Studies of conformation and configuration using crystallographic methods

Pritchford, Nigel Aaron

Studies of Conformation and Configuration using Crystallographic Methods

by

Nigel Aaron Pitchford

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

Doctor of Philosophy

at the

University of Durham

Department of Chemistry

September 1994

Studies of Conformation and Configuration using Crystallographic Methods

Submitted for the degree of Doctor of Philosophy, September 1994, by

Nigel A. Pitchford, University of Durham

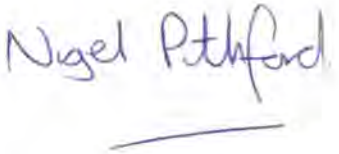
ABSTRACT

This Thesis demonstrates the use of the Cambridge Crystallographic Database for structure correlation studies in two very different fields. The first part of the Thesis (Chapters 2 and 3) is concerned with the systematic conformational analysis of medium-sized rings and satisfies the objectives of the study by: (i) applying novel classification techniques to the conformational descriptions of both the seven- and eight-membered rings, (ii) interpreting the results in terms of the relevant conformational hypersurface by locating the highly populated regions of that hypersurface and mapping the interconversion pathways, (iii) studying, modifying and improving the available methodologies for data analysis, and (iv) relating the conformational minima found using these methods to both the chemical environments of the fragments under investigation, and to energetic features of the hypersurface obtained by computational methods.

The second major structure correlation experiment involves the analysis and description of 3-coordinated transition metal complexes using both simple geometrical models and group-theoretically based symmetry deformation coordinates. Non-bonded interactions will be seen to play a significant part in the geometry of the 3-coordinated fragment, and extrapolation of these results leads to the rationalisation of an addition/elimination scheme linking 4- and 2-coordinated fragments through the intermediate 3-coordinated species.

Chapter 5 describes the crystallographic structure determinations of eight novel compounds: 3,5-cycloheptadienyl-3,5 dinitrobenzoate [$C_{14}H_{12}O_6N_2$]; a 34-membered diolide [$C_{32}H_{60}O_4$]; 1-iodo-3-tosyloxy-propan-2-ane [$C_{10}H_{11}O_4IS$)]; 1β, 9β–diacetyl-7α-chloro-cis-hydrindane [$C_{13}H_{19}O_2Cl$]; (R,R)-1,4-bis (2'-chloro-1'-hydoxyethyl) benzene [$C_{10}H_{12}O_2Cl_2$]; a fused penta-cyclic ring compound [$C_{17}H_{14}$]; 1,4 dibenzyl-1,2,4,5-tetraazacyclohexane [$C_{16}H_{20}N_4$]; 1,5-di (2'-chloroacetoxy)-3,3-dimethyl-2,4-diphenyl-3-silapentane [$C_{22}H_{26}O_4Cl_2Si$].

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



N.A. Pitchford

To Dawn

Acknowledgements

Firstly, I would like to thank my "co-supervisors", Professor Judith Howard and Dr. Frank Allen for their considerable help, and invaluable guidance during these last three years. The support and encouragement that they have given to me have helped to make the time that I have spent here in Durham an extremely enjoyable and rewarding experience.

I am deeply indebted to the Cambridge Crystallographic Data Centre for the financial support that they have given me, and would like to thank Dr. Olga Kennard for her interest and support throughout these studies. I would also like to thank Dr. Andy Vinter for his considerable efforts in developing the COSMIC molecular modelling program to suit my needs, and Mrs. Dee Williams for her graphical art skills and the ability to make helical pathways on tori look easy. Also, Drs. Scott Rowland and Owen Johnson are to be thanked for the rapid turnaround of new CSD code that enabled me to carry out the punishing studies that are detailed in this Thesis, and Jason Cole is also thanked for the many hours he has spent helping me with crystallographic problems.

Since moving to Durham I have seen the lab grow from literally nothing to the size it is today, and during this period many people have been / gone / still remain. Too numerous to mention by name I would like to thank them all for making the lab an interesting and cosmopolitan place to work, and one with a shared, if at times a little strange, sense of humour. Without these people my time in Durham could not have been so enjoyable.

Away from work the care and support of my girlfriend, Dawn, is deeply appreciated, as too is the continued support of my parents. The numerous friends that I have made in Durham, and inparticular at the hockey club, will be sorely missed - thanks to you all for making Durham an enjoyable place to live.

Contents

List of Tables

List of Illustrations

Chapter One

Introduction

1.1 Introduction

"There is no more basic enterprise in chemistry than the determination of the geometrical structure of a molecule. Such a determination, when it is well done, ends all speculation as to the structure and provides us with the starting point for the understanding of every physical, chemical and biological property of the molecule." (Hoffmann, 1983).

Determinations of molecular structure have played an enormous role in the advancement of chemistry this century, with the greatest contribution to these investigations coming from diffraction studies. X-ray crystallography has now become the most powerful method for deriving three-dimensional structural information about a compound at the atomic level. The technique is applicable to every chemical class of compounds, provided that a suitable crystalline sample can be obtained. The purpose of many of the earlier studies of matter was to determine a single and important molecular structure, and then to relate its geometrical structure to the properties of the material in question. Early X-ray analyses concentrated on rather simple ionic and molecular compounds, due to the time-consuming nature of the studies. However, the results established clear definitions for atomic connectivity and for the conformational, configurational, and stereochemical features of molecules. Information was also derived concerning the precise geometrical arrangements of the atoms, and of their aggregation through intermolecular interactions in the crystalline state.

However, it soon became apparent that universal properties could also be derived from a collection of unrelated structural data. Pauling (1939) was one of the first exponents of this source of information, when he stated that : "It has been found that the values of interatomic distances corresponding to covalent bonds can be correlated in a simple way in terms of a set of values of covalent bond radii of atoms." Since that time structure determinations of over 100,000 different molecules have been

performed, and virtually all confirm Pauling's initial postulate. Pauling and his colleagues also noted that the standard distances based on covalent radii sums needed to be slightly adjusted for certain types of bonds (*e.g.* those involving large differences in electronegativity between the atoms), and also depended on the number and type of other atoms surrounding the bond in question. Pauling's work marked the start of an era when it was realised that information could be found by analysing a large number of results together. Work in this field continues to this day, and the results of studies performed with large collections of structural data will be detailed below.

Over the past 30 years the advent of automated data collection and the development of powerful direct-phasing techniques, together with the ever-increasing power of computers, has led to a surge in the number of structure solutions reported: the number of published structures rising from just a few hundred per year in the 1960's, to currently well over 15,000 per year. Computerized databases and sophisticated software systems have been developed over the years to cope with this overwhelming increase in data, so that the four major databases (see *e.g.* Allen, Bergerhoff & Sievers, 1987), shown in Figure 1.1, now comprise more than 200,000 separate crystal structure determinations in total.

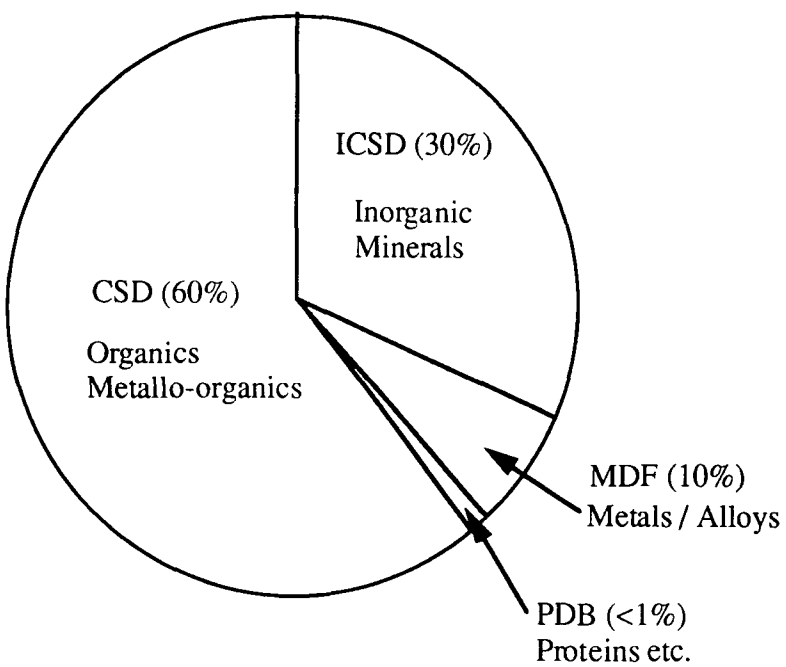


Figure 1.1

Approximate breakdown of the crystallographic databases in terms of their relative number of independent structures, and their content

Early database development involved definition of information content, and the acquisition and organisation of existing and newly published data. The second phase of this development then concentrated on the creation of software systems for the search, retrieval, display and analysis of database content. This is an ongoing concern, as computers change rapidly, and the expectations of database users develop. In particular, the growing awareness that a vast amount of chemical and structural knowledge can be acquired through systematic analysis and correlation of the accumulated data, has led to the rapid development of software designed to facilitate this type of research. The types of studies detailed below rely on rapid and easy access to the large amounts of information held in the databases, the importance of which will become apparent throughout this Thesis.

1.2 Structure correlation

From the studies of Pauling (1947) it was noted that the examination of standardised fragments, and the investigation of their correlation with particular environments, could lead to the determination of specific properties for a general chemical system (in this case the derivation of covalent radii). It is this underlying belief that supports the many reported studies that utilise "structure correlation" techniques (Bürgi & Dunitz, 1994).

Structure correlation studies are founded on the concept that each observation of a specific fragment in a distinct crystal structure can be regarded as a static snapshot of that fragment within a specific chemical and crystallographic environment. The combination of data for the same fragment, but in different environments can, however, provide dynamic information about the fragment in question.

This approach has been succinctly described by Auf der Heyde & Bürgi (1989) in the following manner: "Although direct observation of a molecule along the reaction pathway does not seem feasible, its visualisation at least does. According to the structure correlation hypothesis, the gradual distortion or static deformation that a molecular fragment of interest manifests collectively over a large variety of crystalline

frameworks may be assumed to mirror the distortion which that fragment would undergo along a given reaction coordinate. The various crystal or molecular structures are considered to constitute a series of 'frozen-in' points, or snapshots, taken along the reaction pathway, which, when viewed in the correct order, yield a cinematic film of the reaction."

It is immediately evident that the results of structure correlation studies are not solely limited to describing statistical relationships between geometrical properties, and that information can also be found concerning energetical features of the fragment's chemistry. More formally, the relationship between structural (geometrical) parameters and features of the relevant energy surface can be summarised in the principle of structure correlation (Murray-Rust, Bürgi & Dunitz, 1975): "If a correlation can be found between two or more independent parameters describing the structure of a given structural fragment in a variety of environments, then the correlation function maps a minimum energy path in the corresponding parameter space"; which can also be interpreted as (Bürgi & Dunitz, 1983): "Observed structures tend to concentrate in low lying regions of the potential energy surface".

In the last 10-20 years such approaches have led to a greater understanding of fragment flexibility, conformational diversity, reaction pathways or mechanisms, and molecular design, especially when combined with other physico-chemical quantities; the apparent scope for the technique only being limited by the availability of suitable crystal structure data. The breadth of the topics studied is considerable, and some of the more important ones will now be illustrated as examples of the power and validity of structure correlation techniques.

1.2.1 Mean molecular dimensions

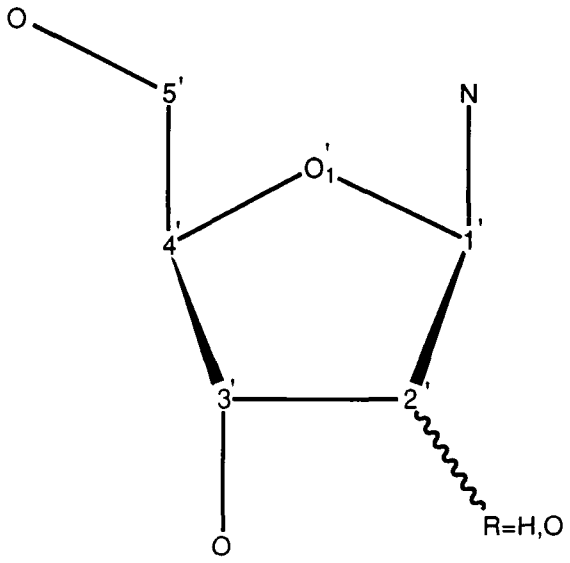
The primary results of X-ray diffraction experiments are high-precision data for the 3D coordinates of the atoms that comprise the crystal structure. From these coordinates it is possible to calculate bond lengths and angles, torsion angles, and any other geometrical descriptors of the shape of the structure. Bond lengths are relatively

"hard" parameters, *i.e.* they have high force constants and are relatively invariant from structure to structure. Hence, it is of value to compile detailed lists of mean bond lengths involving different chemical elements in different chemical environments. The crystallographic databases are ideal for this work and two such studies list the typical interatomic distances found in organic (Allen *et al.*, 1987) and organometallic compounds, and coordination complexes of the d- and f-block metals (Orpen *et al.*, 1989), using almost 50,000 crystal structures as its source of information. Examples of other studies include the determination of peptide chain structure parameters, bond angles and conformational angles (Ashida *et al.*, 1987), and the analysis of the flexible furanose ring (Bartenev *et al.*, 1987). Information from such studies provide structural "yardsticks" that are useful both to the crystallographer in trying to solve new crystal structures, and also to the molecular modeller who relies on precise parameters for accurate calculation purposes.

1.2.2 Conformational analysis

The conformation of a molecule is defined as an arrangement of atoms of given configuration corresponding to a potential energy minimum. At a certain temperature, one or more different conformations may be observed depending on the relative energies involved and the energy barriers that must be overcome to accomplish interconversion. These barriers are influenced by both intra- and inter-molecular forces. X-ray analysis determines the preferred conformation of a molecule, with distortions from the minimum energy form occurring as a result of environmental factors, both chemical and structural. The relative frequency of incidence of similar conformations of the same fragment in different crystal structures can give a qualitative idea about the relative stabilities of the different conformations, as those with lower potential energies are expected to occur more often than those of higher energy. However (Bürgi & Dunitz, 1988), it is formally invalid to try to infer quantitative energy relationships from conformational occupancy ratios: the basic data can never remotely approximate the closed system required by the Boltzmann equation.

Various structure correlation studies have drawn upon these principles in order to perform the conformational analyses of a number of different chemical systems. One of the earliest such studies concerned the conformational analysis of the β-1'-aminofuranoside fragments **[F1]** located in nucleosides and nucleotides, found in the CSD (Murray-Rust & Motherwell, 1978a,b; Murray-Rust & Bland, 1978). A principal component analysis (PCA: see Section 2.3.3 for full details) was carried out on conformations of the five-membered ring, and its direct substituents. This showed that three factors only were primarily responsible for the conformational variation: the first two being ring-puckering factors, and the third the side-chain [C(3')-C(4')-C(5')-O] rotation, with the side-chain rotation apparently being completely independent of the first two factors. The PCA scatterplots produced by this analysis showed two distinct major clusters of "conformational" density: the C(2') *endo* conformation and the C(3') *endo* conformation.



β-1'-aminofuranoside fragment **[F1]**

Conformational analysis can be performed using structure correlation principles on a host of chemical systems. There are, however, certain limitations which can affect the interpretation and visualisation of the results: each parameter used in the analysis adding one more dimension to the problem. Thus, for the conformational analysis of six-membered rings (Allen, Doyle & Taylor, 1991a,b,c) three parameters were required to describe the system (Pickett & Strauss, 1971; Cremer & Pople, 1975), and the resulting representation was a three-dimensional sphere. For seven-membered rings

(Allen, Howard & Pitchford, 1993) the additional parameter results in a four-dimensional toroidal representation of conformational space. For the analysis of the fragment $M(PPh_3)_2$ (Norskov-Lauritsen & Bürgi, 1985) there are eight conformational degrees of freedom, two about the M-P bonds, and six about the P-Ph bonds. The resulting analysis is thus quite complex, but shows a gearing motion of the two PPh_3 groups alternating with helicity inversion of one and then the other PPh_3 group. More information concerning conformational analysis will be given in Chapters 2 and 3 of this Thesis.

1.2.3 Reaction pathways

In the field of organic chemistry there was considerable interest in the nature of the non-bonded intramolecular interaction between nitrogen and carbonyl groups: N...C=O interactions being particularly prevalent in structures of natural products. However, it was not until Bürgi, Dunitz & Shefter (1973a,b) examined six suitable crystal structures that evidence that these interactions mapped the reaction coordinate for the nucleophilic addition to the carbonyl group could be shown.

The structural parameters for intramolecular N...C=O interactions obtained from the structural analyses could be fitted to a simple quantitative model displaying conservation of bond order, defined as a logarithmic function of the interatomic

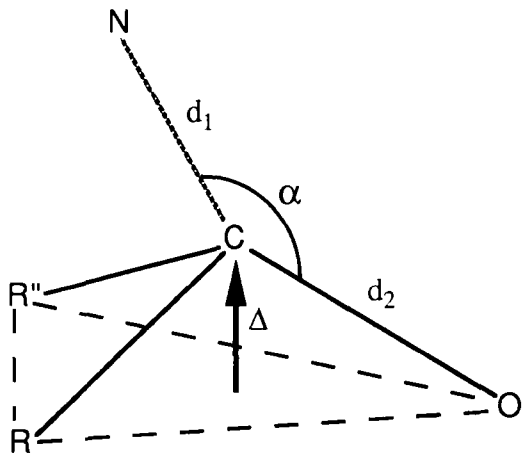


Figure 1.2

Definition of the geometrical parameters that describe the disposition of the nitrogen nucleophile relative to the carbonyl RR"C=O group

distance (Pauling, 1947; Johnston, 1960; Johnston & Parr, 1963). The model successfully relates the order of the C-N bond (logarithmic function of d_1) to the degree of pyramidalisation of the carbonyl carbon (measured by the out-of-plane deformation, Δ), and suggests that even the earliest stages of approach of reagent to the carbonyl group are governed by strict stereoelectronic requirements: even at the van der Waals distance the trigonal carbon is noticeably pyramidalized, while the angle subtended by the nucleophilic nitrogen at the C=O group (α) roughly equals the tetrahedral valence angle, and remains constant along the reaction coordinate (see Figure 1.2). An excellent survey of this system and other early reaction pathway studies is given by Dunitz (1979).

The results of this structure correlation study were immensely influential and proved to be a cornerstone for a large number of other theories, *e.g.* Baldwin's approach to vector analysis (1976), asymmetric induction (Houk *et al.*, 1986), and also diastereoselection in aldol condensation (Seebach & Golinski, 1981). Initial scepticism about the validity of the technique used to arrive at these results, and hence also in the results themselves, was allayed by supportive *ab initio* calculations (Bürgi, Lehn & Wipff, 1974), and the consistency of the model with other models of reactivity and

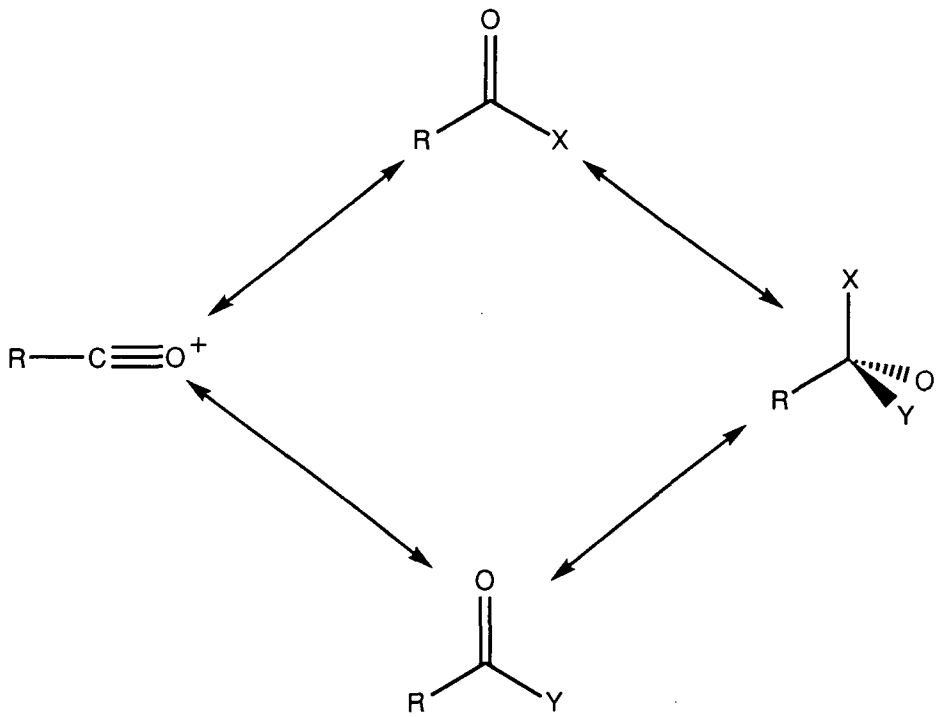


Figure 1.3

The complete addition / elimination scheme for the carbonyl group

ground-state electronic structure of the carbonyl group. More recently, the structure correlation analysis of the reactions at carbonyl sites has also been able to indicate a pathway for substitution via a tetrahedral acylium intermediate (Ferretti *et al.*, 1992), thus completing the cycle of carbonyl addition-elimination reactions (see Figure 1.3) in terms of the geometrical coordinates of the reaction pathways.

Reaction mechanisms involving coordinated metal complexes can also be described using structure correlation techniques, and a brief survey of these can be found in Section 4.1, together with a complete description of the symmetry deformation coordinate approach generally used in this field.

1.2.4 Studies of general non-bonded interactions

Using structure correlation it is possible to gather important information about any non-bonded interactions in a particular fragment system. Primarily this information manifests itself in the form of non-bonded distances, together with a measure of the directionality of the contact, with the significance of the results being apparent from the study of the intramolecular N...C=O interactions illustrated above.

Crystal structure analysis is unique in providing direct experimental observations of molecular aggregation in the solid state. Thus, crystallographic data is of immense value in studies of general non-bonded systems. For example, the X...X contacts among the fragments C-C≡N, $(C)_2C=O$, C=S, C=Se, C-F, C-Cl, C-Br and C-I have been analysed in terms of the orientation-dependent van der Waals surfaces (Nyburg, 1979; Nyburg & Faerman, 1985). For O and N it was found that these surfaces were essentially spherical, whereas for the occurrences of S, Cl, Br, I the surface adopts the shape of a rotation ellipsoid, with its short radius along the C-X bond vector. In a similar fashion, Murray-Rust & Motherwell (1979) found that non-bonded distances were direction dependent, with short distances more likely for linear arrangements of the C-I...I (or O) entity, than for perpendicular approaches.

The acquisition of systematic knowledge concerning hydrogen bonding is another important area in which structure correlation has proved to be a valuable

exploratory device, through an examination of the role, and geometry, of hydrogen bonding in crystal structures. The existence of a strong non-bonded interaction between electropositive H atoms and electronegative "acceptor" atoms was first proposed by Latimer & Rodebush (1920) on the basis of spectroscopic evidence. However, crystal structures provide a direct observation of the hydrogen bond geometry and a large body of knowledge concerning H-bonds was established in the 1950's and 1960's (*e.g.* Pimentel & McClellan, 1960; Hamilton & Ibers, 1968). The advent of crystallographic databases accelerated these studies since the 1970's and, at the same time the determination of hydrogen-atom positions in crystal structures via accurate neutron analysis, became a more accessible technique, leading to a vast collection of structural data on the subject. One particular field of interest concerned the application of hydrogen-bonded motifs in crystal structures to probe the existence of C-H...O and C-H...N hydrogen bonds. These hydrogen bonds, although weak, do have a significant effect on molecular conformation and packing. The question asked by the many groups involved in researching this topic, was whether the motif involving the C-H...O contact repeats itself in different structures, and whether the molecule in question could not have adopted a more favourable arrangement or conformation but for the C-H...O contact (*e.g.* Berkovitch-Yellin & Leiserowitz, 1980, 1982; Leiserowitz, 1976).

Through a systematic statistical analysis of the neutron diffraction data stored in the CSD, Taylor & Kennard (1982, 1984) found three important results: (i) that the C-H groups tend to form short intermolecular contacts to oxygen atoms rather than to carbon or hydrogen atoms, (ii) that C-H...O contacts for which the H...O distance is significantly less (by ≥ 0.30Å) than the sum of the van der Waals radii for hydrogen and oxygen atoms, are common, and (iii) the C-H vector in the majority of the short C-H...O (carbonyl) contacts lies within 30° of the plane containing the >C=O system. In a separate study, Gavezzotti (1991) demonstrated the presence of short-range O...H and O...C attractive interactions by analysing the deviations from average atomic contributions to calculated packing energies, for a sample of 590 oxyhydrocarbon crystal structures. It can be seen from both studies that C-H...O and C-H...N hydrogen-bonded interactions are significant, and following on from this, that the formation of

hydrogen-bonds leads to characteristic one-, two- and three-dimensional molecular patterns (*e.g.* Leiserowitz, 1976; Leiserowitz & Tuval, 1976; Jeffrey, 1982) which can be defined and utilised in predicting crystal packing or in designing crystals with predetermined structural properties: a topic recently reviewed by Desiraju (1991).

In addition, recent studies have focused on the role of hydrogen-bonds in biologically important systems, with a thorough review of recent advances being given by Jeffrey & Saenger (1991). Structure correlation studies on small fragments are shown to give reliable models for describing larger biological structures as a whole, and therefore illustrate once more the potential power of this technique.

1.2.5 Molecular modelling and rational drug design

Numerical and graphical computational techniques are now extensively used to speed the process of design of new bio-active molecules, pharmaceuticals and agrochemicals. Knowledge derived from crystallographic data has a key role to play in determining; (a) the likely conformations adopted by a novel putative target molecule, and (b) how that molecule might interact with a receptor site. This field is a new one but is rapidly advancing. One recent study (Lam *et al.*, 1994) details the rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors, while Johnson (1993) has used knowledge of the detailed structure of glycogen phosphorylase to design inhibitors that perturb the glycogen metabolism in favour of glycogen synthesis: a process which may be relevant in the treatment of diabetes.

1.3 Crystallographic databases

The largest of the crystallographic databases (see Figure 1.1) is the Cambridge Structural Database (CSD: Allen, Davies *et al.*, 1991; Allen & Kennard, 1993), which currently contains information on more than 120,000 organic and metallo-organic small molecules. The CSD was used exclusively for the structure correlations detailed in this Thesis, and so some of the basic features of this database will be described below.

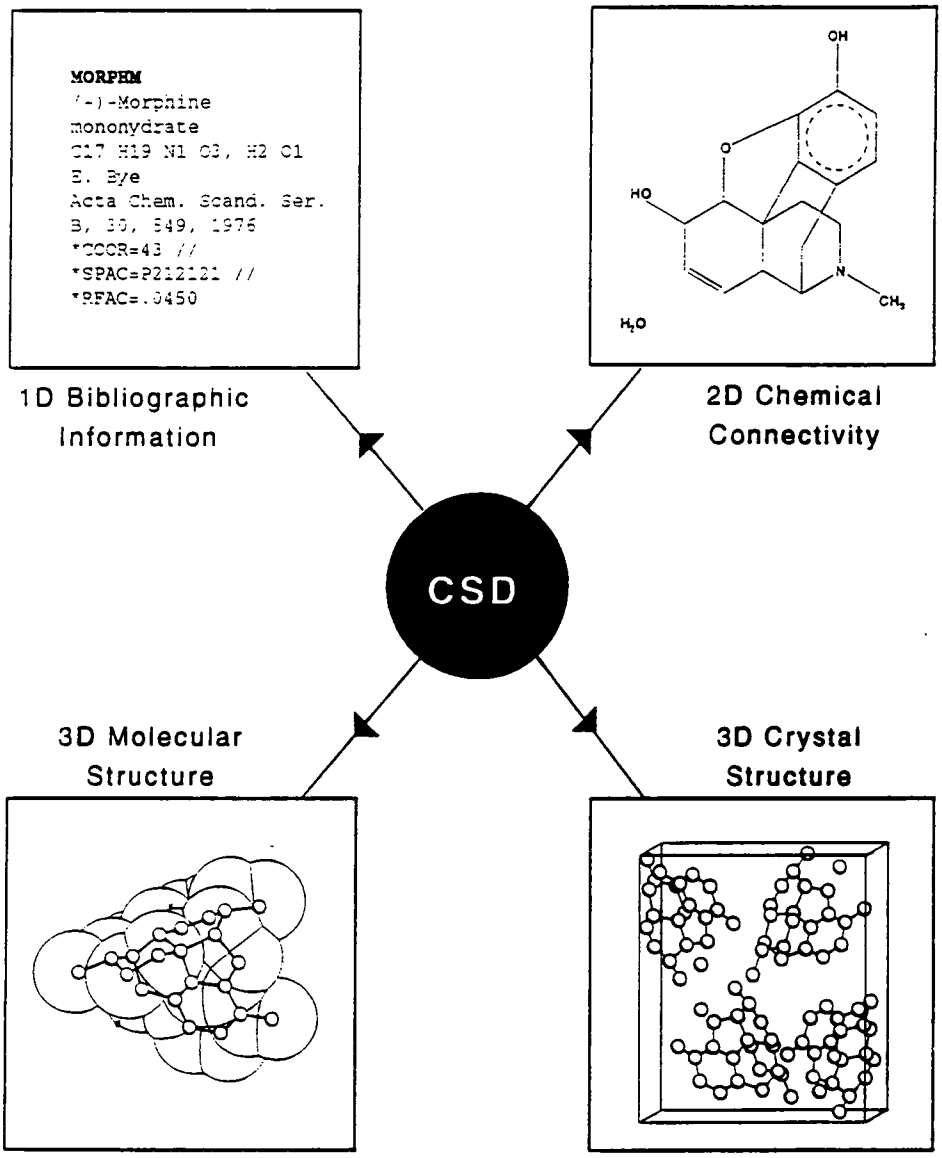


Figure 1.4

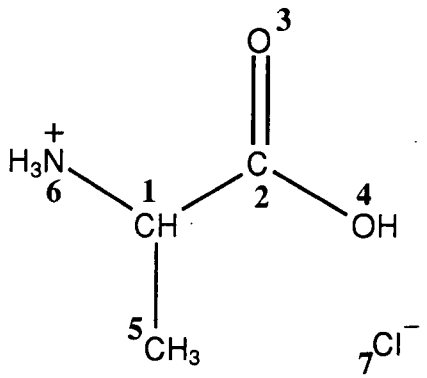
The Cambridge Structural Database System information breakdown (reproduced with kind permission of the CCDC)

1.3.1 Information content of the CSD

The information stored in the CSD can be most appropriately categorised in terms of the dimensionality of the data items. Figure 1.4 provides a pictorial representation of this breakdown of the information content.

1D Text and Numerical Information

This category includes all of the stored information concerning: chemical text strings (compound names, molecular formula, *etc.*), bibliographic information (authors names, literature reference), simple numerical items (chemical class, precision indicators, *etc.*), and text comment (describing special experimental findings, *etc.*).



Atom properties							
Atom number	1	2	3	4	5	6	7
Element number	C	C	O	O	C	N	Cl
No. connected non-H atoms	3	3	1	1	1	1	0
No. terminal H-atoms	1	0	0	1	3	3	0
Net charge	0	0	0	0	0	1	-1

Bond properties					
Atom 1 of bond	1	1	1	2	2
Atom 2 of bond	2	5	6	3	4
Bond type 1-2	1	1	1	2	1

Figure 1.5

Atomic connection table for a 2D chemical fragment

2D Chemical Structural Diagram

The 2D diagram of the structure is encoded in the form of a connection table: a simple graph (Wilson, 1985; Harary, 1972) that describes the atoms as the "nodes", and the bonds as the "edges". An example of the connection table for a particular chemical fragment is given in Figure 1.5. The properties of the atoms are recorded as: the atom sequence number, the element type, the numbers of connected non-hydrogen and terminal hydrogen atoms and the net charge. Finally, the 2D display coordinates, that allow the simple graph to be displayed "on-screen" as a 2D chemical diagram, are also stored. The bond properties are then stored as a pair of atom numbers and a statement of the formal bond type that connects those atoms. Bond cyclicity indicators are added automatically via a ring perception algorithm developed from that of Wippke & Dyott (1975). Bond types used are: 1 = single, 2 = double, 3 = triple, 4 = quadruple (metal-metal), 5 = aromatic, 6 = "catena link", 7 = delocalised double, 9 = π-bond

3D Structural Data

This final category comprises the 3D atomic coordinates, together with the crystallographic space group and symmetry operators for the molecule in question. A crystallographic connectivity is established using standard covalent radii and this is mapped, using graph-theoretical procedures, onto the chemical connection table, so as to relate the "chemical" atoms with their 3D coordinate attributes.

All of the information stored in this way has been extracted from the primary literature, with over 500 journals represented in the database. All of the input data is subject to rigorous computerised and visual checks, in order to preserve the consistency and accuracy of the recorded information. The information for each individual structure determination forms a single entry in the CSD master ASER file: a series of fixed length and directory-controlled variable-length records that are compacted as far as possible and structured for efficient search and retrieval operations.

1.3.2 The CSD software system: an overview

The current version of the CSD system contains menu-driven interactive graphics software which operates according to the flowchart shown in Figure 1.6. QUEST3D is the major system program for search and retrieval of fragments or molecules to be investigated, and incorporates a graphical display of hits together with limited data analysis capabilities. The files saved by QUEST3D can be used in the program VISTA, for detailed data analysis and visualisation functions, and the program PLUTO, which carries extensive display facilities for both the individual molecules and the extended crystal structures. Full details of the CSD System software is to be found in the CSD User Manuals (1992), but a summary will now be given here.

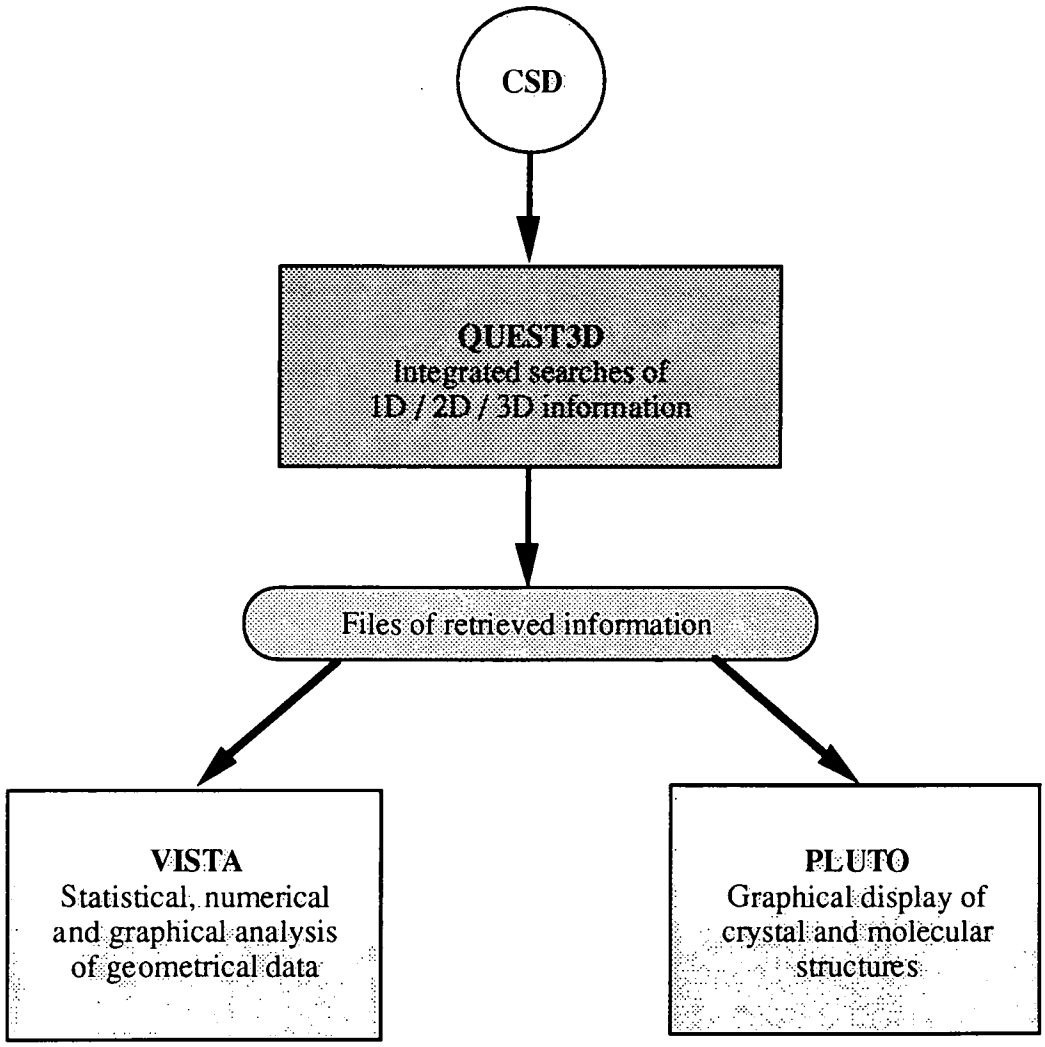


Figure 1.6

The interactive graphics system of CSD

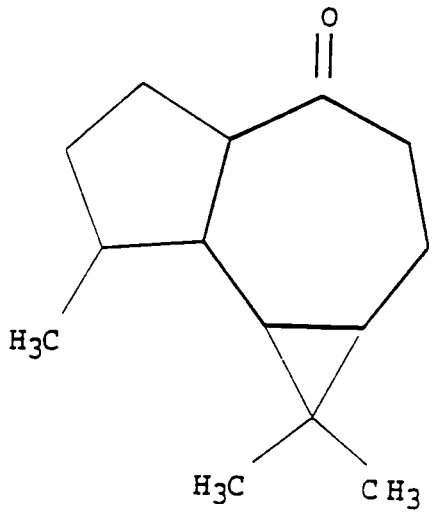
1.3.3 Search and retrieval using QUEST3D

QUEST3D incorporates an extensive menu system that allows the user to perform many database search operations. Search queries can be constructed from any of the 1D, 2D or 3D information fields described above, and individual queries can be linked together using Boolean logic, to form a more complicated, but more exact search query. Upon encountering a hit for a specific query, the program allows the user to display this hit as either text only, 2D frame, 3D frame, or a composite of both 2D and 3D representations. Interaction is then permitted with the search results, enabling the user to, for example, manipulate the 3D image by rotation/translation/zoom, or to display selected geometrical parameters on screen, *etc.* Summary tables of specified geometrical parameters can be generated for all of the saved hits from the search, as well as QUEST3D offering the possibility to display this information as simple histograms or scattergrams. Finally, the files saved by the program can then be used by the programs VISTA and PLUTO, or imported into other external software packages, including many molecular modelling programs.

Although the simple bibliographic search facilities are much used, it is the extensive facilities developed for 2D and 3D substructure searching that are most useful in trying to extract data for structure correlation and knowledge acquisition applications. For instance, in many of the structure correlation studies we are interested in the variations in the geometry of a substructure (chemical fragment) with variations in its chemical environment. Also, in drug design applications, we may need to know if a specific 3D pharmacophoric pattern has been observed in existing crystallographic results, with the implication that a successful search may indicate novel structure types as potential lead compounds in a rational drug design programme.

