)-C(7)	114.5(2)
C(15)-C(14)-C(13)	118.7(2)	C(7)-C(14)-C(13)	88.2(2)
C(16)-C(15)-C(14)	112.2(2)	C(16)-C(15)-C(20)	107.9(2)
C(14)-C(15)-C(20)	110.2(2)	C(17)-C(16)-C(15)	120.0(2)
C(17)-C(16)-Cl(16)	122.1(2)	C(15)-C(16)-Cl(16)	118.0(2)
C(16)-C(17)-C(18)	119.0(3)	C(16)-C(17)-Cl(17)	122.5(2)
C(18)-C(17)-Cl(17)	118.3(2)	C(19)-C(18)-C(17)	119.9(3)
C(19)-C(18)-Cl(18)	122.7(2)	C(17)-C(18)-Cl(18)	117.4(2)
C(18)-C(19)-C(20)	120.2(2)	C(18)-C(19)-Cl(19)	121.2(2)
C(20)-C(19)-Cl(19)	118.5(2)	C(1)-C(20)-C(19)	114.5(2)
C(1)-C(20)-C(15)	111.1(2)	C(19)-C(20)-C(15)	108.2(2)

Table 3A.7.4. Anisotropic displacement parameters (Å^2) for (7).

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
C(1)	0.032(1)	0.035(1)	0.032(1)	0.005(1)	-0.004(1)	0.002(1)
C(2)	0.041(2)	0.048(2)	0.036(1)	0.007(1)	0.007(1)	0.005(1)
C(3)	0.055(2)	0.039(2)	0.043(2)	0.013(1)	-0.001(1)	-0.006(2)
C(4)	0.060(2)	0.034(2)	0.042(2)	0.004(1)	-0.008(2)	0.000(2)
C(5)	0.049(2)	0.040(2)	0.036(1)	0.002(1)	0.001(1)	0.009(1)
C(6)	0.031(1)	0.035(1)	0.031(1)	0.002(1)	-0.004(1)	0.006(1)
C(7)	0.024(1)	0.041(1)	0.036(1)	0.000(1)	0.000(1)	0.003(1)
C(8)	0.031(1)	0.042(2)	0.036(1)	-0.001(1	0.007(1)	0.000(1)
C(9)	0.046(2)	0.043(2)	0.035(1)	0.002(1)	0.000(1)	-0.010(1)
C(10)	0.055(2)	0.060(2)	0.034(1)	0.000(1)	-0.005(1)	-0.013(2)
C(11)	0.053(2)	0.056(2)	0.035(2)	0.011(1)	-0.001(1)	0.000(2)
C(12)	0.056(2)	0.041(2)	0.041(2)	0.008(1)	0.005(1)	-0.003(2)
C(13)	0.034(2)	0.043(2)	0.036(1)	0.002(1)	0.006(1)	-0.010(1)
C(14)	0.031(2)	0.037(1)	0.036(1)	0.001(1)	-0.003(1)	-0.007(1)
C(15)	0.031(1)	0.033(1)	0.032(1)	0.002(1)	-0.005(1)	0.000(1)
C(16)	0.043(2)	0.035(1)	0.044(2)	0.006(1)	-0.002(1)	0.005(1)
Cl(16)	0.081(1)	0.043(1)	0.062(1)	0.011(1)	-0.012(1)	0.018(1)
C(17)	0.054(2)	0.035(2)	0.051(2)	-0.005(1)	0.001(2)	-0.001(1)
Cl(17)	0.124(1)	0.038(1)	0.081(1)	-0.015(1)	-0.010(1)	-0.002(1)
C(18)	0.045(2)	0.054(2)	0.036(1)	-0.010(1)	-0.005(1)	0.003(2)
Cl(18)	0.065(1)	0.081(1)	0.059(1)	-0.023(1)	-0.018(1)	-0.007(1)
C(19)	0.040(2)	0.043(2)	0.030(1)	-0.001(1)	0.000(1)	0.010(1)
Cl(19)	0.062(1)	0.063(1)	0.034(1)	0.004(1)	-0.008(1)	0.019(1)
C(20)	0.030(1)	0.036(1)	0.032(1)	0.001(1)	0.000(1)	0.005(1)

Table 3A.7.5. Hydrogen coordinates and isotropic displacement parameters (Å^2) for (7).

Atom	x	y	z	U(eq)
H(2)	1.0630(36)	0.4349(7)	0.5575(9)	0.050
H(3)	1.0561(37)	0.5728(11)	0.5751(7)	0.055
H(4)	0.7608(7)	0.6260(19)	0.6498(3)	0.055
H(5)	0.4970(38)	0.5379(7)	0.7061(10)	0.050
H(7)	0.2799(53)	0.3888(3)	0.6963(3)	0.041
H(8)	0.3583(42)	0.4008(7)	0.8073(7)	0.043
H(9B)	0.7519(6)	0.4658(11)	0.8149(2)	0.050
H(9A)	0.8897(26)	0.3978(2)	0.7766(6)	0.050
H(10B)	1.0073(37)	0.3845(4)	0.8872(3)	0.060
H(10A)	0.7224(25)	0.3848(4)	0.9076(8)	0.060
H(11)	0.9406(36)	0.2452(11)	0.8979(10)	0.058
H(12)	0.6949(7)	0.1847(19)	0.8204(2)	0.055
H(13)	0.3222(52)	0.2668(5)	0.7918(4)	0.045
H(14)	0.3708(43)	0.2583(9)	0.6820(6)	0.041
H(15)	0.9012(40)	0.2774(4)	0.7021(9)	0.039
H(20)	0.9716(49)	0.3002(4)	0.5916(5)	0.039

Table 3A.8.1. Atomic coordinates and equivalent isotropic displacement parameters [Å^2] for (8).

Atom	x	y	z	U(eq)
Cl(2)	0.9050(1)	0.1616(1)	0.5681(1)	0.045(1)
Cl(21)	0.8448(1)	0.1473(1)	0.3323(1)	0.053(1)
Cl(12)	0.4887(1)	0.0230(1)	0.3458(1)	0.054(1)
Cl(22)	0.5875(1)	0.0703(1)	0.1946(1)	0.057(1)
C(1)	0.6876(2)	0.1210(3)	0.4817(2)	0.035(1)
C(2)	0.7919(2)	0.0830(3)	0.4825(2)	0.032(1)
C(3)	0.7953(2)	-0.0546(2)	0.5053(2)	0.029(1)
C(4)	0.8633(2)	-0.1406(3)	0.4797(2)	0.031(1)
C(5)	0.8492(2)	-0.2688(3)	0.5155(2)	0.034(1)
C(6)	0.9089(3)	-0.3835(3)	0.5053(3)	0.047(1)
C(7)	0.8393(3)	-0.4866(3)	0.5095(3)	0.056(1)
C(8)	0.7409(3)	-0.4397(3)	0.4171(3)	0.050(1)
C(9)	0.7334(2)	-0.3077(3)	0.4561(2)	0.033(1)
C(10)	0.6690(2)	-0.2060(3)	0.3802(2)	0.032(1)
C(11)	0.6786(2)	-0.0930(3)	0.4445(2)	0.029(1)
C(12)	0.6250(2)	0.0296(3)	0.3931(2)	0.033(1)
C(21)	0.7598(2)	0.1051(3)	0.3736(2)	0.034(1)
C(22)	0.6611(2)	0.0747(3)	0.3212(2)	0.037(1)
C(41)	0.8189(2)	-0.1458(3)	0.3675(2)	0.034(1)
C(42)	0.7204(2)	-0.1790(3)	0.3174(2)	0.035(1)
C(61)	0.8830(3)	-0.3968(3)	0.3972(3)	0.053(1)
C(62)	0.7848(3)	-0.4289(3)	0.3458(3)	0.052(1)
C(91)	0.6996(3)	-0.3233(3)	0.5348(2)	0.044(1)
O(92)	0.7804(2)	-0.2929(2)	0.6287(2)	0.053(1)
C(93)	0.8701(3)	-0.2635(3)	0.6224(2)	0.043(1)
O(1)	0.6725(2)	0.2442(2)	0.4618(2)	0.044(1)
C(10)	0.5747(3)	0.2942(3)	0.4463(3)	0.064(1)
O(2)	0.6762(2)	0.0849(2)	0.5629(2)	0.042(1)
C(20)	0.7304(3)	0.1573(3)	0.6524(2)	0.057(1)
O(91)	0.6183(2)	-0.3565(2)	0.5260(2)	0.059(1)
O(93)	0.9484(2)	-0.2403(2)	0.6961(2)	0.063(1)

Table 3A.8.2. Bond lengths [Å] and angles [°] for (8).