A 2D substructural query can be sketched in the BUILD menu of QUEST3D, and refined by specifying further chemical constraints (using options also available in the 2D-CONSTRAINTS sub-menu), as and where applicable, *e.g.* the variability of element types and bond types, the degrees of atomic substitution, or the cyclic or acyclic environments of atoms or bonds, *etc.* The query can also be represented as a graph, and for a hit to occur, this query graph must be located by the QUEST3D search

(a)



(b)

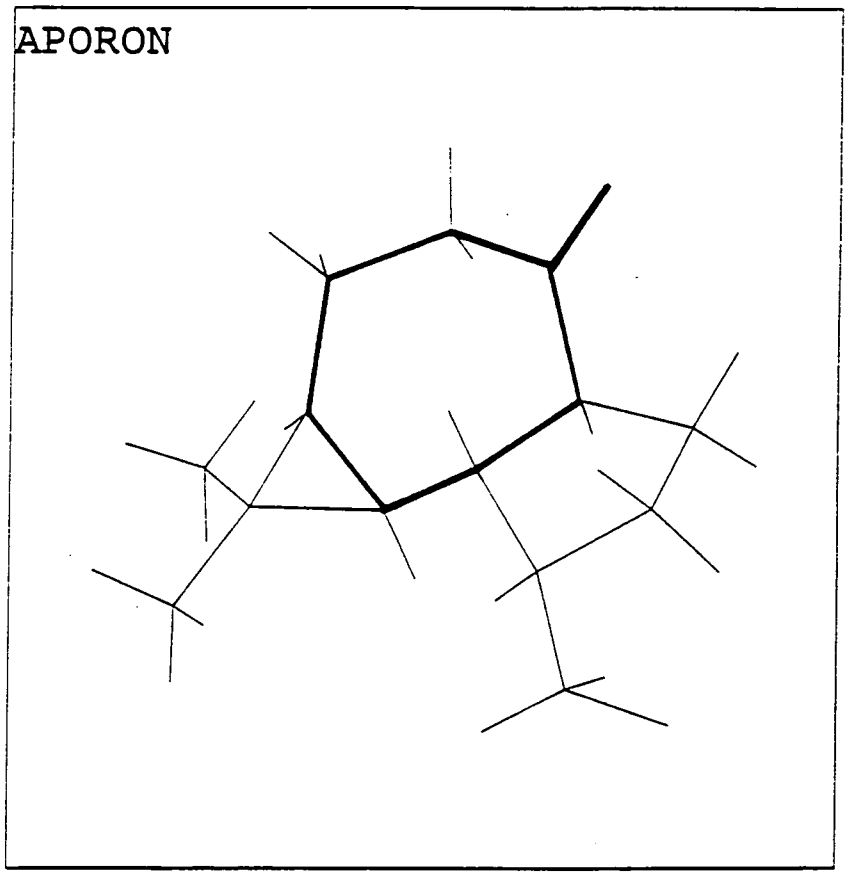


Figure 1.7

Search for cycloheptanone fragments gives: (a) 2D representation of the hit, and (b) the relevant 3D representation for the same hit, with the search query fragment highlighted in both cases.

algorithms within the larger structural graphs of the molecules held in the CSD, *i.e.* we must locate a subgraph isomorphism (Ash *et al.*, 1985). An example of this process is shown in Figure 1.7, where the query substructure is highlighted in both 2D and 3D representations.

These subgraph isomorphism search mechanisms prove to be very cpu-intensive since they involve atom-by-atom and bond-by-bond matching of the query to each of the >120,000 potential hits in the CSD. Hence, an additional search heuristic is usually employed: bit-screen mechanisms (Allen, Davies *et al.*, 1991). For each database entry a bit-screen record contains encoded information on the presence or absence of specific pieces of chemical information, for example the presence or absence of a C=O bond, or of a four-membered ring. There are over 600 bits encoded within each CSD entry. Each 2D substructural query defined by the user is first analysed in terms of these bit-screens, and each search then begins with a comparison of the query bit-screens against those stored for each database entry, with the full subgraph isomorphism algorithm only being invoked if all of the query screens are present in a particular entry.

In addition to the 2D searches that can be constructed in QUEST3D, 3D searching can also be invoked by placing geometrical constraints onto a sketched 2D chemical framework (3D-CONSTRAIN sub-menu). There is no requirement for the atoms to be bonded, and so queries can be constructed that will search the CSD for non-bonded inter-molecular contacts within a specified distance, for instance. The non-bonded interaction search can be restricted to either intramolecular or intermolecular situations, or alternatively encompass both possibilities. 3D searches operate on the geometric structure, since geometrical parameters represent a set of origin independent and rotationally invariant internal coordinates, unlike the set of crystallographic 3D atomic coordinates from which they are derived. The variety of geometrical search parameters that are allowed by QUEST3D is quite wide, and include constraints on: distances, angles and torsion angles involving both bonded and non-bonded atoms, distances from mean planes, special symmetry coordinates that describe ring conformations, spherical polar angles that describe hydrogen-bonded directionality, *etc.* The bit-screen mechanism has also been expanded to incorporate 3D geometrical

information and thus promote efficient 3D searching of the database, using techniques developed by Jakes & Willett (1986), and Poirette, Willett & Allen (1991 & 1993).

1.3.4 Data analysis using GSTAT or VISTA

On completion of the QUEST3D search a data matrix may be created containing N_p chosen parameters for each of the N_f instances of the search fragment hit. This data matrix, either as individual columns or as a whole, then forms the basis for a variety of statistical or numerical analyses, the results of which must be interpreted in structural chemical terms. The analyses may be univariate (on one parameter), bivariate (two parameters) or multivariate (more than two) depending on the nature of the dataset, and examples of all of these techniques will be used in this Thesis.

GSTAT was the original statistical package released with Version 4 of the CSD system, and was used for all of the ring conformation studies presented in Chapter 3. This was a non-interactive program, operating via an alphanumeric instruction set. The current Version 5 CSD system supports the VISTA interactive statistical analysis software, and the studies of changes in the metal coordination sphere described in Chapter 4 of this Thesis are done using this new software package.

The CSD will be seen to be an invaluable tool in the development of the type of structure correlation studies to be detailed in the later Chapters of this Thesis, and with the continued rapid growth in the number of crystal structures published each year, the scope for further acquisition of knowledge from the database will also grow, as the types of systems available for study become more diversified.

1.4 Example of a typical structure correlation experiment

There are a number of distinct stages to any structure correlation experiment: (a) selection and definition of the search fragment, (b) consideration of secondary search criteria, (c) definition of geometry for tabulation, and (d) establishment of a geometry

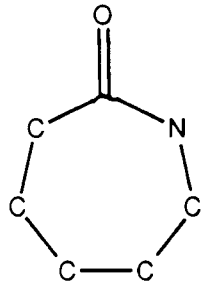


Figure 1.8

The 2D search query for a seven-membered ring lactam

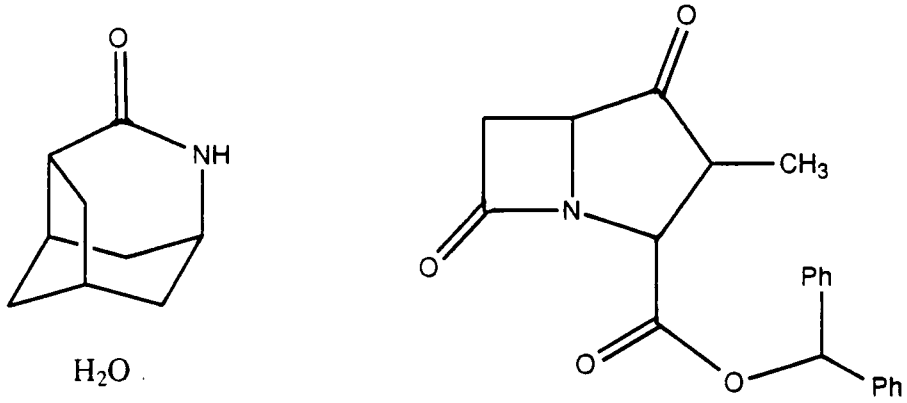


Figure 1.9

Examples of hits eliminated by the 2D-constraints (a) no cyclic routes [NOCR], and (b) no link [NOLN]

table (data matrix) and the application of descriptive statistics to that matrix. To illustrate these procedures let us consider a search of the CSD for the seven-membered ring lactam shown in Figure 1.8. This fragment can be drawn in the BUILD menu of QUEST3D, defined, and searched for in the CSD (Version 5.1: April 1994). With no constraints in place the total number of hits found is 113. Visual inspection of the hits as they are located is important as this may identify any types of fragment, or fragment environment, that are not required for the subsequent analysis. Examples of these are given in Figure 1.9, where in (a) the seven-membered ring has been located in a complex cyclic system, and (b) locates a seven-membered ring with a constraining bond across it. These types of fragments can be eliminated using the (a) no cyclic routes [NOCR], and (b) no links [NOLN] commands present in the 2D-CONSTRAIN sub-menu. On their own these constraints reduce the number of hits found in a complete

```
COMM These are comments in the QUEST initialisation file. This file may
COMM contain QUEST commands, such as terminal type, that are always entered.
COMM For more information enter "HELP INITIALISATION FILES" within QUEST.
COMM More detailed release notes now distributed with the software.
COMM Type "HELP RELEASE NOTES" for more information.
COMM
COMM Type "HELP LATEST NEWS" for a summary of new features. April 94 Highlights
COMM ************************************************************************
COMM * 1828 PDB Sequence Entries. VISTA for Interactive Graphical Statistics *
COMM ************************************************************************
COMM
COMM Set better PRINT style:
PRINT 10
T1 *CONN
NFRAG  1
AT1 C 2                          :XY  383  297
AT2 C 2                          :XY  228  372
AT3 C 2                          :XY  190  539
AT4 C 2                          :XY  297  674
AT5 N 2                          :XY  469  674
AT6 C 3                          :XY  576  539
AT7 C 2                          :XY  538  372
AT8 O 1                          :XY  810  561
BO 1 2 1
BO 2 3 1
BO 3 4 1
BO 4 5 1
BO 5 6 1
BO 6 7 1
BO 1 7 1
BO 6 8 2
NOCR
NOLN
GEOM
DEFINE T1 7 1 2 3
DEFINE T2 2 1 7 6
DEFINE T3 1 2 3 4
DEFINE T4 2 3 4 5
DEFINE T5 3 4 5 6
DEFINE T6 4 5 6 7
DEFINE T7 5 6 7 1
DEFINE *CREME Q2 PHI2 Q3 PHI3 6 7 1 2 3 4 5
C Overlap of crystal fragments permitted
C Search for ALL crystal fragments
NFRAG -99
SYMCHK ON
C Rejection of symmetry equivalent CRYSTAL fragments is ON
ENANT NOIN
END
COMMENT Turning ON "INSIST-ON-COORDS"
SCREEN  153
COMMENT Turning ON "INSIST-RFACTOR<=10%"
SCREEN  88
COMMENT Turning ON "INSIST-ERROR-FREE"
SCREEN  33
SCREEN  57
QUEST T1
```

Figure 1.10

The alphanumeric query file for the seven-membered ring lactam, with 2D-constraints, bit-screens, and 3D-constraints applied. This search finds 17 hits.

search from the original 113 down to 45 hits (with NOCR only), or 77 hits (with NOLN only), and when combined the search locates just 23 hits (NOLN and NOCR).

Other secondary search criteria can also be invoked to speed up searching (using the bit-screen mechanism described above) and further homogenise the dataset. Typical constraints on the search can be seen at the end of the query file shown in Figure 1.10. These include: ensuring that the crystallographic R-factor ≤ 10%, that the structure is error-free, and that 3D-coordinates are available in the CSD (essential for subsequent statistical analysis). With all of these constraints activated the number of hits located is reduced to 17.

The next step is to define the geometrical parameters, and any other parameters based on these, that need to be tabulated. This is done in the 3D-CONSTRAIN sub-menu of QUEST3D. In the query file the intra-annular torsion angles of the ring have been defined, as too have the Cremer-Pople (1975) puckering parameters for the seven

Table 1.1

Geometry table generated by the search of the CSD for the lactam defined by the query file in Figure 1.10

Nfrag	Refcode	T1	T2	T3	T4	T5	T6	T7
1	BAJZOB10	-61.233	83.226	57.152	-74.015	64.143	0.808	-69.085
2	BAJZOB10	-61.966	82.845	57.619	-78.169	71.531	-5.262	-64.526
3	BILJOV	58.791	-86.721	-53.084	80.034	-75.301	6.999	66.079
4	CAPLAC	-63.458	81.951	60.658	-77.000	67.783	-4.094	-63.238
5	CAPLAC01	-63.283	82.163	59.851	-76.132	65.863	-1.947	-64.453
6	CAPRES	-63.630	77.823	63.708	-79.729	67.376	-3.133	-60.633
7	CAPRES	64.736	-79.858	-63.136	76.697	-66.626	4.362	61.090
8	DIKVAU	-54.367	72.479	62.237	-76.276	63.168	0.313	-67.521
9	DIKVAU	-59.638	75.316	65.451	-78.695	61.228	3.145	-68.041
10	DOKMUL	59.100	-84.240	-57.607	80.227	-68.258	-2.238	72.079
11	DOKMUL	-60.644	85.505	58.656	-78.710	63.147	5.848	-72.554
12	FEFSUE	58.929	-81.540	-58.828	81.276	-67.156	-1.508	68.724
13	FULMUU	-64.733	82.172	61.592	-77.160	66.607	-0.817	-65.197
14	FULNAB	64.153	-77.799	-59.814	73.597	-68.083	7.069	57.274
15	FULNEF	-66.524	84.415	63.728	-75.706	63.319	-0.040	-66.858
16	FULNEF	-62.423	82.439	60.234	-77.064	67.569	-0.819	-67.561
17	JORDAV	-56.941	81.172	60.797	-83.484	60.410	13.275	-76.898
18	JUNVUJ	58.975	-79.864	-55.882	73.979	-66.398	-0.364	67.271
19	KOTPAK	-59.480	81.590	59.389	-79.231	67.768	-1.500	-66.665
20	VIMXIY	62.330	-84.242	-54.961	76.562	-72.490	4.587	64.818
21	VUKNOE	71.043	-88.035	-66.624	80.390	-71.442	5.059	67.476
22	WAFYEH	60.095	-80.896	-60.497	81.112	-67.993	0.795	66.656
Nent		17	17	17	17	17	17	17
Nobs		22	22	22	22	22	22	22
Mean		-10.917	14.086	11.847	-13.977	10.280	1.388	-12.807
SDSample		62.213	82.396	60.366	78.550	67.845	4.435	67.001
SDMean		13.264	17.567	12.870	16.747	14.465	0.945	14.285
Minimum		-66.524	-88.035	-66.624	-83.484	-75.301	-5.262	-76.898
Maximum		71.043	85.505	65.451	81.276	71.531	13.275	72.079

Nfrag	Refcode	Q2	PHI2	Q3	PHI3
1	BAJZOB10	0.451	57.018	0.628	74.317
2	BAJZOB10	0.466	56.292	0.641	76.874
3	BILJOV	0.520	237.999	0.632	257.217
4	CAPLAC	0.420	55.260	0.644	76.520
5	CAPLAC01	0.426	55.335	0.641	75.464
6	CAPRES	0.409	48.954	0.651	77.682
7	CAPRES	0.390	234.740	0.654	257.217
8	DIKVAU	0.438	44.914	0.610	77.981
9	DIKVAU	0.417	45.981	0.646	76.131
10	DOKMUL	0.499	231.719	0.645	253.632
11	DOKMUL	0.465	51.687	0.644	72.206
12	FEFSUE	0.469	230.606	0.637	255.608
13	FULMUU	0.429	51.531	0.660	74.711
14	FULNAB	0.396	236.204	0.630	256.806
15	FULNEF	0.416	54.239	0.670	74.032
16	FULNEF	0.481	51.217	0.655	74.826
17	JORDAV	0.501	40.556	0.644	71.818
18	JUNVUJ	0.480	230.344	0.616	254.037
19	KOTPAK	0.469	49.664	0.638	74.986
20	VIMXIY	0.499	238.483	0.639	254.711
21	VUKNOE	0.451	237.662	0.738	254.988
22	WAFYEH	0.459	231.240	0.646	256.243
Nent		17	17	17	17
Nobs		22	22	22	22
Mean		0.452	125.984	0.646	149.000
SDSample		0.036	92.372	0.025	90.804
SDMean		0.008	19.694	0.005	19.359
Minimum		0.390	40.556	0.610	71.818
Maximum		0.520	238.483	0.738	257.217

atoms of the ring (a complete description of these puckering coordinates is given in Chapter 2). On completion of the search a list of parameters is generated (Table 1.1) for each hit. The sequence of the seven torsion angles around the ring describe the conformation adopted, as will be shown in Chapter 3. For the purposes of this experiment these are the only geometrical parameters that need to be specified, however, it has already been shown that QUEST3D can carry out searches for any kind of geometrical information, including non-bonded interactions, and this facility underpins the structure correlation studies performed for the coordinated metals in Chapter 4. The number of entries often differs from the number of observations (here 17 and 22 respectively) because the search algorithm will locate all independent occurrences of the query fragment in any one particular database entry.

Once the search is complete the data matrix can be stored in a file and read into the program VISTA (or GSTAT) and the parameters analysed with the statistical and numerical tools available there. Similar search procedures were adopted for all of the structure correlation studies to be described in this Thesis. The basic geometry table is

always accompanied by descriptive statistics, *e.g.* the mean, the sample standard deviation, *etc*, for each variable, as illustrated at the end of Table 1.1. Other, more complex statistical and numerical techniques (see Allen & Taylor (1994) for a recent overview of these techniques) used in this work will be described in detail at the relevant points in Chapters 2, 3 and 4.

1.5 Subject matter of this Thesis

The CSD will be used for structure correlation experiments in the areas of: (i) conformational analyses of medium rings, and (ii) coordination geometries of transition metal compounds. In Chapters 2 and 3 conformational studies are performed on a wide variety of seven- and eight-membered carbocyclic and heterocyclic ring systems. It is the aim of these Chapters to elucidate the major conformational features in these datasets, using a variety of group-theoretical descriptions and multivariate statistical techniques, together with appropriate visualisations of the relevant (n-3)-dimensional hypersurface. The relationship between chemical environment and the adopted conformation will be drawn, and comparisons of these results will be made with energy values for the fragment obtained from the literature or from original calculations.

Chapter 4 of this Thesis describes the variety of configurations adopted by 3-coordinated transition metal complexes, using both simple geometrical models and the group-theoretically based symmetry deformation coordinates . Chemical environmental effects are investigated, and intermolecular non-bonded interactions with the 3-coordinated metal centre are analysed in terms of their perturbing effect on the overall configuration. The relationships between the ideal configurations of 2-, 3- and 4-coordinated transition metal compounds are also analysed.

At the heart of all structure correlation studies, of course, lies the need for accurate geometrical data. This information is provided by crystal structure analyses using mainly X-ray diffraction techniques. In Chapter 5 of this Thesis a brief résumé of the experimental and computational procedures and theoretical background of structure solution by X-ray diffraction is given, together with descriptions of crystal structures

determined as part of this research program. Special attention will be drawn in Chapter 5 to original determinations that are related to the structure correlation results described in earlier Chapters.

1.6 References

Allen, F.H., Bergerhoff, G. & Sievers, R. (Eds.), *Crystallographic Databases*, **1987**, International Union of Crystallography, Chester, U.K.

Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C.F., Mitchell, G.F., Smith, M.J. & Watson, D.G., *J. Chem. Inf. Comput. Sci.*, **1991**, 31, 187-204.

Allen, F.H., Doyle, M.J. & Taylor, R., *Acta Cryst.*, **1991a**, B47, 29-40.

Allen, F.H., Doyle, M.J. & Taylor, R., *Acta Cryst.*, **1991b**, B47, 41-49.

Allen, F.H., Doyle, M.J. & Taylor, R., *Acta Cryst.*, **1991c**, B47, 50-61.

Allen, F.H., Howard, J.A.K. & Pitchford, N.A., *Acta Cryst.*, **1993**, B49, 910-928.

Allen, F.H. & Kennard, O., *Chem. Des. Auto. News*, **1993**, 8, 1.

Allen, F.H., Kennard, O., Watson, D.G., Brammer, L., Orpen, G. & Taylor, R., *J. Chem. Soc.*, **1987**, S1.

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

Ashida, T., Tsunogae, Y., Tanaka, I & Yamane, T., *Acta Cryst.*, **1987**, B43, 212-218.

Auf der Heyde, T.P.E., Bürgi, H.B., *Inorg. Chem.*, **1989**, 28, 3960-3969.

Baldwin, J.E., *J. Chem. Soc. Chem. Comm.*, **1976**, 738.

Bartenev, V.N., Kameneva, N.G. & Lipanov, A.A., *Acta Chem.*, **1987**, B43, 275-280.

Berkovitch-Yellin, Z., & Lieserowitz, L., *J. Am. Chem. Soc.*, **1980**, 102, 7677-7690.

Berkovitch-Yellin, Z., & Lieserowitz, L., *J. Am. Chem. Soc.*, **1982**, 104, 4052-4064.

Bürgi, H.B. & Dunitz, J.D., *Acc. Chem. Res.*, **1983**, 16, 153-161.

Bürgi, H.B. & Dunitz, J.D., *Acta Cryst.*, **1984**, B44, 445-448.

Bürgi, H.B. & Dunitz, J.D., *Structure Correlation*, **1994**, VCH, Weinheim, Germany.

Bürgi, H.B., Dunitz, J.D. & Shefter, E., *Cryst. Struct. Comm.*, **1973a**, 3, 667.

Bürgi, H.B., Dunitz, J.D. & Shefter, E., *J. Am. Chem. Soc.*, **1973b**, 95, 5065-5067.

Bürgi, H.B., Lehn, J.M. & Wipff, G., *J. Am. Chem. Soc.*, **1974**, 96, 1956-1957

Cremer, D. & Pople, J.A., *J. Am. Chem. Soc.*, **1975**, 97, 1354-1358.

Cambridge Structural Database, **1992**. *CSD User's Manual.* Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England.

Cremer, D. & Pople, J.A., *J. Am. Chem. Soc.*, **1975**, 97, 1354-1358.

Desiraju, G., *Acc. Chem. Res.*, **1991**, 24, 290-296.

Dunitz, J.D., *X-ray Analysis and the Structure of Organic Molecules*, **1979**, Cornell University Press, Ithaca, New York.

Ferretti, V., Dubler-Steudle, K.C. & Bürgi, H.B., in: *Accurate Molecular Structures*, **1992**, eds. Domenicano, A. & Hargittai, I., OUP, London.

Gavezzotti, A., *J. Am. Chem. Soc.*, **1991**, 95, 8948

Hamilton, W.C. & Ibers, J.A., *Hydrogen-Bonding in Solids*, **1968**, Benjamin, New York.

Harary, F., *Graph Theory*, **1972**, Addison-Wesley, London.

Hoffmann, R., Foreword to Vilkov, L.V., Mastryukov, V.S. & Sadova, N.I., *Determination of the Geometrical Structure of Free Molecules*, **1983**, MIR Publishers, Moscow.

Houk, K.N., Paddon-Row, M.N., Rondan, N.G., Wu, Y.-D., Brown, F.K., Spellmayer, D.C., Metz, J.T., Li, Y. & Loncharich, R.J., *Science*, **1986**, 231, 1108.

Jakes, S.E. & Willett, P., *J. Mol. Graphics*, **1986**, 4, 12-20.

Jeffrey, G.A., in: *Structure and Biological Activity*, **1982**, Griffin, J.F. & Duax, W.L. (eds.), Elsevier Science, New York.

Jeffrey, G.A. & Saenger, W., *Hydrogen Bonding in Biological Structures*, **1991**, Springer, Berlin.

Johnson, L.N., *Structure Based Drug Design*, **1993**, Plenary lecture, IUCrXVI Congress and General Assembly, Beijing.

Johnston, H.S., *Adv. Chem. Phys.*, **1960**, 3, 131-170.

Johnston, H.S. & Parr, C., *J. Am. Chem. Soc.*, **1963**, 85, 2544.

Lam, P.Y.S., Jadhav, P.K., Eyermann, C.J., Hodge, C.N., Ru, Y., Bachelor, L.T., Meek, J.L., Otto, M.J., Rayner, M.M., Wong, Y.N., Chang, C-H, Weber, P.C., Jackson, D.A., Sharpe, T.R. & Erickson-Viitanen, S., *Science*, **1994**, 263, 380-384.

Latimer, W.H. & Rodebush, W.H., *J. Am. Chem. Soc.*, **1920**, 42, 1419-1433.

Leiserowitz, L., *Acta Cryst.*, **1976**, B32, 775-802.

Leiserowitz, L. & Tuval, M., *Acta Cryst.*, **1978**, B34, 1230-1247.

Murray-Rust, P. & Bland, R., *Acta Cryst.*, **1978**, B34, 2527-2533.

Murray-Rust, P., Bürgi, H.B.& Dunitz, J.D., *J. Am. Chem. Soc.*, **1975**, 96, 921-922

Murray-Rust, P. & Motherwell, W.D.S., *Acta Cryst.*, **1978a**, B34, 2518-2526.

Murray-Rust, P. & Motherwell, W.D.S., *Acta Cryst.*, **1978b**, B34, 2534-2546.

Murray-Rust, P. & Motherwell, W.D.S., *J. Am. Chem. Soc.*, **1979**, 101, 4374-4376.

Norskov-Lauritsen, L. & Bürgi, H.B., *J. Comp. Chem.*, **1985**, 6, 216-228.

Nyburg. S.C., *Acta Cryst.*, **1979**, A35, 641-645.

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

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

Pauling, L., *The Nature of the Chemical Bond*, **1939**, 1st edition, Cornell University Press, Ithaca, New York.

Pauling, L., *J. Am. Chem. Soc.*, **1947**, 69, 542-553.

Pickett, H.M. & Strauss, H.L., *J. Chem. Phys.*, **1971**, 55, 324-334.

Pimentel, G.C. & McClellan, A.L., *The Hydrogen Bond*, **1960**, Freeman, San Francisco

Poirette, A.R., Willett, P. & Allen, F.H., *J. Mol. Graphics*, **1991**, 9, 203-217.

Poirette, A.R., Willett, P. & Allen, F.H., *J. Mol. Graphics*, **1993**, 11, 2-14.

Seebach, D. & Golinski, J., *Helv. Chim. Acta*, **1981**, 64, 1413-1423.

Taylor, R. & Kennard, O., *J. Am. Chem. Soc.*, **1982**, 104, 5063-5070.

Taylor, R. & Kennard, O., *Acc. Chem. Res.*, **1984**, 17, 320-326.

Wilson, R.J., *Introduction to Graph Theory*, **1985**, Longman, London.

Wippke, W.T. & Dyott, T.M., *J. Chem. Inf. Comp. Sci.*, **1975**, 15, 140.

Chapter Two

Conformational Analysis using Crystallographic Data

2.1 Introduction

This Chapter details the statistical and mathematical tools that are employed in the subsequent systematic analyses found in Chapters 3 and 4. Throughout this Thesis, correlations will be established between the descriptions that are derived from group theoretical considerations and those found by mathematical methods. The techniques described below are invaluable in applying structure correlation principles to symmetry-expanded, multi-dimensional systems, and provide an accessible tool for viewing the conformational hyperspace in more understandable, reduced dimensions.

The conformations of the different ring systems will be described in depth, and the nomenclature established which will be used explicitly in the following Chapters. Each conformation has discerning characteristics that set it aside from the other conformations of the particular system, and which place it in certain regions of the complete hyper-dimensional potential energy surface. The calculated energies of the different conformational variants are important quantities, and are used to describe the observations encountered during the experimental structure correlation studies. Hence, the calculation of ring energies for those molecules which have not already been determined, is a vital component in guiding the complete analysis.

2.2 Conformations of medium rings

2.2.1 Overview of geometrical descriptors

The conformations of the n-membered rings ($n \geq 4$) are usually analysed in terms of their n intra-annular torsion angles. These parameters are well understood and structurally informative, and have been used extensively as the characteristic descriptors of conformation, since the landmark studies of Hendrickson (1967). One of

the major problems with using torsion angles as conformational identifiers, however, is that we require all n values for a complete description: for the cycloheptanes therefore, classification or mappings of conformational interconversion requires a knowledge of relationships between these parameters in seven-dimensional space, which is not an easy concept to grasp.

Alternative descriptions have therefore tried to reduce the dimensionality of the overall problem. Recent studies using multivariate statistical techniques have proved very successful in reducing the dimensionality (see Section 2.3), but the measure of their success is how well they correlate against the results derived from group theoretical considerations of the degree of pucker present in a certain ring conformation. These methods reduce the dimensionality of the problem down to manageable levels, and can be used to give a good account of the macroscopic properties of the system as a whole.

The two descriptions used in this Thesis centre on the work of Pickett & Strauss (1970), and of Cremer & Pople (1975). Both use similar ideas concerning descriptions of ring conformation in terms of the out of plane displacements of each ring atom, but use slightly different basis sets as the planar representation from which these displacements are measured. Both of these descriptions will be summarized below and the relationship between them established. Their relevance to the problems of the 7- and 8-membered ring systems which are studied later in this thesis, will also be detailed, and their usefulness explored.

The method proposed by Pickett & Strauss (1970) describes a ring conformation in terms of the atomic displacements (z_j) from a planar basis conformation of D_{nh} symmetry. The z_j are then transformed according to the irreducible representations of the D_{nh} point group that do not involve rotations or translations, yielding (n-3) symmetry-adapted normal coordinates. These coordinates can then be used to map the potential-energy hypersurface and to locate possible pseudorotational pathways in the system.

Cremer & Pople (1975) find their (n-3) puckering coordinates by measuring atomic displacements from a unique mean plane that is different for each ring

conformation encountered. In this respect their derivation is different from that of Pickett & Strauss. However, Evans & Boeyens (1985) have recently shown that the two (n-3)-dimensional descriptions of conformation are in fact self-consistent. This is because the conditions stipulated by Cremer & Pople in their definition of the unique mean plane arise naturally from group-theoretical considerations.

Other descriptions of ring pucker can also be found by using the torsion angles themselves as the out-of-plane deformation coordinates in a Fourier transformation (Altona, Geise & Romers, 1968; Altona & Sundaralingam, 1972). These methods were not used in the studies reported here.

2.2.2 Cycloheptane conformations and puckering coordinates

Four major, archetypal, symmetrical conformations exist for the cycloheptane fragment. These are the C_s symmetry chair (C) and boat (B) conformations, together with the C_2 symmetry twist-chair (TC) and twist-boat (TB), as illustrated in Figure 2.1. The mathematical models described below have to be able to map each of these variants independently, as well as provide a description of any other unsymmetrical

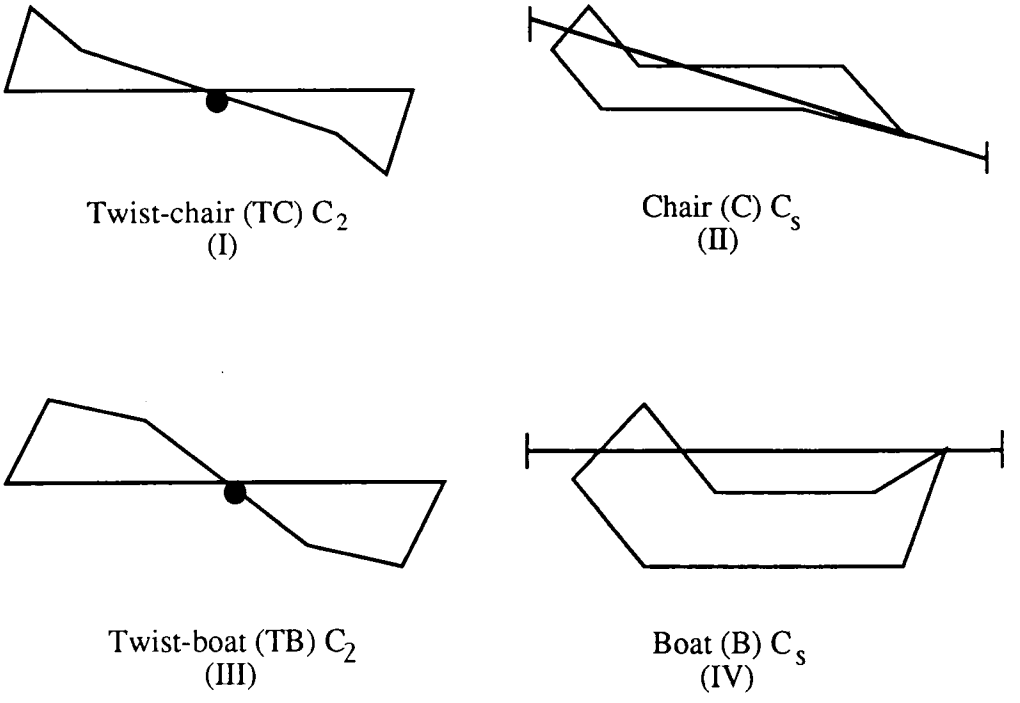


Figure 2.1

The four symmetrical conformations of the cycloheptane ring.

conformations observed.

The model outlined by Pickett & Strauss (1970) specifies the conformation of the ring in terms of the out-of-plane coordinates (z_j) of all of the ring atoms. These displacements occur in sets determined by the symmetry of the ring system, with the number of independent sets being equal to (n-3), where n is the number of ring atoms. The four independent sets for cycloheptane can be written as:

$$z_j = \xi_2 \cos(2\pi 2j/7) + \eta_2 \sin(2\pi 2j/7) + \xi_3 \cos(2\pi 3j/7) + \eta_3 \sin(2\pi 3j/7) \qquad \textbf{[1]}$$

$$j = 1, 2, ...7.$$

The j-values correspond to the label of the ring carbon atom, and the ξ_2, η_2, ξ_3, and η_3 are coordinates which correspond to the sets of z_j that transform the planar D_{7h} entity into the conformations described as being boat, chair, twist-boat and twist-chair, and as characterised in Figure 2.1 above.

There are two alternative forms to this equation as proposed by Bocian, Pickett, Rounds & Strauss (1975). The first of these details the relationships of two amplitude-phase pairs: ρ_2, ϕ_2 transforming as the E_2" representation, and ρ_3, ϕ_3 transforming as the E_3" representation, where the representations are given by the formula **[2]**:

$$z_j = \rho_2 \cos(2\pi 2j/7 + \phi_2) + \rho_3 \cos(2\pi 3j/7 + \phi_3) \qquad \textbf{[2]}$$

where

$$\rho_2 = \sqrt{\xi_2^2 + \eta_2^2} \qquad \tan\phi_2 = \eta_2/\xi_2$$

$$\rho_3 = \sqrt{\xi_3^2 + \eta_3^2} \qquad \tan\phi_3 = \eta_3/\xi_3$$

The second description is in terms of a set of coordinates ρ, θ, ϕ_2, ϕ_3 (hereinafter BPRS coordinates) which are derived from the equation **[3]**:

$$z_j = \rho\left[\cos\theta\cos(2\pi 2j/7 + \phi_2) + \sin\theta\cos(2\pi 3j/7 + \phi_3)\right] \qquad \textbf{[3]}$$

with

$$\rho = \sqrt{\rho_2^2 + \rho_3^2} \qquad \tan\theta = \rho_3/\rho_2 \qquad 0 \le \theta \le \pi/2$$

This last set is the most useful, with the energy of the conformation in question

being able to be written in terms of these variables, as the seven-fold nature of the symmetry group is satisfied. In equation **[3]**, ρ gives a measure of the total distortion of the observed conformation from planarity, *i.e.* the amount of ring pucker, while θ shows the relative contributions of the boat-like or chair-like conformations to the observed geometry. Thus ρ and θ are constant for any of the possible symmetry related conformations of an individual ring. The two ϕ's are used to specify which of the possible identical conformations is being observed, and also give some measure of the degree of twist present in the fragment conformation.

In the Cremer & Pople (1975) description, the geometry of the puckering relative to the unique mean plane is also described in terms of amplitude and phase coordinates, which are generalisations of those derived by Kilpatrick *et al.* (1947) for the cyclopentane rings. For the cycloheptane fragment the two relevant phase-amplitude pairs are q_2, ϕ_2, and q_3, ϕ_3 (hereinafter known as the CP coordinates). The phases ϕ_2, ϕ_3 are identical to the BPRS values and the q_2, q_3 are related to ρ, θ by the following relationships (Evans & Boeyens, 1989; Boessenkool & Boeyens, 1980):

$$q_2 = \rho\cos\theta \qquad \text{and} \qquad q_3 = \rho\sin\theta \qquad \textbf{[4]}$$

The BPRS coordinates can also be related to the CP coordinates by the relationships (Bocian *et al.*, 1975):

$$\rho = \sqrt{q_2^2 + q_3^2} \qquad \tan\theta = {}^{q_3}\!/_{q_2} \qquad 0 \le \theta \le {}^{\pi}\!/_{2} \qquad \textbf{[5]}$$

2.2.3 Conformational space and conformational variants of cycloheptane

The four major conformational variants (I-IV in Figure 2.1) can be described by the sequence of their intra-annular torsion angles (Hendrickson, 1967), as well as in terms of the BPRS and CP coordinates derived above. Typical values are collected in Table 2.1.

The conformers I-IV have C_S (chair, boat), or C_2 (twist-chair, twist-boat) symmetry. For classification purposes, the symmetry elements of these are chosen to run through the ring carbon atom C1, and the mid-point of the C4-C5 bond. Thus, for

Table 2.1

Sequence of intra-annular torsion angles that describe the symmetrical conformations, together with the relevant BPRS and CP parameters, and the calculated energy differences between each conformation.

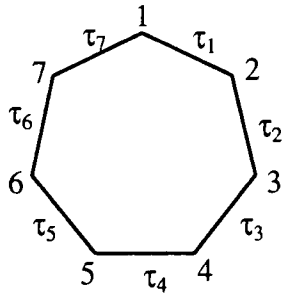
Conf.	TC	C	TB	B
τ_1	-39.1	63.8	45.4	-57.5
τ_2	88.1	-83.5	-64.4	-30.9
τ_3	-72.3	66.1	-17.9	69.9
τ_4	54.3	0.0	74.6	0.0
τ_5	-72.3	-66.1	-17.9	-69.9
τ_6	88.1	83.5	-64.4	30.9
τ_7	-39.1	-63.8	45.4	57.5
θ (°)	50	54	0	0
ρ	0.80	0.74	1.15	1.15
q_2	0.52	0.43	1.15	1.15
ϕ_2	90.0	0.0	270.0	0.0
q_3	0.61	0.60	0.0	0.0
ϕ_3	90.0	180.0	----	----
ΔE	0.0	5.86	10.04	11.30

each of the conformational possibilities (I-IV) there are seven conformational equivalents that can be generated by the cyclic permutation of the atom labels. This number of equivalents must be increased to 14 when the enantiomorphic equivalents are also considered. When a general asymmetric cycloheptane fragment is considered therefore, there are 14 possible cyclic permutations of the atom labels (Table 2.2), which lead to 28 possible equivalent conformations, each of which has a discrete sequence of torsion angles, together with a discrete set of four-dimensional BPRS and CP coordinates, which define the conformational hypersurface.

In order to get a clear idea of how the symmetry equivalents are related to each other on the conformational hypersurface, it is imperative to have a clear understanding of the four-dimensional nature of this surface. The best representation of a four-dimensional system is given by the geometrical properties of a torus, as proposed by Bocian *et al.* (1975) and Bocian & Strauss (1977a,b,c). A pictorial representation of

Table 2.2

The 14 cyclic permutations of atom labels around the cycloheptane ring



Forward								Reverse						
1	2	3	4	5	6	7		7	6	5	4	3	2	1
2	3	4	5	6	7	1		6	5	4	3	2	1	7
3	4	5	6	7	1	2		5	4	3	2	1	7	6
4	5	6	7	1	2	3		4	3	2	1	7	6	5
5	6	7	1	2	3	4		3	2	1	7	6	5	4
6	7	1	2	3	4	5		2	1	7	6	5	4	3
7	1	2	3	4	5	6		1	7	6	5	4	3	2

how a torus can be used to map the conformational equivalents of a given cycloheptane ring, is shown in Figure 2.2.

Using the BPRS and CP coordinate systems, the major and minor axes of the torus can be defined as:

$$c\rho\sin\theta = cq_3 \quad \text{(major)} \qquad \rho\cos\theta = q_2 \quad \text{(minor)} \qquad \textbf{[6]}$$

and the ϕ_2 and ϕ_3 values will position the conformational equivalents on the surface of the torus. An arbitrary constant c is employed (Bocian *et al.*, 1975) which aims to avoid the re-entrant tori seen when tan θ > c. The ϕ_3-dependent torus is the best representation for visualisation of the pseudorotation pathways. However, it is also possible to create the ϕ_2-dependent torus using exactly the same equations **[6]**, but switching the axial labels so that the major radius is given by $c\rho\cos\theta = cq_2$, and the minor radius by $\rho\sin\theta = q_3$. This torus is also constrained not to be re-entrant.

In such a way, the cycloheptane rings can be depicted as a family of concentric tori, in exactly the same way that cyclohexanes can be represented as a family of concentric spheres (Allen & Taylor, 1991): the properties of the conformations

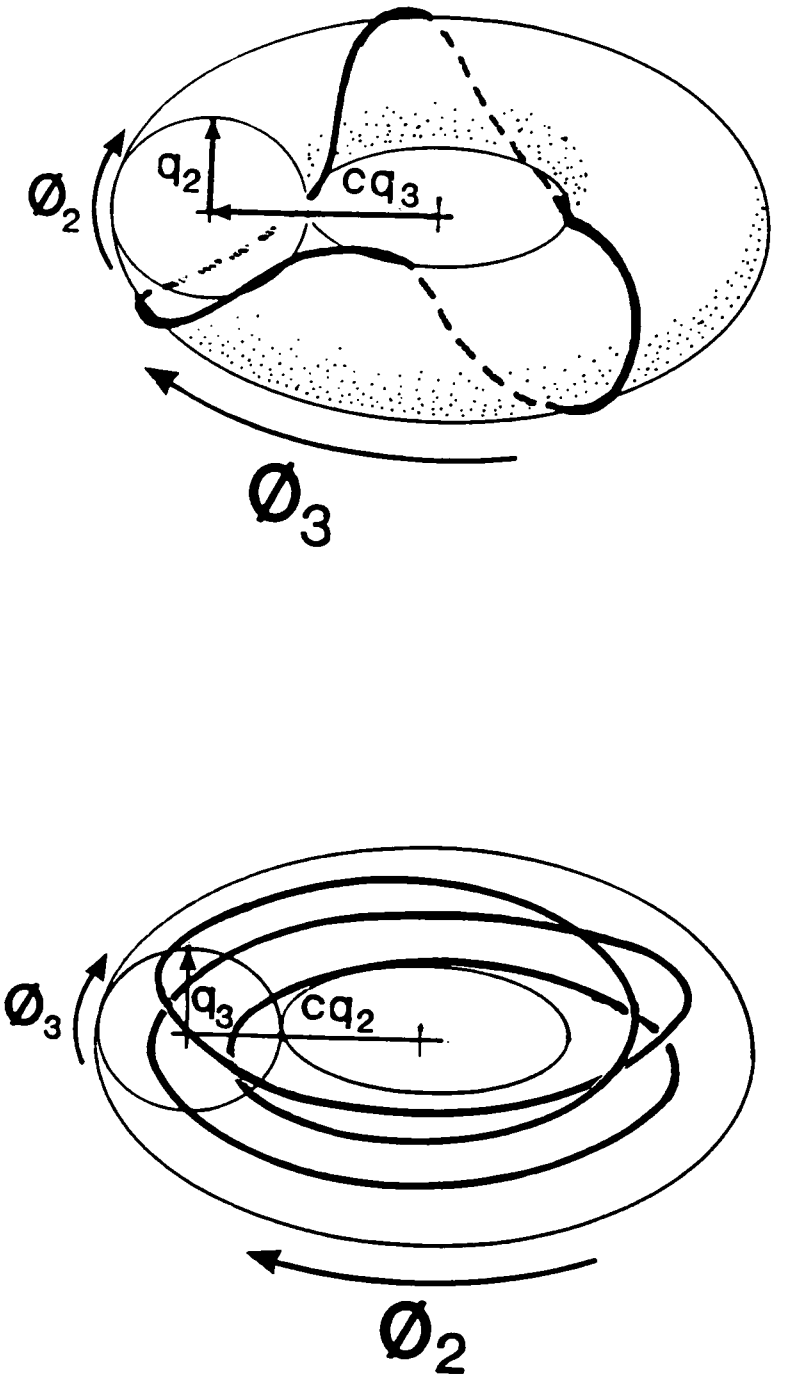


Figure 2.2

Toroidal representations of the 4-D hyperspace for cycloheptane

determining the values of the four-dimensional coordinate representations (CP or BPRS), thereby determining the properties of the tori, and of the helical paths that weave across their surfaces. The tori can thus be used to break the conformational hyperspace down into distinct areas, each set being defined by its ρ and θ values.

The chair-like conformations (I and II) both have $\rho = 0.75$, and $\theta = 60°$, and so they can be described together on the ϕ_3-dependent torus in terms of their different ϕ_2 and ϕ_3 values alone. There is a clearly defined low energy pathway that follows the chair - twist-chair interconversion process that links the symmetry-equivalent conformations. The pathway follows the line of minimum geometrical movement between conformations. Bocian *et al.* (1975), and Bocian & Strauss (1977), show that the chair - twist-chair pseudorotation itinerary follows a helical path on the surface of the torus such that if the "origin" chair (C) and twist-chair (TC) conformations are defined as being at the positions $\varphi_2^o(C) = 0$, $\varphi_3^o(C) = \pi$ and $\varphi_2^o(TC) = \pi/2$, $\varphi_3^o(TC) = \pi/2$, then the permutationally isomeric conformations (k) will occur at positions determined by the equations:

$$\varphi_2^k(C,TC) = \varphi_2^o(C,TC) + 3\pi k/7 \qquad \varphi_3^k(C,TC) = \varphi_3^o(C,TC) + \pi k/7$$

$$\text{for k = 0 to 13} \qquad [7]$$

This then places the TC conformers exactly midway between the C conformers at:

$$\varphi_2^k(TC) = \varphi_2^k(C) + 3\pi/7 \qquad \varphi_3^k(TC) = \varphi_3^k(C) + \pi/14 \qquad [8]$$

However, from Table 2.1 it should be noted that this C-TC pathway is not an isoenergetic one, since the E(TC) is found to be 5.86 kJmol^{-1} below that of the E(C) (Hendrickson, 1967).

The boat family of conformers (III and IV) are found on the torus with $\rho = 1.10$, but θ very close to zero. The resulting torus therefore has a major radius close to zero, and the pseudorotation pathway is reduced to being essentially a circle of radius ρ. The pathway is ϕ_2-dependent only, since the ϕ_3 variable becomes indeterminate as the value of θ approaches zero. For the "origin" boat and twist-boat conformers chosen to lie at $\varphi_2^o(B) = 0$ and $\varphi_2^o(TB) = \pi k/7$, the permutationally isomeric conformers (k) will be

found to occur at:

$$\varphi_2^k(B,TB) = \varphi_2^o(B,TB) + \pi k/7 \qquad \text{for k = 0 to 13} \qquad [9]$$

The value of θ is found to deviate slightly from zero in practice and, as a result, there is a small ϕ_3-dependency in the pseudorotation itinerary between the minimum-energy forms. The pathway could therefore be described as being a helix, but one that wraps itself around a very thin torus of major radius ρ, that is normal to the chair - twist-chair helix in the four-dimensional hyperspace. The boat - twist-boat (B-TB) pathway is almost isoenergetic and exists some 10-11 $kJmol^{-1}$ above that of the chair - twist-chair (C-TC) pathway.

2.2.4 Cycloheptane: conformational nomenclature and definition of an asymmetric unit

Each of the conformations I-IV can be written as a set of 14 isomers, as defined by the equations [7] and [9]. These isomers correspond to the 14 unique permutations and inversions of the underlying torsion angle sequences and as such can be given a unique identification. This classification is based on the positioning of the C_s or C_2 symmetry element within the fragment. Thus, for the fixed atomic and bond enumeration shown in Table 2.2, the torsional sequences given correspond to the origin, or reference, conformers in each case. These conformers have the plane or axis of symmetry passing through C1 and the mid-point of the bond C4-C5: the bond that carries the unique asymmetric torsion angle in each conformation. These conformers have been denoted as C^4, TC^4, B^4 and TB^4 respectively. The enantiomorphs of these isomers are found at $\phi_2+\pi$, $\phi_3+\pi$ and are denoted as $C^{\bar{4}}$, $TC^{\bar{4}}$, $B^{\bar{4}}$ and $TB^{\bar{4}}$. Permutation of the torsional sequence moves the unique torsion angle to the six other possible bonds in turn, leading to general descriptors of the form C^n, TC^n, B^n and TB^n, together with their enantiomers.

For the TC and TB forms which possess C_2 symmetry there can be no ambiguity: the positive descriptors of TC^n and TB^n arising from a positive value of the

Table 2.3

The column headed $\tau = 0$ indicates the position of the zero-valued torsion angle in the τ_1-τ_7 sequence for an ideal chair (see Table 2.1), and also gives the signs of the two torsion angles which flank this zero value; this yields the designators in the "code" column which are then used in the visual display of pseudorotation itineraries in Figure 2.3.

Chairs				Twist-chairs		
ϕ_2	ϕ_3	$\tau = 0$	Code	ϕ_2	ϕ_3	Code
0	14	+4-	C^4	3	15	TC^1
6	16	-5+	$C^{\bar{5}}$	9	17	$TC^{\bar{2}}$
12	18	+6-	C^6	15	19	TC^3
18	20	-7+	$C^{\bar{7}}$	21	21	$TC^{\bar{4}}$
24	22	+1-	C^1	27	23	TC^5
2	24	-2+	$C^{\bar{2}}$	5	25	$TC^{\bar{6}}$
8	26	+3-	C^3	11	27	TC^7
14	0	-4+	$C^{\bar{4}}$	17	1	$TC^{\bar{1}}$
20	2	+5-	C^5	23	3	TC^2
26	4	-6+	$C^{\bar{6}}$	1	5	$TC^{\bar{3}}$
4	6	+7-	C^7	7	7	TC^4
10	8	-1+	$C^{\bar{1}}$	13	9	$TC^{\bar{5}}$
16	10	+2-	C^2	19	11	TC^6
22	12	-3+	$C^{\bar{3}}$	25	13	$TC^{\bar{7}}$

unique torsion angle, and the enantiomorphs being described by the presence of a negative value for this unique torsion. However, for the C and B conformations the C_s symmetry means that this definition is not so direct. Instead the identification of the given isomer relies on the sense of the torsion angles which flank the unique torsion, *i.e.* whether τ_3 or τ_5 is positive or negative. For isomers where τ_3 is positive (and hence τ_5 is negative) the positive descriptor C^n or B^n is used, and similarly when τ_5 is positive (and τ_3 negative) the negative description is invoked and the isomer is classified as $C^{\bar{n}}$ or $B^{\bar{n}}$.

The ϕ_2, ϕ_3 correspondences for the C/TC descriptors are presented in Table 2.3 and illustrated in Figure 2.3, which serves to show for example, how one permutational isomer of the chair conformation transforms into the geometrically closest next chair

isomer, via one of the conformers of the twist-chair variety. The nomenclature established above and used throughout this Thesis differs from that proposed by Boessenkool & Boeyens (1980), who classify the different isomers by virtue of the "symmetry atom", rather than the "symmetry bond" as used here. The bond-centred description is preferred since it seems more natural, particularly since we are classifying torsion angles, which themselves are bond-centred parameters.

It will be seen later that the most important pseudorotation itinerary is that which maps transformations between C and TC conformers. This pathway is easily visualized by the planar projection in Figure 2.3. Here the lengths of the ϕ_2, ϕ_3 axes are respectively proportional to $c\rho\sin\theta = cq_3$ (major) and $\rho\cos\theta = q_2$ (minor). This projection could be obtained by cutting the ϕ_3-dependent torus of Figure 2.2 along the $\phi_3 = \pm 180$ axis, straightening to form a cylinder, and then cutting the cylinder lengthwise along $\phi_2 = \pm 180$ and flattening. In fact to get a better appreciation of four-dimensional space and the helical pathways that can be observed for the C/TC

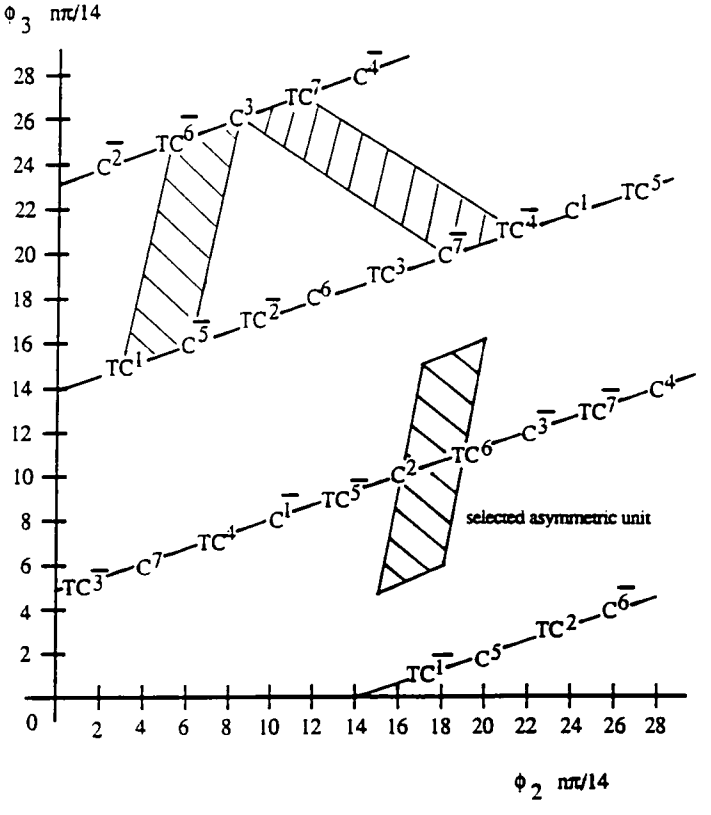


Figure 2.3

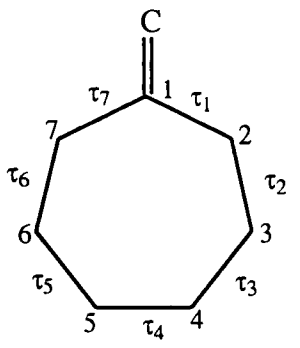
A 2D schematic representation of the C/TC pseudorotation itinerary for cycloheptane, with three possible asymmetric units marked

pseudorotation, it is advisable to physically perform this operation in reverse with paper or plastic tubing, and thus build up a model of the torus.

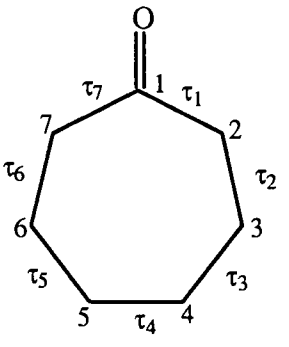
The asymmetric unit of this conformational space must contain one of the possible 28 permutations / inversions for each general (asymmetric) ring, *i.e.* it must represent 1/28th of the total volume of the conformational space. Conversely, it should be possible to map the entire conformational space by application of the 28 symmetry elements of the D_{7h} topological point group to any one of the chosen asymmetric units. In this four-dimensional space ρ is a size parameter, whilst θ must be allowed to vary over all of the permitted values in its range ($0 \leq \theta \leq \pi/2$). The definition of the asymmetric unit relies on the choice of suitable values for ϕ_2, ϕ_3 such that the chair, twist-chair, and inverse conformers, are fully represented. Three of the possible choices are shown in Figure 2.3, with each of the parallelograms highlighted having an area of $\pi^2 / 7$ rad^2, or 1/28th of the complete area ($4\pi^2$ rad^2) of the complete ϕ_2, ϕ_3 map.

2.2.5. Exo-unsaturated and heterocyclic seven-membered rings

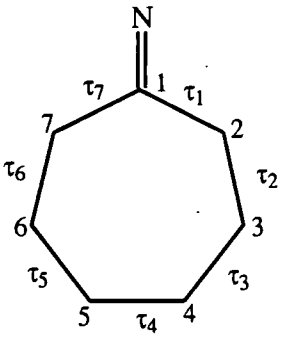
The conformational diversity observed in crystal structures that contain these seven-membered ring systems are also studied later in this Thesis. Examples of the parent structures for each of the datasets are given in Figure 2.4. From the outline given above, it will be obvious that there is conformational and energetic equivalence between the 14 isomers of each canonical form of cycloheptane, with only one symmetry-independent conformation in each case. The low energy barriers between the C and TC, and the B and TB allow pseudorotation to occur. For the structures V-IX there are significant reductions in the topological symmetry of the fragment in question, from the D_{7h} of the planar cycloheptane species, to the C_{2v} of the planar species of V-IX. The reduction in symmetry leads to four possible symmetry-independent conformations in each sub-class: C^1-C^4, TC^1-TC^4, B^1-B^4 and TB^1-TB^4. This can be seen in the revised ϕ_2, ϕ_3 plot for the C/TC pathway (Figure 2.5) in which, for example, C^1 is equivalent to $C^{\bar{1}}$, C^7, and $C^{\bar{7}}$. This effect can be repeated for the general C^n and TC^n conformers.



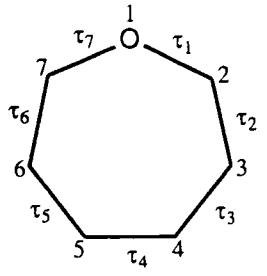
V



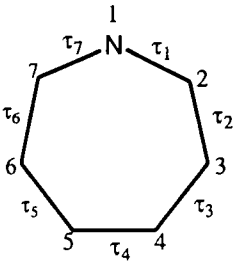
VI



VII



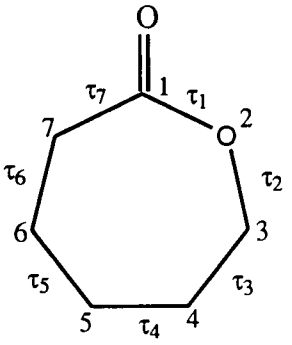
VIII



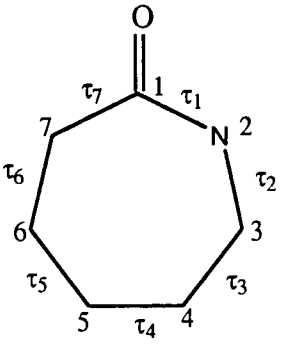
IX

Atomic permutational symmetry group for **V - IX**

1 2 3 4 5 6 7 1 7 6 5 4 3



X



XI

Atomic permutational symmetry group for **X - XI**

1 2 3 4 5 6 7

Figure 2.4

The parent fragments for the datasets V-XI

The asymmetric unit marked in Figure 2.5 now has an area of $\pi^2 rad^2$. Each of the symmetry independent conformers occupies a different energy minimum and pseudorotation can only occur if the energy barriers between these different conformers are suitably low. The interconversion of conformations between the C/TC or B/TB minima is restricted by the energetical properties of each different conformer.

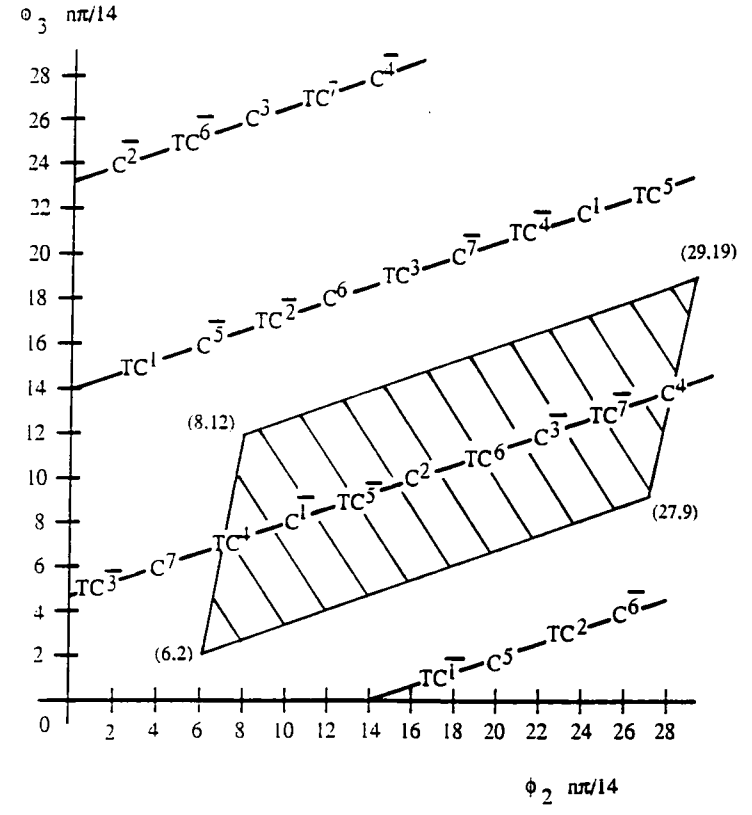
The fragments X-XI are completely topologically asymmetric, and as such each conformer is distinctly different from any of the others. All seven forms in each

canonical subclass are symmetry-independent and therefore energetically unique. The corresponding asymmetric unit would therefore have to incorporate all of these non-equivalents, and would have an area $2\pi^2 rad^2$ and occupy half of the ϕ_2, ϕ_3 plot, the other half being occupied by enantiomorphic conformers.

2.2.6. Energetically preferred conformations of exo-unsaturated and hetero-substituted seven-membered rings

Cycloheptanone (VI) has been the subject of potential energy calculations by Allinger, Tribble & Miller (1972) as well as by Bocian & Strauss (1977b). These studies cite the TC^3 and TC^4 conformers as the minimum energy forms: Allinger *et al.* (1972) having the TC^3 lower than TC^4 by 1.05 $kJmol^{-1}$, while Bocian & Strauss (1977b) find the TC^4 to be the most stable by 0.84 $kJmol^{-1}$. The latter authors also cite the C^1 conformation at 0.53 $kJmol^{-1}$ above the TC^4, and then calculate all other C/TC forms to be at least 6.7 $kJmol^{-1}$ higher in energy than the TC^4 global minimum. All of the B/TB conformers are determined to lie at least 9.2 $kJmol^{-1}$ higher in energy than the TC^4, in agreement with the results found by Allinger *et al.* (1972). The C/TC interconversion pathway can therefore be seen to have broad low-energy regions encompassing the conformational sequences TC^3 - C^1 - TC^4 - C^1 - TC^3 (Figure 2.5), and bounded by high energy barriers across that track > 6.7 $kJmol^{-1}$, and separated from the B/TB forms by even higher potential barriers of > 9.2 $kJmol^{-1}$. No similar force-field calculations appear in the literature for the methylene (V) or azo (VII) analogues of cycloheptanone.

Studies on the oxacycloheptane [oxepane (VIII)] molecule (Bocian & Strauss, 1972a, c), have incorporated vibrational spectra and force-field calculations in arriving at the conclusion that isolated energy minima occur for the TC^2 and TC^3 conformations, with the TC^3 being preferred by just 0.05 $kJmol^{-1}$. All other C/TC forms are found to lie at least 8.1 $kJmol^{-1}$ above the TC^3, and the B/TB forms are at least 10.4 $kJmol^{-1}$ higher in energy than the global minimum TC^3 conformer. Similar calculations for the azacycloheptane (IX) could not be found in the literature.



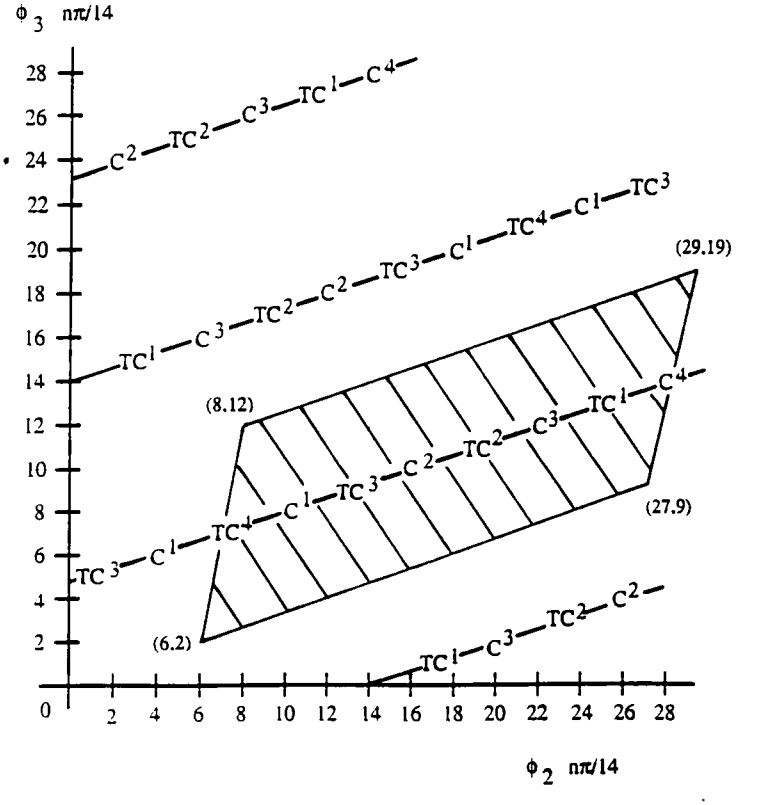


Figure 2.5

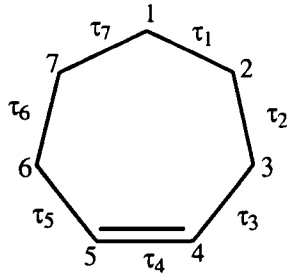
(a) ϕ_2 - ϕ_3 plot of the C/TC pseudorotation itinerary with a full set of conformational descriptors, and (b) the ϕ_2 - ϕ_3 plot of the C/TC itinerary with a reduced (C_{2v}) set of conformational descriptors

Conformational studies of the 5,5-difluoro-ε-caprolactone and its lactam analogue (X) and (XI) have been performed by Noe & Roberts (1971) using ^{19}F NMR. The barrier to ring inversion was determined to be 42.0 kJmol^{-1} and 43.7 kJmol^{-1} for (X) and (XI) respectively, with the NMR data suggesting that the lowest-energy conformation is a chair. Force-field calculations on the lactone support this view (Allinger, 1982), locating four energy minima. The global minimum is determined to be the C^1 conformation with the B^1 in a local minimum 11.4 kJmol^{-1} higher. Two other asymmetric intermediate conformations, described as half-chair and *trans*-forms, are some 17.8 kJmol^{-1} and 22.3 kJmol^{-1} higher in energy than the chair form, respectively.

2.2.7. Endo-unsaturated seven-membered rings

For the endo-unsaturated seven-membered rings studied in this Thesis (see Figure 2.6), the presence of one or more double bonds imposes a considerable torsional constraint on the conformation of the fragment. Thus, despite having C_{2v} topological symmetry, the conformational flexibility for the structures XII - XV has been severely reduced by comparison with the saturated, single-bonded analogues.

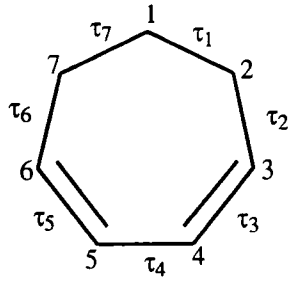
Cycloheptene was originally assumed to exist in the C^4 or B^4 conformations (Pauncz & Ginsburg, 1960; Eliel, Allinger, Morrison & Angyal, 1965) but spectroscopic evidence (Allinger & Sprague, 1972; Ermer & Lifson, 1973) determines the C^4 chair as being the lowest-energy conformation for this species. Force-field calculations on the cycloheptene molecule (Allinger & Sprague, 1972; Ermer & Lifson, 1973) also indicate the C^4 chair to be the global minimum, but with a C_2 symmetric twist form (1-T in Table 2.4) only 2.1 kJmol^{-1} higher in energy. Analysis of the sequence of intra-annular torsion angles for this fragment shows that the C_2 form is a symmetric intermediate in the TC-TB interconversion of I (Hendrickson, 1967). Ermer & Lifson (1973) cite an asymmetric and heavily distorted boat (1-A in Table 2.4 at 11.1 kJmol^{-1}) as the next lowest-energy minimum, whereas Allinger & Sprague (1972) locate the B^4 conformer at 14.2 kJmol^{-1}.



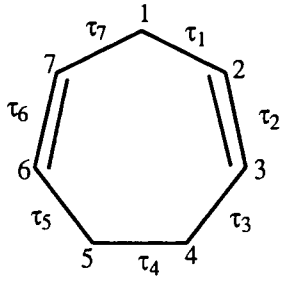
XII

Atomic permutational
symmetry group for **XII-XV**

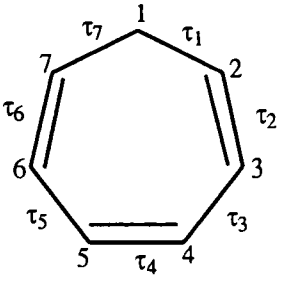
1 2 3 4 5 6 7 1 7 6 5 4 3 2



XIII



XIV



XV

Figure 2.6

The parent fragments for the datasets XII-XV

The 1, 3-cycloheptadiene (XIII) has been studied using a wide variety of experimental and theoretical techniques. From electron diffraction (Chiang & Bauer, 1966; Hagen & Traetteberg, 1972) and microwave (Avirah, Molloy & Cook, 1978) studies, the lowest-energy form has been determined unequivocally to be the conformation with C_S symmetry, which holds all of the ring atoms except C1 coplanar. A distorted, twist conformation which preserves C_2 symmetry, has been proposed on the basis of NMR data (Crews, 1971), but molecular mechanics calculations (Burkert & Allinger, 1982; Favini, Maggi & Todeschini, 1983) support the view that the C_S form is in fact the lowest in energy by some 7-8 $kJmol^{-1}$. This result is further supported by the *ab initio* calculations performed by Saebo & Boggs (1982). The torsion angles for the C_S "envelope" conformation (1,3-E) and the C_2 twist form (1,3-T) are given in Table 2.4, together with the data for the asymmetric form (1,3-A) which was identified uniquely by Favini, Maggi & Todeschini (1983). Comparison of the results for the 1,3-E and 1,3-T forms with the results found by Hendrickson (1967) show them to

correspond to the half-boat and twist-chair/twist-boat intermediate forms of cycloheptane respectively.

Force-field calculations appear to be the only source of information in the literature about the preferred conformations of the 1,4-cycloheptadienes. In this case a

Table 2.4

Intra-annular torsional sequences for the ideal conformations of datasets XII-XV, and the parent cycloheptane (I), for comparison. Results of various energy calculations for each of the conformations are also shown.

Conf.	Sym	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	ΔE	Ref
(I) Cycloheptane										
TC	C_2	-39.1	88.1	-72.3	54.3	-72.3	88.1	-39.1	0.0	(a)
C	C_1	63.8	-83.5	66.1	0.0	-66.1	83.5	-63.8	5.9	(a)
TB	C_2	45.4	-64.4	-17.9	74.6	-17.9	-64.4	45.4	10.0	(a)
B	C_s	-57.5	-30.9	69.0	0.0	-69.9	30.9	57.5	11.3	(a)
(XII) Cycloheptene										
C_4	C_s	67.6	-74.9	58.5	0.0	-58.5	74.9	-67.6	0.0	(b)
1-T	C_2	-43.1	77.9	-27.9	-7.8	-27.9	77.9	-43.1	1.8	(b)
1-A	1	86.7	-46.6	-10.4	-4.8	63.2	-49.0	-30.0	11.1	(b)
(XIII) Cyclohepta-1,3-diene										
1,3-E	C_s	70.5	-30.4	1.4	0.0	-1.4	30.4	-70.5	0.0	(c)
1,3-T	C_2	40.6	-71.4	3.2	46.4	3.2	-71.4	40.6	10.5	(c)
1,3-T	C_2	40.0	-73.8	8.3	38.6	8.3	-73.8	40.0	7.6	(d)
1,3-A	1	-49.9	1.1	2.2	22.7	-1.8	-57.9	85.8	3.4	(d)
(XIV) Cyclohepta-1,4-diene										
1,4-A1	1	64.7	-0.4	-69.0	56.1	-4.1	0.5	-49.0	0.0	(d)
1,4-A2	1	0.0	0.0	44.7	-80.0	57.8	1.6	-24.4	3.8	(d)
1,4-T	C_2	-10.3	1.7	49.1	-74.9	49.1	1.7	-10.3	4.6	(c)
1,4-C	C_s	-70.2	4.2	-46.0	-0.1	46.0	-4.2	70.2	12.6	(d)
1,4-C	C_s	-60.1	9.2	-22.1	0.0	22.1	-9.2	60.1	42.8	(c)
(XV) Cyclohepta-1,3,5-triene										
1,3,5-B	C_s	-62.5	-4.9	32.7	0.0	-32.7	4.9	62.5	0.0	(c)

References: (a) Hendrickson, 1967; (b) Ermer & Lifson, 1973; (c) Saebo & Boggs, 1982; (d) Favini, Maggi & Todeschini, 1983.

flattened C_2 twist form (1,4-T) appears to be the most energetically favourable, being preferred to the flattened C_s chair form (1,4-C). The cited energy differences between the two forms range from 12.6 $kJmol^{-1}$ (force-field: Favini, Maggi & Todeschini, 1983) to 42.8 $kJmol^{-1}$ (*ab initio*: Saebo & Boggs, 1982). These two symmetric forms were the only ones considered in the *ab initio* study, but the force-field calculations also show a diplanar conformer, and an asymmetric conformer to be possible (1,4-A1, and 1,4-A2 respectively), which are lower in energy than the preferred symmetric form, 1,4-T.

The conformation of the severely constrained 1,3,5-cycloheptatriene molecule is shown to be that of a flattened boat (1,3,5-B), with the typical intra-annular torsion angles given in Table 2.4. The experimental evidence for this comes from the electron diffraction data (Traetteberg, 1964), and the microwave spectroscopy results (Butcher, 1965), while two theoretical *ab initio* studies (Saebo & Boggs, 1982; Schulman, Disch & Sabio, 1982) reinforce this view.

2.2.8. Puckering coordinates for eight-membered ring systems

For the seven-membered rings, the usual geometrical descriptors of conformation are the seven directly measurable independent intra-annular torsion angles. A full torsional representation of the complete conformational hypersurface would therefore be seven-dimensional in nature. We have seen how the application of suitable mathematical techniques (*e.g.* to give the CP and BPRS coordinates) can reduce the overall dimensionality of this problem down to only (n-3) parameters, with the resulting four-dimensional space being much more easily visualized here in terms of the geometrical properties of a torus.

These mathematical techniques are generally applicable, so reducing an n-dimensional torsional description of that ring, to an (n-3) dimensional representation in terms of symmetry-adapted deformation coordinates. In this Thesis, the conformational space of eight-membered rings (cyclooctane and simple analogues) is also explored. These rings can be described by their 8 intra-annular torsion angles, but, using the

equations set out by Cremer & Pople (1975), this eight-dimensional conformational hypersurface can be reduced to a five-dimensional problem. The five symmetry-adapted puckering coordinates so generated describe the symmetrical deformations of the observed fragment away from the mean plane, and therefore relate to the deformations away from the parent fragment with D_{8h} topological symmetry.

For cyclooctanes, the eight atoms of the ring can each be described in terms of their z_j puckering coordinates away from the unique mean plane, for which $\sum_{j=1}^{j=8} z_j = 0$.

Two conditions are applied in order to avoid trivial rotation:

$$\sum z_j \cos 2\pi(j-1)/n = 0$$
$$\sum z_j \sin 2\pi(j-1)/n = 0 \qquad \textbf{[10]}$$

The five (n-3) puckering coordinates occur as two amplitude-phase pairs (q_2, ϕ_2), (q_3, ϕ_3) and a unique amplitude, q_4. These coordinates define two q, ϕ-pseudorotational pathways, together with a single inversional subspace corresponding, in the case of the eight-membered rings, to the crown form of D_{4d} symmetry (see Figure 2.7). For even-membered rings the puckering coordinates are defined by **[11]**:

$$q_m \cos\phi_m = \sqrt{(2/n)} \sum z_j \cos\left\{2\pi m(j-1)/n\right\}$$
$$q_m \sin\phi_m = \sqrt{(2/n)} \sum z_j \sin\left\{2\pi m(j-1)/n\right\}$$

where: $m = 2, 3 \ldots (n/2 - 1)$, $q_m > 0$, **[11]**

and: $0 < \phi_m < 2\pi$, $\qquad q_{n/2} = 1/\sqrt{n} \sum (-1)^j z_j$

An excellent and detailed description of the features of conformational space for eight-membered rings has recently been presented by Evans & Boeyens (1988). They show that a suitable 3D representation of the five-dimensional hypersurface generated by the CP coordinates, is provided by a set of tori on a sphere, as illustrated in Figure 2.8. The distinct tori are seen from Figure 2.8, to lie on the surface of a sphere. The radius of this sphere (Q) can be determined using the following expression:

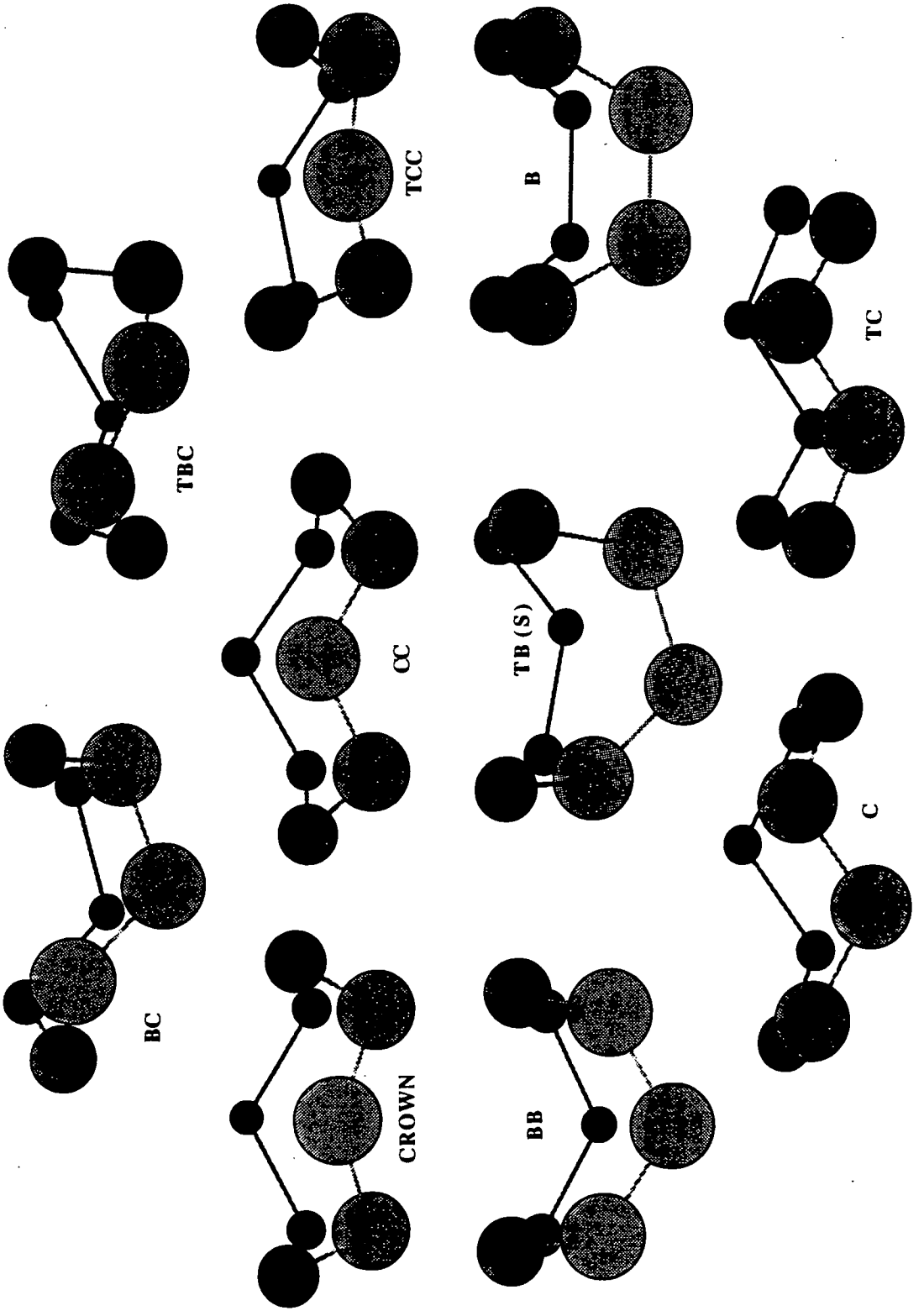


Figure 2.7

The ten canonical forms of cyclooctane

$$Q = \sqrt{q_2^2 + q_3^2 + q_4^2}$$

where the spherical polar angle (θ) lies in the range $0 < \theta < \pi$, and is given by:

$$\theta = \cos^{-1}\left(q_4 / Q\right)$$

The equations given by Evans & Boeyens (1988) imply that Q^2 is given by a summation of the q_m^2 values for m = 2, 3 ... ($n/2 - 1$) for all of the n = 8 atoms in the ring,

$$i.e.\ Q^2 = q_2^2 + q_3^2.$$

Their equation should contain a summation of the q_i^2 for i = 2 to $n/2$, so that

$$Q^2 = q_2^2 + q_3^2 + q_4^2$$

In this construction, for each value of θ a torus can be defined in terms of the q_2, ϕ_2 and q_3, ϕ_3 amplitude-phase pairs, so that θ defines a plane that runs through the central track of the associated torus. By doing this q_2 becomes redundant, as it is related to θ, and the major toroidal radius q_3 can be selected so as to avoid any overlap.