Cl(2)-C(2)	1.769(3)	Cl(21)-C(21)	1.701(3)
Cl(12)-C(12)	1.770(3)	Cl(22)-C(22)	1.704(3)
C(1)-O(2)	1.382(3)	C(1)-O(1)	1.382(4)
C(1)-C(2)	1.576(4)	C(1)-C(12)	1.578(4)
C(2)-C(21)	1.512(4)	C(2)-C(3)	1.547(4)
C(3)-C(4)	1.550(4)	C(3)-C(11)	1.563(4)
C(3)-H(3)	0.93(3)	C(4)-C(41)	1.512(4)
C(4)-C(5)	1.559(4)	C(4)-H(4)	0.95(3)
C(5)-C(93)	1.508(4)	C(5)-C(9)	1.553(4)
C(5)-C(6)	1.582(4)	C(6)-C(61)	1.509(5)
C(6)-C(7)	1.544(5)	C(6)-H(6)	0.91(3)
C(7)-C(8)	1.545(5)	C(7)-H(71)	1.00(2)
C(7)-H(72)	1.00(2)	C(8)-C(62)	1.506(5)
C(8)-C(9)	1.591(4)	C(8)-H(8)	0.95(3)
C(9)-C(91)	1.507(4)	C(9)-C(10)	1.561(4)
C(10)-C(42)	1.500(4)	C(10)-C(11)	1.547(4)
C(10)-H(10)	0.97(3)	C(11)-C(12)	1.568(4)
C(11)-H(11)	0.98(3)	C(12)-C(22)	1.507(4)
C(21)-C(22)	1.318(4)	C(41)-C(42)	1.320(4)
C(41)-H(41)	0.96(3)	C(42)-H(42)	0.96(3)
C(61)-C(62)	1.315(5)	C(61)-H(61)	0.91(3)
C(62)-H(62)	0.90(3)	C(91)-O(91)	1.189(4)
C(91)-O(92)	1.395(4)	O(92)-C(93)	1.398(4)
C(93)-O(93)	1.188(4)	O(1)-C(1O)	1.442(4)
C(1O)-H(1O1)	0.95(2)	C(1O)-H(1O2)	0.95(2)
C(1O)-H(1O3)	0.95(2)	O(2)-C(2O)	1.446(4)
C(2O)-H(2O1)	0.96(2)	C(2O)-H(2O2)	0.96(2)
C(2O)-H(2O3)	0.96(2)		

278

Table 3A.8.3. Bond Angles [Å] for (8)

O(2)-C(1)-O(1)	113.7(2)	O(2)-C(1)-C(2)	116.4(2)
O(1)-C(1)-C(2)	108.4(2)	O(2)-C(1)-C(12)	108.3(2)
O(1)-C(1)-C(12)	118.0(2)	C(2)-C(1)-C(12)	90.2(2)
C(21)-C(2)-C(3)	111.3(2)	C(21)-C(2)-C(1)	98.9(2)
C(3)-C(2)-C(1)	100.9(2)	C(21)-C(2)-Cl(2)	115.0(2)
C(3)-C(2)-Cl(2)	113.1(2)	C(1)-C(2)-Cl(2)	116.0(2)
C(2)-C(3)-C(4)	119.8(2)	C(2)-C(3)-C(11)	102.8(2)
C(4)-C(3)-C(11)	109.7(2)	C(2)-C(3)-H(3)	108.01(14)
C(4)-C(3)-H(3)	108.01(14)	C(11)-C(3)-H(3)	108.0(2)
C(41)-C(4)-C(3)	108.9(2)	C(41)-C(4)-C(5)	107.6(2)
C(3)-C(4)-C(5)	104.9(2)	C(41)-C(4)-H(4)	111.7(2)
C(3)-C(4)-H(4)	111.70(14)	C(5)-C(4)-H(4)	111.70(14)
C(93)-C(5)-C(9)	104.1(2)	C(93)-C(5)-C(4)	110.4(2)
C(9)-C(5)-C(4)	109.5(2)	C(93)-C(5)-C(6)	107.4(2)
C(9)-C(5)-C(6)	102.9(2)	C(4)-C(5)-C(6)	121.0(2)
C(61)-C(6)-C(7)	98.4(3)	C(61)-C(6)-C(5)	108.6(3)
C(7)-C(6)-C(5)	100.1(2)	C(61)-C(6)-H(6)	115.8(2)
C(7)-C(6)-H(6)	115.8(2)	C(5)-C(6)-H(6)	115.8(2)
C(6)-C(7)-C(8)	93.7(3)	C(6)-C(7)-H(71)	113.0(2)
C(8)-C(7)-H(71)	113.0(2)	C(6)-C(7)-H(72)	113.0(2)
C(8)-C(7)-H(72)	113.0(2)	H(71)-C(7)-H(72)	110.4
C(62)-C(8)-C(7)	98.7(3)	C(62)-C(8)-C(9)	109.0(3)
C(7)-C(8)-C(9)	99.5(2)	C(62)-C(8)-H(8)	115.8(2)
C(7)-C(8)-H(8)	115.8(2)	C(9)-C(8)-H(8)	115.8(2)
C(91)-C(9)-C(5)	104.3(2)	C(91)-C(9)-C(10)	111.1(2)
C(5)-C(9)-C(10)	110.0(2)	C(91)-C(9)-C(8)	107.4(2)
C(5)-C(9)-C(8)	102.6(2)	C(10)-C(9)-C(8)	120.1(2)
C(42)-C(10)-C(11)	108.7(2)	C(42)-C(10)-C(9)	107.7(2)
C(11)-C(10)-C(9)	105.4(2)	C(42)-C(10)-H(10)	111.6(2)
C(11)-C(10)-H(10)	111.59(14)	C(9)-C(10)-H(10)	111.59(14)
C(10)-C(11)-C(3)	109.7(2)	C(10)-C(11)-C(12)	120.1(2)
C(3)-C(11)-C(12)	102.1(2)	C(10)-C(11)-H(11)	108.14(14)
C(3)-C(11)-H(11)	108.14(14)	C(12)-C(11)-H(11)	108.1(2)
C(22)-C(12)-C(11)	112.3(2)	C(22)-C(12)-C(1)	98.6(2)
C(11)-C(12)-C(1)	99.9(2)	C(22)-C(12)-Cl(12)	115.7(2)
C(11)-C(12)-Cl(12)	111.5(2)	C(1)-C(12)-Cl(12)	117.2(2)
C(22)-C(21)-C(2)	107.6(2)	C(22)-C(21)-Cl(21)	128.8(2)
C(2)-C(21)-Cl(21)	123.3(2)	C(21)-C(22)-C(12)	107.7(2)
C(21)-C(22)-Cl(22)	127.8(2)	C(12)-C(22)-Cl(22)	124.1(2)
C(42)-C(41)-C(4)	115.3(3)	C(42)-C(41)-H(41)	122.3(2)
C(4)-C(41)-H(41)	122.3(2)	C(41)-C(42)-C(10)	115.3(3)
C(41)-C(42)-H(42)	122.4(2)	C(10)-C(42)-H(42)	122.4(2)
C(62)-C(61)-C(6)	108.2(3)	C(62)-C(61)-H(61)	125.9(2)
C(6)-C(61)-H(61)	125.9(2)	C(61)-C(62)-C(8)	108.1(3)
C(61)-C(62)-H(62)	125.9(2)	C(8)-C(62)-H(62)	125.9(2)
O(91)-C(91)-O(92)	120.1(3)	O(91)-C(91)-C(9)	129.4(3)
O(92)-C(91)-C(9)	110.5(3)	C(91)-O(92)-C(93)	110.5(2)
O(93)-C(93)-O(92)	119.9(3)	O(93)-C(93)-C(5)	129.6(3)
O(92)-C(93)-C(5)	110.5(3)	C(1)-O(1)-C(10)	116.9(2)
O(1)-C(10)-H(101)	109.5(2)	O(1)-C(10)-H(102)	109.5(2)
H(101)-C(10)-H(102)	109.5	O(1)-C(10)-H(103)	109.5(2)
H(101)-C(10)-H(103)	109.5	H(102)-C(10)-H(103)	109.5
C(1)-O(2)-C(20)	116.1(2)	O(2)-C(20)-H(201)	109.5(2)
O(2)-C(20)-H(202)	109.5(2)	H(201)-C(20)-H(202)	109.5
O(2)-C(20)-H(203)	109.5(2)	H(201)-C(20)-H(203)	109.5
H(202)-C(20)-H(203)	109.5		

Table 3A.8.4. Anisotropic displacement parameters ($\mathring{A}^2$) for 1.

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
Cl(2)	0.032(1)	0.041(1)	0.050(1)	-0.006(1)	0.009(1)	-0.008(1)
Cl(21)	0.042(1)	0.068(1)	0.052(1)	0.022(1)	0.025(1)	0.000(1)
Cl(12)	0.023(1)	0.053(1)	0.075(1)	0.001(1)	0.016(1)	0.001(1)
Cl(22)	0.050(1)	0.068(1)	0.034(1)	0.012(1)	0.004(1)	-0.004(1)
C(1)	0.032(2)	0.028(2)	0.044(2)	0.001(1)	0.019(1)	0.001(1)
C(2)	0.026(1)	0.031(2)	0.033(2)	-0.001(1)	0.010(1)	-0.002(1)
C(3)	0.030(1)	0.031(2)	0.025(1)	0.002(1)	0.013(1)	0.000(1)
C(4)	0.021(1)	0.035(2)	0.034(2)	0.000(1)	0.011(1)	0.000(1)
C(5)	0.029(1)	0.035(2)	0.037(2)	0.004(1)	0.014(1)	0.007(1)
C(6)	0.036(2)	0.039(2)	0.063(2)	0.004(2)	0.021(2)	0.012(2)
C(7)	0.056(2)	0.033(2)	0.082(3)	0.007(2)	0.036(2)	0.011(2)
C(8)	0.047(2)	0.030(2)	0.071(2)	-0.010(2)	0.028(2)	-0.004(2)
C(9)	0.032(1)	0.030(2)	0.041(2)	0.000(1)	0.020(1)	0.002(1)
C(10)	0.025(1)	0.032(2)	0.035(2)	-0.003(1)	0.011(1)	-0.002(1)
C(11)	0.024(1)	0.033(2)	0.031(1)	0.001(1)	0.013(1)	0.001(1)
C(12)	0.020(1)	0.035(2)	0.041(2)	0.002(1)	0.012(1)	0.002(1)
C(21)	0.029(2)	0.035(2)	0.039(2)	0.012(1)	0.016(1)	0.005(1)
C(22)	0.033(2)	0.033(2)	0.035(2)	0.009(1)	0.009(1)	0.005(1)
C(41)	0.033(2)	0.038(2)	0.036(2)	0.002(1)	0.020(1)	0.002(1)
C(42)	0.036(2)	0.039(2)	0.030(2)	-0.001(1)	0.016(1)	0.003(1)
C(61)	0.054(2)	0.041(2)	0.074(2)	-0.010(2)	0.038(2)	0.009(2)
C(62)	0.053(2)	0.039(2)	0.067(2)	-0.021(2)	0.030(2)	0.002(2)
C(91)	0.045(2)	0.037(2)	0.052(2)	0.012(2)	0.026(2)	0.007(2)
O(92)	0.063(2)	0.059(2)	0.044(1)	0.013(1)	0.031(1)	0.006(1)
C(93)	0.047(2)	0.040(2)	0.040(2)	0.014(2)	0.018(2)	0.010(2)
O(1)	0.040(1)	0.028(1)	0.065(1)	0.002(1)	0.026(1)	0.004(1)
C(10)	0.051(2)	0.043(2)	0.101(3)	0.004(2)	0.039(2)	0.017(2)
O(2)	0.051(1)	0.034(1)	0.049(1)	-0.009(1)	0.031(1)	-0.005(1)
C(20)	0.077(2)	0.048(2)	0.051(2)	-0.011(2)	0.035(2)	-0.005(2)
O(91)	0.055(2)	0.062(2)	0.077(2)	0.016(1)	0.044(1)	0.002(1)
O(93)	0.059(2)	0.075(2)	0.038(1)	0.013(1)	0.009(1)	0.007(1)

Table 3A.8.5. Hydrogen coordinates and isotropic displacement parameters ($\mathring{A}^2$) for (8).

Atom	x	y	z	U(eq)
H(3)	0.8156(6)	-0.0635(4)	0.5731(20)	0.034
H(4)	0.9342(22)	-0.1161(8)	0.5119(10)	0.037
H(6)	0.9775(26)	-0.3885(3)	0.5506(17)	0.056
H(71)	0.8327(3)	-0.4850(3)	0.5718(15)	0.067
H(72)	0.8613(6)	-0.5689(19)	0.4992(3)	0.067
H(8)	0.6805(23)	-0.4892(19)	0.3952(9)	0.060
H(10)	0.5970(22)	-0.2300(8)	0.3395(12)	0.038
H(11)	0.6540(7)	-0.1153(7)	0.4922(14)	0.035
H(41)	0.8593(13)	-0.1264(7)	0.3351(10)	0.041
H(42)	0.6844(11)	-0.1853(3)	0.2460(22)	0.042
H(61)	0.9277(18)	-0.3846(6)	0.3720(10)	0.064
H(62)	0.7500(13)	-0.4421(6)	0.2797(25)	0.063
H(101)	0.5853(5)	0.3740(19)	0.4727(18)	0.096
H(102)	0.5470(12)	0.2447(17)	0.4789(17)	0.096
H(103)	0.5275(12)	0.2965(22)	0.3770(15)	0.096
H(201)	0.8002(15)	0.1277(16)	0.6906(12)	0.086
H(202)	0.6951(13)	0.1516(18)	0.6916(12)	0.086
H(203)	0.7320(17)	0.2406(18)	0.6344(4)	0.086

Appendix 4A.

Tables of structural data for compounds in Chapter 4

Table 4A.1.1. Atomic coordinates and equivalent isotropic displacement parameters [Å^2] for (1).

Atom	x	y	z	U(eq)
N(1)	-0.6560(3)	0.4289(3)	0.7485(1)	0.044(1)
C(2)	-0.4926(3)	0.3955(3)	0.7431(1)	0.031(1)
C(3)	-0.2921(3)	0.3543(3)	0.7346(1)	0.026(1)
C(4)	-0.2012(3)	0.3548(3)	0.8147(2)	0.028(1)
N(5)	-0.1275(3)	0.3526(3)	0.8803(1)	0.038(1)
C(6)	-0.1875(3)	0.3181(3)	0.6496(1)	0.023(1)
C(7)	0.0133(3)	0.2803(3)	0.6404(2)	0.025(1)
C(8)	0.1122(3)	0.2434(3)	0.5588(1)	0.024(1)
C(9)	0.0194(3)	0.2440(3)	0.4791(1)	0.021(1)
C(10)	-0.1811(3)	0.2824(3)	0.4883(2)	0.025(1)
C(11)	-0.2802(3)	0.3169(3)	0.5700(1)	0.025(1)
C(12)	0.1241(3)	0.2062(3)	0.3938(1)	0.023(1)
C(13)	0.3287(3)	0.1608(3)	0.3929(1)	0.025(1)
N(14)	0.4915(3)	0.1275(3)	0.3877(1)	0.036(1)
C(15)	0.0497(3)	0.2129(3)	0.3104(1)	0.025(1)
C(16)	0.1641(3)	0.1600(3)	0.2328(1)	0.025(1)
N(17)	0.1101(2)	0.1667(2)	0.1499(1)	0.027(1)
C(18)	0.2486(3)	0.1031(3)	0.0751(2)	0.033(1)
C(19)	0.3151(4)	0.2689(4)	0.0045(2)	0.043(1)
C(20)	-0.0854(3)	0.2442(3)	0.1259(2)	0.032(1)
C(21)	-0.2162(4)	0.0955(4)	0.1575(2)	0.044(1)

Table 4A.1.2. Bond lengths [Å] for (1).

N(1)-C(2)	1.156(3)	C(2)-C(3)	1.418(3)
C(3)-C(6)	1.419(3)	C(3)-C(4)	1.422(3)
C(4)-N(5)	1.157(3)	C(6)-C(7)	1.419(3)
C(6)-C(11)	1.424(3)	C(7)-C(8)	1.364(3)
C(7)-H(7)	0.92(2)	C(8)-C(9)	1.422(3)
C(8)-H(8)	0.99(2)	C(9)-C(10)	1.417(3)
C(9)-C(12)	1.427(3)	C(10)-C(11)	1.361(3)
C(10)-H(10)	0.98(2)	C(11)-H(11)	0.97(2)
C(12)-C(15)	1.389(3)	C(12)-C(13)	1.452(3)
C(13)-N(14)	1.151(3)	C(15)-C(16)	1.397(3)
C(15)-H(15)	0.97(2)	C(16)-N(17)	1.315(2)
C(16)-H(16)	1.01(2)	N(17)-C(18)	1.477(3)
N(17)-C(20)	1.478(3)	C(18)-C(19)	1.507(3)
C(18)-H(181)	1.00(2)	C(18)-H(182)	0.99(2)
C(19)-H(191)	1.03(3)	C(19)-H(192)	1.01(3)
C(19)-H(193)	1.00(3)	C(20)-C(21)	1.519(4)
C(20)-H(201)	1.02(2)	C(20)-H(202)	1.00(2)
C(21)-H(211)	0.98(3)	C(21)-H(212)	1.07(3)
C(21)-H(213)	1.02(3)		