2.2.9. Conformational geometries of the eight-membered rings

Evans & Boeyens (1988) also identify the ten canonical symmetrical forms of cyclooctane (Figure 2.7), and present idealized Cartesian coordinates (Table 2.5) from which the torsion angles and puckering parameters of Table 2.6 may be derived.

The unique D_{4d}-symmetric crown form occurs at $\theta = 0°$ and its enantiomorph at $\theta = 180°$, a situation analogous to that of the unique chair form of the cyclohexane (Pickett & Strauss, 1970; Cremer & Pople, 1975). The tori describing canonical forms have major and minor radii q_2, q_3, and the positions of the individual conformers, and of their symmetry-related permutational and inversional variants, occur on these tori at specific values of ϕ_2, ϕ_3. The tori themselves occur at specific values of θ (Table 2.6). In this construction (and using the mnemonics of Table 2.6), the TCC/CC and BB/B/S pseudorotational pathways occur at $\theta = 30°$, and $\theta = 90°$ respectively, with their

Table 2.5

Idealized Cartesian coordinates for the 10 symmetrical conformations of cyclooctane rings

	B			CROWN	
-0.729800	-0.398400	1.316000	-0.681400	-0.069400	1.403400
0.140000	-0.758600	2.548500	-0.918500	-1.490700	1.977400
0.874000	-2.125000	2.547400	0.254200	-2.100400	2.789000
0.693400	-3.047600	1.313800	1.317100	-2.877100	1.969000
1.419400	-2.657200	0.000000	2.481800	-2.031200	1.391500
2.281800	-1.368100	0.000000	2.233600	-1.392700	0.000000
1.551000	0.000000	0.000000	1.551000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	BB			TB(S)	
-0.732000	-0.852900	-1.077000	-0.735900	-0.522800	-1.408600
-0.362000	-0.631000	-2.566000	0.054700	-0.328100	-2.728700
1.129000	-0.713000	-2.986000	1.358900	-1.144700	-2.923600
1.938000	-1.968000	-2.566000	1.431000	-2.531700	-2.233200
1.913000	-2.410000	-1.079000	1.537400	-2.570400	-0.686300
2.290000	-1.361000	0.000000	2.257800	-1.380600	0.000000
1.554000	0.004000	0.004000	1.551000	0.000000	0.000000
0.003000	-0.003000	-0.007000	0.000000	0.000000	0.000000
	BC			TBC	
-0.671300	-0.982900	-0.977000	-0.728300	-0.946800	1.125200
-0.459700	-0.745000	-2.495000	-1.090400	-2.400000	0.721800
0.994400	-0.901700	-3.011500	0.019800	-3.461800	0.935700
1.769800	-2.141300	-2.494000	0.934800	-3.743900	-0.288400
2.796300	-1.877100	-1.361700	1.628600	-2.422200	-0.989800
2.233600	-1.392700	0.000000	2.233600	-1.392700	0.000000
1.551000	0.000000	0.000000	1.551000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	C			TC	
-0.703800	-1.388700	-0.007000	-0.684200	0.447500	-1.323400
-0.697400	-2.126000	-1.371500	0.017400	0.014100	-2.637000
0.105000	-3.506000	-1.367600	0.654900	-1.399800	-2.632900
1.561500	-3.501800	-1.363900	2.205900	-1.392500	-2.630500
2.261900	-2.117900	-1.364500	2.885900	-1.849700	-1.313600
2.257800	-1.380600	0.000000	2.184600	-1.415700	0.000000
1.551000	0.000000	0.000000	1.551000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	CC			TCC	
-0.657700	0.360600	1.365000	-0.674700	0.581300	1.256400
-0.413100	-0.661700	2.505500	-0.817400	-0.381100	2.464300
1.061000	-0.782200	2.972700	0.511500	-1.002200	2.968000
1.796400	-2.062700	2.498200	0.995300	-2.245500	2.177000
2.821900	-1.844100	1.355300	2.322700	-2.078100	1.392400
2.209200	-1.404400	0.000000	2.209200	-1.404400	0.000000
1.551000	0.000000	0.000000	1.551000	0.000000	0.000000
0.000000	-0.000000	0.000000	0.000000	0.000000	0.000000

enantiomorphic conformations occurring at $\theta = 150°$ and $\theta = 90°$. Since $q_3 = 0$ for the idealized forms of all these conformers, the tori contract to latitudinal circles in both cases, with symmetry variants defined by ϕ_2.

The C and TC forms have $q_2 = 0$ and their pseudorotational pathway follows a meridian on the sphere, with symmetry variants at positions defined by ϕ_3. Finally, the important, energetically-preferred BC conformers have $q_2, q_3 \neq 0$, and the permutational variants lie on helical tracks, on a torus at $\theta = 75°$, with a set of inverse

conformations on a torus at $\theta = 105°$. The TBC conformers, approximately 8.4 $kJmol^{-1}$ higher in energy than the BC form (Hendrickson, 1967), lie on these helical tracks midway between the BC variants. The BC/TBC representation is in effect the only true torus, as it is the only one with non-zero values for both q_2 and q_3 and so the toroidal properties that were observed for the conformational hyperspace of the seven-membered rings, are only truly mimicked by the BC/TBC pseudorotation itinerary in the conformational hyperspace of the eight-membered rings. When the ϕ_2, ϕ_3 coordinates are plotted for the BC conformers therefore (Figure 2.9), the resulting itinerary shows the "four-fold" nature of the conformational space: a direct consequence of the inherent symmetry in the D_{8h} system.

Table 2.6

Archetypal torsion angles for the ten canonical forms as found by minimum energy calculations (Hendrickson, 1967), corresponding puckering amplitudes (q_m) and polar angles (θ) (Evans and Boeyens, 1988), and differences in energies ($kJmol^{-1}$) (Anet and Krane, 1973).

Conf.	BC	CC	BB	TC	C	B	Crown	TCC	TBC	S
τ_1	65.0	66.0	52.5	37.3	119.9	-73.5	87.5	56.2	88.0	64.9
τ_2	44.7	-105.2	52.5	-109.3	-76.2	0.0	-87.5	-82.4	-93.2	37.6
τ_3	-102.2	105.2	-52.5	109.3	0.0	73.5	87.5	114.6	51.9	-64.9
τ_4	65.0	-66.0	-52.5	-37.3	76.2	0.0	-87.5	-82.4	44.8	-37.6
τ_5	-65.0	66.0	52.5	-37.3	-119.9	-73.5	87.5	56.2	-115.6	64.9
τ_6	102.2	-105.5	52.5	109.3	76.2	0.0	-87.5	-82.4	44.8	37.6
τ_7	-44.7	105.2	-52.5	-109.3	0.0	73.5	87.5	114.6	51.9	-64.9
τ_8	-65.0	-66.0	-52.5	37.3	-76.2	0.0	-87.5	-82.4	-93.2	-37.6
θ (°)	75,105	30,150	90	90	90	90	0,180	30,150	75,105	90
q_2 (Å)	1.07	0.46	1.54	0.0	0.0	1.64	0.0	0.46	0.67	1.63
q_3 (Å)	0.61	0.0	0.0	0.96	0.96	0.0	0.0	0.0	0.82	0.0
q_4 (Å)	0.33	± 0.80	0.0	0.0	0.0	0.0	± 0.87	± 0.80	0.28	0.0
ΔE	0.0	7.5	11.7	32.2	31.4	46.9	6.3	3.3	7.1	11.7

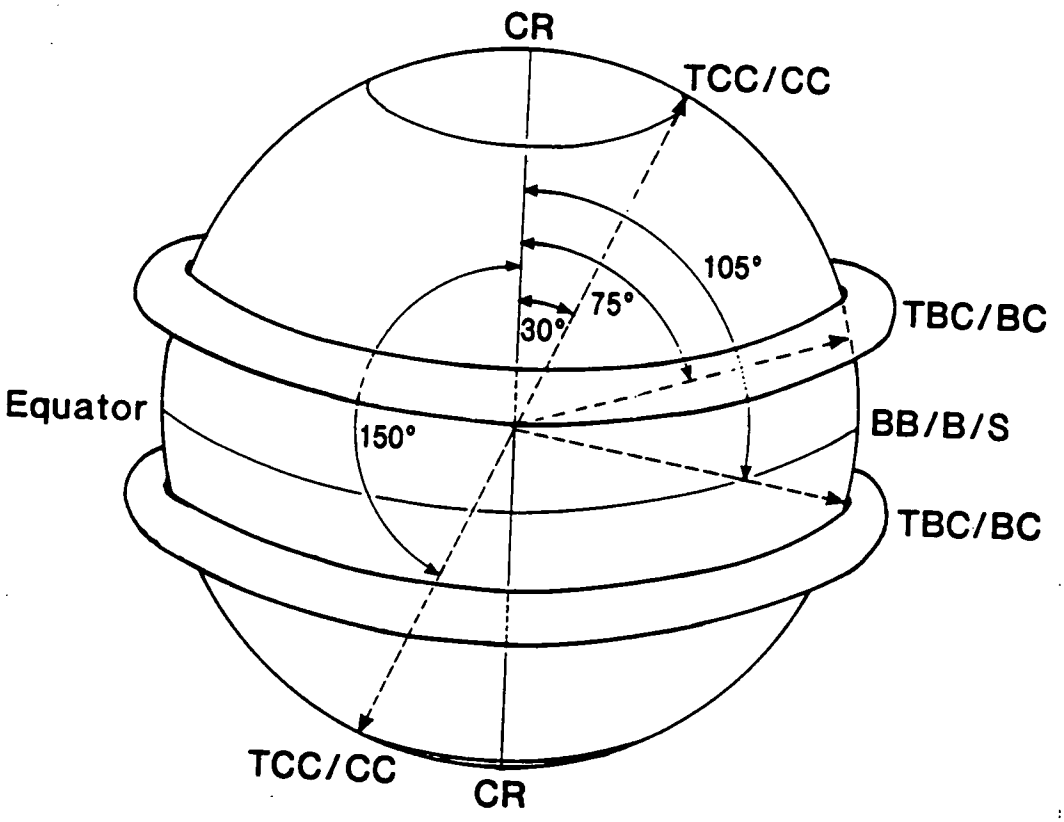


Figure 2.8

Tori on a sphere: a pictorial representation of the 5D-hyperspace for cyclooctanes

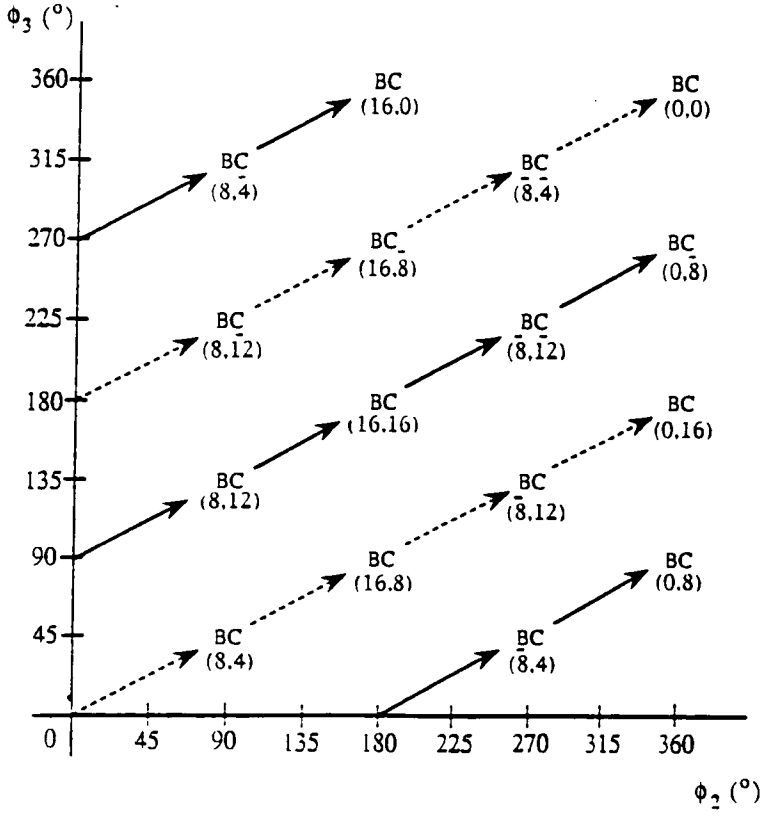
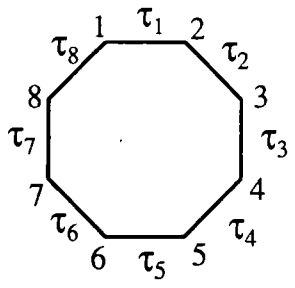


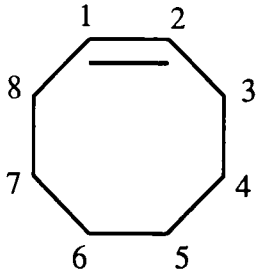
Figure 2.9

The pseudorotation itinerary for BC conformers of cyclooctane

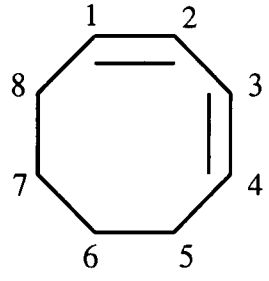
8C1-cyclooctane (**XVI**)



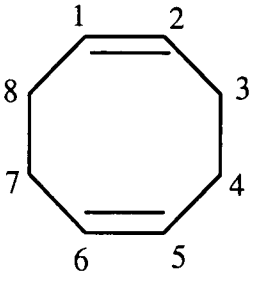
8C2-cyclooctene (**XVII**)

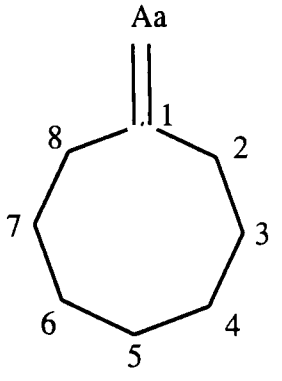


8C3-cycloocta-1,3-diene (**XVIII**)



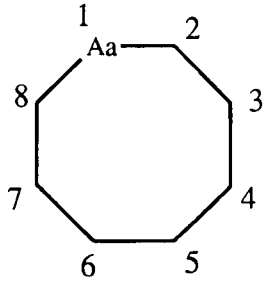
8C4-cycloocta-1,5-diene (**XIX**)



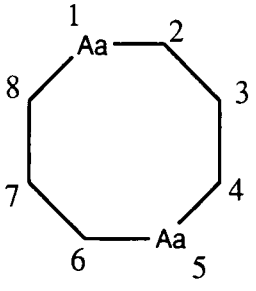


8C5 - exo-unsaturated rings (**XX**)

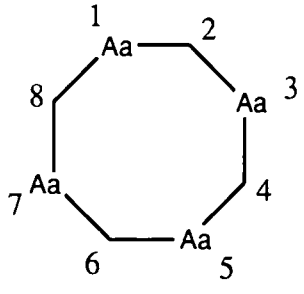
(Aa = any atom)



8A1 (**XXI**)



8A2 (**XXII**)



8A3 (**XXIII**)

Figure 2.10

Eight-membered rings: parent fragments for the datasets (XVI)-(XXIII)

Evans & Boeyens (1988) define a nomenclature for eight-membered ring conformations, that specifies the angular positions of the canonical conformations and their symmetry variants. This nomenclature involves the three indices h, k, l in angular definitions of the form :

$$\phi_2 = h\pi / 16 \qquad \phi_3 = k\pi / 16 \qquad \theta = l\pi / 16 \qquad \textbf{[12]}$$

where $-16 < h, \quad k < 16, \quad (\bar{h}, \bar{k} = 32\text{-}h, k), \quad$ and, $\quad 0 < l < 16.$

The canonical forms are distinguished by only one or two integers, and the enantiomeric conformations occur at :

$$\phi_2' = \phi_2 + \pi = \phi_2 (h + 16) \qquad \textbf{[12a]}$$

$$\phi_3' = \phi_3 (k + 16) \qquad \textbf{[12b]}$$

$$\theta' = \pi - \theta = \theta (16 - l) \qquad \textbf{[12c]}$$

Evans & Boeyens (1988) apply their constructs and definitions to a small subset of 11 carbocyclic and heterocyclic eight-membered rings. This Thesis will extend this preliminary study to include all of the available examples of cylooctane, cyclooctene, and related fragments retrieved from organic molecules stored in the Cambridge Structural Database (see Figure 2.10 for a complete identification of the datasets analysed). As was the case for the different varieties of the seven-membered rings, when other structural elements are introduced into the parent fragment for the eight-membered ring datasets, the topological symmetry is reduced and the resulting symmetry expansion necessary to describe the complete conformational hypersurface is also reduced. These effects are analogous to those described for the seven-membered rings and need not be further discussed here.

2.3 Multivariate statistics applied to crystallographic data

2.3.1. Overview of multivariate methods

When the number of variables required to completely describe an object is larger than two, then multivariate methods of analysis are needed that will allow us to handle the potentially large number of parameters that are generated for many instances of that object. The chemical fragments that will be studied in this Thesis fall into this category: the n-membered rings being best described by their n intra-annular torsion angles, and the metal n-coordinated fragments by the n(n-1)/2 valence angles generated

at the metal centre by the bonded ligands. Hence, the successful application of multivariate techniques to these systems should be of great advantage in the understanding and interpretation of the mechanisms and trends that we hope to observe, and impart the statistical confidence that we need to guarantee the validity of the results (Taylor and Allen, 1994).

In terms of the parameters that are actually retrieved from the Cambridge Crystallographic Database by systematic searching, we derive a data matrix $\mathbf{G}(N_f, N_p)$ comprising the N_p parameters that describe the geometry of each of the N_f examples of the fragment. For the multivariate analysis to have any meaning the parameters chosen to represent the data must arise on an equal footing, *i.e.* it is formally invalid to analyse valence angles together with bond length displacements as the relative terms involved would render any results seriously biased.

Multivariate techniques basically aim to simplify the wealth of data by the use of two different approaches. The first is through a variable-directed approach, which looks at the interdependencies that may exist between the observed variables, and tries to find ways of better describing the system in terms of a set of fewer variables. The second technique analyses the data with an object-oriented approach, treating each of the N_f objects as a data point in an N_p-dimensional parameter space, so as to identify clusters of like data density and thereby allowing classification and pattern-recognition in the system as a whole. Examples of both of these methods are used for data analysis in this Thesis and will be described in Sections 2.3.3 and 2.3.4. However, before any use of these multivariate statistical methods can be made, it is imperative that the topological symmetry of the system under investigation is understood and rigorously accounted for.

2.3.2. Symmetry Considerations

Relatively few of the molecular structures to be found in the CSD exhibit molecular symmetry. However, many substructural fragments (*e.g.* the cycloheptane unit to be studied later in this Thesis) are small and symmetric, and the fragment symmetry has consequences for the search process itself, and hence, on the relative

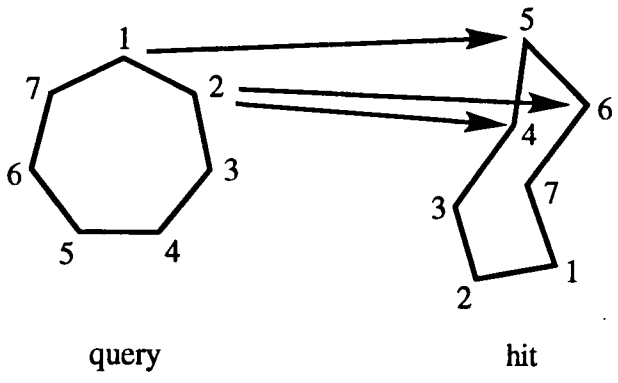


Figure 2.11

The D_{7h} planar query fragment can map equivalently to 28 isomers of the lower-symmetry 3D hit, due to atomic enumeration properties and CSD search procedures

ordering of the N_p geometrical parameters recorded for each fragment in the $\mathbf{G}(N_f, N_p)$ multivariate data matrix (Taylor & Allen, 1994).

The cycloheptane fragment in Figure 2.11 is specified as a 2D chemical connectivity query, with the 7 carbon atoms arbitrarily enumerated. This fragment has plane symmetry of D_{7h}. Substructure searching in the CSD is accomplished by an atom-by-atom, bond-by-bond matching of the query sketch against a succession of graphs representing complete molecules (targets). For cycloheptane, atom 1 of the query fragment can be mapped equivalently onto any of the atoms of a hit fragment in a CSD molecule. If for instance atom 1 (query) maps to atom 5 (hit), then atom 2 (query) can then map to either atom 4 (hit) or atom 6 (hit), arbitrarily and equivalently. If this approach is expanded to account for all of the possible arrangements, then a total of 14 equivalent and independent mappings can be generated (Table 2.7), but the fragment search routines in the CSD system will select only one of these as satisfying the search query.

In describing the conformational properties of the cycloheptane fragments we employ the sequence of seven intra-annular torsion angles (τ_1-τ_7 : see Table 2.2). Only in the special case of a planar ring (where all of the torsion angles are zero) will the D_{7h} symmetry of the query be matched by that of the hit, and the sequence of torsion angles will be equivalent. In more general cases the 3D symmetry is lower than D_{7h}, *e.g.* C_s (chair and boat) and C_2 (twist-chair and twist-boat). Here the alternative atomic

enumerations obviously lead to alternative orderings of the cyclic sequence of torsion angles.

In 3D it is also necessary to realise that each fragment also has an enantiomorph of equal interest. Torsion angles are enantiomorph-sensitive: a change in the enantiomorph results in a reversal of the signs of the torsion angles for each of the 14 possible enumerations. This means that there are now 28 equivalent torsion angle sequences possible for the cycloheptane fragment. Since it is impossible to know which sequence has been selected by the search routine, the only way to avoid discrepancies is to expand the original multivariate data matrix $\mathbf{G}(N_f, N_p)$ to include all of the 28 possible symmetry-equivalent representations.

When the symmetry properties of the search query change, then so too do the requirements for symmetry expansion of the resulting hits. For instance, if the search criteria were now altered so that we were searching for a cycloheptanone fragment,

Table 2.7

The 14 possible equivalent permutations of the atomic labels for the cycloheptane fragment

Frag	atomic labels : permutational possibilities						
Query	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	1	2	3	4	5	6	7
	2	3	4	5	6	7	1
	3	4	5	6	7	1	2
	4	5	6	7	1	2	3
	5	6	7	1	2	3	4
	6	7	1	2	3	4	5
Hit	7	1	2	3	4	5	6
	7	6	5	4	3	2	1
	6	5	4	3	2	1	7
	5	4	3	2	1	7	6
	4	3	2	1	7	6	5
	3	2	1	7	6	5	4
	2	1	7	6	5	4	3
	1	7	6	5	4	3	2

instead of the cycloheptane fragment, the search query would require the carbonyl group to be at one specific ring carbon atom, *e.g.* atom 1 (query). When a hit is located amongst the entries in the CSD the carbonyl carbon of the hit fragment must now be labelled as atom 1, and there are no other acceptable possibilities. Atom 2 of the query can still map to either atom 7 or atom 2 of the hit fragment, and the enantiomorphic properties of the torsional angles must still be considered, but when all of these are accounted for, the resulting symmetry expansion is just four-fold.

The symmetry expansion necessary to describe the D_{3h} symmetry of the 3-coordinated metal problem also studied later, is six-fold. Here the parameters defining the geometry are the angles subtended by the ligands at the metal, and these are always positive. The labelling on each ligand can change, but not on the metal, so the resulting expansion that fully accounts for atomic permutation is six-fold. The full atomic permutations used are shown above (Table 2.8).

Care must be taken to accurately and completely account for all of the possible permutations so that tools such as PCA will give more meaningful results in the later analyses. For the majority of the work carried out in this Thesis it was necessary to manually input all of the possible permutations into the GSTAT routines, but a developmental modification to QUEST3D (Johnson, 1994) will now allow all of the

Table 2.8

Atomic permutations allowed for the mapping of the 2D, D_{3h} query fragment, to the 3D hit, for three-coordinated metal complexes

Fragment	atomic labels : allowed permutations			
	metal	**ligand1**	**ligand2**	**ligand3**
Query	**1**	**2**	**3**	**4**
Hit	1	2	3	4
	1	3	4	2
	1	4	2	3
	1	2	4	3
	1	3	2	4
	1	4	3	2

permutational isomers to be located automatically in the search, and their geometries to be included in the data matrix. Hence, complete data analysis with VISTA is now possible (this new feature is described in Chapter 4.2).

2.3.3. Principal Component Analysis (PCA)

The aim of principal component analysis (PCA; see *e.g.* Chatfield & Collins, 1980; Murray-Rust & Bland, 1978; Auf der Heyde, 1990) is to transform a set of correlated variables into a new set of uncorrelated variables, or principal components. Since the PC's are uncorrelated, they are mutually orthogonal. PCA is a mathematical technique for which there is no underlying statistical model. The principal components are linear combinations of the original variables in N_p dimensional space and are derived in decreasing importance, with the first component attempting to account for the majority of the variance found in the original data. Successive components are derived until all of the variance in the original set of data is accounted for. PCA is therefore an example of a variable-directed analytical technique, with the derived principal components being given by a transformation that amounts to an orthogonal rotation in N_p-space.

PCA has one major objective: dimension reduction. If the analysis of an N_p dimensional problem can generate N_c principal components (where $N_c < N_p$) which account for >95% of the variance in the original data, then it is possible to use these new axes (components) to describe the system completely. In such a way the dimensionality of the problem can be reduced. For instance, if a chemical system requires a set of seven correlated variables in order to describe the geometry of the member fragments, and PCA can account for the vast majority (*e.g.* >95%) of the variance in the dataset as a whole, with just four components, then it would seem logical to use these new components to describe the original data, and thereby simplify the analysis of the problem.

If dimension reduction is achieved, then PCA can assist significantly the visualisation of a dataset. Quite often the results of a PCA will have no directly

interpretable chemical meaning, but when used in subsequent analyses these results can be most informative in terms of pattern recognition. For instance, when two components are plotted against one another, the resulting scattergram may show definite signs of data agglomeration, or clustering, in that 2-dimensional representation, that would not have been visible by taking two of the original parameters and plotting them. So the use of principal components should allow for easier visualisation of any trends that are inherent in the original data, but which may not be visible through use of the original parameters.

The derivation of the principal components from the original parameters is *via* eigenanalysis and is described fully in Chatfield & Collins (1980). A simple interpretation of the way that the technique works however, is that the first component is calculated in N_p-dimensional space as the vector for which the sum of the squares of the displacements of the (N_p x N_f) points in that N_p-space is minimised. This then satisfies the criterion that the first component vector accounts for the majority of the variance in the dataset as a whole. The second derived component must be orthogonal to the first and also have a minimised sum of square displacements from it to each of the (N_p x N_f) points, and this vector should then account for the next largest amount of variance. Every subsequently derived principal component must satisfy these requirements, and the analysis continues until the variance in the original N_p-parameter space is accounted for by N_c principal components.

The principal component analysis technique performs an eigenanalysis on either the covariance or correlation matrix, with the generated components being the eigenvectors of these matrices. For original variables in the form [x_i (i=1, 2, ...N_p)] the principal components are found to be linear combinations that satisfy:

$$(PC)_j = a_{1j}x_1 + a_{2j}x_2 + a_{3j}x_3 + ... + a_{Npj}x_{Np}$$

where the coefficients a_{ij} are known as the loadings, and are a measure of the relative contribution from each of the original variables to the derived $(PC)_j$. The eigenvalues of the matrix l_j (j=1, 2, ...N_p) represent the proportion of the total variance in the original dataset for each PC, and the PC "scores" that are generated are the coordinates

of the original fragment that are relocated in the N_c-dimensional PC-space. It is these scores that are plotted against one another to generate some of the scatterplots that will be seen to be significant visual aids in the interpretation of each dataset (Chapters 3 and 4).

One important point to emphasise is that PCA does require the original variables to occur in a consistent fashion. For the analyses performed in this Thesis, all of the original variables have to be symmetry-expanded before any PCA is run. Essentially, this expansion fills all asymmetric units of the hyperdimensional parameter space and removes the effects of the localisations of fragments in random asymmetric units that result from the QUEST search process, as described above.

Much information can be derived from the generated N_c components. Firstly, it should be noted that sometimes $N_c = N_p$ (Chapter 4), and the overall dimensionality is not reduced by doing a PCA. In this case the original N_p parameters are already orthogonal representations of the data, and so performing a PCA simply re-locates these N_p parameters in the equivalent N_c parameter space. Secondly, the PC's are required to be orthogonal to one another so the correlation, Cor (PC_n, PC_m), always equals zero. However, the correlation tables can be used to provide information about how well the PC axes are correlated to parameter axes generated in other ways (*e.g.* the symmetry deformation coordinates calculated in Chapter 4, or the CP descriptions used for ring conformational analysis in Chapter 3). These results are potentially very useful in assessing the chemical meaning of the PC's. Finally, plots of the generated PC scores can aid the interpretation of the data in chemical terms, since they "explain" the PC axes in chemical structural terms, *e.g.* the "observation" of pseudorotation itineraries in the ring systems. Such chemical "visualisations" are of great significance in the application of the PCA technique in structure correlation studies.

2.3.4. Cluster Analysis

The aim of all clustering algorithms is to try to find the natural groupings in a set of multivariate data, *i.e.* to allocate an individual data point to one of a set of

mutually exclusive groups, such that the points within each group are closely similar, but the points in different groups are highly dissimilar (see *e.g.* Everitt, 1980; Murray-Rust & Raftery, 1985; Taylor, 1986). Humans have a particular aptitude for this form of pattern recognition and can classify data very accurately, however attempts to get machines to perform these visual processes are fraught with potential pitfalls.

All clustering algorithms are based on the need to measure the degree of similarity (or more accurately dissimilarity) between every unique pair of data points in a dataset, and to then group points together on the basis of their mutual dissimilarities. The commonest measure of dissimilarity is the "distance" between points in an N_p-dimensional space, as can be found using the following general definition, the Minkowski metric **[13]**:

$$D(r,s) = \left\{ \sum_{i=1}^{N_p} \left[G(r,i) - G(s,i) \right]^n \right\}^{1/n} \qquad \textbf{[13]}$$

for two fragments r and s, each described by N_p parameters, where **G**(r, i) and **G**(s, i) are values of the i'th parameter, and the integer power factor, n, is user-specified (1 = "city-block" metric, 2 = Euclidean metric). The result is a square dissimilarity matrix, **D**(N_f, N_f), which is symmetrical about a zero diagonal.

There are problems associated with the computation of dissimilarity matrices for fragments that show topological symmetry, such as the cycloheptanes that are studied in Chapter 3. It has already been shown that topological symmetry can be accounted for by the relevant symmetry expansion, however when using this approach it is easy for the dissimilarity matrix to grow very rapidly in size. For cycloheptanes, the 28-fold increase in the number of data prior to generating the dissimilarity matrix, yields a dissimilarity matrix D($28N_f$, $28N_f$) whose calculation and storage is a tremendous drain on computational resources. An alternative approach has been developed (Allen, Doyle & Taylor, 1991a, b, c; Allen & Taylor, 1991) that tackles the problem of symmetry-expansion in a slightly different fashion. Their solution involves comparing all possible symmetry-equivalent permutations of the torsion angle sequence for fragment *s* with a fixed sequence for fragment *r* in **[13]**. The value of D(r, s) is computed for each

permutation of *s* and the lowest value of D(r, s) is stored in the final dissimilarity matrix, which then never exceeds the maximum dimensions of $D(N_f, N_f)$.

The CSD program GSTAT supports single-linkage, complete-linkage, and Jarvis-Patrick clustering algorithms (see Everitt, 1980; Jarvis & Patrick, 1973) for which the dissimilarity matrices computed in the above manner, can be used. These programs produce clusters with the fragments optimally overlaid on one another, and also try to ensure that the different clusters are placed in a single asymmetric unit of the conformational space. This asymmetric unit can be expanded after the clustering routines are finished to generate the full permutational (and inversional) symmetry of the relevant hyperspace, by application of the symmetry operations appropriate for that chemical fragment.

As an example of the use of the clustering algorithms we consider again the case of the seven-membered cycloheptane fragments (Allen, Howard & Pitchford, 1993). The symmetry-modification is an integral component of the GSTAT program and is utilised here. The algorithm used almost exclusively throughout this Thesis is the Jarvis-Patrick method (Jarvis & Patrick, 1973; Allen, Doyle & Taylor, 1991b), which has a proven ability to handle the complex chemical data of the type generated here. This routine calculates the dissimilarities $\left(D_{pq}^{n}\right)$ between two fragments p and q using the torsion angles τ_1-τ_7. The number of parameters (N_t) is therefore seven. This leads us to the dissimilarity equation given below:

$$D_{pq}^{n} = \left[\sum_{i=1}^{N_t} \left(\Delta\tau_i\right)_{pq}^{n}\right]^{1/n} \qquad \textbf{[14]}$$

where $\left(\Delta\tau_i\right)_{pq}$ is the minimum value obtained from one of the following two equations:

$$\left(\Delta\tau_i\right)_{pq} = \left|\left(\tau_i\right)_p - \left(\tau_i\right)_q\right| \Big/ 180N_t \qquad \textbf{[15a]}$$

or

$$\left(\Delta\tau_i\right)_{pq} = \left[360 - \left|\left(\tau_i\right)_p - \left(\tau_i\right)_q\right|\right] \Big/ 180N_t \qquad \textbf{[15b]}$$

which account for the phase change in the torsion angles as they pass through $\pm 180°$. As above, the value of n determines whether the routine uses the "city-block" metric or the Euclidean metric to establish dissimilarities. For the symmetric fragments studied $\left(D_{pq}^{n}\right)$ is the minimum value obtained by holding the $(\tau_i)_p$ constant, while allowing the values $(\tau_i)_q$ to adopt all of the permutations, as well as their inversions, that are allowed by the topological symmetry of the parent fragment.

In the Jarvis-Patrick method the $\left(D_{pq}^{n}\right)$ are then used to determine the K_{NN} nearest neighbours for each of the N_f fragments in the dataset, where K_{NN} is a user-supplied value. In order to qualify, these nearest neighbours must lie within a limiting value of the $\left(D_{pq}^{n}\right)$ which is specified by the user as $\left(D_{\lim}\right)$. The Jarvis-Patrick clustering criteria will then assign two fragments (p, q) to the same cluster if: (a) p occurs in the NN table of q and *vice versa*, and (b) if a further K_{JP} fragment are common to the NN tables of both p and q, where again, K_{JP} is supplied by the user. This method is an agglomerative hierarchical clustering algorithm that operates in a single pass, which makes it both computationally economical and renders it less susceptible to some of the more common problems of clustering routines, such as the "chaining" effect which can cause clusters to coalesce, particularly in the single-linkage method.

The Jarvis-Patrick clustering routine, as illustrated above, is by no means an automatic one. The user must specify values for n, $\left(D_{\lim}\right)$, the nearest neighbour table length value K_{NN}, as well as the commonality criterion K_{JP}, all of which can have distinct effects on the number of clusters and their composition that are generated by the routine. In practice the $\left(D_{\lim}\right)$ value is usually found to lie between 0.08 and 0.10 and can be regarded as a constant, and the value of n (1 or 2) does not appear to have a particularly significant effect on the cluster formation. The composition and number of clusters formed is, however, highly dependent on the K_{JP} parameter that is chosen for a particular value of K_{NN}: the ratio K_{JP} / K_{NN} appears to be critical, and values close to 0.5 are generally found to give chemically sensible clustering.

The clustering parameters used for each dataset are given in Chapter 3 and vary for each system. The selection of the optimum criteria was guided in all cases by *a*

priori chemical knowledge, as well as by examination of the various statistical descriptors that summarise the distributions of τ_1-τ_7 in each cluster. The resulting clusters are consistent and well defined and give a good conformational classification of the datasets in question.

2.4 Energy calculations

The principal aim of the database analyses to be detailed in the Chapter 3, is to classify the crystallographic observations for a given system into conformational subgroups where possible, and to locate any interconversion pathways that link these subgroups. Thus, we may gain an insight into the nature of the conformational energy hypersurface itself. It is very useful to be able to correlate the results obtained by database analyses, with the details of the energy hypersurface that are found by independent methods, such as force-field or *ab initio* studies. Unfortunately, the systems studied in the literature do not completely cover those to be examined in this Thesis, and so the need arose to perform our own independent calculations of the energetical properties of some of these ring systems. A consistent force-field method was required that could first give good correlation to the existing results, and then go on to analyse the systems for which no detailed data could be located.

The software used to gather and analyse the accessible conformations of the medium rings was based on the COSMIC molecular modelling package (Vinter, Davis & Saunders, 1987) and the extensions embodied in COSMIC90, which incorporates an updated force-field for hydrocarbons and conjugated systems (Morley, Abraham, Haworth, Jackson, Saunders & Vinter, 1991). The ring conformation hunter attempts to locate all of the local energy minima without first breaking the ring and then re-forming it after minimisation. These ring systems are not greatly affected by Coulombic interactions and so partial charges could be ignored. The program allows the user to define the number of randomly generated starting conformations, the total number of conformations to be held during a run, and also the energy range within which this number of conformations must lie. The complete procedure is summarized in Figure

2.12. The final output conformations can then be viewed and analysed using separate graphical modules. The results of these calculations will be in Chapter 3, together with their significance in terms of the conformational variety observed for the given structural systems.

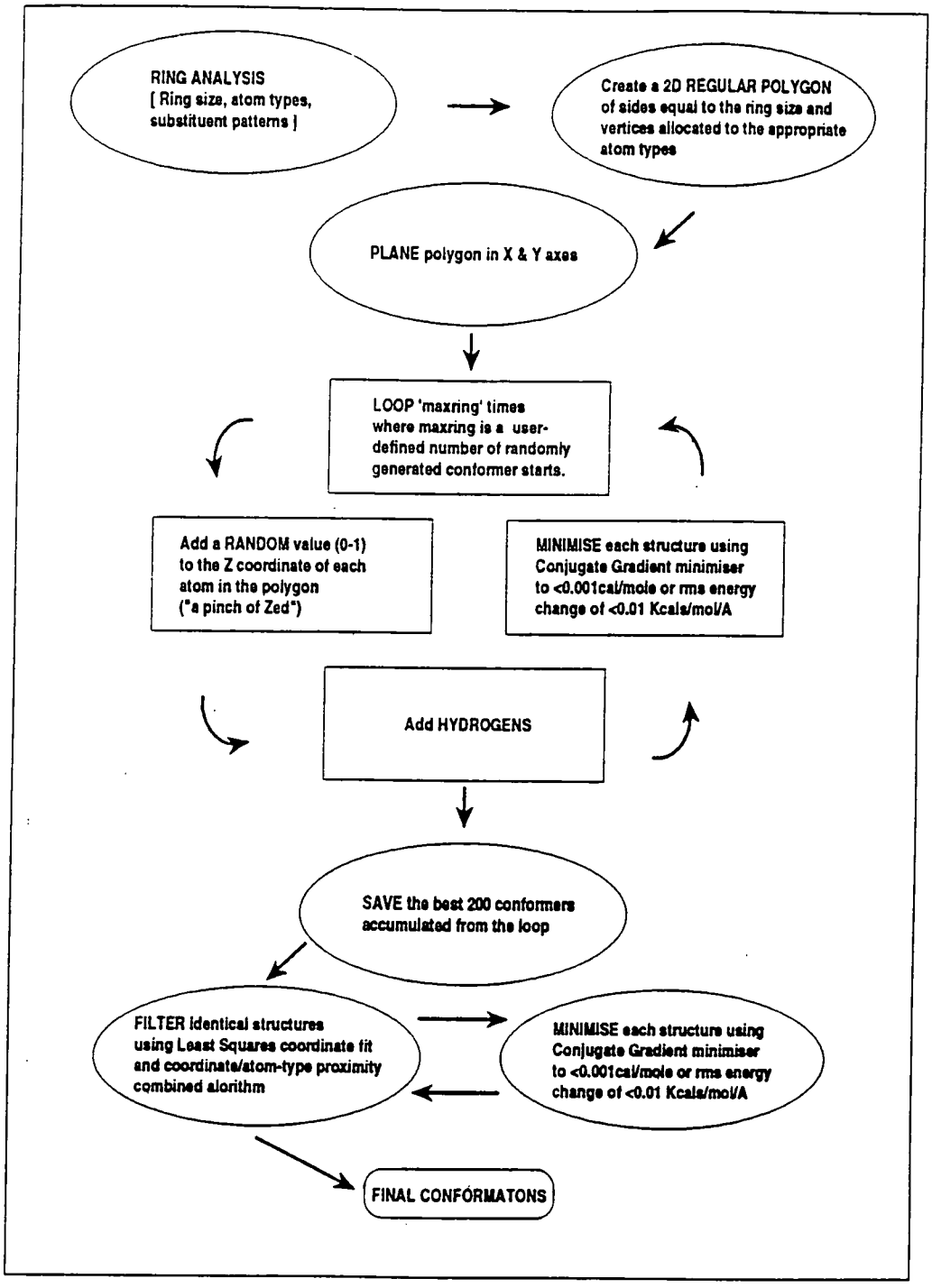


Figure 2.12

Flowchart detailing the operations involved in a COSMIC energy calculation

2.5 References

Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C.F., Mitchell, G.F., Smith, M.J. & Watson, D.G., *J. Chem. Inf. Comput. Sci.*, **1991**, 31, 187-204.