Table 4A.1.3. Bond angles [°] for (1)

N(1)-C(2)-C(3)	178.9(2)	C(2)-C(3)-C(6)	121.6(2)
C(2)-C(3)-C(4)	116.7(2)	C(6)-C(3)-C(4)	121.7(2)
N(5)-C(4)-C(3)	179.0(2)	C(7)-C(6)-C(3)	121.7(2)
C(7)-C(6)-C(11)	117.1(2)	C(3)-C(6)-C(11)	121.2(2)
C(8)-C(7)-C(6)	121.1(2)	C(8)-C(7)-H(7)	120.2(13)
C(6)-C(7)-H(7)	118.6(13)	C(7)-C(8)-C(9)	121.7(2)
C(7)-C(8)-H(8)	120.8(12)	C(9)-C(8)-H(8)	117.5(12)
C(10)-C(9)-C(8)	117.1(2)	C(10)-C(9)-C(12)	121.5(2)
C(8)-C(9)-C(12)	121.3(2)	C(11)-C(10)-C(9)	121.2(2)
C(11)-C(10)-H(10)	118.7(12)	C(9)-C(10)-H(10)	120.1(12)
C(10)-C(11)-C(6)	121.7(2)	C(10)-C(11)-H(11)	119.2(12)
C(6)-C(11)-H(11)	119.1(12)	C(15)-C(12)-C(9)	126.3(2)
C(15)-C(12)-C(13)	116.4(2)	C(9)-C(12)-C(13)	117.2(2)
N(14)-C(13)-C(12)	176.9(2)	C(12)-C(15)-C(16)	121.5(2)
C(12)-C(15)-H(15)	118.4(12)	C(16)-C(15)-H(15)	120.0(12)
N(17)-C(16)-C(15)	126.6(2)	N(17)-C(16)-H(16)	114.2(11)
C(15)-C(16)-H(16)	119.1(11)	C(16)-N(17)-C(18)	120.1(2)
C(16)-N(17)-C(20)	122.5(2)	C(18)-N(17)-C(20)	117.4(2)
N(17)-C(18)-C(19)	112.3(2)	N(17)-C(18)-H(181)	105.6(13)
C(19)-C(18)-H(181)	111.0(13)	N(17)-C(18)-H(182)	107.0(12)
C(19)-C(18)-H(182)	111.2(12)	H(181)-C(18)-H(182)	110(2)
C(18)-C(19)-H(191)	111(2)	C(18)-C(19)-H(192)	110(2)
H(191)-C(19)-H(192)	109(2)	C(18)-C(19)-H(193)	112.3(14)
H(191)-C(19)-H(193)	108(2)	H(192)-C(19)-H(193)	107(2)
N(17)-C(20)-C(21)	112.1(2)	N(17)-C(20)-H(201)	106.1(13)
C(21)-C(20)-H(201)	111.5(13)	N(17)-C(20)-H(202)	107.8(12)
C(21)-C(20)-H(202)	112.8(12)	H(201)-C(20)-H(202)	106(2)
C(20)-C(21)-H(211)	111(2)	C(20)-C(21)-H(212)	112(2)
H(211)-C(21)-H(212)	107(2)	C(20)-C(21)-H(213)	107(2)
H(211)-C(21)-H(213)	110(2)	H(212)-C(21)-H(213)	109(2)

Table 4A.1.4. Anisotropic displacement parameters [$Å^2$] for (1).

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
N(1)	0.032(1)	0.066(2)	0.037(1)	-0.020(1)	-0.001(1)	-0.003(1)
C(2)	0.034(1)	0.035(1)	0.023(1)	-0.007(1)	-0.002(1)	-0.003(1)
C(3)	0.027(1)	0.026(1)	0.024(1)	-0.005(1)	-0.003(1)	-0.004(1)
C(4)	0.030(1)	0.027(1)	0.026(1)	-0.006(1)	0.001(1)	-0.004(1)
N(5)	0.040(1)	0.049(1)	0.027(1)	-0.013(1)	-0.005(1)	-0.005(1)
C(6)	0.026(1)	0.017(1)	0.026(1)	-0.002(1)	-0.005(1)	-0.005(1)
C(7)	0.027(1)	0.027(1)	0.023(1)	-0.005(1)	-0.009(1)	-0.005(1)
C(8)	0.022(1)	0.024(1)	0.026(1)	-0.002(1)	-0.006(1)	-0.003(1)
C(9)	0.024(1)	0.018(1)	0.022(1)	-0.002(1)	-0.004(1)	-0.004(1)
C(10)	0.024(1)	0.027(1)	0.026(1)	-0.003(1)	-0.007(1)	-0.006(1)
C(11)	0.021(1)	0.030(1)	0.026(1)	-0.005(1)	-0.004(1)	-0.006(1)
C(12)	0.024(1)	0.020(1)	0.025(1)	-0.002(1)	-0.004(1)	-0.005(1)
C(13)	0.031(1)	0.025(1)	0.021(1)	-0.004(1)	-0.005(1)	-0.006(1)
N(14)	0.028(1)	0.046(1)	0.034(1)	-0.008(1)	-0.005(1)	-0.006(1)
C(15)	0.024(1)	0.027(1)	0.024(1)	-0.004(1)	-0.005(1)	-0.002(1)
C(16)	0.026(1)	0.022(1)	0.027(1)	-0.003(1)	-0.003(1)	-0.004(1)
N(17)	0.028(1)	0.030(1)	0.022(1)	-0.006(1)	-0.004(1)	0.000(1)
C(18)	0.036(1)	0.034(1)	0.027(1)	-0.011(1)	0.000(1)	0.003(1)
C(19)	0.046(2)	0.042(2)	0.031(1)	-0.004(1)	0.008(1)	0.003(1)
C(20)	0.034(1)	0.039(1)	0.022(1)	-0.008(1)	-0.008(1)	0.006(1)
C(21)	0.034(1)	0.059(2)	0.041(2)	-0.011(1)	-0.011(1)	-0.007(1)

Table 4A.1.5. Hydrogen coordinates and isotropic displacement parameters for (1).

Atom	x	y	z	U(eq)
H(7)	0.0766(29)	0.2852(29)	0.6902(15)	0.027(6)
H(8)	0.2515(30)	0.2131(29)	0.5534(14)	0.026(5)
H(10)	-0.2513(28)	0.2830(29)	0.4354(14)	0.027(5)
H(11)	-0.4175(31)	0.3380(30)	0.5747(14)	0.031(6)
H(15)	-0.0849(31)	0.2533(31)	0.3071(14)	0.030(6)
H(16)	0.3022(29)	0.1046(29)	0.2394(13)	0.024(5)
H(181)	0.1820(32)	0.0272(34)	0.v446(16)	0.041(7)
H(182)	0.3557(31)	0.0181(32)	0.1065(15)	0.033(6)
H(191)	0.2046(40)	0.3544(42)	-0.0293(19)	0.066(9)
H(192)	0.4138(39)	0.2190(39)	-0.0438(19)	0.063(8)
H(193)	0.3761(34)	0.3508(36)	0.0342(17)	0.045(7)
H(201)	-0.0759(31)	0.2916(33)	0.0546(17)	0.042(7)
H(202)	-0.1296(30)	0.3609(33)	0.1537(15)	0.035(6)
H(211)	-0.2327(34)	0.0572(37)	0.2262(19)	0.053(8)
H(212)	-0.1633(40)	-0.0329(45)	0.1313(20)	0.074(9)
H(213)	-0.3428(39)	0.1552(38)	0.1324(18)	0.057(8)

Table 4A.2.1. Atomic coordinates and equivalent isotropic displacement parameters [Å^2] for (2).