Allen, F.H., Doyle, M.J. & Auf der Heyde, T.P.E., *Acta Cryst.*, **1991**, B47, 412-424.

Allen, F.H., Doyle, M.J. & Taylor, R., *Acta Cryst.*, **1991a**, B47, 29-40.

Allen, F.H., Doyle, M.J. & Taylor, R., *Acta Cryst.*, **1991b**, B47, 41-49.

Allen, F.H., Doyle, M.J. & Taylor, R., *Acta Cryst.*, **1991c**, B47, 50-61.

Allen, F.H., Howard, J.A.K. & Pitchford, N.A., *Acta Cryst.*, **1993**, B49, 910-928.

Allen, F.H. & Taylor, R., *Acta Cryst.*, **1991**, B47, 404-412.

Allinger, N.L., *Pure Appl. Chem.*, **1982**, 54, 2515-2522.

Allinger, N.L. & Sprague, J.T., *J. Am. Chem. Soc.*, **1972**, 94, 5734-5747.

Allinger, N.L., Tribble, M.T. & Miller, M.A., *Tetrahedron*, **1972**, 28, 1173-1190.

Altona, C., Geise, H.J. & Romers, C., *Acta Cryst.*, **1968**, 24, 13-32.

Altona, C. & Sundaralingam, M., *J. Am. Chem. Soc.*, **1972**, 94, 8205-8212.

Anet, F.A.L. & Krane, J., *Tetrahedron Letters*, **1973**, 50, 5029-5032.

Auf der Heyde, T.P.E., *J. Chem. Educ.*, **1990**, 67, 461-469.

Auf der Heyde, T.P.E., Bürgi, H.B., *Inorg. Chem.*, **1989a**, 28, 3960-3969.

Auf der Heyde, T.P.E., Bürgi, H.B., *Inorg. Chem.*, **1989b**, 28, 3970-3981.

Auf der Heyde, T.P.E., Bürgi, H.B., *Inorg. Chem.*, **1989c**, 28, 3982-3989.

Avirah, T.K., Molloy, T.B. & Cook, R.L., *J. Chem. Phys*, **1979**, 71, 2194-2201.

Bocian, D.F., Pickett, H.M., Rounds, T.C. & Strauss, H.L., *J. Am. Chem. Soc.*, **1975**, 97, 687-695.

Bocian, D.F. & Strauss, H.L., *J. Am. Chem. Soc.*, **1977a**, 99, 2876-2882.

Bocian, D.F. & Strauss, H.L., *J. Am. Phys.*, **1977b**, 67, 1071-1081.

Bocian, D.F. & Strauss, H.L., *J. Am. Chem. Soc.*, **1977c**, 99, 2866-2876.

Boessenkool, I.K. & Boeyens, J.C.A., *J. Cryst. Mol. Struct.*, **1980**, 10, 11-18.

Burkert, R. & Allinger, N.L., *Molecular Mechanics*, **1982**, ACS Monograph, No. 148, Washington D.C.: American Chemical Society.

Butcher, S.S., *J. Chem. Phys.*, **1965**, 42, 1833-1836.

Chatfield, C. & Collins, A.J., *Introduction to Multivariate Analysis*, **1980,** London, Chapman and Hall.

Chiang, J.F. & Bauer, S.H., *J. Am. Chem. Soc.*, **1966**, 88, 420-425.

Cremer, D. & Pople, J.A., *J. Am. Chem. Soc.*, **1975**, 97, 1354-1358.

Crews, P., *Chem. Commun.*, **1971**, 583-584.

Eliel, E.L., Allinger, N.L., Morrison, G.A. & Angyal, S.J., *Conformational Analysis*, **1965**, New York, Wiley-Interscience.

Ermer, O. & Lifson, S., *J. Am. Chem. Soc.*, **1973**, 95, 4121-4132.

Evans, D.G. & Boeyens, J.C.A., *Acta Cryst.*, **1988**, B44, 663-671.

Everitt, B., *Cluster Analysis*, **1980**, 2nd ed., New York, Wiley.

Favini, G., Maggi, A. & Todeschini, R., *J. Mol. Struct.*, **1983**, 105, 17-29.

Hagen, K. & Traetteberg, M., *Acta Chem. Scand.*, **1972**, 26, 3643-3648.

Hendrickson, J.B., *J. Am. Chem. Soc.*, **1967**, 89, 7047-7054.

Jarvis, R.A. & Patrick, E.A., *IEEE Trans. Comput.*, **1973**, 22, 1025-1034.

Kilpatrick, J.E., Pitzer, K.S. & Spitzer, R., *J. Am. Chem. Soc.*, **1947**, 69, 2483-2488.

Morley, S.D., Abraham, R.J., Haworth, I.S., Jackson, D.E., Saunders, M.R. & Vinter, J.G., *J. Comput. Aided Mol. Des.*, **1991**, 5, 475-504.

Murray-Rust, P. & Bland, R., *Acta Cryst.*, **1978**, B34, 2527-2533.

Murray-Rust, P. & Raftery, J., *J. Mol. Graphics*, **1985**, 3, 50-59.

Noe, E.A. & Roberts, J.D., *J. Am. Chem. Soc.*, **1971**, 93, 7261-7288.

Norskov-Lauritsen, L. & Bürgi, H.B., *J. Comp. Chem.*, **1985**, 6, 216-228.

Pauncz, R. & Ginsburg, D., *Tetrahedron*, **1960**, 9, 40-52.

Pickett, H.M. & Strauss, H.L., *J. Am. Chem. Soc.*, **1970**, 92, 7281-7288.

Saebo, S. & Boggs, J.E., *J. Mol. Struct.*, **1982**, 87, 365-373.

Schulman, J.M., Disch, R.L. & Sabio, M.L., *J. Am. Chem. Soc.*, **1982**, 104, 3785-3788.

Taylor, R., *J. Mol. Graphics*, **1986**, 4, 123-131.

Taylor, R. & Allen, F.H., in*Structure Correlation*, **1994**, 111-161, ed. Bürgi, H.B. & Dunitz, J., Weinheim, VCH.

Traetteberg, M., J. *Am. Chem. Soc.*, **1964**, 86, 4265-4270.

Vinter, J.G., Davis, A. & Saunders, M.R., *J. Comput. Aided Mol. Des.*, **1987**, 1, 31-50.

Chapter Three

Conformational studies of medium rings

3.1 Introduction

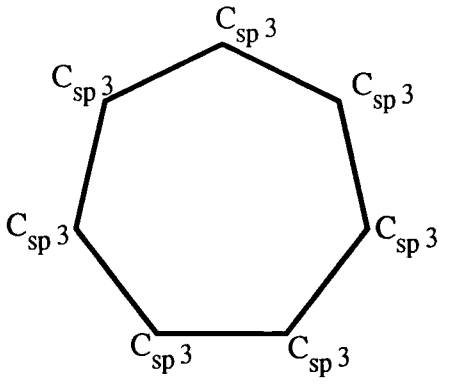
The results to be detailed in this Chapter demonstrate the chemical insights that can be obtained through studies of systematic molecular properties using crystallographic structural data. Here, the principles of structure correlation, as detailed in Chapter 1, are used to analyse the conformational variety, and observed pseudorotation effects for the seven- and eight-membered rings (Allen, Howard & Pitchford, 1993; Allen, Howard, Pitchford & Vinter, 1994; Allen, Garner, Howard & Pitchford, 1994; Allen, Howard & Pitchford, *in press*), using the techniques and approaches detailed in Chapter 2. This work is a continuation of that performed for five- and six-membered rings. (Allen, Doyle & Taylor, 1991a,b,c; Allen, Doyle & Auf der Heyde, 1991).

3.2 Conformational analysis of cycloheptane

3.2.1 Database search and retrieval

The results reported in this Chapter were all gathered using Version 4.5 (1 January, 1992) of the Cambridge Structural Database System, which was used throughout for substructure search, coordinate retrieval, and data analysis using the programs QUEST and GSTAT (Allen, Davis *et al.*, 1991).

The chemical substructure studied here was defined to consist of a ring of seven, sp^3 hybridised carbon atoms, connected by single bonds. The fragment is shown in Figure 3.1 below. The search was further constrained, using the CSD bit-screen search mechanisms (CSD User Manuals, 1992), to locate only entries for which (a) atomic coordinates are available, (b) there is no reported disorder in the structure, (c) there are



(a)

(b)

```
FRAG
AT1  C  2
AT2  C  2
AT3  C  2
AT4  C  2
AT5  C  2
AT6  C  2
AT7  C  2
BO  1  2
BO  2  3
BO  3  4
BO  4  5
BO  5  6
BO  6  7
BO  1  7
END
EXPAND INV
APERM  1 2 3 4 5 6 7
APERM  2 3 4 5 6 7 1
APERM  3 4 5 6 7 1 2
APERM  4 5 6 7 1 2 3
APERM  5 6 7 1 2 3 4
APERM  6 7 1 2 3 4 5
APERM  7 1 2 3 4 5 6
APERM  7 6 5 4 3 2 1
APERM  6 5 4 3 2 1 7
APERM  5 4 3 2 1 7 6
APERM  4 3 2 1 7 6 5
APERM  3 2 1 7 6 5 4
APERM  2 1 7 6 5 4 3
APERM  1 7 6 5 4 3 2
DEF %T1  1  2  3  4
DEF %T2  2  3  4  5
DEF %T3  3  4  5  6
DEF %T4  4  5  6  7
DEF %T5  5  6  7  1
DEF %T6  6  7  1  2
DEF %T7  7  1  2  3
FAC 4
DEF *CREM Q2 P2 Q3 P3 7
TRA #PP=Q3/Q2
TRA THETA=ATAN #PP
```

```
SEL THETA 30.0 90.0
BOX P2 P3 192.86 64.29 231.43 77.14 257.14 205.71 218.57 192.86
SCAT P2  P3
SCAT P2  THETA
SCAT P3  THETA
```

Figure 3.1

(a) the cycloheptane search fragment and (b) an associated GSTAT instruction set which defines the cycloheptane fragment, applies the required symmetry expansion, performs PCA on the seven intra-annular torsion angles of the cycloheptane fragment, then generates CP puckering parameters. Results of the CP analysis are reported for θ *= 30 - 90°, and reported for one asymmetric unit of conformational space defined by the four pairs of* ϕ_2, ϕ_3 *values in the BOX command*

no residual numerical errors after all of the CSD check procedures, (d) the structure solution reported a crystallographic R-factor ≤ 0.10 (10%), and finally, (e) that the compounds are "organic" by CSD definitions. Preliminary searches using all of these constraints showed that this procedure also allows entries with very complex ring systems to be hit. These were removed using the CSD bit-screen setting -620, which ignores those entries which fail the CSD ring-identification procedures (less than 3% of the entire database), due to major complexities in the chemical ring assemblies.

Searches of the CSD showed the necessity to further subdivide the resulting list of entries: a large proportion of the hits found by the search query outlined above contained bridged rings. Entries containing bridged rings can be avoided using the SCREEN -622 command, and conversely those entries specifically containing bridges, can be searched for exclusively with SCREEN 622. In such a way two datasets were extracted from the CSD: 7C1 containing non-bridged rings, and 7C2 with only bridged rings. However, this bit-screen is a non-specific mechanism and it does not indicate which of the rings in a given molecule is actually bridged. Upon close inspection of the chemical diagrams for the subset 7C2 two molecules were found which did not have bridging in the seven-membered ring itself, and so these entries were relocated into the 7C1 dataset. As a result, dataset 7C1 comprised 86 CSD entries containing 101 unbridged cycloheptane fragments, and 7C2 is comprised of 276 entries containing 310 bridged cycloheptane fragments. The full list of CSD refcodes for these datasets is given in Appendix 1.

3.2.2 Data analysis

Analysis of the conformational features of cycloheptane in these separate datasets was carried out using the GSTAT program. The cycloheptane fragments of 7C1 and 7C2 are relocated in the crystallographic connection tables by this program using bond-length constraints (1.44Å - 1.70Å) to approximate a C_{sp3} - C_{sp3} single bond length. It was further required that no atomic overlap should exist between multiple occurrences of the cycloheptane fragment in the same molecule. The conformation of each fragment which passed these criteria was described in terms of its seven intra-annular torsion angles, τ_1 - τ_7 (Table 2.2).

The fragment location procedures of GSTAT place each fragment in some arbitrary asymmetric unit of conformational space, and so it is necessary to fill this space according to the D_{7h} topological symmetry of the parent planar search fragment, together with the enantiomorphic conformations in three dimensions. This procedure was described fully in Chapter 2. For cycloheptane the total symmetry-expansion is 28-fold, and the initial datasets of 101 (7C1) and 310 (7C2) fragments, are therefore expanded up to 2828 (7C1) and 8680 (7C2) fragments using the EXPAND feature of GSTAT.

The CP parameters (q_2, ϕ_2 and q_3, ϕ_3: Section 2.2.2) are generated by GSTAT, for each fragment in the expanded dataset, and the BPRS coordinates (ρ, θ: Section 2.2.2) are derived using the TRAnsform facility. A typical GSTAT instruction set for those operations is shown in Figure 3.1(b). Classification of the conformations was carried out using the symmetry-modified cluster-analysis algorithms described in Section 2.3.

In many of the examples that follow, it has also proved useful to display and examine the conformational relationships that exist between fragments, that fall within a single asymmetric unit of the four-dimensional space. This unit can be isolated out of the expanded set of data by specifying appropriate limiting values for the pertinent conformational descriptors (*e.g.* values of CP or BPRS coordinates) as described in Chapter 2.

3.2.3 Conformational mapping

The results from each of the separate procedures used for the conformational mapping of the two datasets, 7C1 and 7C2, will now be presented, and the correlations between each of them determined.

BPRS coordinate plots

The conformational diversity of datasets 7C1 and 7C2 was initially explored using plots derived directly from the BPRS (ρ, θ, ϕ_2, ϕ_3) coordinates, where these are related to the degree of ring pucker by the equation (Chapter 2, **[3]**):

$$z_j = \rho\left[\cos\theta\cos\left(2\pi 2j/7 + \phi_2\right) + \sin\theta\cos\left(2\pi 3j/7 + \phi_3\right)\right] \qquad \textbf{[1]}$$

A great deal of information can be extracted from the resulting plots shown in Figures 3.2 and 3.3. The symmetry-expanded ϕ_2-ϕ_3 map shown in Figure 3.2(a) is of the form shown earlier in Figure 2.3, with the C/TC pseudorotation pathway being illustrated by the major continuous feature in Figure 3.2(a). The density along this pathway rises to maxima at the positions which correspond to those occupied by the TC conformers in Figure 2.3. The ϕ_2-θ map shown in Figure 3.2(b) shows that there is a clear separation in θ between the two itineraries: B/TB for θ in the range 0-30°, and C/TC for θ values between 30° and 90°. This is very much in accordance with the results of Bocian *et al.* (1975), which place the C/TC at θ = 50-54°, and the B/TB pathway at θ = 0°. The limiting values of θ observed for the 7C1 dataset can be used to segment Figure 3.2(a) into two distinct components: Figure 3.2(c) showing the C/TC conformers only (θ =30-90°), and Figure 3.2(d) showing the B/TB conformers (θ =0-30°). The C/TC pseudorotation pathway is now clearly depicted.

In Figure 3.3 the scope of the mapping has been restricted to that of a single asymmetric unit, defined to lie within ϕ_2, ϕ_3 limits (in units of π/14) of 15, 5; 18, 6; 20, 16; 17, 15. This "box" is shown in Figure 2.3. The GSTAT program was adapted so as to generate scatterplots within non-orthogonal axial limits. Thus Figures 3.3(a), (c), and (e) are respectively ϕ_2-ϕ_3 and ϕ_2-θ plots and the θ histogram for the asymmetric unit of dataset 7C1, with the comparable plots for the dataset 7C2 being shown in

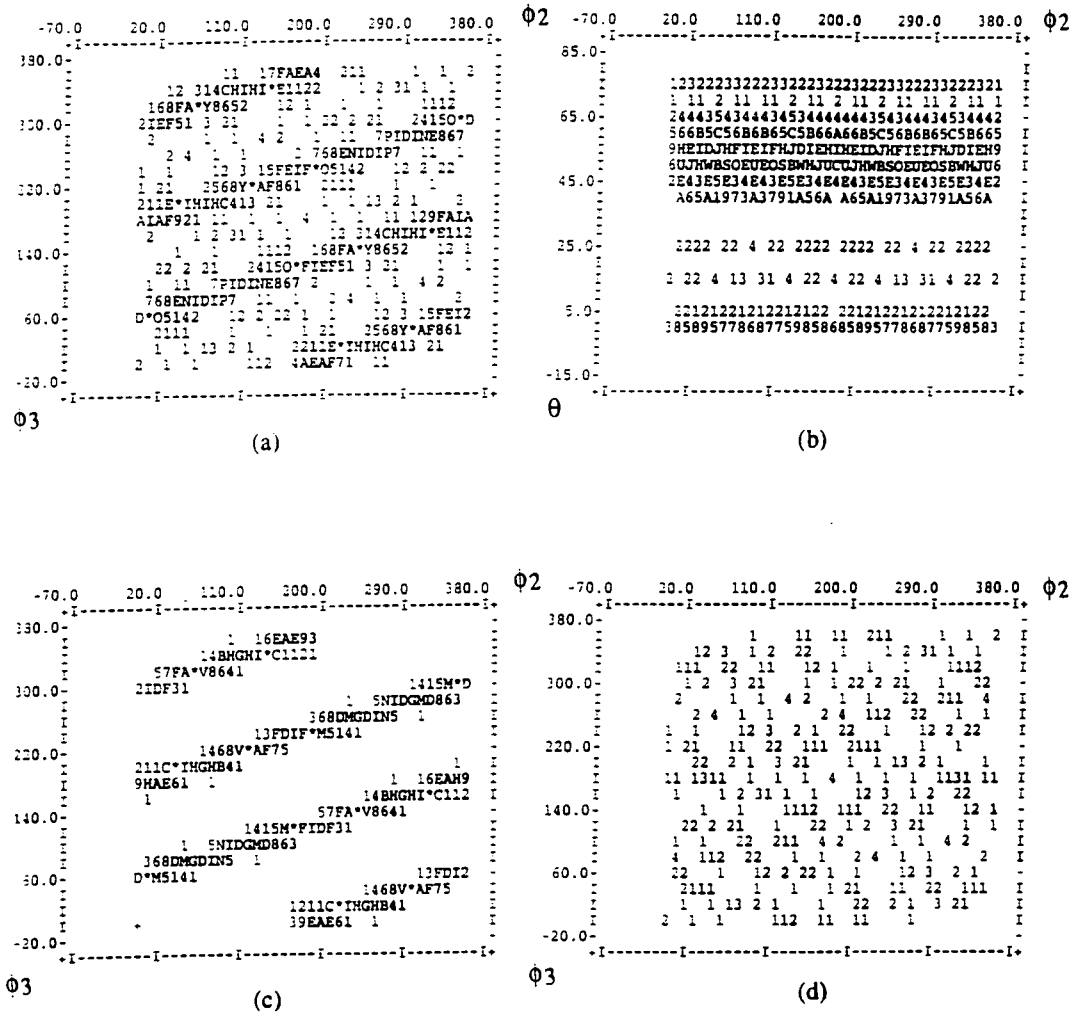


Figure 3.2

Conformational plots from symmetry-expanded BPRS coordinates for dataset 7C1. The ϕ_2-ϕ_3 plot (a) and the ϕ_2-θ plot (b) include all data points. The ϕ_2-ϕ_3 plot (c) covers rings having $30 \leq \theta \leq 90°$, i.e. chairs and twist-chairs, whilst the ϕ_2-ϕ_3 plot (d) covers rings with $0 \leq \theta \leq 30°$, i.e. boats and twist-boats.

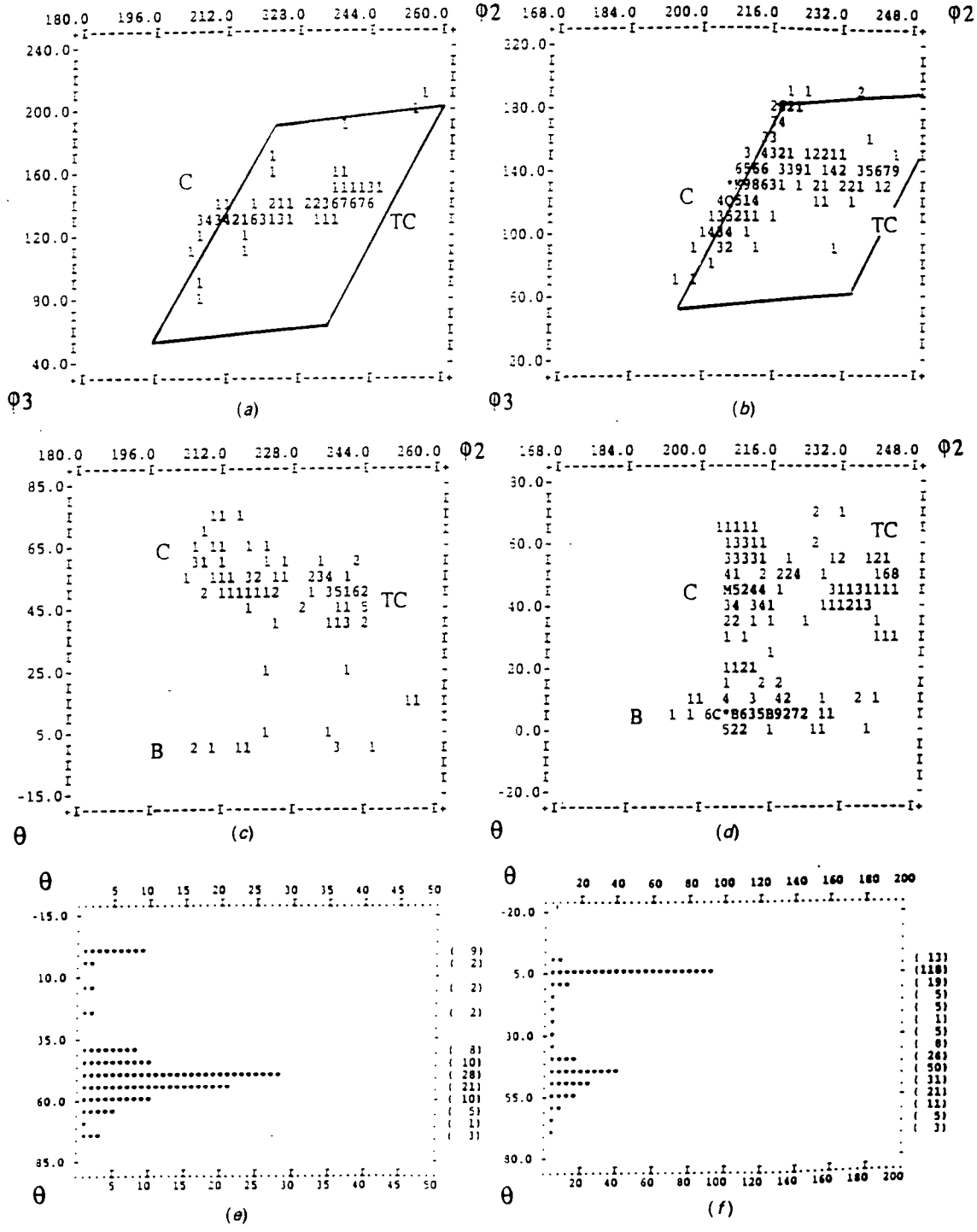


Figure 3.3

Conformational plots (BPRS coordinates) of the asymmetric unit of conformational space for cycloheptane. (a), (c) and (e) for dataset 7C1; (b), (d) and (f) for dataset 7C2.

Figures 3.3(b), (d), and (f). Perhaps the most informative results arise from the ϕ_2-θ plots (Figures 3.3c,d). For the 7C1 dataset (Figure 3.3c), this plot shows a marked concentration of density at the position corresponding to the TC form ($\theta \simeq 50°$, $\phi_2 \simeq 240°$), and a continuation of this density through less prevalent intermediate forms, until the density corresponding to the C conformers is reached ($\theta \simeq 60°$, $\phi_2 \simeq 212°$). There is a small break in the otherwise continuous transformation from the TC into the C forms, and that can be seen to occur at $\phi_2 \simeq 228\text{-}234°$. This should correspond to the highest energy point in the C/TC pseudorotation itinerary. Very few fragments (15 out of 101 = 14.9%) do not fall on the C/TC pathway of Fig. 3.3(c), and only 11 of these (10.9%) can be truly classed as being B/TB conformers ($\theta = 0\text{-}10°$), as borne out by the histogram of Fig. 3.3(e).

For the bridged rings of dataset 7C2 the situation is clearly different. In Figure 3.3(d) for example, the ϕ_2-θ plot is somewhat different to that observed for 7C1. The higher energy C form is now the most dominant conformation present in the C/TC pathway, but with a reduced θ value ($\simeq 45\text{-}50°$) than was found for the comparable C forms in the unbridged rings of dataset 7C1. The TC conformers of 7C2 are still observed to occur at the expected θ value of approximately 50°. Another prominent feature of the 7C2 plots is the emergence of the B/TB conformers, with a far greater proportion of the bridged conformers lying on this pathway than was observed for the unbridged seven-membered rings. 149 of the 310 fragments (48.1%) have θ values in the range 0-10°. The plots of Figure 3.3 proved particularly useful in assessing the cluster analysis results described later.

Cremer-Pople (CP) coordinate plots

The two CP phase-amplitude pairs (q_2, ϕ_2) and (q_3, ϕ_3) describe the two circular pseudorotation pathways. These coordinates can be re-expressed in Cartesian form as CP1 = $q_3 \sin \phi_3$ (the set of points along the CP1 axis mapping a family of perfect TC conformers), CP2 = $q_3 \cos \phi_3$ (mapping a family of perfect chairs), CP3 = $q_2 \sin \phi_2$ (mapping a family of perfect TB conformers), and CP4 = $q_2 \cos \phi_2$ (mapping the

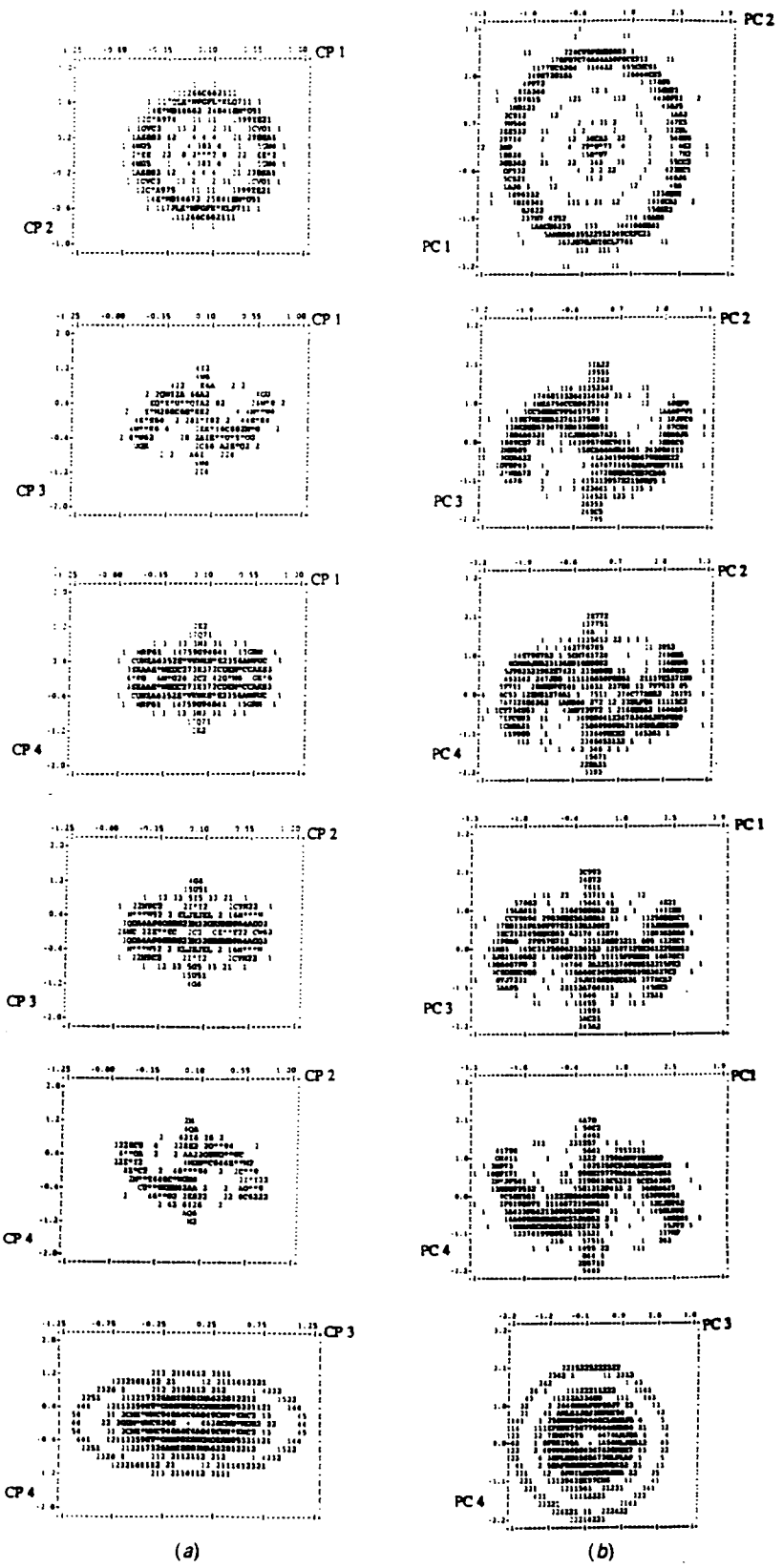


Figure 3.4

Comparative 2D plots of (a) CP coordinates, and (b) PC coordinates (or "scores")

perfect boats). The six 2-D scatterplots that can be generated by these four parameters are shown in Figure 3.4, for the symmetry -expanded 7C1 dataset.

When viewing the plots in Figure 3.4, it is necessary to recall that each of the axes represented is mutually orthogonal to the other three. The true four-dimensional nature of the conformational space of cycloheptane can thus begin to be discerned. The CP1-CP2 plot (top) can be obtained by projection along either of CP3 or CP4. The C/TC pathway forms the dense outer circle with the B/TB examples clustered in the centre of the plot. Two fragments with θ in the range 10-30°, form an intermediate circle between these two extremes. In terms of the toroidal model introduced in Chapter 2, the CP1-CP2 plot represents a view along the axis of a cylinder, obtained by cutting and straightening the ϕ_2 torus of Figure 2.2(b): the ϕ_3-dependent C/TC helical pathway now appears as a circle. This situation is reversed in the CP3-CP4 plot (Fig. 3.4 bottom), which is obtained by projection along either CP1 or CP2. The few B/TB conformers present in this dataset manifest themselves as a narrow band of density surrounding the dense central region defined by the C/TC conformers. The B/TB examples do not define a perfect circle because of a "tailing effect" caused by angles close to 0°, as discussed by Bocian *et al.* (1975) and by Bocian & Strauss (1977).

The four remaining plots of Figure 3.4(a) show different side views of helical tracks on both the ϕ_2-torus and the ϕ_3-torus; two side views of each torus, with each view of each pair being at 90° to the other. In each plot, the B/TB pathway is represented by a vertical band of density, which is crossed by orthogonal views of the C/TC helix.

Symmetry-modified principal-component analysis (PCA)

PCA was performed on the seven torsion angles of the symmetry-expanded data matrices for 7C1 and 7C2, with the results shown in Table 3.1. It can be seen from the table that the variance in each seven-dimensional dataset can be fully accounted for by the four PC's. Furthermore the 4 PC's occur as two degenerate pairs, as predicted by the group-theoretical analysis of Bocian *et al.* (1975).

Table 3.1

Results of the PCA on datasets 7C1 and 7C2

Variance analysis

Dataset	No. of fragments	Variance(%)				
		PC1	PC2	PC3	PC4	Total
7C1	2828	40.08	40.08	9.88	9.88	99.92
7C2	8680	26.94	26.94	22.86	22.86	99.60

PC loadings

	Dataset 7C1				Dataset 7C2			
	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
τ_1	55.4	11.1	3.8	27.8	-9.2	-51.5	4.7	-47.9
τ_2	-54.7	14.0	26.2	-9.8	-14.0	50.4	-47.8	6.1
τ_3	43.2	-36.4	-15.4	-23.4	34.5	-39.3	16.6	45.2
τ_4	-23.2	51.5	-19.4	20.3	-48.1	20.4	40.4	-26.2
τ_5	-1.5	-56.5	24.1	14.4	52.2	2.5	-34.5	-33.5
τ_6	25.8	50.2	8.7	-26.7	-46.0	-24.9	-25.0	41.1
τ_7	-45.1	-34.0	-27.9	-2.5	30.6	42.4	45.7	15.2
Sym.	C	TC	B/TB	B/TB	TC	C	B/TB	B/TB

The PC loadings for 7C1 show that values for PC1 have approximate C_s symmetry, leading to the interpretation that this axis is closely associated with the chair conformers. PC2 has approximate C_2 symmetry and can be considered to be associated with a TC conformer at 90° in ϕ_3 to the chair mapped by PC1. The loadings of both PC3 and PC4 are essentially asymmetric, and can be taken to represent intermediate conformations on the B/TB pathway, separated by 90° in ϕ_2. The PCA data would tend to suggest that 19.8% of the total variance is related to conformations on the B/TB pathway, which compares with the 15.9% that can be attributed to this itinerary from the BPRS plots of Figure 3.2. However, C/TC conformations do have some degree of B/TB character (and *vice versa*) and this factor is not taken into account by the BPRS plots, which are based on θ alone.

PCA of the symmetry-expanded dataset 7C2, again shows that just four PC's are required to describe the total variance. Here however, the distribution in the variance and the PC loadings are somewhat different. The PC's still occur in pairs, with PC1 and PC2 corresponding to the C/TC conformations, but PC2 now maps a chair conformation, and PC1 the TC; PC3 and PC4 are seen to once again describe the B/TB intermediates on the pseudorotation pathway. Most noticeable, however, is the fact that 45.7% of the total variance is now accounted for by PC3 and PC4 and hence by the B/TB conformers. This value is again very close to that found using the BPRS coordinate approach (48.1%). Thus it can be seen that bridging of the cycloheptane ring increases the proportion of B/TB conformers very significantly.

Comparison of Cremer-Pople (CP) and principal-component (PC) mappings

The PC scatterplots derived for the 7C1 symmetry-expanded dataset, are shown opposite the CP scatterplots in Figure 3.4. The axial assignments in Fig. 3.4(b) are chosen so as to align the chair, twist-chair, *etc.* conformations most closely with their CP equivalents. Interchange of the PC1/PC2 and PC3/PC4 labels is permissible due to the degeneracy of each pair. There is an obvious functional equivalence between the four-dimensional PC description, and the 4-dimensional CP and BPRS descriptions. Similar equivalence's were observed for the two- and three-dimensional conformational spaces of the five- and six-membered carbocycles (Allen, Doyle & Auf der Heyde, 1991), and arise from the close mathematical analogies between PCA and normal coordinate analysis.

The PC plots of Fig. 3.4(b) are slightly rotated with respect to the CP plots in Fig. 3.4(a). The pairwise degeneracy of PC1/PC2 and PC3/PC4, allows their respective eigenvectors to adopt any orientation within the plane that they define, and therefore the axial directions obtained in a given PCA are essentially arbitrary, and depend upon round-off and similar numerical effects. This is, of course, an explanation of the association of PC1 with chair conformations in dataset 7C1, whereas the association belongs to the degenerate PC2 in dataset 7C2.

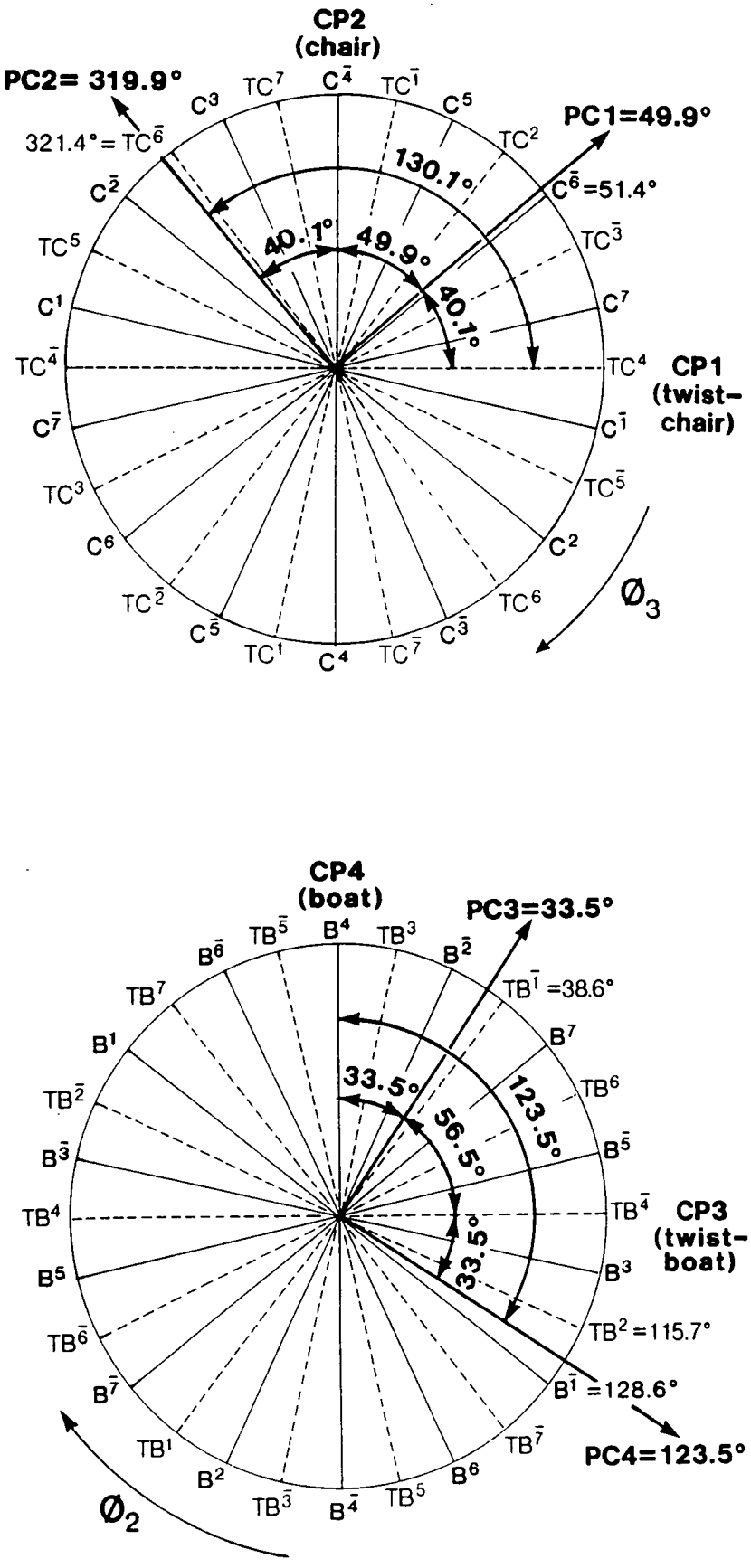


Figure 3.5

Correlation of the PC axes with the CP axes for (a) the chair - twist-chair pseudorotation pathway, and (b) the boat - twist-boat pseudorotation pathway

The degree of rotation of the PC axes with respect to the CP axes, can be deduced from the correlation coefficients, R(CP, PC), generated by GSTAT. These R values are given in Table 3.2 for 7C1 (GSTAT limits prevent calculation of values for the 8680 symmetry-expanded fragments of 7C2), together with the inverse cosines, $\cos^{-1}$[R(CP, PC)], which indicate the relative orientations of both sets of axes to each other. The correlations of the form $R(CP_m, PC_n)$ for m=1, 2 and n=3,4 (and *vice versa*) are all zero, as expected, indicating the orthogonality of the axial pairs. However, it can be seen from Figure 3.5 that PC1 is rotated by 1.52° from a CP chair axis at $\phi_3 = 2\pi/7$ (51.4°), and PC2 is rotated by an identical amount from the orthogonal twist-chair axis at $\phi_3 = 25\pi/14$ (321.4°). For the B/TB ϕ_2 pathway of Fig. 3.5(b), the PC3 axis is rotated by 5.12° from a CP twist-boat axis, with PC4 similarly rotated by 5.12° from an orthogonal boat axis. It is these arbitrary axial rotations that explain the asymmetry of the PC loadings of Table 3.1, and of the minor discrepancies found in Figure 3.4.

Rotation of the PC1, PC2 axes by 1.52°, or the PC3, PC4 axes by 5.12°, would therefore bring the PC axes into direct alignment with one of the directions of cokernel symmetry for the D_{7h} point group (Murray-Rust, Bürgi & Dunitz, 1979).*i.e.* into alignment with the C_2 or C_s symmetry groups, that describe the chair and twist-chair

Table 3.2

Correlation coefficients for the PC and CP axial representations of dataset 7C1

PC_n		CP_m			
n	function	m=1	2	3	4
1	R	0.765	0.644	0.0	0.0
	$\cos^{-1}$ R (°)	40.1	49.9	90.0	90.0
2	R	-0.644	0.765	0.0	0.0
	$\cos^{-1}$ R (°)	130.1	40.1	90.0	90.0
3	R	0.0	0.0	0.552	0.834
	$\cos^{-1}$ R (°)	90.0	90.0	56.5	33.5
4	R	0.0	0.0	0.834	-0.552
	$\cos^{-1}$ R (°)	90.0	90.0	33.5	123.5

conformations. There are obviously 14 such directions, each separated by rotations of $2\pi/14$ radians in Figure 3.5. Rotation of PC1, PC2 by (1.52 + 360n/14)°, and PC3, PC4 by (5.12 + 360n/14)°, will bring these axes into direct coincidence with the CP axes. In such an event the correlation coefficients R(CP1, PC2) and R(CP3, PC3) will both be 1.0, and the R(CP1, PC1) and R(CP3, PC4) will both be exactly 0.0. The rotation of the PC eigenvectors to any of the cokernel directions will also generate PC loadings (eigenvalues) that have perfect C_2 and C_s symmetry.

Three-dimensional scatterplots of principal-component scores

The 2-dimensional scatterplots shown in Figure 3.4 show how the PC axes are related in the 4-dimensional conformational space of cycloheptane, but to gain an even clearer insight into the nature of these relationships, it proved worthwhile to produce three-dimensional models of the PC's. Two fragments were isolated; one an almost perfect chair, and the other an almost perfect boat. When the symmetry of these fragments is taken into consideration fourteen doubly degenerate representations are generated. The PC scores for these equivalents can be regarded as coordinates and input into a suitable molecular modelling program. For the analyses performed here the plotting program MacMoMo (Dobler, 1990) was chosen, for use on an Apple Macintosh computer. Once the PC coordinates have been input, a pseudomolecule can be created by connecting the adjacent equivalents for the two conformers, so that various views of the pseudorotation pathway can be generated. The PC axes in this analysis of 56 fragments (28 isomers for each of the two rings) were very closely aligned with the cokernel directions, so that further PC-axis rotation was considered unnecessary.

In Figure 3.6(a) the scores for PC1, PC2 and PC3 are used as pseudocoordinates. The first five frames show how the PC2, PC1 plot of Fig. 3.4(b) transforms into the PC2, PC3 plot of Fig. 3.4(b). The chair pseudomolecule forms the outer 'circle' in the 0, 0, 0 position: a top view of the ϕ_3 helix of Figure 2.2(a). The boat pseudomolecule forms the inner 'circle' at 0, 0, 0: a top view of the ϕ_2 helix of Figure 2.2(b). Rotation through 90° about the PC2 axis brings the original projection axis,

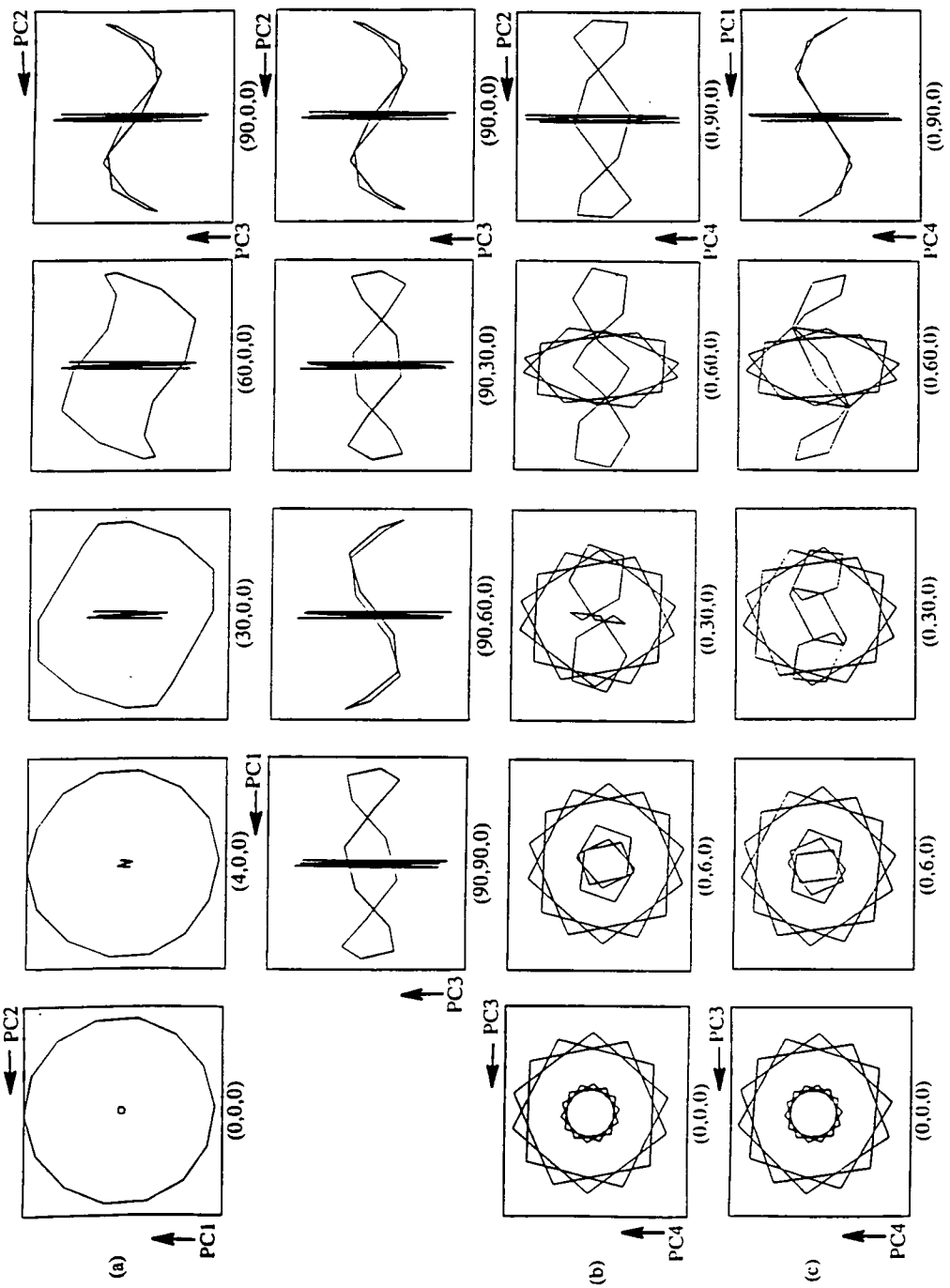


Figure 3.6

Views of the chair - twist-chair and boat - twist-boat helical pseudorotation pathways in the four-dimensional principal component (PC) space

PC3, into the viewing plane to generate side views of both helices in frame five of Fig. 3.6(a). The PC2 rotations of frames 2, 3, 4 show how the B/TB helix in ϕ_2 rapidly elongates to an apparently vertical straight line as a result of the minimal dependency of this helix on ϕ_3 variations. Frames 6-9 of Fig. 3.6(a) show (right to left) the transformation of the PC2, PC3 plot to the PC1, PC3 plot *via* a 90° rotation about the PC3 axis. This is equivalent to a rotation of the edge-on views of the ϕ_3 (C/TC) and the ϕ_2 (B/TB) helices, which are mutually perpendicular in this four-dimensional hyperspace.

Figures 3.6(b) and (c) use the PC scores for PC2, PC3, PC4 and PC1, PC3, PC4, respectively. Both begin at (0, 0, 0) with identical projections onto the PC3, PC4 plane along the mutually orthogonal PC2 (Fig. 3.6b) and the PC1 (Fig. 3.6c) axes. The boat pseudomolecule forms the outer helical projection, and the chair pseudomolecule forms the inner helical projection, in these first frames. A rotation of 90° about the PC4 axis in Fig. 3.6(b) then generates the PC2, PC4 plot (comparable with Fig. 3.4b) and a rotation of 90° about PC4 in Fig. 3.6(c) generates the PC1, PC4 plot (which is again comparable with Fig. 3.4b). Thus, Figure 3.6 provides a graphic illustration of the inter-relationships between the four mutually orthogonal axes which describe the conformational hypersurface of the cycloheptane fragment.

3.2.4 Classification of Conformations

The conformational maps generated for the single asymmetric unit in Figure 3.3, showed quite clearly that the chair and twist-chair conformers are the major constituents of dataset 7C1: the unbridged cycloheptanes. In dataset 7C2 the boat and twist-boat conformations are more evident, together with indications of some splitting of both the C and TC peaks into subgroups. Thus, we would expect to observe at least three conformational subgroups for dataset 7C1, and probably more than three in dataset 7C2, using the cluster-analysis algorithms incorporated into the program GSTAT.

Of the clustering algorithms available, the symmetry-modified Jarvis-Patrick

Table 3.3

Results of the Jarvis-Patrick symmetry-modified clustering analysis of dataset 7C1

				n=1		n=2	
run	K_{NN}	K_{JP}	K_{JP} / K_{NN}	N_c	ΣN_p(%)	N_c	ΣN_p(%)
1	11	7	0.636	6	82 (81.2)	6	82 (81.2)
2	11	6	0.546	4	89 (88.1)	4	87 (86.1)
3	11	5	0.455	2	91 (90.1)	3	56 (55.4)
4	10	7	0.700	9	57 (56.4)	8	56 (55.4)
5	10	6	0.600	6	84 (83.2)	7	83 (82.2)
6	10	5	0.500	5	90 (89.1)	4	88 (87.1)
7	10	4	0.400	3	95 (94.1)	3	94 (93.1)
8	9	6	0.666	8	65 (64.4)	8	64 (63.4)
9	9	5	0.555	6	83 (82.2)	6	82 (81.2)
10	9	4	0.444	3	91 (90.1)	3	92 (91.1)
11	8	4	0.500	6	85 (84.2)	5	82 (81.2)
12	8	3	0.375	4	91 (90.1)	3	92 (91.1)
13	7	4	0.571	9	68. (67.3)	9	75 (74.3)
14	7	3	0.428	4	83 (82.1)	8	92 (91.1)
15	7	2	0.286	4	92 (91.1)	4	93 (92.1)

algorithm (Jarvis & Patrick, 1973; Allen, Doyle & Taylor, 1991b) has been shown to have the most success in clustering trial datasets of known conformational complexity (Allen, Doyle & Taylor, 1991c). This procedure was described fully in Chapter 2.

For cycloheptane, the Jarvis-Patrick clustering was performed with a variety of values for K_{JP} and K_{NN} using the normal values (1, 2) of the power factor n, and with D_{lim} = 0.06. The relevant results are gathered in Table 3.3 and show that high values of the K_{JP} / K_{NN} ratio result in a large number of clusters (N_c) of low population, whilst lowering this ratio permits the coalescence of these original clusters into larger units. For dataset 7C1, the results for n=1 and n=2 are quite similar, but N_c values are somewhat higher for n=1 at a given value of K_{JP} / K_{NN}. The shaded areas in Table 3.3 indicate those clustering results that are deemed to be chemically sensible.

The selection of an optimum clustering structure was guided by *a priori* chemical knowledge, and by examination of the various statistical descriptors (Allen & Johnson, 1991) that summarise the distributions of τ_1-τ_7 in each cluster. The results for K_{NN} = 10, K_{JP} = 5, n = 2, and D_{lim} = 0.06 were selected as the optimum parameters for

Table 3.4

Optimum results for the Jarvis-Patrick clustering of datasets 7C1 and 7C2

(a) Dataset 7C1

N_c	1*	2*	3	4
N_p	76	60	10	10
τ_1	-84.8 (1.0)	-62.9 (1.4)	-81.7 (0.8)	-73.6 (1.0)
τ_2	71.2 (0.6)	80.9 (1.1)	57.7 (1.0)	-2.8 (1.9)
τ_3	-54.2 (1.0)	-63.8 (1.1)	-37.3 (1.2)	66.7 (1.0)
τ_4	71.2 (0.6)	63.8 (1.1)	60.8 (1.6)	-23.3 (1.7)
τ_5	-84.8 (1.0)	-80.9 (1.1)	-92.5 (1.5)	-62.8 (1.6)
τ_6	37.3 (1.5)	62.9 (1.4)	46.6 (1.0)	50.8 (1.4)
τ_7	37.3 (1.5)	0.0	36.1 (1.2)	40.3 (1.4)
$\bar{R}_{min}$	0.988	0.988	0.996	0.994
$\bar{R}_{max}$	0.988	0.995	0.999	0.999
D_{max}	31.1	30.7	13.7	16.2
Conformer	TC	C	TC^d	B^d

(b) Dataset 7C2

N_c	1	2	3	4	5
N_p	116	90	39	50	30
τ_1	0.0	0.0	-14.9 (1.2)	0.0	-44.4 (0.8)
τ_2	-87.9 (0.3)	-86.9 (1.6)	-74.7 (1.0)	-103.9(0.1)	-44.4 (0.8)
τ_3	53.1 (0.5)	104.6 (1.0)	56.8 (0.5)	127.5 (0.2)	124.6 (1.0)
τ_4	42.5 (0.9)	-59.6 (1.9)	38.0 (0.6)	-52.5 (0.2)	-88.1 (1.2)
τ_5	-42.5 (0.9)	59.6 (1.9)	-40.5 (0.6)	52.5 (0.2)	57.4 (1.6)
τ_6	-53.1 (0.5)	-104.6 (1.0)	-51.1 (0.6)	-127.5 (0.2)	-88.1 (1.2)
τ_7	87.9 (0.3)	86.9 (1.6)	100.1 (1.0)	103.9(0.1)	124.6 (1.0)
$\bar{R}_{min}$	0.993	0.969	0.992	1.000	0.994
$\bar{R}_{max}$	0.999	0.993	0.998	1.000	0.999
D_{max}	35.2	53.6	24.9	3.4	25.1
Conformer	B	C+	B^{+d}	C^{++}	TC^+

N_c	6	7	8	9	10
N_p	30	30	14	9	7
τ_1	-34.8 (0.7)	0.0	-33.5 (1.7)	-13.2 (0.3)	-31.9 (2.5)
τ_2	-60.6 (0.7)	-62.7 (1.8)	-48.3 (1.4)	-90.7 (0.4)	-59.2 (2.6)
τ_3	53.5 (0.7)	79.8 (1.2)	85.0 (2.3)	88.2 (2.2)	121.6 (1.4)
τ_4	53.5 (0.7)	-65.4 (1.0)	-62.2 (1.9)	33.6 (2.1)	-69.8 (0.8)
τ_5	-60.6 (0.7)	65.4 (1.0)	57.4 (1.3)	-48.8 (1.2)	42.6 (1.9)
τ_6	-34.8 (0.7)	-79.8 (1.2)	-81.4 (2.1)	-46.3 (1.3)	-80.6 (2.4)
τ_7	115.0 (0.5)	62.7 (1.8)	88.9 (2.0)	119.4 (0.4)	115.8 (1.9)
$\bar{R}_{min}$	0.999	0.989	0.989	0.993	0.993
$\bar{R}_{max}$	0.999	0.998	0.996	1.000	0.999
D_{max}	14.3	25.7	27.7	19.1	20.1
Conformer	TB	C	TC	B^{++d}	TC^{++d}

In Table 3.4(a) and (b): N_c is the cluster number, N_p is the cluster population, $\bar{\tau}_i$ (i = 1-7) are the (circular) mean torsion angles (°) with their circular e.s.d's in parentheses, $\bar{R}_{min}$ and $\bar{R}_{max}$ are the upper and lower limits of the circular concentration values for the individual $\bar{\tau}_i$ distributions, and D_{max} is the maximum intracluster dissimilarity (deg^2) of any fragment from the cluster centroid defined by the $\bar{\tau}_i$. Symmetry-coalesced clusters are marked (*). Full details of circular statistical descriptors are to be found in Allen & Johnson (1991). In the description of conformation the superscripted + or ++ indicates increased puckering from the minimum-energy form, and superscript d *indicates an asymmetric distortion.*

dataset 7C1. The mean torsion angles and statistical descriptors for the four major clusters generated by this clustering run, are given in Table 3.4(a).

From the results generated by the "optimum" clustering run it can be seen that of the 101 fragments in the asymmetric unit of 7C1, 38 can be classified as twist-chairs (cluster 1), 30 as being chairs (cluster 2), ten distorted chairs (3), and ten distorted boats (4). However, the centroids of clusters 1 and 2 are sufficiently close to (respectively) a two-fold axis, and a mirror plane in conformational space, so that these clusters coalesce with their symmetry-related counterparts using the limiting distance criterion

(D_{max} = 1.0) described by Allen & Taylor (1991). The results shown in Table 3.4(a) relate to the coalesced clusters of 76 and 60 fragments, respectively, for which symmetrized torsion-angle means can be reported. The twist-chair and boat conformations of clusters 3 and 4 are sufficiently distorted to prevent coalescence with their symmetry-related counterparts. Of the remaining 13 fragments, only one adopts a true twist-boat conformation, whilst two fragments having 1, 3, 5-fusion to three-membered rings adopt a conformation with zero torsion angles about the three fusion bonds (these rings are related to the 1, 3, 5-cycloheptatrienes discussed in Section 3.4). The remaining ten fragments adopt conformations that are heavily distorted variants of the C/TC forms described in Table 3.4(a).

A survey of the cluster composition for the trial runs of Table 3.3 shows that results from runs 2 (n=2) and 14 (n=1) are almost identical to the selected optimal results of run 6 (n=2). In run 10 (n=1) twist-chair clusters 1 and 3 form a single large cluster. In other runs with increasing K_{JP} / K_{NN} ratios, the increasing N_c values result from the progressive splitting of the two main C/TC clusters (1 and 2) into smaller units. A feature of all of these runs, however, is the omnipresence of the distorted twist-chair cluster 3 with N_p= 10 fragments. Lowering the K_{JP} / K_{NN} ratio from the optimum value of around 0.5 merely leads to a coalescence of the boat and twist-chair conformations.

Another simple way to look at the conformational variety in a dataset is to use a "conformational spectrum" (Allen, 1990) as shown in Figure 3.7. The D^n_{pq} values used in the calculation of clusters (Chapter 2, and below) are plotted with n=1. The values are minimised as before by allowing fragments q to adopt all permutations/inversions of the torsion angles. Now though, we use the torsion angles of the most representative chair conformation from Table 3.4(a) as the invariant static fragment p in the equation,

$$D^n_{pq} = \left[\sum_{i=1}^{N_t} \left(\Delta\tau_i \right)^n_{pq} \right]^{1/n} \qquad [2]$$

Thus D^n_{pq} =DSIM, plotted in Figure 3.7, represents the minimum distance of each fragment from a fixed chair position in conformational space. For 7C1 Figure 3.7(a) shows a "chair" peak close to DSIM=0.0, followed by a broad "twist-chair" peak at

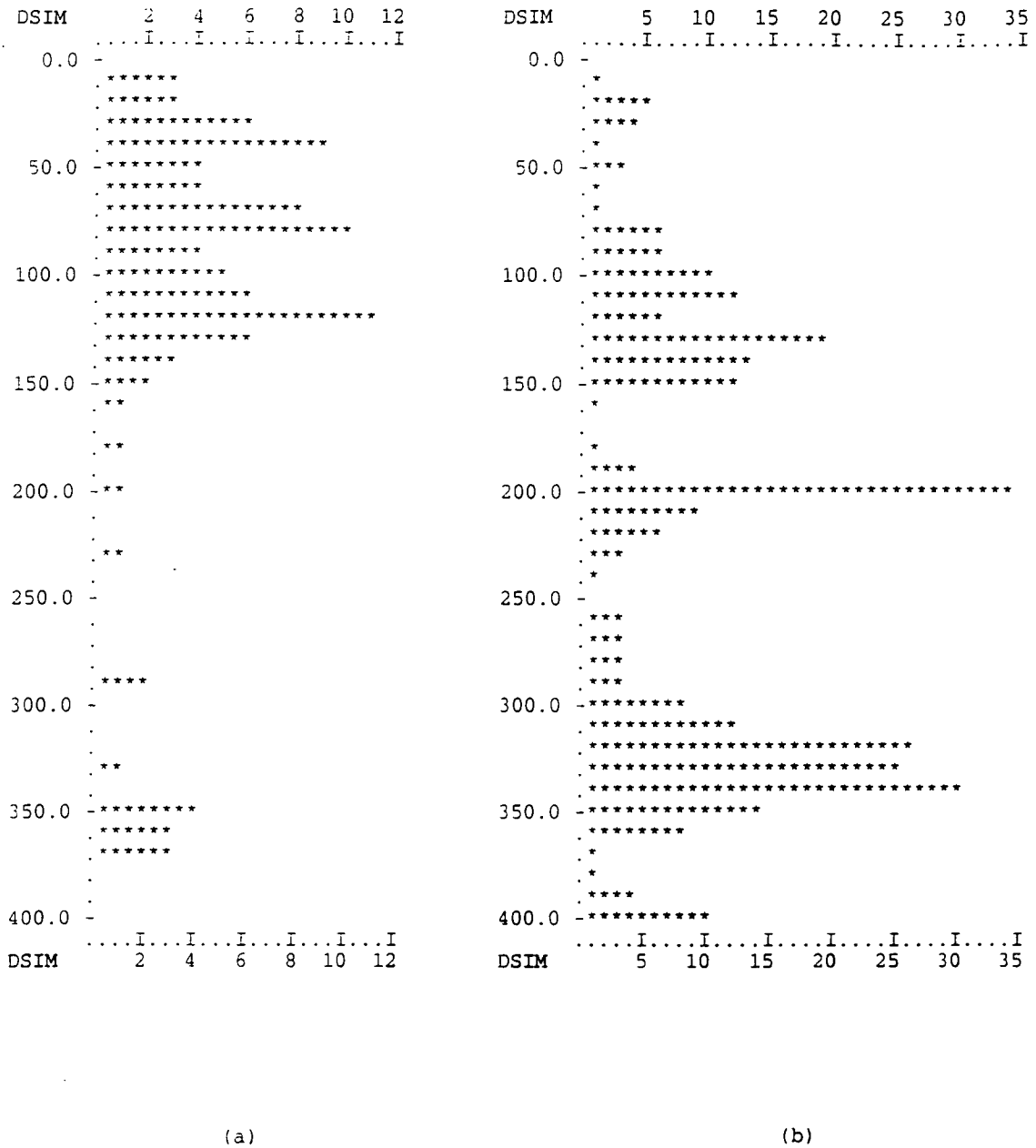


Figure 3.7

Histograms of the minimized torsional dissimilarities (DSIM) of fragments in (a) dataset 7C1, and (b) dataset 7C2, from the torsion angles of the most representative chair conformation taken from cluster 2 of dataset 7C1 (Table 3.4a)

DSIM in the 80-150° range. Cluster 3 is, in fact, a small shoulder of ten distorted twist-chairs at the high-end of this range. The "boat" peak at DSIM > 300.0° is clearly demarcated.

The conformational spectrum of the 310 fragments of the larger bridged-ring dataset of 7C2 is shown in Figure 3.7(b). Again this spectrum is with respect to a fixed chair conformation, as defined above. This spectrum has very few points close to the normal chair position at DSIM=0.0°. Major peaks occur as a doublet at 80-150°, a singlet at approximately 200°, a broad band between 300-360°, and finally a sharp peak at 400°. A brief inspection of the torsion angle listing shows that the doublet arises because of normal twist-chairs, and highly puckered chair conformations. The sharp 200° singlet is primarily due to highly puckered twist-chairs, while the singlet at 400° is due to twist-boats. The dominant feature, the broad band between 300-360° is due to boat conformations exhibiting different degrees of ring pucker. Thus, the spectrum provides a rapid visual overview of the conformational complexity of 7C2, in terms of both the conformations present, and their relative abundancies, prior to any clustering runs. For 7C2, we would therefore expect the clustering runs to dissect the data into at least six major clusters.

A number of Jarvis-Patrick clusterings of 7C2 were performed using the Euclidean metric (n=2), D_{lim} = 0.06, and with values of K_{JP} and K_{NN} chosen such that the ratio K_{JP} / K_{NN} is approximately 0.5. Runs at K_{JP} / K_{NN} = 5/10 and 6/12 led to co-clustering of some of the chair and twist-chair forms, whilst the ratio 7/12 yielded 13 clusters with $N_p \geq 4$, of which some pairs of clusters had almost identical geometries. The ten-cluster structure (all values of $N_p \geq 7$) obtained with a 7/13 ratio was selected as the optimal solution for this dataset. The mean torsion angles and statistical descriptors are given in Table 3.4(b). The ten clusters encompass 242 (78.1%) fragments. Additionally there were three clusters with N_p= 3, two with N_p= 2 and 55 singletons. Cluster coalescence across symmetry positions, with a D_{max} = 1.0 criterion (Allen & Taylor, 1991), is observed for six of the clusters. Thus, although many of the conformations of Table 3.4(b) represent highly puckered forms, the puckering retains the C_2 or C_s symmetry of the basic archetype.

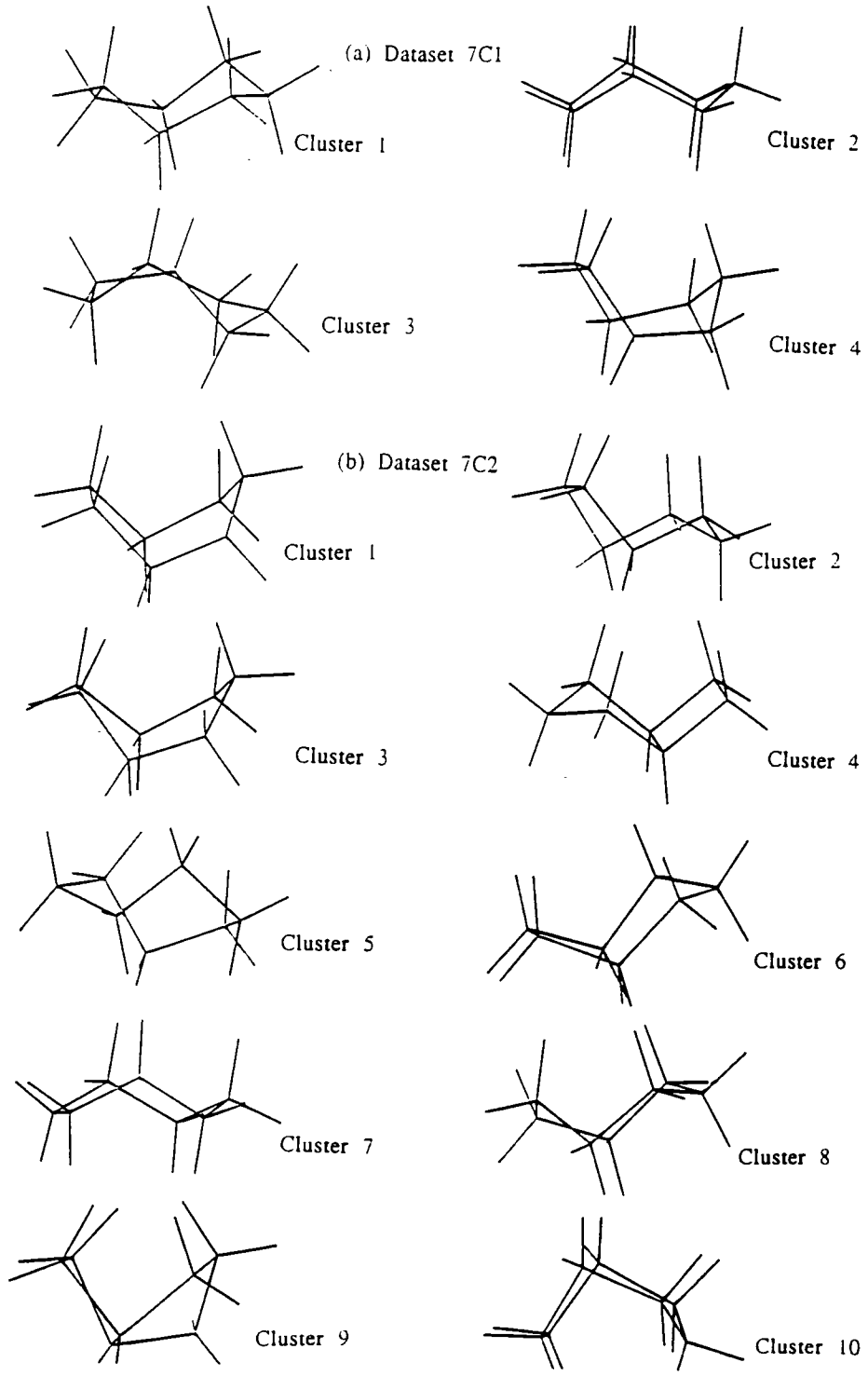


Figure 3.8

Plots of the most representative fragments from conformational clusters identified for (a) dataset 7C1, and (b) dataset 7C2. Cluster numbers correspond to those in Table 3.4

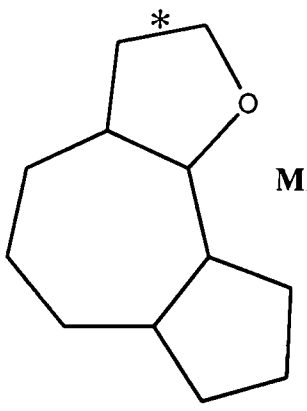
The cluster analysis package within GSTAT will locate the chemical fragment in each cluster which is closest to the centroid of the cluster. This is termed the "most representative fragment" (MRF) of the cluster, and the orthogonal coordinates for this fragment are of use in modelling applications, since they represent the typical TC, C, B, TB or distorted conformations as observed in the accumulated crystallographic results. The coordinates of the MRF's are shown in Appendix 2 for the clusters summarised in Table 3.4, and the conformations are illustrated graphically in Figure 3.8.

3.2.5 Conformational variation and chemical environment

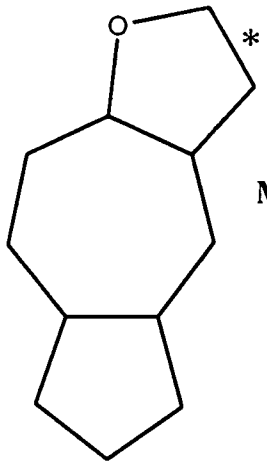
Having successfully classified the ring conformations, it is now of interest to study the relevant chemical environments to see if any particular factors are responsible for the observation of the higher-energy conformations. The conformations of 7C2 are difficult to assess in this respect, because of the complexity of the bridged structures that were observed, ranging from simple bridging types to complicated cages. However, the conformations of 7C1 can be related to chemical environment, and some generalities can be observed.

Non-bridged rings: dataset 7C1

The chemical environments of the cycloheptane fragments in this dataset can be divided into three broad classes: (a) fused to small rings of size three or four (17 fragments), (b) fused to one or more five-membered rings (66 fragments), and (c) non-fused fragments or fragments fused to flexible rings of size greater than five atoms (18 fragments). The small rings of group (a) force the intra-annular cycloheptane torsion angle at the fusion bond to be close to zero. Twelve of these fragments are classified as chairs (cluster 2 of Table 3.4a), whilst a further three are distorted chairs with τ (fusion) = 20°. The last two have the 1, 3, 5 tri-fusion to three-membered rings referred to earlier. Two structure types, (**M1**), (**M2**), dominate subclass (b). In both cases, the O-containing five-ring contains points of unsaturation in the form of exocyclic keto or methylene groups, or has an endocyclic C=C double bond at (*); the other five-ring is



M1



M2

usually, but not always, saturated. The seven-ring itself carries a variety of additional single substituents, *e.g.* Me, OH, halogen. The 24 fragments of type (**M1**) are classified into 15 TC conformers and eight C conformers. It might have been expected that planarity in one (or both) of the five-rings would force a chair-conformation. However, this is not the case, since the seven-ring is flexible enough to adopt *cis* [τ (fusion) = 0°] or *trans* [τ (fusion) = 80-90°] arrangements with respect to the planar five-ring. Analysis of fragments of type (**M2**) is even less predictive: of the 28 fragments, 14 are in the TC class (clusters 1 and 3 of Table 3.4a), three are chairs (cluster 2), and six are boats (cluster 4), the remaining four fragments are some intermediate conformation between C and TC. Here, however, all of the boat-form seven-rings are *cis*-fused to planar five-rings.

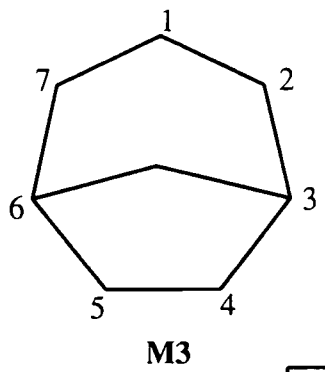
The 18 fragments of subclass (c), either non-fused or fused to rings of size six or more, divide into twelve TC conformers and four C conformers. The remaining two are sufficiently distorted so as to fall between TC and C forms.

As an exercise in human pattern recognition and correlation, this visual analysis has not produced rules of thumb by which we may yet predict the conformation of a seven-ring in a given chemical environment, except for the case of small-ring fusion. However, if we ignore the possible chemical environmental effects in subsets (b) and (c), but exclude the C conformers enforced by small-ring fusion, then the ratio of TC:C:B:TB conformations arising on an apparently statistical basis is 48:18:10:1. These results show a 62% preference for the TC form over all other forms, and a 73% preference for the TC form in the C/TC pathway. These data compare well with the

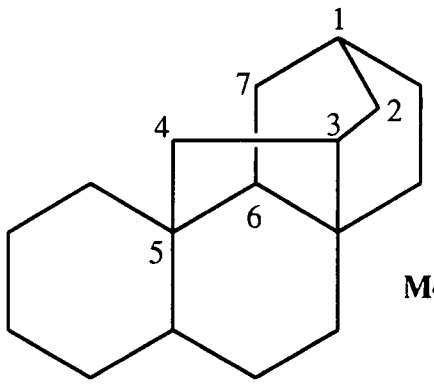
gas-phase electron-diffraction results for cycloheptane (Dillen & Geise, 1979) which were interpreted in terms of a TC/C mixture having a 76 (6)% abundance of the TC form.

Bridged rings: dataset 7C2

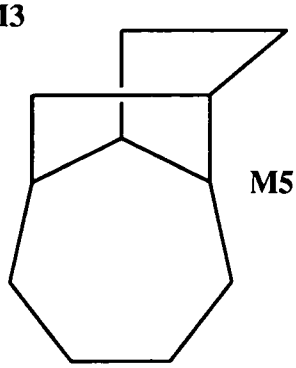
Table 3.4(b) shows that exactly 50% (121) of the rings in clusters 1-10 adopt C/TC conformations and 50% are B/TB. The complete ratios C:TC:B:TB are 85 (35.1%):36 (14.8%):106 (43.8%):15 (6.2%). These ratios cannot of course, be related to free-molecule structures, since they are simply a manifestation of the types of rigidified bridged rings studied so far by the crystallographic method.



M3



M4



M5

The major cluster 1 contains normal boats, whilst the additional boat clusters 3 and 9 show increasing degrees of both twist and pucker. Cluster 1 is dominated by non-fused rings with simple single-atom 1,4-bridging, (**M3**). The deformations observed in clusters 3 and 9 may be ascribed to additional ring fusion(s) at cycloheptane bonds adjacent to the 1, 4-bridge. The puckered twist-boats of cluster 6 occur in simple cages, of which the structures of type (**M4**) are common. The normal chairs of cluster 7 are generally similar to (**M3**) but with two or more atoms in the bridge; fusion and/or cage formation then give rise to the deformations observed in clusters 2 and 4, or in some situations (*e.g.* **M5**) leads to the adoption of a normal twist-

chair conformation. Further objective correlation with chemical environment is impossible, because of the wide variety of bridged/fused and cage-like situations that arise in dataset 7C2.

3.2.6 Mean fragment dimensions

Mean bond lengths and valence angles for the 76 fragments of clusters 1-4 (dataset 7C1, Table 3.4) are collected in Table 3.5. In clusters 1 and 2, where conformational clusters coalesce about symmetry elements, averages are taken over both contributing permutational isomers to yield symmetrical bond length and valence-angle sequences. The 12 fragments of cluster 2 that have chair conformations enforced by small-ring fusion are omitted. The mean geometry of Table 3.5, together with the

Table 3.5

Mean bond lengths (D) and valence angles (A) for the clusters of dataset 7C1. H1-H7 are valence angles computed by Hendrickson (1967) for TC, C, and B forms using fixed bond lengths (C-C 1.533Å, C-H 1.109Å)

N_c	1	2	3	4	Mean
N_p	76	36	10	10	132
D1	1.533 (2)	1.526 (3)	1.522 (5)	1.514 (4)	1.529 (2)
D2	1.530 (2)	1.533 (3)	1.534 (5)	1.548 (3)	1.532 (2)
D3	1.528 (2)	1.521 (2)	1.533 (4)	1.532 (3)	1.527 (2)
D4	1.530 (2)	1.521 (2)	1.529 (4)	1.545 (3)	1.529 (2)
D5	1.533 (2)	1.533 (3)	1.540 (3)	1.548 (5)	1.535 (2)
D6	1.532 (3)	1.526 (3)	1.532 (5)	1.532 (4)	1.530 (2)
D7	1.532 (3)	1.546 (7)	1.530 (3)	1.543 (3)	1.536 (2)
[D1-D7]	1.531 (2)	1.531 (4)	1.531 (4)	1.537 (4)	1.531 (2)
A1	116.1 (3)	117.6 (5)	116.9(5)	113.4 (3)	116.4 (3)
A2	114.5 (3)	115.3 (4)	116.7 (8)	117.7 (4)	115.1 (4)
A3	115.3 (3)	115.6 (5)	121.4 (6)	114.9 (4)	115.8 (4)
A4	115.3 (3)	114.4 (4)	116.6 (9)	115.4 (5)	115.1 (4)
A5	114.5 (3)	115.6 (5)	112.2 (8)	115.0 (4)	114.5 (4)
A6	116.1 (3)	115.3 (4)	116.8 (11)	117.3 (6)	116.0 (5)
A7	116.5 (4)	117.6 (5)	113.2 (5)	112.2 (6)	116.2 (5)
[A1-A7]	115.5 (3)	115.9 (5)	116.3 (7)	115.1 (4)	115.6 (4)
H1	115	118	115	115	
H2	113	115	113	116	
H3	115	114	115	116	
H4	115	115	115	115	
H5	113	114	113	115	
H6	115	115	115	115	
H7	116	118	116	115	

mean torsion angles of Table 3.4, do not necessarily correspond to a feasible (closed) ring in three dimensions (Murray-Rust, Bürgi & Dunitz, 1978) by contrast to the coordinates of Appendix 2.