Atom	x	y	z	U(eq)
N(1)	0.3662(2)	1.0663(2)	0.7942(1)	0.042(1)
C(2)	0.3527(2)	0.9159(2)	0.7896(1)	0.029(1)
C(3)	0.3320(2)	0.7286(2)	0.7854(1)	0.024(1)
C(4)	0.3573(2)	0.6563(2)	0.7012(1)	0.026(1)
N(5)	0.3785(2)	0.5993(2)	0.6333(1)	0.036(1)
C(6)	0.2958(2)	0.6230(2)	0.8622(1)	0.022(1)
C(7)	0.2773(2)	0.4359(2)	0.8600(1)	0.023(1)
C(8)	0.2497(2)	0.3398(2)	0.9366(1)	0.023(1)
C(9)	0.2358(2)	0.4199(2)	1.0220(1)	0.021(1)
C(10)	0.2512(2)	0.6061(2)	1.0231(1)	0.024(1)
C(11)	0.2805(2)	0.7036(2)	0.9474(1)	0.024(1)
C(12)	0.2091(2)	0.3177(2)	1.1010(1)	0.021(1)
C(13)	0.2018(2)	0.1333(2)	1.0893(1)	0.023(1)
N(14)	0.2021(2)	-0.0096(2)	1.0772(1)	0.031(1)
C(15)	0.1979(2)	0.3885(2)	1.1876(1)	0.024(1)
C(16)	0.1666(2)	0.2991(2)	1.2686(1)	0.022(1)
N(17)	0.1535(2)	0.3788(2)	1.3475(1)	0.024(1)
C(18)	0.1414(2)	0.2874(2)	1.4345(1)	0.028(1)
C(19)	-0.0742(3)	0.1658(2)	1.4507(1)	0.034(1)
C(20)	0.1718(2)	0.5711(2)	1.3556(1)	0.026(1)
C(21)	0.3879(3)	0.7243(2)	1.3859(1)	0.036(1)
C(22)	0.1412(2)	0.1072(2)	1.2746(1)	0.028(1)
N(23)	0.1201(2)	-0.0436(2)	1.2845(1)	0.041(1)

Table 4A.2.2. Bond lengths [Å] for (2).

N(1)-C(2)	1.152(2)	C(2)-C(3)	1.423(2)
C(3)-C(6)	1.404(2)	C(3)-C(4)	1.427(2)
C(4)-N(5)	1.148(2)	C(6)-C(7)	1.428(2)
C(6)-C(11)	1.438(2)	C(7)-C(8)	1.361(2)
C(7)-H(7)	0.97(2)	C(8)-C(9)	1.435(2)
C(8)-H(8)	0.96(2)	C(9)-C(12)	1.418(2)
C(9)-C(10)	1.431(2)	C(10)-C(11)	1.356(2)
C(10)-H(10)	0.97(2)	C(11)-H(11)	0.95(2)
C(12)-C(15)	1.411(2)	C(12)-C(13)	1.448(2)
C(13)-N(14)	1.148(2)	C(15)-C(16)	1.393(2)
C(15)-H(15)	0.95(2)	C(16)-N(17)	1.347(2)
C(16)-C(22)	1.454(2)	N(17)-C(20)	1.471(2)
N(17)-C(18)	1.481(2)	C(18)-C(19)	1.517(2)
C(18)-H(181)	0.98(2)	C(18)-H(182)	0.98(2)
C(19)-H(191)	1.02(2)	C(19)-H(192)	0.96(2)
C(19)-H(193)	1.01(2)	C(20)-C(21)	1.518(2)
C(20)-H(201)	0.98(2)	C(20)-H(202)	0.98(2)
C(21)-H(211)	0.99(2)	C(21)-H(212)	0.97(2)
C(21)-H(213)	0.99(2)	C(22)-N(23)	1.150(2)

Table 4A.2.3. Bond angles [°] for (2)

N(1)-C(2)-C(3)	178.5(2)	C(6)-C(3)-C(2)	121.14(12)
C(6)-C(3)-C(4)	121.64(12)	C(2)-C(3)-C(4)	117.18(13)
N(5)-C(4)-C(3)	179.53(14)	C(3)-C(6)-C(7)	122.11(12)
C(3)-C(6)-C(11)	120.27(12)	C(7)-C(6)-C(11)	117.61(12)
C(8)-C(7)-C(6)	120.80(12)	C(8)-C(7)-H(7)	120.9(10)
C(6)-C(7)-H(7)	118.3(10)	C(7)-C(8)-C(9)	121.90(13)
C(7)-C(8)-H(8)	119.1(10)	C(9)-C(8)-H(8)	119.0(10)
C(12)-C(9)-C(10)	122.31(12)	C(12)-C(9)-C(8)	120.80(12)
C(10)-C(9)-C(8)	116.89(12)	C(11)-C(10)-C(9)	121.61(12)
C(11)-C(10)-H(10)	117.3(9)	C(9)-C(10)-H(10)	121.1(9)
C(10)-C(11)-C(6)	121.19(12)	C(10)-C(11)-H(11)	121.1(10)
C(6)-C(11)-H(11)	117.7(10)	C(15)-C(12)-C(9)	124.02(12)
C(15)-C(12)-C(13)	120.31(12)	C(9)-C(12)-C(13)	115.62(12)
N(14)-C(13)-C(12)	176.89(14)	C(16)-C(15)-C(12)	128.84(13)
C(16)-C(15)-H(15)	114.2(9)	C(12)-C(15)-H(15)	116.9(9)
N(17)-C(16)-C(15)	123.95(12)	N(17)-C(16)-C(22)	113.80(12)
C(15)-C(16)-C(22)	122.24(12)	C(16)-N(17)-C(20)	122.34(12)
C(16)-N(17)-C(18)	122.14(11)	C(20)-N(17)-C(18)	115.22(11)
N(17)-C(18)-C(19)	112.27(13)	N(17)-C(18)-H(181)	105.4(10)
C(19)-C(18)-H(181)	111.1(10)	N(17)-C(18)-H(182)	108.2(9)
C(19)-C(18)-H(182)	111.3(10)	H(181)-C(18)-H(182)	108.3(14)
C(18)-C(19)-H(191)	107.7(11)	C(18)-C(19)-H(192)	109.7(12)
H(191)-C(19)-H(192)	110(2)	C(18)-C(19)-H(193)	109.6(10)
H(191)-C(19)-H(193)	109.3(14)	H(192)-C(19)-H(193)	110(2)
N(17)-C(20)-C(21)	111.89(13)	N(17)-C(20)-H(201)	106.3(9)
C(21)-C(20)-H(201)	110.5(9)	N(17)-C(20)-H(202)	109.4(9)
C(21)-C(20)-H(202)	110.1(10)	H(201)-C(20)-H(202)	108.6(13)
C(20)-C(21)-H(211)	110.0(11)	C(20)-C(21)-H(212)	110.7(12)
H(211)-C(21)-H(212)	108(2)	C(20)-C(21)-H(213)	108.7(11)
H(211)-C(21)-H(213)	108(2)	H(212)-C(21)-H(213)	111(2)
N(23)-C(22)-C(16)	176.0(2)		

Table 4A.2.4. Anisotropic displacement parameters (Å^2) for (2).

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
N(1)	0.067(1)	0.036(1)	0.032(1)	0.007(1)	0.012(1)	0.032(1)
C(2)	0.039(1)	0.033(1)	0.019(1)	0.004(1)	0.006(1)	0.020(1)
C(3)	0.026(1)	0.024(1)	0.021(1)	0.001(1)	0.003(1)	0.013(1)
C(4)	0.026(1)	0.028(1)	0.022(1)	0.003(1)	0.001(1)	0.013(1)
N(5)	0.045(1)	0.043(1)	0.024(1)	0.001(1)	0.005(1)	0.025(1)
C(6)	0.020(1)	0.024(1)	0.021(1)	0.000(1)	0.002(1)	0.010(1)
C(7)	0.024(1)	0.022(1)	0.021(1)	-0.002(1)	0.003(1)	0.010(1)
C(8)	0.022(1)	0.020(1)	0.025(1)	-0.002(1)	0.003(1)	0.009(1)
C(9)	0.020(1)	0.020(1)	0.021(1)	0.000(1)	0.003(1)	0.009(1)
C(10)	0.028(1)	0.023(1)	0.021(1)	-0.001(1)	0.004(1)	0.012(1)
C(11)	0.029(1)	0.022(1)	0.023(1)	0.000(1)	0.004(1)	0.014(1)
C(12)	0.018(1)	0.019(1)	0.024(1)	-0.001(1)	0.001(1)	0.008(1)
C(13)	0.022(1)	0.024(1)	0.021(1)	0.002(1)	0.003(1)	0.009(1)
N(14)	0.037(1)	0.026(1)	0.032(1)	0.001(1)	0.005(1)	0.016(1)
C(15)	0.029(1)	0.021(1)	0.022(1)	0.002(1)	0.004(1)	0.013(1)
C(16)	0.024(1)	0.022(1)	0.022(1)	0.001(1)	0.002(1)	0.012(1)
N(17)	0.033(1)	0.022(1)	0.019(1)	0.003(1)	0.004(1)	0.013(1)
C(18)	0.040(1)	0.028(1)	0.018(1)	0.004(1)	0.002(1)	0.018(1)
C(19)	0.043(1)	0.036(1)	0.030(1)	0.011(1)	0.012(1)	0.022(1)
C(20)	0.036(1)	0.024(1)	0.022(1)	0.001(1)	0.006(1)	0.017(1)
C(21)	0.041(1)	0.026(1)	0.036(1)	0.004(1)	0.007(1)	0.012(1)
C(22)	0.033(1)	0.028(1)	0.023(1)	0.002(1)	0.003(1)	0.015(1)
N(23)	0.060(1)	0.030(1)	0.035(1)	0.005(1)	0.006(1)	0.024(1)

Table 4A.2.5. .Hydrogen coordinates and isotropic displacement parameters [Å^2] for (2).

Atom	x	y	z	U(eq)
H(7)	0.2851(25)	0.3792(24)	0.8025(12)	0.035(4)
H(8)	0.2418(25)	0.2156(24)	0.9337(11)	0.033(4)
H(10)	0.2446(23)	0.6694(21)	1.0791(11)	0.026(4)
H(11)	0.2892(24)	0.8267(24)	0.9486(11)	0.031(4)
H(15)	0.2155(23)	0.5152(23)	1.1930(11)	0.027(4)
H(181)	0.2165(25)	0.3932(24)	1.4829(12)	0.035(4)
H(182)	0.2119(24)	0.2111(22)	1.4329(11)	0.027(4)
H(191)	-0.0702(28)	0.1232(26)	1.5150(14)	0.047(5)
H(192)	-0.1414(29)	0.0566(27)	1.4057(14)	0.045(5)
H(193)	-0.1470(26)	0.2447(25)	1.4462(12)	0.039(5)
H(201)	0.0868(22)	0.5692(20)	1.4009(11)	0.023(4)
H(202)	0.1191(24)	0.5963(22)	1.2966(12)	0.030(4)
H(211)	0.3956(27)	0.8522(27)	1.3902(13)	0.044(5)
H(212)	0.4733(29)	0.7241(26)	1.3420(13)	0.044(5)
H(213)	0.4346(28)	0.7020(26)	1.4477(14)	0.045(5)

286

Table 4A.3.1. Atomic coordinates and equivalent isotropic displacement parameters [Å^2] for (4).

Atom	x	y	z	U(eq)
N(1)	1.0147(10)	0.3833(3)	0.4356(9)	0.055(2)
C(2)	0.8633(12)	0.3901(3)	0.3663(12)	0.037(2)
C(3)	0.6730(10)	0.3984(3)	0.2833(11)	0.032(2)
C(4)	0.5736(12)	0.3526(3)	0.2208(11)	0.040(2)
N(5)	0.4942(9)	0.3145(2)	0.1673(10)	0.056(2)
C(6)	0.5901(10)	0.4496(3)	0.2693(11)	0.031(2)
C(7)	0.3978(10)	0.4579(3)	0.1911(11)	0.036(2)
C(8)	0.3198(10)	0.5073(3)	0.1761(12)	0.035(2)
C(9)	0.4222(9)	0.5541(3)	0.2400(10)	0.024(2)
C(10)	0.6127(10)	0.5452(3)	0.3206(10)	0.031(2)
C(11)	0.6934(10)	0.4956(3)	0.3333(11)	0.031(2)
C(12)	0.3323(9)	0.6040(3)	0.2221(12)	0.032(2)
C(13)	0.1325(12)	0.6052(3)	0.1427(11)	0.036(2)
N(14)	-0.0240(10)	0.6057(3)	0.0777(10)	0.052(2)
N(15)	0.3992(8)	0.6526(2)	0.2637(8)	0.035(2)
C(16)	0.2767(9)	0.7018(2)	0.2504(11)	0.046(2)
C(17)	0.4148(11)	0.7444(4)	0.3423(15)	0.039(3)
C(18)	0.6153(13)	0.7267(3)	0.2732(15)	0.039(3)
C(18')	0.5866(45)	0.7253(5)	0.3697(43)	0.077(12)
C(17')	0.4293(27)	0.7473(8)	0.2311(39)	0.043(9)
C(19)	0.6046(9)	0.6661(3)	0.3130(11)	0.042(2)

Table 4A.3.2. Bond lengths [Å] for (4).

N(1)-C(2)	1.146(8)	C(2)-C(3)	1.427(11)
C(3)-C(4)	1.401(10)	C(3)-C(6)	1.405(9)
C(4)-N(5)	1.159(9)	C(6)-C(11)	1.421(8)
C(6)-C(7)	1.426(8)	C(7)-C(8)	1.351(8)
C(7)-H(7A)	0.96	C(8)-C(9)	1.435(9)
C(8)-H(8A)	0.96	C(9)-C(12)	1.397(9)
C(9)-C(10)	1.424(8)	C(10)-C(11)	1.363(8)
C(10)-H(10A)	0.96	C(11)-H(11A)	0.96
C(12)-N(15)	1.333(9)	C(12)-C(13)	1.461(11)
C(13)-N(14)	1.153(8)	N(15)-C(19)	1.474(7)
N(15)-C(16)	1.491(7)	C(16)-C(17')	1.557(10)
C(16)-C(17)	1.559(9)	C(16)-H(16A)	0.96
C(17)-C(18)	1.550(9)	C(16)-H(16B)	0.96
C(17)-H(17B)	0.97	C(17)-H(17A)	0.97
C(17)-H(18D)	1.53(3)	C(17)-H(17C)	1.06(2)
C(18)-C(19)	1.551(7)	C(18)-H(18A)	0.97
C(18)-H(18B)	0.97	C(18)-H(18C)	0.88(3)
C(18')-C(19)	1.551(9)	C(18')-C(17')	1.550(10)
C(18')-H(18C)	0.97	C(18')-H(18D)	0.97
C(17')-H(17C)	0.97	C(17')-H(17D)	0.97
C(19)-H(19A)	0.96	C(19)-H(19B)	0.96

N(1)-C(2)-C(3)	178.9(9)	C(4)-C(3)-C(6)	122.4(7)
C(4)-C(3)-C(2)	116.1(7)	C(6)-C(3)-C(2)	121.5(7)
N(5)-C(4)-C(3)	178.8(9)	C(3)-C(6)-C(11)	121.6(6)
C(3)-C(6)-C(7)	121.6(7)	C(11)-C(6)-C(7)	116.8(6)
C(8)-C(7)-C(6)	121.3(7)	C(8)-C(7)-H(7A)	121.4(4)
C(6)-C(7)-H(7A)	117.2(4)	C(7)-C(8)-C(9)	122.6(7)
C(7)-C(8)-H(8A)	119.1(4)	C(9)-C(8)-H(8A)	118.3(4)
C(12)-C(9)-C(10)	124.8(7)	C(12)-C(9)-C(8)	119.7(6)
C(10)-C(9)-C(8)	115.5(7)	C(11)-C(10)-C(9)	122.1(7)
C(11)-C(10)-H(10A)	118.4(4)	C(9)-C(10)-H(10A)	119.4(4)
C(10)-C(11)-C(6)	121.6(7)	C(10)-C(11)-H(11A)	120.1(4)
C(6)-C(11)-H(11A)	118.2(4)	N(15)-C(12)-C(9)	130.7(6)
N(15)-C(12)-C(13)	112.2(7)	C(9)-C(12)-C(13)	117.1(7)
N(14)-C(13)-C(12)	178.9(9)	C(12)-N(15)-C(19)	125.4(6)
C(12)-N(15)-C(16)	123.8(6)	C(19)-N(15)-C(16)	110.6(5)
C(17')-C(16)-N(15)	103.6(10)	N(15)-C(16)-C(17)	102.2(5)
C(17')-C(16)-H(16A)	133.4(11)	N(15)-C(16)-H(16A)	111.2(4)
C(17)-C(16)-H(16A)	109.3(5)	C(17')-C(16)-H(16B)	84.4(10)
N(15)-C(16)-H(16B)	111.3(4)	C(17)-C(16)-H(16B)	113.2(5)
H(16A)-C(16)-H(16B)	109.5	C(18)-C(17)-C(16)	100.8(7)
C(18)-C(17)-H(17A)	112.5(5)	C(16)-C(17)-H(17A)	112.2(5)
C(18)-C(17)-H(17B)	110.1(6)	C(16)-C(17)-H(17B)	111.4(5)
H(17A)-C(17)-H(17B)	109.7	C(16)-C(17)-H(18D)	122.7(11)
C(16)-C(17)-H(17C)	105.4(10)	H(18D)-C(17)-H(17C)	131.9(14)
C(17)-C(18)-C(19)	99.6(6)	C(17)-C(18)-H(18A)	113.6(5)
C(19)-C(18)-H(18A)	112.8(5)	C(17)-C(18)-H(18B)	110.1(6)
C(19)-C(18)-H(18B)	110.9(5)	H(18A)-C(18)-H(18B)	109.6
C(19)-C(18)-H(18C)	112.8(14)	C(19)-C(18')-C(17')	103(2)
C(19)-C(18')-H(18A)	105(2)	C(19)-C(18')-H(18C)	107.4(14)
C(17')-C(18')-H(18C)	105(2)	C(19)-C(18')-H(18D)	114.5(14)
C(17')-C(18')-H(18D)	117(2)	H(18C)-C(18')-H(18D)	109.3
C(16)-C(17')-C(18')	97(2)	C(16)-C(17')-H(17A)	107(2)
C(16)-C(17')-H(18B)	117(2)	H(17A)-C(17')-H(18B)	134.5(9)
C(16)-C(17')-H(17C)	110.9(11)	C(18')-C(17')-H(17C)	109.4(13)
C(16)-C(17')-H(17D)	112.4(11)	C(18')-C(17')-H(17D)	117(2)
H(17C)-C(17')-H(17D)	109.0	N(15)-C(19)-C(18')	101.5(12)
N(15)-C(19)-C(18)	103.3(6)	N(15)-C(19)-H(19A)	111.1(4)
C(18')-C(19)-H(19A)	133(2)	C(18)-C(19)-H(19A)	109.2(4)
N(15)-C(19)-H(19B)	110.7(4)	C(18')-C(19)-H(19B)	88.2(12)
C(18)-C(19)-H(19B)	112.8(5)	H(19A)-C(19)-H(19B)	109.6

288

Table 4A.3.4. Anisotropic displacement parameters [Å^2] for (4).