The mean bond length across all ring bonds is 1.531(2)Å, almost identical to the 1.530(2)Å cited by Allen *et al.* (1987) for 5777 Csp3-Csp3 bonds averaged over all bond-substitution patterns. The value is also very close to fixed value of 1.533Å used by Hendrickson (1967) in his energy-minimisation studies of the medium rings. Indeed, the results of Table 3.5 show that his predicted intra-annular valence angles (H1-H7) for the TC and C forms are in close agreement with the experimental data currently available.

3.3 *exo*-Unsaturated and heterocyclic seven-membered rings

3.3.1 Database search and retrieval

Version 4.5 (January 1992) of the Cambridge Structural Database was used throughout for sub-structure search, coordinate retrieval and data analysis, using the programs QUEST and GSTAT (Allen *et al.*, 1991).

Two general searches were performed initially in order to assess the nature of the rings present in the CSD: (i) for a seven-membered cycle of ring atoms, X, connected by single bonds, and where X = any non-H atom, and (ii) for this same substructure in which one of the atoms carries an *exo* double bond. As described in Chapter 3.2.1 for the cycloheptanes, the searches were restricted using CSD bit-screen mechanisms to locate only entries with (a) atomic coordinates available, (b) no residual numerical errors following CSD evaluation procedures, (c) no reported disorder in the crystal structure, (d) a crystallographic R-factor less than 12%, (e) classified as an 'organic' compound according to CSD definitions, and (f) contained no bridged or highly complex ring systems (SCREEN -620 -622). As a result of these searches fragments V-XI (Figure 2.4) were selected as chemically coherent subsets for full conformational analysis. The individual searches for the selected fragments have C_{sp2}

Table 3.6

Number of entries and fragments, together with the maximum R-value for datasets V-XI. Optimum clustering criteria are also shown.

	search results			Jarvis-Patrick criteria			
dataset	N_e	N_f	R_{max}	n	K_{NN}	K_{JP}	D_{max}
V	56	64	0.095	2	10	5	0.08
VI	30	32	0.096	1	5	3	0.08
VII	15	16	0.091	1	5	3	0.08
VIII	33	50	0.103	2	7	3	0.08
IX	22	26	0.110	2	5	3	0.08
X	10	10	0.096	2	5	3	0.08
XI	16	23	0.096	1	8	4	0.10

as identified in Figure 2.4, with the other ring carbon atoms all being sp^3 hybridised. The number of CSD entries retrieved (N_e), number of fragments located (N_f), and the maximum R-factor found for each dataset are reported in Table 3.6.

Analysis of the data was as before for the cycloheptane fragments, and is based on the geometrical description using seven intra-annular torsion angles of the ring. Cremer-Pople puckering coordinates, together with BPRS and PCA descriptions were generated in GSTAT, and the datasets were conformationally classified using the clustering algorithms also available in this program.

3.3.2 Energy calculations

As reported in Section 2.4, the need for reliable energy calculations on the conformational variants of each dataset, arises from the incomplete coverage of this topic in the current literature. The molecular modelling package COSMIC, (Vinter, Davis & Saunders, 1987) was locally modified (by Dr. J.G. Vinter) for better application to ring conformation hunting, and used here to verify results already present in the literature, and then to determine energies for those that are not well represented.

The calculated energies (E, $kJmol^{-1}$), and energy differences (ΔE) from the relevant global minimum, are reported in Table3.7. Literature data for cycloheptanone

Table 3.7

Calculated energy values for the different C and TC conformations of the seven-membered rings in datasets V-IX. Literature values are cited where available.

dataset	function	C^1	C^2	C^3	C^4	TC^1	TC^2	TC^3	TC^4
V	E	15.62	24.50	27.00	28.99	23.55	23.09	16.86	16.83
	ΔE	0.00	8.88	11.38	13.37	7.93	7.47	1.24	1.21
VI	E	12.64	22.62	25.02	26.30	20.71	21.57	15.04	13.09
	ΔE	0.00	9.98	12.38	13.66	8.07	8.93	2.40	0.45
	ΔE^a	0.54	10.56	12.45	13.50	8.39	6.72	0.84	0.00
	ΔE^b					6.80	7.56	0.00	1.05
VII	E	8.03	17.62	19.52	22.29	16.62	15.91	10.22	9.75
	ΔE	0.00	9.59	11.49	14.26	8.59	7.88	2.19	1.72
VIII	E	44.01	38.47	40.42	39.44	36.43	37.49	33.85	40.71
	ΔE	10.16	4.62	6.57	5.59	2.58	3.64	0.00	6.86
	ΔE^c	(12.68)	8.14	(11.00)	(16.45)	(8.90)	0.05	0.00	(11.42)
IX	E	31.87	32.30	35.65	36.45	32.33	31.44	28.04	32.68
	ΔE	3.83	4.26	7.61	8.41	4.29	3.40	0.00	3.64

References: (a) Bocian & Strauss (1977b); (b) Allinger, Tribble & Miller (1972); (c) Bocian & Strauss (1977a), all other data were calculated here using COSMIC

[(VI); Bocian & Strauss, 1977b; Allinger, Tribble & Miller, 1972], and for oxepane [(VIII); Bocian & Strauss, 1977a, c], are also included for comparison.

The general features of the energy hypersurface for the exo-unsaturated (V)-(VII) are very similar: a broad minimum encompasses the TC3-C1-TC4 areas of the pseudorotation itinerary (Figure 2.5b), with symmetry-equivalent minima separated by an energy barrier of approximately 13-14 $kJmol^{-1}$. There exist, however, minor differences between the energy levels of the individual TC^3, C^1 and TC^4 minima, both between substructures and, for (VI), between the different calculational methods (each of which employed a different force field). Thus, modified COSMIC calculations consistently yield an energy ordering of $TC^3 > TC^4 > C^1$, but Bocian & Strauss (1977b) indicate a TC^4 global minimum in contrast with the TC^3 minimum calculated by

Allinger, Tribble & Miller (1972). However, these fluctuations are probably not predictive since they only span energy variations at the 2.4 $kJmol^{-1}$ level.

The COSMIC results obtained here for the hetero-rings (VIII) and (IX) are again consistent in predicting a TC^3 global minimum in agreement with results found by Bocian & Strauss (1977a, c) for (VIII). However, their calculations then indicate an almost isoenergetic TC2 form separated from the TC^3 by a C^2 energy barrier of 8.14 $kJmol^{-1}$, with other conformations of (VIII) being at higher energies. COSMIC results for (VIII) and (IX) suggest TC^3 as a clear global minimum, but with much lower energy barriers between TC^3 and the other conformers, than those predicted by Bocian & Strauss (1977a, c).

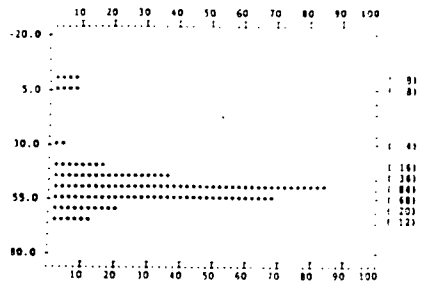
Energy calculations for the lactone (X) have been reported by Allinger (1982), and show that the dominant C^1 conformer lies some 11.34 $kJmol^{-1}$ below the next local minimum, which is a boat conformation. The presence of a single chair conformer for both (X), and the lactam (XI). is also indicated by the NMR results (Noe & Roberts, 1971).

3.3.3 Conformational Mapping

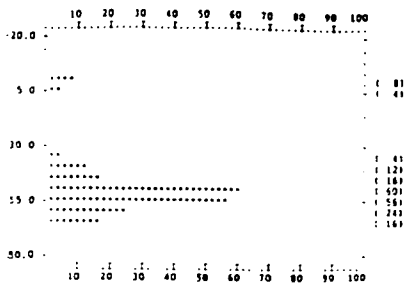
BPRS coordinate plots

The conformational diversity present in the datasets of (V)-(XI) was initially investigated using the histograms of the BPRS θ-coordinate (Figure 3.9), and the ϕ_2-ϕ_3 scatterplots of Figure 3.10. Histograms for the datasets (V)-(IX) are generated from the symmetry-expanded datasets, and so the numbers present in each of the angular bins represent all permutational isomers and enantiomers, and are thus four times the actual number of fragments. The histograms for (X) and (XI) are generated from the raw datasets.

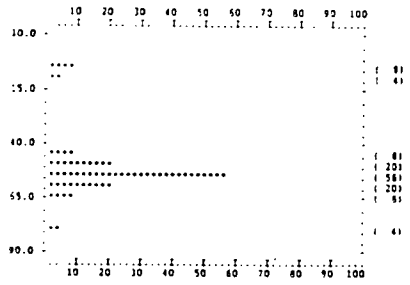
From the histogram data it can be seen that the C/TC conformers are once again the dominant structural forms for all of the subsets, *i.e.* the majority of all of the data is found for $\theta \geq 40°$. Table 3.8 summarises the results observed in these histograms, and



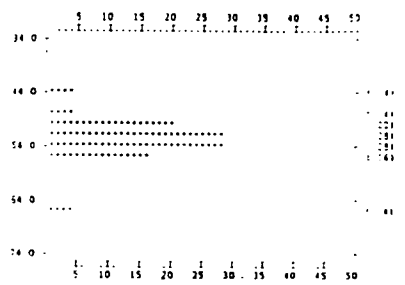
(a) methylenecycloheptanes (V)



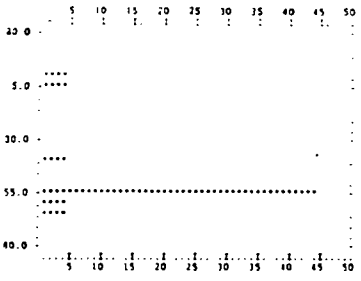
(d) oxepanes (VIII)



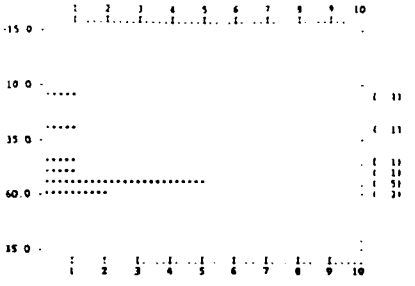
(b) cycloheptanones (VI)



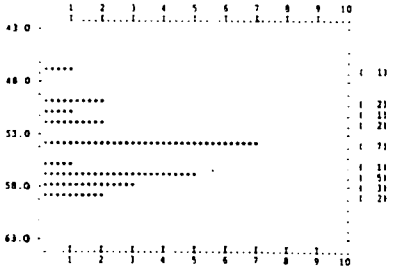
(e) azepanes (IX)



(c) azocycloheptanes (VII)



(f) lactones (X)



(g) lactams (XI)

Figure 3.9

θ–histograms for the datasets (V)-(XI)

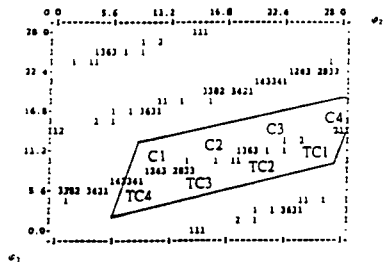
Table 3.8

Percentage of variance accounted for by each of the 4 PC's and their related conformational variants. Percentage variance for C, TC conformations (%C/TC) is compared to the percentage of conformers with $\theta > 40°$

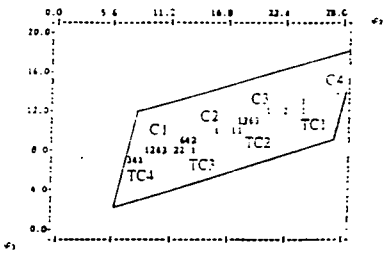
Fragment	TC	C	TB	B	%C/TC	$\%\theta_h$
V	54.7	30.4	8.6	6.2	85.1	92.2
VI	60.5	29.5	5.3	4.7	90.0	90.6
VII	69.5	23.3	4.2	2.9	92.8	87.5
VIII	45.6	42.1	5.5	6.8	87.7	86.0
IX	59.1	31.3	3.2	6.4	90.4	100
X	5.7	75.1	2.5	16.6	78.8	81.8
XI	2.1	74.1	19.4	4.5	76.2	100

shows that 80% of the observations have $\theta \geq 40°$ and therefore lie on the TC/C pathway (θ_h corresponding to $\theta \geq 40°$).

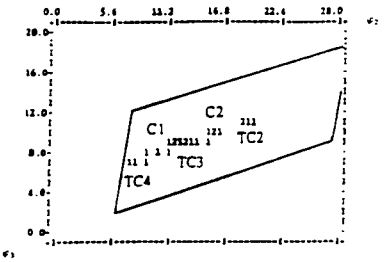
The ϕ_2-ϕ_3 scatterplots of Figure 3.10 are all generated using the symmetry-expansion for the fragments with $\theta \geq 40°$. The complete C/TC pseudorotation itinerary is exemplified for (V) in Figure 3.10(a1) but the other plots for (V)-(IX) show only the relevant asymmetric unit. Complete maps are shown for (X) and (XI). The ϕ_2-ϕ_3 plots for (V)-(VII) all show high density in the TC^3-C^1-TC^4 area, as predicted by the energy calculations (Table 3.7). However, methylenecycloheptane (V) and, to a lesser extent, cycloheptanone (VI), both show density in the TC^2 area. TC^4 conformers appear to be absent for (VII). The heterocyclic fragments (VIII) and (IX) show rather different ϕ_2-ϕ_3 patterns. For oxepanes (VIII), the TC^4 and C^1 conformers are not represented, as suggested by the energy calculations (Table 3.7). The C^4 conformer is also absent but C^3, one of the higher energy conformers, shows a small population density. The azepanes (IX) show a broad conformational distribution, perhaps a reflection of the rather similar ΔE values (Table 3.7) for six of the possible symmetric conformers. One of the two higher energy forms (C^4) is also present. The lactones (X) and lactams (XI) are predominantly C^1 conformers, as expected.



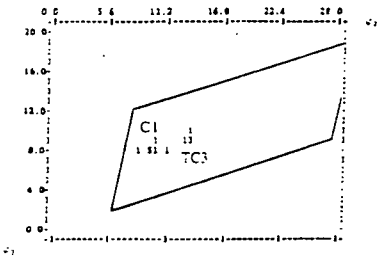
(a1) methylenecycloheptanes (V)



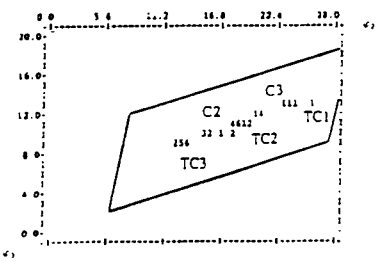
(a2) methylenecycloheptanes (V)



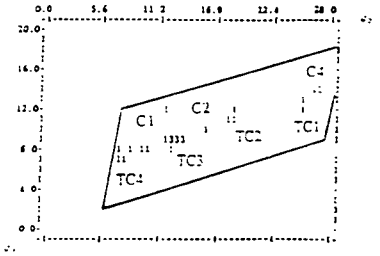
(b) cycloheptanones (VI)



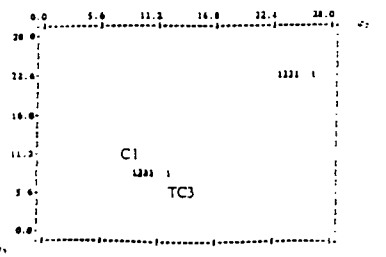
(c) azocycloheptanes (VII)



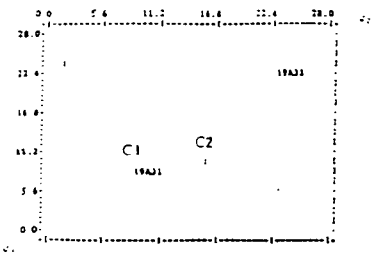
(d) oxepanes (VIII)



(e) azepanes (IX)



(f) lactones (X)



(g) lactams (XI)

Figure 3.10

ϕ_2-ϕ_3 plots for substructures (V)-(XI). Part (a1) shows all fragments for (V); parts (a2) to (e) include only those fragments with $\theta \geq 40°$

Principal Component Analysis (PCA)

PCA was applied to the symmetry-expanded torsional datasets of (V)-(XI), and the results treated in a manner analogous to the previous study on cycloheptane fragments (Chapter 3.2). Here the four non-degenerate principal components (PC's) were found to account for >99.9% of the variance in each of the datasets (Table 3.8). The symmetry of the PC loadings indicated that each of the PC's maps to one of the symmetric conformations, and the percentage of the total variance accounted for by each PC is also given in Table 3.8.

These results represent a quantification of the visual interpretation of Figures 3.9 and 3.10, and can also be correlated with the energy calculations of Table 3.7. Obviously, the C/TC conformations dominate all datasets and the ratios (C + TC) : (B + TB), given as %C/TC in Table 3.8, closely mimic the $\%\theta_h$ values deduced from Figure 3.9. For the exo-unsaturated rings (V)-(VII), the TC^3, C^1 and TC^4 conformers are of approximately equal (minimum) energies and, in any random sample of rings, we might expect a TC:C ratio of 2:1. The TC:C ratios for (V)-(VII) (Table 3.8) are very close to this value. The very high C:TC ratios for (X) and (XI) are also to be expected energetically, and are a restatement of Figures 3.10(f) and 3.10(g) in numerical terms. The energy values of Table 3.7 would suggest TC:C ratios > 1.0 for the heterocyclic rings (VIII) and (IX), since TC energies are at a generally lower level than C energies. This is evident from the TC:C variances of Table 3.8, but the correlation is less convincing than for the other substructures.

The six PC-scatterplots were each prepared for (V)-(XI), as described previously in Chapter 3.2, but only two of these plots are shown in Figure 3.11. The PC1-PC2 scatterplots for the oxepanes (Figure 3.11a) and the azepanes (Figure 3.11b) both show a perpendicular, hence circular, view of the C/TC toroidal pseudorotation itinerary (compare with the plots obtained for cycloheptane - Figure 3.4). In Figure 3.12 circular pseudorotation itineraries are presented, which are equivalent to the two-dimensional ones determined previously (*e.g.* Figure 2.5). When Figure 3.11(a) is compared to Figure 3.12(a) it can be seen that there is a marked absence of density in the regions C^1-TC^4, and C^4, while the greatest concentration appear in the TC^2-C^2-TC^3

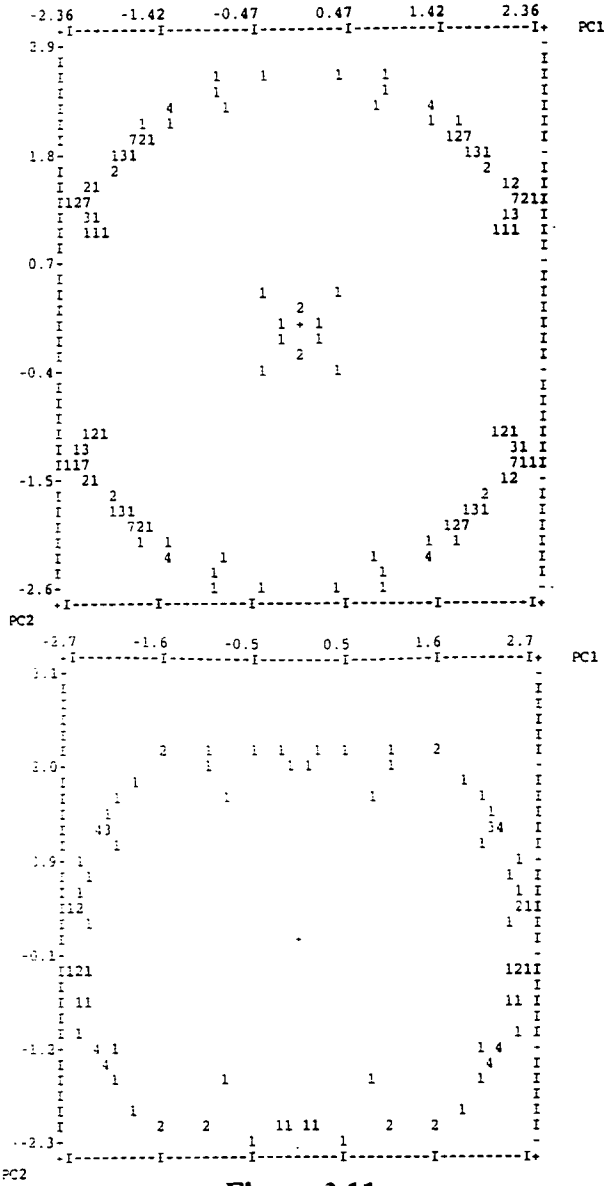


Figure 3.11

Scatterplots of the principal component scores referred to PC1, PC2 axes for (a) fragment (VIII) and (b) fragment (IX)

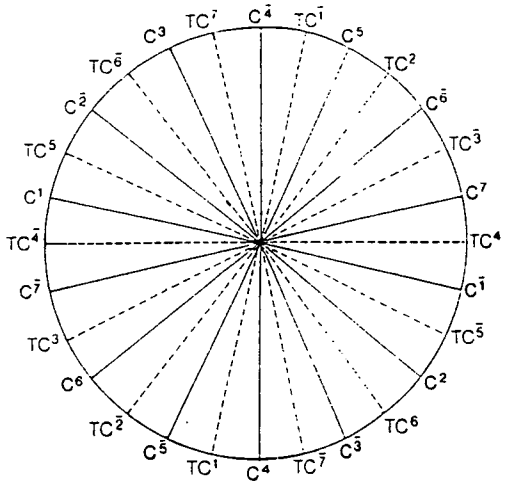


Figure 3.12

Idealized conformational mapping for the chair - twist-chair ϕ_3-pseudorotation itinerary of cycloheptane

areas (see also Figure 3.10d). The central density in Figure 3.11(a) represents three B/TB conformers. Since five of the oxepanes had $\theta < 40°$. (Figure 3.9d and Table 3.8), it would appear that $\theta < 25$-$30°$ might have been a more appropriate discriminator of B/TB conformations. In Figure 3.11(b) we see a more complete distribution of conformations along the C/TC itinerary, and the complete absence of B/TB conformers, in agreement with the BPRS analysis of Figures 3.9(e) and 3.10(e).

3.3.4 Conformational classification

Classification of the conformations present in each dataset was carried out using the Jarvis-Patrick clustering techniques detailed in Chapter 2, and for cycloheptanes in Section 3.2.4. The results of the symmetry-modified clustering on the torsional datasets are summarised in Table 3.9 below, in the form of mean torsion angles for each cluster with a population (N_p) ≥ 2 fragments. The appropriate conformational descriptor is assigned in each case. For the smaller datasets, *i.e.* all except (V) (64 fragments) and (VIII) (50 fragments), a number of variations in the clustering criteria K_{NN} and K_{JP} (Table 3.6) were required to generate a chemically acceptable classifications.

For cycloheptanone (VI) however, none of the clustering algorithms described earlier (Chapter 2) were able to generate satisfactory results. Upon closer examination of the various Jarvis-Patrick runs it was noticed that small twist deformations in the C^1 examples towards both TC^3 and TC^4 conformations were preventing the formation of discrete clusters in automated processes. For this reason, the manual results of Table 3.9(b) are preferred for (VI).

Despite some of the difficulties, the results of the Jarvis-Patrick clustering are again informative in the selection of the parameters, K_{NN} and K_{JP}. The ratio K_{JP}/K_{NN} has been in the region of 0.5 for previous experiments (Allen, Doyle & Taylor, 1991a, b; Allen, Howard & Pitchford, 1993), resulting in clusters that satisfied the chemical sensibility of the system. In Table 3.6 it can be seen that the K_{JP}/K_{NN} ratios for the datasets vary between 0.43 and 0.60 for a series of small integer values.

Table 3.9(a)

Jarvis-Patrick cluster analysis results for the seven datasets. The mean torsion angles and e.s.d. for each cluster are given, together with the MRF.

Frag.	Cluster no.	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
II	1	-86.0(17)	70.2(14)	-52.7(14)	71.0(11)	-88.4(11)	43.2(15)	35.3(7)
	2	-67.6(20)	42.9(25)	-65.5(13)	94.1(13)	-45.8(12)	-35.1(12)	88.4(7)
	3	0.9(15)	64.6(13)	-84.9(8)	64.5(15)	-62.0(10)	77.8(14)	-62.8(17)
	4	-28.5(16)	81.7(17)	-74.9(14)	56.0(17)	-68.6(21)	85.4(15)	-45.7(17)
	5	69.5(13)	-10.4(0)	-57.2(4)	88.4(20)	-71.2(27)	59.3(26)	-75.1(16)
III	1	19.7(33)	-81.3(9)	82.5(26)	-61.2(17)	65.9(11)	-85.4(8)	55.1(24)
	2	-84.2(15)	75.1(12)	-59.6(19)	70.0(13)	-79.8(21)	39.2(19)	33.7(10)
	3	-97.4(21)	69.0(9)	-54.6(10)	76.2(9)	-85.2(15)	28.7(30)	52.8(24)
	4	-18.5(57)	-48.0(44)	77.6(18)	-65.6(17)	61.5(10)	-79.5(08)	74.6(40)
	5	-63.8(20)	0.3(37)	70.9(79)	-90.5(65)	62.4(28)	-62.4(26)	83.2(1)
IV	1	1.3(21)	64.2(24)	-84.6(29)	65.6(13)	-62.4(14)	80.8(14)	-65.9(11)
	2	94.6(10)	-70.7(8)	57.2(11)	-74.8(4)	79.4(13)	-25.8(14)	-51.7(8)
V	1	-98.9(7)	67.1(17)	-46.6(21)	67.1(10)	-84.3(14)	32.9(11)	52.8(9)
	2	-82.7(6)	54.9(11)	-67.5(11)	86.3(9)	-43.9(13)	-33.1(12)	94.0(9)
	3	-59.2(41)	-3.5(41)	73.6(18)	-90.3(21)	63.7(7)	-63.8(20)	79.1(14)
	4	-75.7(13)	74.2(7)	-15.1(17)	-58.6(6)	100.3(6)	-78.6(4)	59.8(15)
VI	1	86.8(15)	-71.8(16)	53.8(14)	-68.3(7)	87.0(16)	-43.4(27)	-36.3(26)
	2	14.7(42)	55.5(36)	-86.3(18)	65.9(18)	-56.1(18)	76.7(24)	-72.7(24)
VII	1	-0.9(23)	64.5(20)	-77.1(17)	59.1(14)	-58.8(14)	79.8(13)	-66.4(20)
VIII	1	-2.1(8)	68.1(9)	-78.2(5)	59.5(8)	-61.3(8)	81.2(9)	-65.1(8)

Frag.	Cluster no.	Conformation	N_p	MRF
II	1	TC^3	15	DHYGUA
	2	TC^2	14	SAYZOH
	3	C^1	12	MIKTIN
	4	TC^4	8	CELPEO
	5	C^2	2	
III	1	C^1/TC^3	4	HONGUA
	2	TC^3	4	JILJIX
	3	TC^3	4	DOTJOL
	4	C^1/TC^4	3	
	5	C^2	3	
IV	1	C^1	8	BUVVAP
	2	TC^3	5	CACBEN
V	1	TC^3	17	BUJRED
	2	TC^2	12	JAMVUI
	3	C^2	8	GETKEV
	4	C^3	5	JAMVOI
VI	1	TC^3	11	VANRAD
	2	C^1/TC^4	6	STEMON10
VII	1	C^1	8	JOCCOT
VIII	1	C^1	23	CAPLAC01

Table 3.9(b)

Conformational groupings and populations for dataset VI by visual assessment. The torsion angles are those of a characteristic example of each conformation.

Conf.	N_f	MRF	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
TC^3	12	FUSNEM	87.0	-74.9	58.1	-68.3	82.8	-41.1	-36.5
C^1	5	JIDHUZ	1.4	-59.0	83.4	-74.7	68.4	-67.5	51.3
C^2	5	HAESTD	-72.5	5.8	59.8	-82.9	64.4	-64.2	88.4
TC^2	4	FUFLAT	-80.8	52.1	-68.2	86.1	-40.2	-36.1	95.0

The orthogonal molecular axis coordinates for the most representative fragments are presented in Appendix 2, for the 8 [(V)-(VII)], 7 [(VIII), (IX)] and 9 [(X), (XI)] atoms of each substructure. These coordinates are taken from the individual crystal structures, and are therefore, not related to the mean geometries shown in Table 3.10 below.

Table 3.10

Mean bond lengths (Å) and valence angles (°) with esd's in parentheses, for the seven-membered rings of the datasets II-VIII. D1-D7 refer to the bonds identified by τ_1-τ_7 of Table 2.2, and A1-A7 are the valence angles having atoms 1-7 as the vertex.

	II	III	IV	V	VI	VII	VIII
D1	1.511(1)	1.515(2)	1.506(2)	1.428(1)	1.473(4)	1.340(3)	1.339(3)
D2	1.539(1)	1.546(3)	1.533(2)	1.525(1)	1.501(4)	1.465(3)	1.446(7)
D3	1.533(1)	1.531(2)	1.517(3)	1.523(1)	1.518(3)	1.536(6)	1.530(5)
D4	1.532(2)	1.529(2)	1.523(2)	1.532(2)	1.513(5)	1.554(6)	1.534(5)
D5	1.533(1)	1.531(2)	1.517(3)	1.523(1)	1.518(3)	1.549(5)	1.530(4)
D6	1.539(1)	1.546(3)	1.533(2)	1.525(1)	1.501(4)	1.526(14)	1.541(4)
D7	1.511(1)	1.515(2)	1.506(2)	1.428(1)	1.473(4)	1.501(3)	1.517(4)
A1	118.7(2)	119.6(3)	119.4(3)	115.9(2)	117.5(2)	121.3(4)	117.8(4)
A2	114.9(1)	113.1(4)	114.0(4)	111.9(2)	114.9(3)	123.9(8)	125.1(4)
A3	114.4(2)	114.6(3)	115.0(3)	114.6(2)	115.9(3)	113.1(6)	113.3(4)
A4	114.7(2)	116.0(2)	115.5(3)	115.0(2)	115.3(4)	118.0(7)	113.6(3)
A5	114.7(2)	116.0(2)	115.5(3)	115.0(2)	115.3(4)	112.6(10)	115.2(5)
A6	114.4(2)	114.6(3)	115.0(3)	114.6(2)	115.9(3)	115.1(10)	114.6(6)
A7	114.9(1)	113.1(4)	114.0(4)	111.9(2)	114.9(3)	115.1(8)	112.9(5)

We now consider the classifications reported in Table 3.9 for subsets of the individual substructures (V)-(XI), in the context of the energy calculations of Table 3.7. Of particular interest are any chemical or structural features that may cause a particular fragment to adopt a higher energy conformation. The mean bond lengths and angles within (V)-(XI) (Table 3.10) are also discussed in these subsections. These data have been calculated across all observations of each substructure, it being unrealistic to generate mean values for each conformer in view of the small datasets involved. Mean values take account of the appropriate topological symmetry in each case.

3.3.5 Discussion

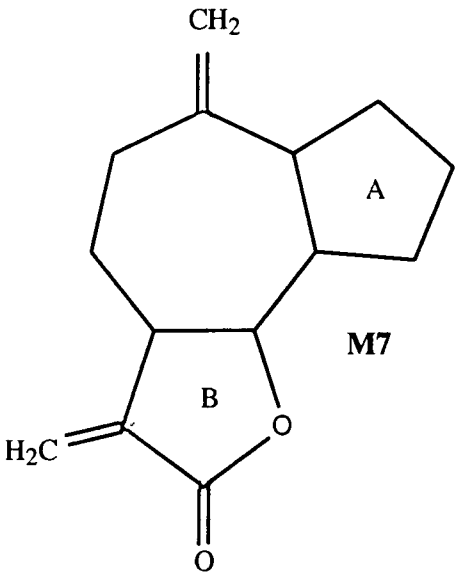
Exo-unsaturated fragments [(V), (VI) and (VII)]

The general dominance of the TC^3 and C^1 conformers in these datasets is as expected from the energy calculations. However, the TC^4 conformer which is calculated to be almost isoenergetic with TC^3 and C^1 is only prominent (N_p=8) for the methylenecycloheptane fragments (V). There are two clear TC^4 conformers for (VI) and none for (VII), although the predicted global minimum for (VII) (C^1) is the major contributor with N_p = 9 of 16 fragments.

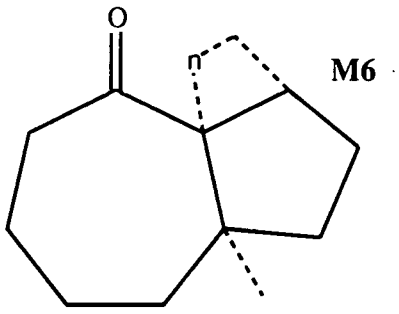
The most interesting feature for (V) and (VI) is the presence of significant populations of higher energy TC^2 and C^2 conformers. One of the five obvious C^2 conformers of (VI) has cyclopropyl fusion on bond 2 (C2-C3), thus forcing a zero τ_2 as in cycloheptane (see Chapter 3.2). The other four cases all exhibit steric overcrowding at C2 and C3. All have quaternary C2 atoms, and C2-C3 is fused to a five-membered ring which. in three cases is further fused (**M6**), placing extra constraints on C2. If the five-membered ring is not further fused, then C3 is also quaternary (**M6**).

The 14 TC^2 conformers of (V) are even more distinct: ten of these contain substructure (**M7**) exactly, and three more are identical except that the exo-methylene of the fused furan ring is replaced by a methyl group. Ring A (**M7**) has a variety of substituents but, in all cases where the ring A bonds are single, the TC^2 conformation is

preserved. The γ-lactone (furanone) ring B is forced to be planar due to stereoelectronic effects at oxygen (Kirby, 1993) and it appears that, in combination with the ring A constraints, the cycloheptanone is forced to adopt a TC^2 conformation.

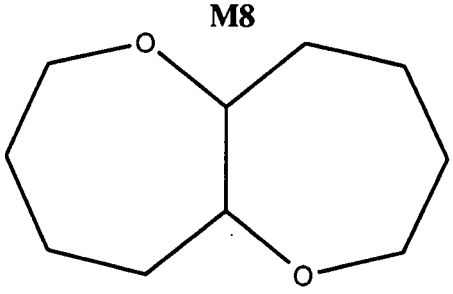


M7



M6

M8



The mean geometries of (V)-(VII) are unremarkable (Table 3.10). The intra-annular $C_{sp}3$-$C_{sp}3$ bond lengths are typical and any differences can be ascribed to the degree of substitution of these C atoms (*e.g.* Allen, Kennard *et al.*, 1987). The $C_{sp}3$-$C_{sp}3$ distances are normal and the intra-annular angle at C1 is only slightly less than 120° in all cases.

Hetero-cycloheptanes [(VIII) and (IX)]

The lower energy TC^2 and TC^3 conformations predominate for (VIII) (30 out of 50 examples), while the occurrence of TC^3 as the major conformer for (IX) is in agreement with our energy calculations (Table 3.7). The major higher energy conformers of (VIII) are C^2 and C^3. Six of the C^2 conformers arise from constraints imposed by small-ring fusion at C2-C3. One of the others is the only example of a free β-D-glucoheptanoside, *i.e.* one in which there is no further cyclisation through adjacent oxygen substituents. The final C^2 conformer is another example of a penta-substituted ring analogous to the sugar substitution pattern. Most of the C^3 examples arise from

the *trans*-isomer of the oxepano-oxepane fragment (**M8**), for which the *cis*-isomer adopts a distorted TC3 conformation.

For (IX), the smaller cluster represents a chaining in the C^1-TC^4 area of conformational space. The presence of a quaternary N^+ bonded to an increasing number of non-H atoms appears to drive the conformation towards C^1. However, the clearest examples of the C^1 conformation occur in two structures involving N lone-pair delocalization in >N-N=C-, and >N-C=N- systems.

Once again, the mean bond lengths and valence angles for (VIII) and (IX) are as expected. The only value which is worthy of comment is the angle A1, which reaches a value of 117.5(2)° in azepanes. This value implies a degree of $N_{sp}2$ involvement as might be expected from the range of conjugative substituents (*e.g.* phenyl, -C=O *etc.*) at the N atom in (IX).

Lactones and lactams [(X) and (XI)]

As expected, the C^1 conformer is completely dominant within the these datasets: for (X) there are 9 out of 10 observations, and for (XI) all 23 observations are C^1. The only outlier is surenolactone (Kraus, Kypke, Grimminger, Sawitzi & Schwinger, 1982), in which a TC^3 conformation is observed for an ε-lactone ring fused to an unsaturated ε-lactone ring [(X) with a double bond at C6=C7].

The geometry of lactones and lactams has been discussed in detail by Norskov-Lauritsen, Bürgi, Hofmann & Schmidt (1985), who present mean geometry for both ε-lactones (X) and ε-lactams (XI). Our results agree with, and extend their averages, which were taken over only four examples of (X) and seven of (XI).

3.4 *endo*-Unsaturated seven-membered rings

3.4.1 Database search and retrieval

Version 4.5 (January 1992) was used throughout this study of the endo-unsaturated seven-membered rings. Searches were carried out for the fragments

Table 3.11

Search results for substructures XII-XV of Figure 2.6. N_e is the number of CSD entries retrieved, N_f is the number of chemically independent fragments located in those entries, and R_{max} is the minimum R-factor in each dataset. Clustering criteria used in the Jarvis-Patrick analyses are also given.

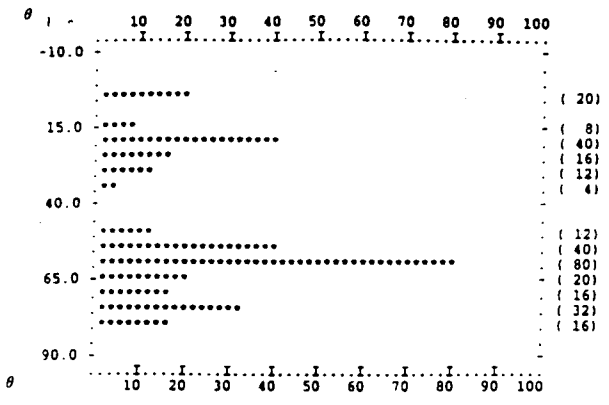
	composition			clustering criteria			
Substructure	N_e	N_f	R_{max}	n	K_{NN}	K_{JP}	D_{max}
XII	67	74	0.151	1	12	6	0.08
XIII	15	14	0.080	2	8	4	0.08
XIV	12	13	0.094	2	4	2	0.10
XV	29	34	0.088	1	10	5	0.10

(XII)-(XV) (Figure 2.6), in which the double bonds indicate either formal double or aromatic bonding (bond type = 2 or 5 in the CSD representations). Atoms not involved in the multiple bonds were required to be sp^3 hybrids. Searches were further constrained by using the CSD bit-screen mechanism to locate entries with (a) atomic coordinates available, (b) no residual numerical errors following CSD check procedures, (c) no reported disorder in the crystal structure, (d) a crystallographic R-factor ≤ 0.12, (e) classified as "organic" by CSD definitions, and (f) containing no bridged or highly complex ring systems (SCREEN -620 -622). The search results are summarised in Table 3.11, and the refcodes for the different datasets can be found in Appendix 1.