Atom	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
N(1)	0.044(5)	0.061(5)	0.058(6)	0.003(4)	-0.016(4)	0.008(4)
C(2)	0.037(6)	0.032(5)	0.041(7)	0.006(5)	0.007(5)	0.000(4)
C(3)	0.034(6)	0.030(5)	0.031(7)	0.000(4)	-0.006(5)	-0.007(4)
C(4)	0.041(5)	0.049(5)	0.031(6)	-0.006(5)	-0.003(4)	0.013(5)
N(5)	0.057(4)	0.038(4)	0.071(6)	-0.012(4)	-0.001(4)	0.004(4)
C(6)	0.029(5)	0.035(5)	0.028(6)	0.007(5)	0.003(4)	0.002(5)
C(7)	0.034(5)	0.024(5)	0.050(7)	0.006(5)	-0.002(5)	-0.007(4)
C(8)	0.026(5)	0.043(5)	0.037(7)	-0.001(5)	-0.001(4)	-0.001(4)
C(9)	0.015(4)	0.037(5)	0.021(6)	0.004(4)	-0.001(4)	0.001(4)
C(10)	0.030(5)	0.042(5)	0.022(6)	-0.010(5)	0.004(4)	-0.004(5)
C(11)	0.026(5)	0.036(5)	0.032(7)	0.001(5)	0.005(4)	-0.006(4)
C(12)	0.014(5)	0.044(6)	0.037(7)	0.007(5)	-0.003(4)	-0.001(4)
C(13)	0.038(6)	0.038(5)	0.033(7)	-0.004(5)	0.002(5)	-0.003(5)
N(14)	0.032(4)	0.065(5)	0.059(6)	-0.002(4)	-0.005(4)	-0.001(4)
N(15)	0.026(4)	0.029(4)	0.052(5)	-0.001(4)	0.006(4)	-0.002(3)
C(16)	0.035(4)	0.030(4)	0.074(6)	0.009(5)	0.006(4)	0.009(4)
C(19)	0.032(4)	0.033(4)	0.061(6)	-0.005(4)	0.005(4)	0.003(4)

Table 4A.3.5. Hydrogen coordinates and isotropic displacement parameters [Å^2] for (4).

Atom	x	y	z	$U(eq)$
H(7A)	0.3271(10)	0.4271(3)	0.1474(11)	0.044
H(8A)	0.1930(10)	0.5117(3)	0.1163(12)	0.043
H(10A)	0.6864(10)	0.5748(3)	0.3703(10)	0.037
H(11A)	0.8254(10)	0.4911(3)	0.3810(11)	0.037
H(16A)	0.1613(9)	0.6982(2)	0.3186(11)	0.055
H(16B)	0.2396(9)	0.7100(2)	0.1266(11)	0.055
H(17A)	0.3799(11)	0.7803(4)	0.3051(15)	0.047
H(17B)	0.4160(11)	0.7417(4)	0.4734(15)	0.047
H(18A)	0.7262(13)	0.7439(3)	0.3358(15)	0.046
H(18B)	0.6195(13)	0.7332(3)	0.1440(15)	0.046
H(18C)	0.7072(45)	0.7431(5)	0.3404(43)	0.092
H(18D)	0.5635(45)	0.7308(5)	0.4969(43)	0.092
H(17C)	0.3794(27)	0.7808(8)	0.2749(39)	0.052
H(17D)	0.4669(27)	0.7518(8)	0.1071(39)	0.052
H(19A)	0.6924(9)	0.6475(3)	0.2373(11)	0.050
H(19B)	0.6359(9)	0.6577(3)	0.4379(11)	0.050

Appendix B

The following research conferences were attended during the period of tuition for this Thesis:

December 1991	Siemens Users Meeting, University of Cambridge
March 1992	British Crystallographic Association Annual Spring Meeting, University of Liverpool
July 1992	European Crystallographic Association Annual Meeting, Enschede
March 1993	British Crystallographic Association Intensive School of Crystallography, University of Aston
March 1993	British Crystallographic Association Annual Spring Meeting, University of Manchester
August 1993	International Union of Crystallography Congress, Beijing, China
August 1993	International Union of Crystallography Small Molecule Symposium, Fuzhou, China: at which the lecture "Systematic Conformational Analysis from Crystallographic Data" was presented.

The following is a list of colloquia, lectures and seminars from invited speakers to the University of Durham's, Department of Chemistry during the period of this research. Those marked (*) were attended by the author of this Thesis.

1991

| October 17 | Dr. J.A. Salthouse(*), University of Manchester |
| | Son et Lumiere - a demonstration lecture |

| October 31 | Dr. R. Keeley (*), Metropolitan Police Forensic Science |
| | Modern forensic science |

| November 6 | Prof. B.F.G. Johnson (*), Edinburgh University |
| | Cluster-surface analogies |

| November 7 | Dr. A.R. Butler (*), St. Andrews University |
| | Traditional Chinese herbal drugs: a different way of treating disease |

| November 13 | Prof. D. Gani (*), St. Andrews University |
| | The chemistry of PLP-dependent enzymes |

| November 20 | Dr. R. More O'Ferrall (*), University College, Dublin |
| | Some acid-catalysed rearrangements in organic chemistry |

| November 28 | Prof. I.M. Ward (*), IRC in Polymer Science, University of Leeds |
| | The SCI lecture: the science and technology of orientated polymers |

| December 4 | Prof. R. Grigg (*), Leeds University |
| | Palladium-catalysed cyclisation and ion-capture processes |

| December 5 | Prof. A.L. Smith (*), ex Unilever |
| | Soap, detergents and black puddings |

| December 11 | Dr. W.D. Cooper, Shell Research |
| | Colloid science: theory and practice |

1992

| January 22 | Dr. K.D.M. Harris, St. Andrews University |
| | Understanding the properties of solid inclusion compounds |

| January 29 | Dr. A. Holmes (*), Cambridge University |
| | Cycloaddition reactions in hte service of the synthesis of piperidine and indolizidine natural products |

| January 30 | Dr. M. Anderson (*), Sittingbourne Research Centre, Shell Research |
| | Recent Advances in the Safe and Selective Chemical Control of Insect Pests |

| February 12 | Prof. D.E. Fenton (*), Sheffield University |
| | Polynuclear complexes of molecular clefts as models for copper biosites |

| February 13 | Dr. J. Saunders (*), Glaxo Group Research Limited |
| | Molecular Modelling in Drug Discovery |

| February 19 | Prof. E.J. Thomas, Manchester University |
| | Applications of organostannanes to organic synthesis |

| February 20 | Prof. E.Vogel (*), University of Cologne |
| | The Musgrave Lecture: Porphyrins: Molecules of Interdisciplinary Interest |

| February 25 | Prof. J.F. Nixon (*), University of Sussex |
| | The Tilden Lecture: Phosphaalkynes: new building blocks in inorganic and organometallic chemistry |

February 26	Prof. M.L. Hitchman, Strathclyde University
	Chemical vapour deposition
March 5	Dr. N.C. Billingham (*), University of Sussex
	Degradable Plastics - Myth or Magic?
March 11	Dr. S.E. Thomas, Imperial College
	Recent advances in organoiron chemistry
March 12	Dr. R.A. Hann (*), ICI Imagedata
	Electronic Photography - An Image of the Future
March 18	Dr. H. Maskill, Newcastle University
	Concerted or stepwise fragmentation in a deamination-type reaction
April 7	Prof. D.M. Knight, Philosophy Department, University of Durham
	Interpreting experiments: the beginning of electrochemistry
May 13	Dr. J-C Gehret (*), Ciba Geigy, Basel
	Some aspects of industrial agrochemical research
October 15	Dr M. Glazer & Dr. S. Tarling, Oxford University & Birbeck College
	It Pays to be British! - The Chemist's Role as an Expert Witness in
	Patent Litigation
October 20	Dr. H. E. Bryndza (*), Du Pont Central Research
	Synthesis, Reactions and Thermochemistry of Metal (Alkyl) Cyanide
	Complexes and Their Impact on Olefin Hydrocyanation Catalysis
October 22	Prof. A. Davies (*), University College London
	The Ingold-Albert Lecture: The Behaviour of Hydrogen as a
	Pseudometal
October 28	Dr. J. K . Cockcroft (*), University of Durham
	Recent Developments in Powder Diffraction
October 29	Dr. J. Emsley (*), Imperial College, London
	The Shocking History of Phosphorus
November 4	Dr. T. P. Kee (*), University of Leeds
	Synthesis and Co-ordination Chemistry of Silylated Phosphites
November 5	Dr. C. J. Ludman (*), University of Durham
	Explosions, A Demonstration Lecture

November 11	Prof. D. Robins, Glasgow University
	Pyrrolizidine Alkaloids: Biological Activity, Biosynthesis and Benefits
November 12	Prof. M. R. Truter (*), University College, London
	Luck and Logic in Host - Guest Chemistry
November 18	Dr. R. Nix, Queen Mary College, London
	Characterisation of Heterogeneous Catalysts
November 25	Prof. Y. Vallee (*), University of Caen
	Reactive Thiocarbonyl Compounds
November 25	Prof. L. D. Quin (*), University of Massachusetts, Amherst
	Fragmentation of Phosphorous Heterocycles as a Route to
	Phosphoryl Species with Uncommon Bonding
November 26	Dr. D. Humber (*), Glaxo, Greenford
	AIDS - The Development of a Novel Series of Inhibitors of HIV
December 2	Prof. A. F. Hegarty (*), University College, Dublin
	Highly Reactive Enols Stabilised by Steric Protection
December 2	Dr. R. A. Aitken, University of St. Andrews
	The Versatile Cycloaddition Chemistry of Bu3P.CS2
December 3	Prof. P. Edwards (*), Birmingham University
	The SCI Lecture - What is Metal?
December 9	Dr. A. N. Burgess, ICI Runcorn
	The Structure of Perfluorinated Ionomer Membranes

1993

January 20	Dr. D. C. Clary (*), University of Cambridge
	Energy Flow in Chemical Reactions
January 21	Prof. L. Hall (*), Cambridge
	NMR - Window to the Human Body
January 27	Dr. W. Kerr, University of Strathclyde
	Development of the Pauson-Khand Annulation Reaction:
	Organocobalt Mediated Synthesis of Natural and Unnatural Products
January 28	Prof. J. Mann (*), University of Reading
	Murder, Magic and Medicine

| February 3 | Prof. S. M. Roberts, University of Exeter |
| | Enzymes in Organic Synthesis |

| February 10 | Dr. D. Gillies, University of Surrey |
| | NMR and Molecular Motion in Solution |

| February 11 | Prof. S. Knox (*), Bristol University |
| | The Tilden Lecture: Organic Chemistry at Polynuclear Metal Centres |

| February 29 | University of Leicester |
| | Oxatrimethylenemethane Metal Complexes |

| February 18 | Dr. I. Fraser (*), ICI Wilton |
| | Reactive Processing of Composite Materials |

| February 22 | Prof. D. M. Grant (*), University of Utah |
| | Single Crystals, Molecular Structure, and Chemical-Shift Anisotropy |

| February 24 | Prof. C. J. M. Stirling, University of Sheffield |
| | Chemistry on the Flat-Reactivity of Ordered Systems |

| March 10 | Dr. P. K. Baker (*), University College of North Wales, Bangor |
| | Chemistry of Highly Versatile 7-Coordinate Complexes |

| March 11 | Dr. R. A. Y. Jones (*), University of East Anglia |
| | The Chemistry of Wine Making |

| March 17 | Dr. R. J. K. Taylor (*), University of East Anglia |
| | Adventures in Natural Product Synthesis |

| March 24 | Prof. I. O. Sutherland, University of Liverpool |
| | Chromogenic Reagents for Cations |

| May 13 | Prof. J. A. Pople (*), Carnegie-Mellon University, Pittsburgh, USA |
| | The Boys-Rahman Lecture: Applications of Molecular Orbital Theory |

| May 21 | Prof. L. Weber, University of Bielefeld |
| | Metallo-phospha Alkenes as Synthons in Organometallic Chemistry |

| June 1 | Prof. J. P. Konopelski, University of California, Santa Cruz |
| | Synthetic Adventures with Enantiomerically Pure Acetals |

| June 2 | Prof. F. Ciardelli, University of Pisa |
| | Chiral Discrimination in the Stereospecific Polymerisation of Alpha Olefins |

June 7	Prof. R. S. Stein (*), University of Massachusetts
	Scattering Studies of Crystalline and Liquid Crystalline Polymers

June 16	Prof. A. K. Covington, University of Newcastle
	Use of Ion Selective Electrodes as Detectors in Ion Chromatography

June 17	Prof. O. F. Nielsen, H. C. 0rsted Institute, University of Copenhagen
	Low-Frequency IR - and Raman Studies of Hydrogen Bonded Liquids

September 13	Prof. Dr. A.D. Schluter, Freie Universitat Berlin, Germany
	Synthesis and Characterisation of Molecular Rods and Ribbons

September 13	Dr. K.J. Wynne, Office of Naval Research, Washington, USA
	Polymer Surface Design for Minimal Adhesion

September 14	Prof. J.M. DeSimone, University of North Carolina, USA
	Homogeneous and Heterogeneous Polymerisations in Environmentally Responsible Carbon Dioxide

September 28	Prof. H. Ila, North Eastern Hill University, India
	Synthetic Strategies for Cyclopentanoids via Oxoketene Dithioacetals

October 4	Prof. F.J. Feher, University of California, Irvine, USA
	Bridging the Gap between Surfaces and Solution with Sessilquioxanes

October 14	Dr. P. Hubberstey (*), University of Nottingham
	Alkali Metals: Alchemist's Nightmare, Biochemist's Puzzle and Technologist's Dream

October 20	Dr. P. Quayle (*), University of Manchester
	Aspects of Aqueous ROMP Chemistry

October 21	Prof. R. Adams (*), University of South Carolina, USA
	Chemistry of Metal Carbonyl Cluster Complexes: Development of Cluster Based Alkyne Hydrogenation Catalysts

October 27	Dr. R.A.L. Jones (*), Cavendish Laboratory, Cambridge
	Perambulating Polymers

November 10	Prof. M.N.R. Ashfoldt, University of Bristol
	High Resolution Photofragment Translational Spectroscopy: A New Way to Watch Photodissociation

November 17	Dr. A. Parker, Rutherford Appleton Laboratory, Didcot
	Applications of Time Resolved Resonance Raman Spectroscopy
	to Chemical and Biochemical Problems

November 24   Dr. P.G. Bruce (*), University of St. Andrews
                        Structure and Properties of Inorganic Solids and Polymers

November 25   Dr. R.P. Wayne (*), University of Oxford
                        The Origin and Evolution of the Atmosphere

December 1   Prof. M.A. McKervey, Queen's University, Belfast
                        Synthesis and Applications of Chemically Modified Calixarenes

December 8   Prof. O. Meth-Cohn (*), University of Sunderland
                        Friedel's Folly Revisited - A Super Way to Fused Pyridines

December 16   Prof. R.F. Hudson, University of Kent
                        Close Encounters of the Second Kind

1994

January 26   Prof. J. Evans (*), University of Southampton
                        Shining Light on Catalysts

February 2   Dr. A. Masters, University of Manchester
                        Modelling Water Without Using Pair Potentials

February 9   Prof. D. Young, University of Sussex
                        Chemical and Biological Studies on the Coenzyme Tetrahydrofolic
                        Acid

February 16   Prof. K.H. Theopold, University of Delaware, USA
                        Paramagnetic Chromium Alkyls: Synthesis and Reactivity

February 23   Prof. P.M. Maitlis (*), University of Sheffield
                        Across the Border: From Homogeneous to Heterogeneous Catalysis

March 2   Dr. C. Hunter, University of Sheffield
                        Noncovalent Interactions between Aromatic Molecules

March 9   Prof. F. Wilkinson, Loughborough University of Technology
                        Nanosecond and Picosecond Laser Flash Photolysis

March 10   Prof. S.V. Ley, University of Cambridge
                        New Methods for Organic Synthesis