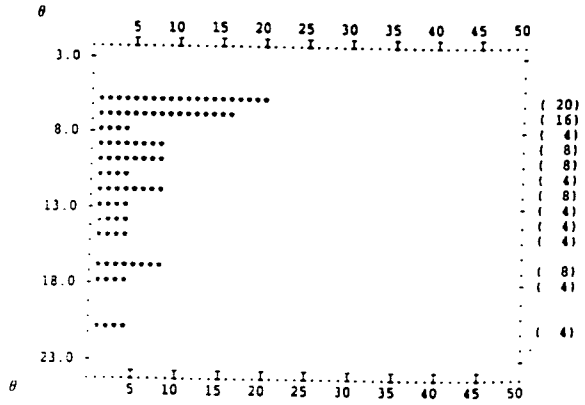
3.4.2 Conformational Mapping

BPRS coordinate plots

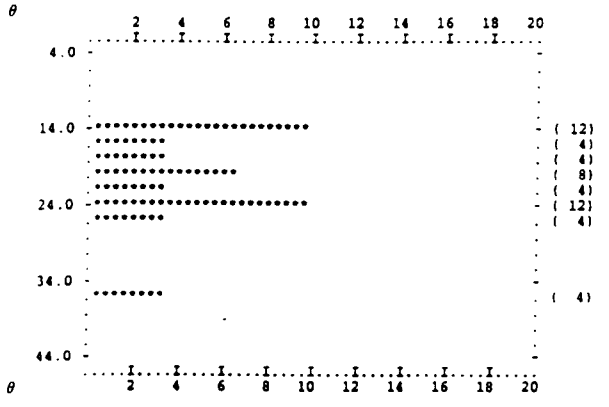
Histograms of the BPRS θ coordinate (Figure 3.13) provide a rapid visual overview of the conformational complexity of the datasets for fragments (XII)-(XV). For the cycloheptene (Fig. 3.13a) fragments the largest peak occurs in the TC/C area ($\theta > 40°$). A smaller grouping is observed at $\theta = 5–35°$ spanning the area of



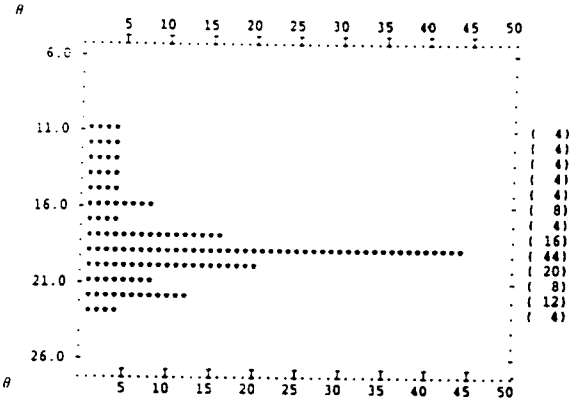
(a) cycloheptene (XII)



(b) cyclohepta-1,3-diene (XIII)



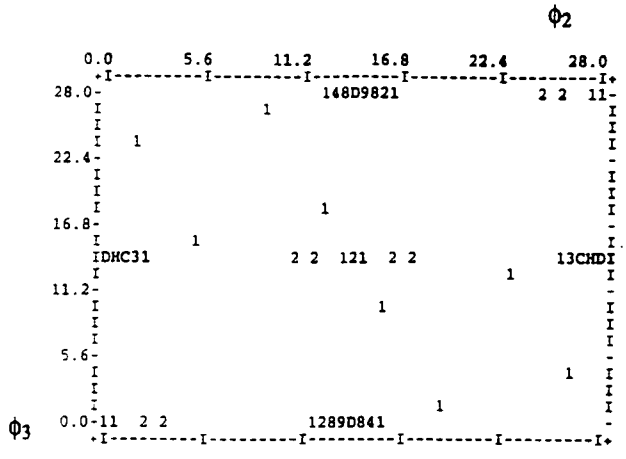
(c) cyclohepta-1,4-diene (XIV)



(d)cyclohepta-1,3,5-triene (XV)

Figure 3.13

Histograms of the BPRS θ coordinate for the fragments (XII)-(XV)



for θ > 40°

Figure 3.14

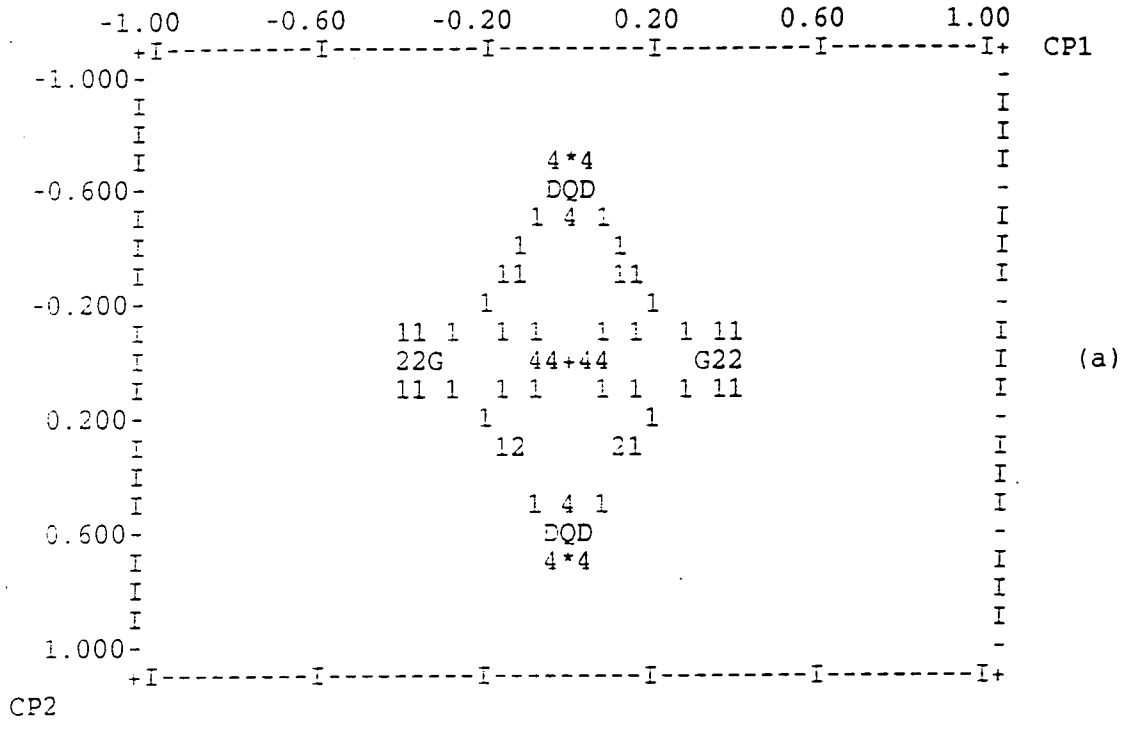
Scatterplot of the BPRS ϕ_2 and ϕ_3 coordinates for (XII) in units $\pi/14$

conformational space between the C/TC forms, and the true B/TB conformers at $\theta = 0°$. The ϕ_2-ϕ_3 plot for instances of (XII) having $\theta > 40°$ (Figure 3.14) shows isolated peaks at ϕ_2, $\phi_3 = \pi, 0; 0, \pi; \pi, 2\pi; 2\pi, \pi$, corresponding to the four permutational isomers and enantiomers of the C^4 conformation. The energy data of Table 2.4 are in agreement with these observations. For the dienes (XIII) and (XIV) and the triene (XV), which have additional conformational constraints, all conformers are intermediate between the chair-like and boat-like conformations, and therefore have θ values in the range 5-35° (Figures 3.13b-d).

Principal component (PC) and Cremer-Pople (CP) plots

PCA on the symmetry-expanded torsional datasets was an important feature in the conformational mapping of cycloheptanes (Section 3.2.3), where the PC's occurred in two sets of degenerate pairs; PC1, PC2 mapping the C/TC conformers, and PC3, PC4 mapping the B/TB conformers. With decreasing fragment symmetry, the degeneracy disappears and PC coordinates related to directions of maximum variance are less and less easy to interpret in relation to the mutually perpendicular toroidal pseudorotational surfaces of the parent cycloheptane ring (Figure 3.1). This is due to axial rotations within the four-dimensional hyperspace that reflect the reduced symmetry and a greater variability in chair-like/boat-like character than exists for the parent ring (Figure 3.1). For fragments (XII)-(XV), we report that, as expected, four PC's account for >99% of the total variance in all cases. The PC's occur in pairs as PC1, PC2 and PC3, PC4 where, for each fragment, one member of each pair maps a conformer of C_2 symmetry and the second member maps a conformer of C_s symmetry.

Plots based upon CP coordinates for (XII)-(XV) are, however, directly comparable to those for the parent cycloheptane, since they refer to a consistent axial frame defining the hyperspace. Thus, the CP1, CP2 plot for cycloheptene (XII) in Figure 3.15(a) shows only two peaks on the C/TC pseudorotation itinerary corresponding to the C^4, $C^{\bar{4}}$ forms. The plot also shows central density due to a number of boat-like conformers, and a pair of (symmetry-related) peaks along the CP1 axis, representing conformations that lie intermediate between the TC^4 and TB^4 forms.



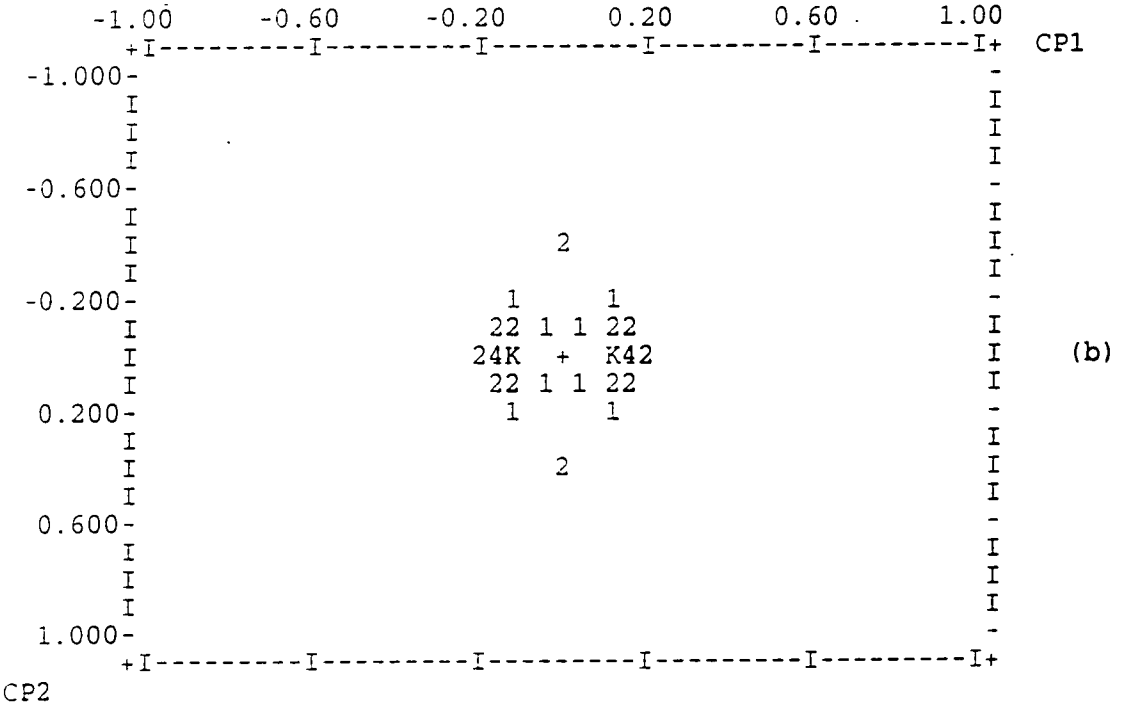


Figure 3.15

Scatterplots of the CP1 and CP2 coordinates for (a) cycloheptenes (XII) and (b) cyclohepta-1,3-dienes (XIII)

By comparison, the CP1, CP2 plot for cyclohepta-1, 3-diene (XIII) (Figure 3.15b) lacks any density on the C/TC circular pathway, and shows only the central boat-like region and some TC^4/TB^4 intermediates parallel to the CP1 direction. Plots similar to Figure 3.15(b) are obtained for (XIV) and (XV), and are not shown here.

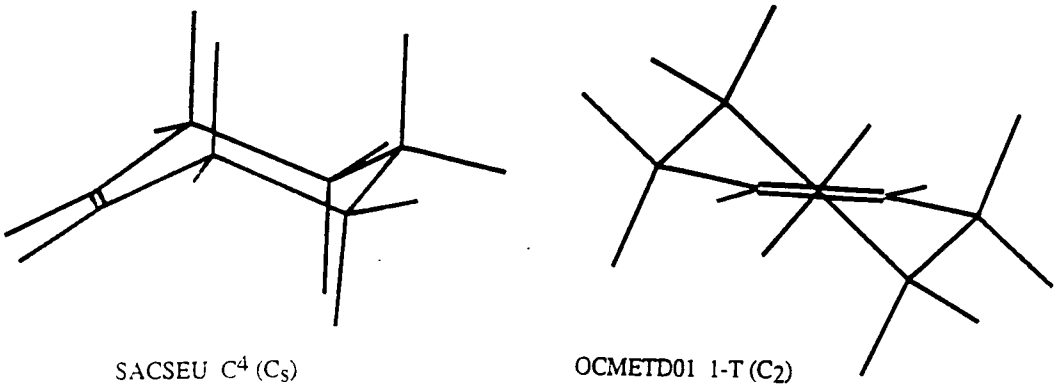
3.4.3 Conformational classification

The results of the symmetry-modified Jarvis-Patrick clustering applied to the torsional datasets (τ_1-τ_7) are shown in Table 3.12 for the fragments (XII)-(XV). Mean torsion angles are given for each discrete cluster having a population (Np) ≥ 3 fragments, and the appropriate conformational descriptor from Table 2.4 is assigned in each case. The Jarvis-Patrick criteria used for clustering the datasets to generate chemically acceptable results are shown in Table 3.11. For the smaller datasets (XIII)-(XV) some experimentation was necessary to optimise these criteria. However, as in earlier parts of this Chapter, it was found that a K_{JP} : K_{NN} ratio of close to 0.5 represented a suitable starting point, and the experimentation was primarily concerned with the selection of a suitable K_{NN} value for the smaller datasets.

In Appendix 2 the orthogonal coordinates, referred to molecular axes, for the most representative fragment (MRF) of each major ($N_p \geq 4$) cluster. The MRF is that fragment whose torsion angles lie closest to those of the centroid of a given cluster, and the conformations represented by these coordinates are shown graphically in Figure 3.16, and discussed below.

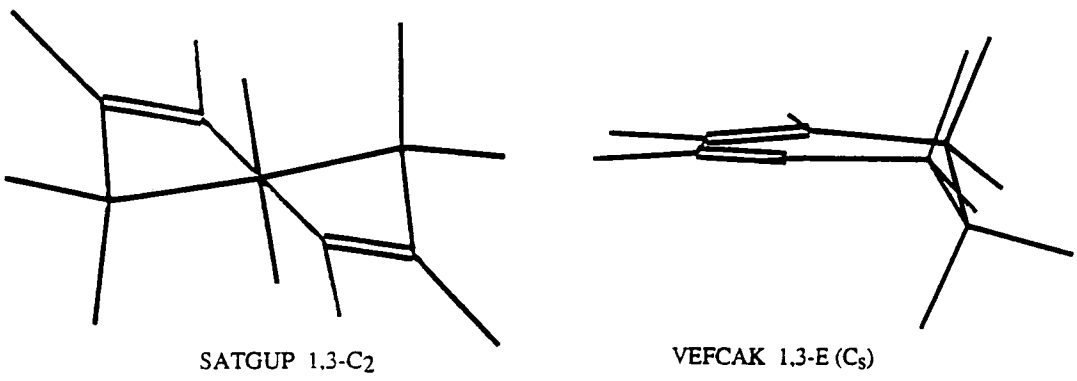
Cycloheptene (XII)

The 74 instances of (XII) are readily clustered as 48 C^4 conformers and 12 1-T conformers (Table 2.4). In both cases the asymmetric clusters formed initially are so close to a special position in conformational space that they coalesce with a symmetry-related counterpart within MULT=1.0 (Allen & Taylor, 1991). As a result, mean torsion angles for these clusters reflect perfect C_s (C^4) and C_2 (1-T) symmetry. Five of the 14 (unclustered) singletons show fusion to three- or four-membered rings and the



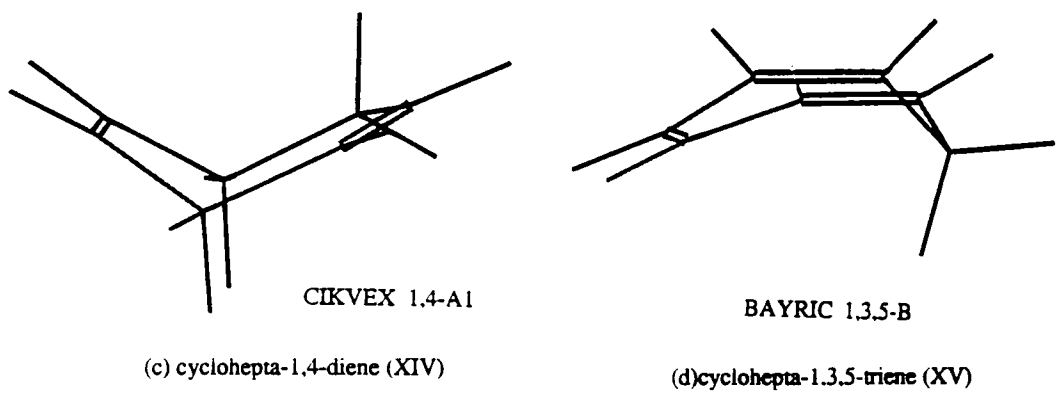
SACSEU C^4 (C_S) OCMETD01 1-T (C_2)

(a) cycloheptene (XII)



SATGUP 1,3-C_2 VEFCAK 1,3-E (C_S)

(b) cyclohepta-1,3-diene (XIII)



CIKVEX 1,4-A1 BAYRIC 1,3,5-B

(c) cyclohepta-1,4-diene (XIV) (d)cyclohepta-1,3,5-triene (XV)

Figure 3.16

The most representative fragments for the major conformational subgroups of fragments (XII)-(XV). The 1,3-C_s form of (XIII) is included for comparison purposes.

cycloheptene ring is effectively converted to a 1, 3- or 1, 4-diene: torsion angles at the fusion bonds are constrained to be close to zero, and the seven-membered rings adopt conformations (*e.g.* 1, 3-T; 1, 3-A; 1, 4-A1 of Table 2.4) which are typical of the appropriate diene. The remaining nine singletons adopt a range of asymmetrically distorted conformations in which multiple fusion to unsaturated five-membered rings is a common environmental problem.

Examination of the local chemical environments of the 12 higher energy 1-T forms shows that seven examples contain substructure (**M9**), in which one or both of the C^* atoms is $C_{sp}2$ and carries an exo double bond. However, the 1-T conformation

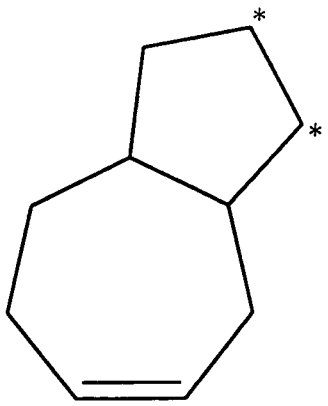
Table 3.12

Jarvis-Patrick cluster analysis results for the four datasets. The mean torsion angles for each cluster are given together with the esd of the means, and the CSD refcode of the most representative fragment in that particular cluster.

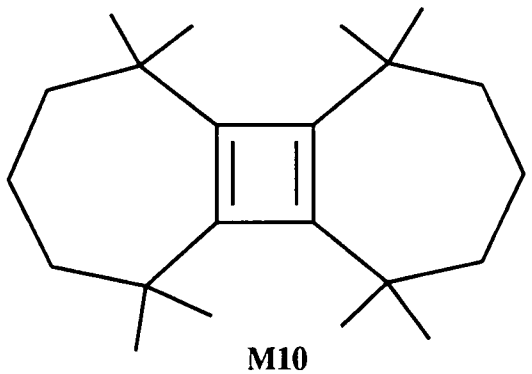
Frag.	Cluster no.	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
XII	1	-67.6(11)	73.7(12)	-56.4(17)	0.0(0)	56.4(17)	-73.7(12)	67.6(11)
	2	-41.2(40)	73.2(13)	-26.9(30)	-6.9(23)	-26.9(30)	73.2(13)	-41.2(40)
XIII	1	-41.5(32)	72.2(17)	-4.3(19)	-44.3(18)	-4.3(19)	72.2(17)	-41.5(32)
XIV	1	40.0(49)	1.1(10)	11.9(54)	-63.5(51)	66.4(14)	3.8(24)	-60.0(22)
XV	1	-62.8(16)	6.9(6)	30.7(6)	0.0(0)	-30.7(6)	-6.9(6)	62.8(16)
	2	-67.6(11)	-2.3(20)	43.9(25)	0.0(0)	-43.9(25)	2.3(20)	67.6(11)
	3	-73.6(24)	11.0(23)	31.1(11)	0.1(6)	-31.1(6)	-11.3(9)	74.1(8)

Fragment	Cluster no.	Conformation	N_P	MRF
XII	1	C_4	48	SACSEU
	2	1-T	12	OCMETD01
XIII	1	1, 3-T	13	SATGUP
XIV	1	1, 4-A1	8	CIKVEX
XV	1	1, 3, 5-B	25	BAYRIC
	2	1, 3, 5-B	5	KEXNIK
	3	1, 3, 5-B	3	

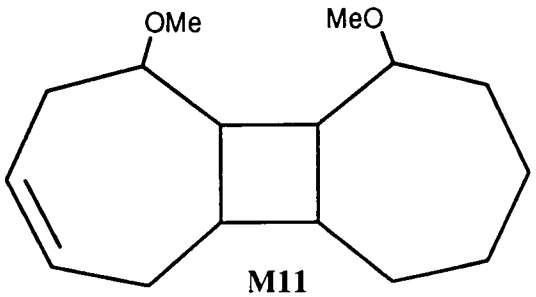
is not exclusive to (**M9**) since there are several clear examples of a C^4 conformation in this case. The remaining five 1-T conformers arise from the tricyclic systems (**M10**) [OCMETD01: Irngartinger *et al.*, 1988 ($P\bar{1}$, room temperature data); OCMETD10: Irngartinger & Nixdorf, 1988 ($P\bar{1}$, data at 100K)], and (**M11**) [MEOCHP: Jendralla, 1980]. however, (**M10**) is also observed in a flattened C^4 conformation in a monoclinic ($P2_1/n$) polymorph, while the cycloheptene ring in (**M12**) [TCTDOL: Courtois, Protas, Fixari & Brunet, 1975] also adopts the lower energy chair form. The chair form of (**M10**) is associated with a planar cyclobutadiene, while the 1-T form is associated with a folded four-membered ring.



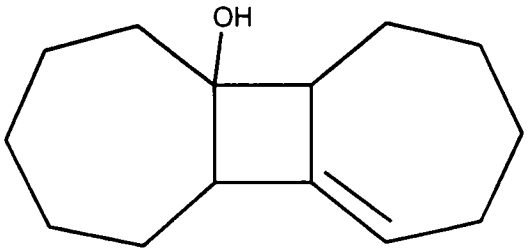
M9



M10



M11

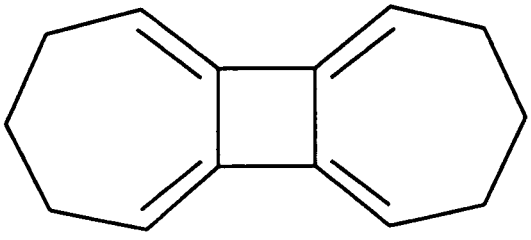


M12

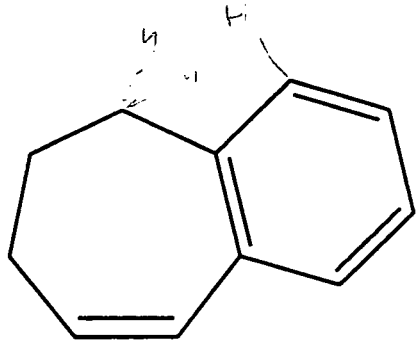
The calculations reported by Allinger & Sprague (1972) and by Burkert & Allinger (1982) would indicate that the C^4 seven-membered ring conformation in benzocycloheptenes is more energetically favoured (with respect to 1-T), than is the case for the parent (XII). Eight of the ten instances of benzocycloheptene in our dataset are, indeed, highly symmetrical C^4 conformers; the other two instances are distorted by three-ring fusion.

Cyclohepta-1, 3-diene (XIII)

This small dataset of 20 fragments is dominated by a single cluster comprising 13 instances of the 1, 3-T conformation of C_2 symmetry. As with (XII), all instances are very close to the ideal symmetry and coalescence of symmetry-related clusters occurs within MULT=1.0 (Allen & Taylor, 1991). Six of the remaining seven singletons have small-ring fusion (three- and five-membered rings) which generate distorted variants of the 1, 3, 5-B conformation typical of (XV) (see Table 2.4). It is interesting that even a fused and saturated cyclopentane ring consistently has this effect. Only one fragment, from structure (**M13**) [VEFCAK: Hashmi, Polborn & Sziemies, 1989]. adopts the C_s 1, 3-E conformation (Figure 3.16b) that is observed in the gas-phase experiments, and which is consistently predicted to be the global minimum for (XIII). Given the fusion of the four-membered ring at the =C-C= single bond, it would be reasonable to regard this 1, 3-E conformation as being forced by the local chemical constraints.



M13



M14

At first sight, then, it would appear that the predominance of the C_2 (1, 3-T) form of (XIII) is at variance not only with the energy calculations, but also with the electron-diffraction (Chiang & Bauer, 1966; Hagen & Traetteberg, 1972) and microwave data (Avirah, Molloy & Cook, 1979). Indeed, the C_2 conformation might have been expected by analogy to the structures of 1, 3-cyclohexadiene (Traetteberg, 1970) and of 1, 3-cyclooctadiene (Oberhammer & Bauer, 1969; Traetteberg, 1968) which show C=C-C=C torsion angles of 18° and 38° respectively. However, examination of the chemical environment in the 13 structures that exhibit the 1, 3-T (C_2) conformation for (XIII) show that all of these instances are unrepresentative of the free parent molecule. In 10 of these structures, one (7) or both (3) of the 1, 3-double bonds arise from benzo-fusion, and in all seven mono-fused rings the ethylenic double

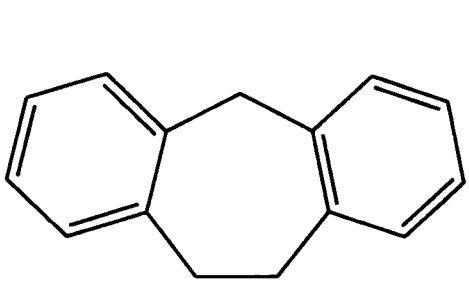
bond is further conjugated to -C=C, -C=O or phenyl substituents. In the three remaining cases, one (2) or both (1) of the ethylenic double bonds are conjugated to -C=C or -C=O groups.

We have investigated some relevant features of the potential energy hypersurface for the parent (XIII) and the mono-benzo fragment (**M14**) using three different force-fields in the program systems: (1) the PCRNGMIN procedures in COSMIC90 (Morley *et al.*, 1991; Allen, Howard, Pitchford & Vinter, 1994); (2) the MM2 force-field within the CaChe system (Tektronix Inc., 1990), and (3) the DISCOVER module of INSIGHT 2.0 (Biosym Technologies, 1992). For the parent (XIII), all three systems show the 1, 3-E (C_s) form as a global minimum, with the 1, 3-T (C_2) form at a consistently higher energy. The ΔE (C_2-C_s) values ($kJmol^{-1}$) are 6.7 (COSMIC), 4.8 (CaChe) and 3.4 (INSIGHT). The results were obtained using a typical C_2 conformer as the starting point. A similar treatment of a typical example of (**M14**) (C_2 starting conformation of the seven-ring) using both CaChe and INSIGHT retains the C_2 conformation as a global minimum, and with torsion angles which differ very little from the X-ray mean values listed in Table 3.12. Thus, the 45° twist about the C=C-C=C single bond minimises H - H interactions in the case of benzo-fusion, a point which is discussed at some length by Allinger & Sprague (1972), with respect to conformational differences between cycloheptene (XII) and its benzo-derivative. In Section 5.5.1 the details of a crystal structure determination for a 1, 3-cycloheptadiene will be given. This molecule has no benzo-fusion and exists as the 1, 3-E (C_s) form. This result would seem to emphasise the strong conformational bias that exists in this dataset for the 1, 3-T (C_2) form, which occurs because of benzo-fusion, but which is not the minimum energy form for this fragment: the 1, 3-E (C_s) form being clearly preferred in molecules which do not involve benzo-fusion.

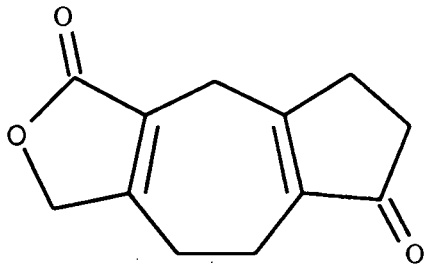
Cyclohepta-1, 4-diene (XIV)

This very limited dataset of just 13 fragments proved particularly difficult to analyse using automated clustering methods. The Jarvis-Patrick analysis with K_{NN}=4, and K_{JP}=2 yielded a coherent grouping of eight fragments (Table 3.16) having the

asymmetric 1, 4-A1 conformation identified by Favini, Maggi & Todeschini (1983) as the minimum-energy form. Two further singleton fragments have significantly flattened versions of this same conformation. All but one of these ten 1, 4-A1 conformers are simple derivatives of the 10, 11-dihydro-5H-dibenzo[a, d]cycloheptene substructure (**M15**). Two further singletons are effectively 1, 3, 5-trienes due to cyclopropyl fusion (and have the appropriate 1, 3, 5-B conformation of Table 2.4). This leaves a single example of the second asymmetric form (1, 4-A2) of Favini, Maggi & Todeschini (1983) in the unique substructure (**M16**) [JASGIT: Tochtermann *et al.*, 1989].



M15



M16

Cyclohepta-1, 3, 5-triene (XV)

All 33 examples of the rigid ring (XV) exhibit the 1, 3, 5-B (C_s) conformation of Table 2.4. The Jarvis-Patrick analysis, run with a number of K_{NN} and K_{JP} variants, consistently dissects the multivariate torsional distribution into three subgroups (Table 3.12) which are highly coherent within themselves. Both clusters 1 (N_p=25) and 2 (N_p=5) are highly symmetrical and coalesce with their symmetry-related counterparts within MULT=1.0 (Allen & Taylor, 1991). The mean torsion angles for the largest cluster (1) are very close to the calculated values of Saebo & Boggs (1982). Cluster 2 differs from cluster 1 primarily in the value of τ_3 (=τ_5). This angle ranges from 20.1° to 35.5° in cluster 1, and from 37.6° to 51.6° in cluster 2. Cluster 3 differs from cluster 1 in having approximately 11° torsion angles about the "outer" double bonds. Whilst these dissections of the data are a mathematical nicety, there are no obvious chemical constraints that can be correlated with the individual groupings.

3.5 Eight-membered rings

3.5.1 Database search and retrieval

Throughout this study Version 4.5 (1st Jan 1992) of the CSD System was used for substructure search, coordinate retrieval and data analysis using the programs QUEST and GSTAT (Allen *et al.*, 1991). The chemical substructure under investigation consists of a ring of $C_{sp}3$ atoms, connected by single bonds, the search being further constrained by the use of CSD bit-screen commands (CSD User Manual, 1992) to locate entries with (a) atomic coordinates available, (b) no residual numerical errors following CSD check procedures, (c) no reported disorder in the crystal structure, and (d) "organic" compounds according to the CSD definitions. Eight-membered rings encountered within very complex ring systems, *i.e.* those failing CSD ring identification procedures, can be eliminated using the SCREEN -620 command. These highly complex systems proved too complicated for analysis, and so the bridged molecules are not examined further here. Subsequently, a dataset 8C1 (fragment (XVI), Figure 2.7a), containing 32 of the unbridged cyclooctane fragments was generated.

Further searches using the same bit-screen restrictions were performed to give datasets for other eight-membered ring systems of interest. Datasets of C-eight-membered ring systems containing unsaturated bonds were generated using variable bond type definitions available in the QUEST program. In this fashion true alkene bonds and those with some degree of aromatic character (due to an adjoining aromatic ring system) can be located simultaneously (bond types 2 or 5 were used in the QUEST procedures). Dataset 8C2 (fragment (XVII) in Figure 2.7a) contains 24 examples of eight-membered C-rings with only one unsaturated bond present using the search procedures given above. Two further datasets, 8C3 (XVIII) and 8C4 (XIX) were also extracted from the CSD. These datasets contain 21 examples of 1,3-dienes (XVIII), and 11 examples of 1,5-dienes (XIX) respectively. Statistically insignificant numbers of the other possible unsaturated rings render them unacceptable for further analysis.

The only other C-eight-membered rings of any note, are those with an external double bond to carbon, oxygen or nitrogen. These form the basis of the dataset 8C5

(XX in Figure 2.7a), with all of the examples grouped together. There are 12 examples of these systems, after all previous search restrictions have been applied.

Eight-membered rings containing one or more hetero-atoms are also moderately prevalent in the CSD. In the case of one such dataset, 8A1, one of the 8 Csp^3 atoms in the singly-bonded ring is replaced in a QUEST search by an "Any Atom" label, thus allowing eight-membered rings with 7 Csp^3 atoms and 1 undefined non-hydrogen atom to be located (XXI in Figure 2.7a). Again, all of the bit-screen search mechanisms used in the previous investigations are maintained here. As a result, 10 entries are encountered with a single hetero-atom in the eight-membered ring. By replacing yet another Csp^3 atom of the ring by an undefined atom, with similar search conditions enforced, 12 entries are encountered for dataset 8A2. All of these entries have the hetero-atoms present in only the 1 and 5 positions of the ring (XXII). Finally, with four hetero-atoms present in the ring at the 1, 3, 5, and 7 positions (XXIII) a total of 24 entries are located in the CSD forming the dataset 8A3. CSD reference codes for all of the datasets are given in Appendix 1.

3.5.2 Data analysis

The eight-membered ring fragments were re-located in the crystallographic connection tables of datasets 8C1 *etc.* by use of the program GSTAT, where the bond length constraints of 1.44Å to 1.70Å are used to approximate a C_{sp3}-C_{sp3} bond length. The conformation of each fragment which passes these criteria was described in terms of the eight intra-annular torsion angles, τ_1 to τ_8, of the ring (see Table 2.6).

Each fragment has then to be symmetry-expanded, in order to account for all of the enumerational possibilities in the eight-dimensional hyperspace (see Chapter 2.1.8) of the topological parent point group symmetry (D_{8h}). For 8C1 therefore, the required expansion is 32-fold, and the number of fragments increases from 32 in the raw dataset, up to 1024 in the symmetry-expanded dataset. In datasets where the parent fragment is of lower topological symmetry, such as the D_{2h} symmetry for the 1,5-di-hetero-cyclooctane fragment, the required symmetry-expansion is only 8-fold.

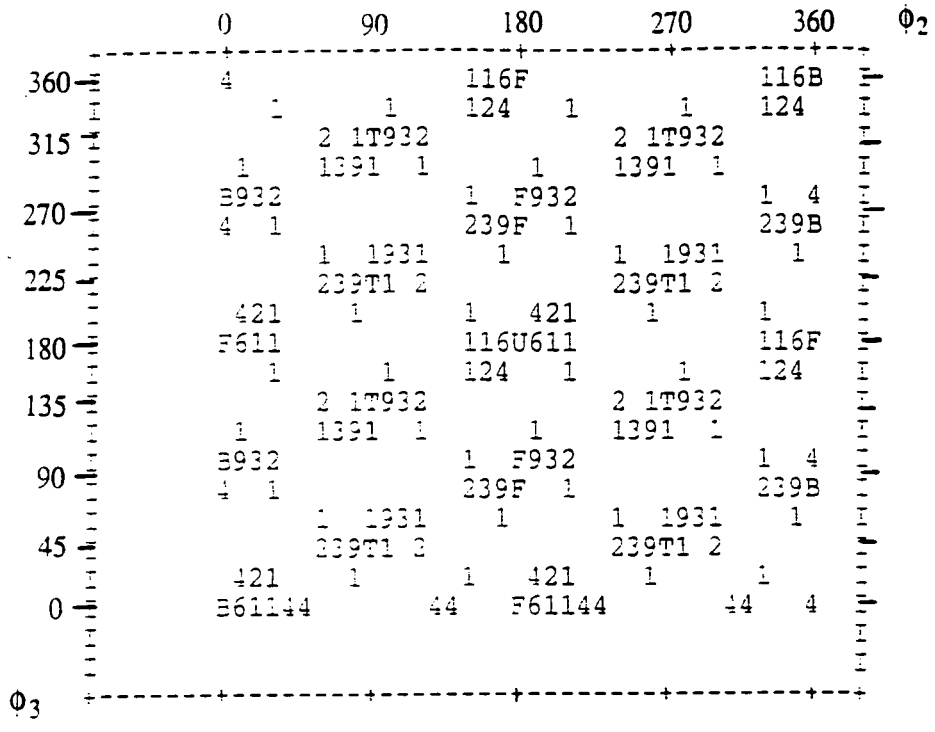


Figure 3.17

The ϕ_2-ϕ_3 scatterplot for dataset 8C1 showing a pathway similar to that postulated in Figure 2.9

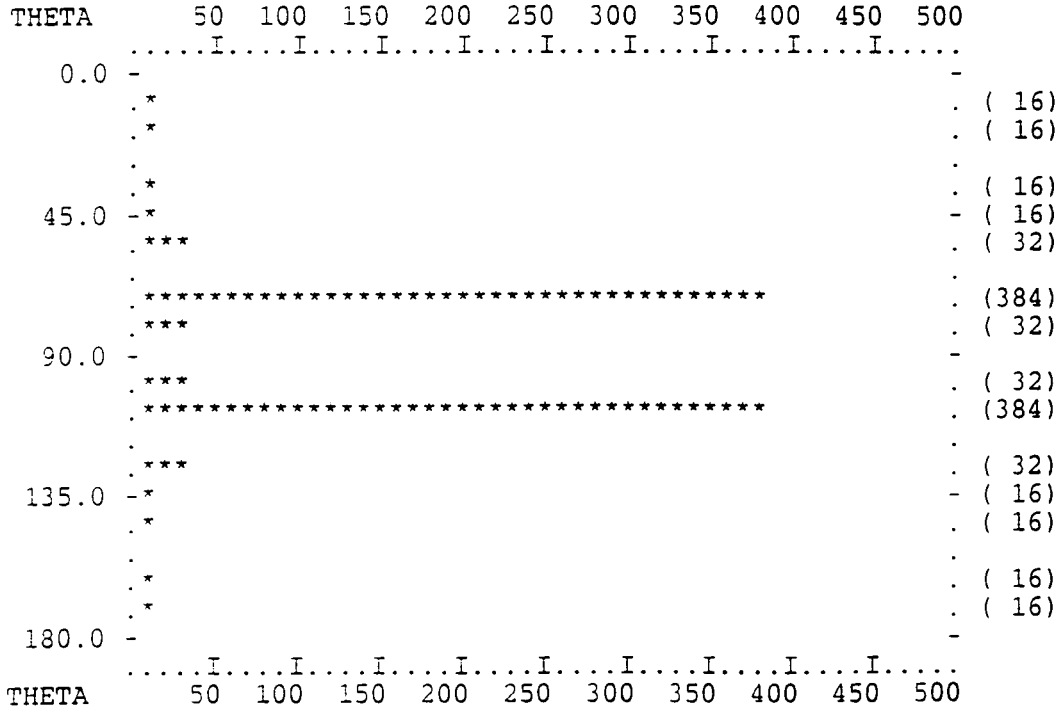


Figure 3.18

θ-histogram for dataset 8C1

For each fragment in the expanded dataset the CP parameters q_2, ϕ_2, q_3, ϕ_3, and q_4 were generated by GSTAT, and the BPRS coordinates ρ, θ were calculated using the GSTAT TRAnsform facility, using an instruction set analogous to that shown in Figure 3.1(b). Classification of conformations was performed using the symmetry-modified cluster analysis algorithms described in detail in Chapter 2.3

3.5.3 Conformational mapping

BPRS Coordinate plots

The BPRS coordinates generated by GSTAT can be used to explore the conformational diversity of the 8C1 dataset. The symmetry-expanded ϕ_2 - ϕ_3 map for 8C1 over the entire conformational hypersurface in θ (0°-180°) gives a plot (Figure 3.17) not dissimilar in structure to those obtained from the study of cycloheptane fragments (Chapter 3.2). The major pseudorotational pathway in evidence here however, is that of the BC / TBC pathway, situated at θ = 75°, 105° (see Figure 2.9). This pathway is not a continuous one though, with all of the density along the path being concentrated at those positions corresponding to the BC conformers, thus highlighting the dominance of the BC conformation in this dataset.

The θ - ϕ_2 plot shown in Figure 3.18 illustrates a situation analogous to the Evans & Boeyens construct (Figure 2.8) with most of the density concentrated along lines equivalent to θ = 75°, 105° positions. Small amounts of density at the other positions, notably θ = 0°, 180°, are representative of possible crown and CC / TCC pathways. The θ-histogram in Figure 3.18 also highlights the preponderance of the BC structures over and above most others with the bands at θ = 75°, 105° being the most highly populated.

Cremer-Pople (CP) coordinate plots

The two phase-amplitude pairs (q_2, ϕ_2) and (q_3, ϕ_3) describe two circular pseudorotation pathways, the fifth factor, q_4, being a unique amplitude corresponding

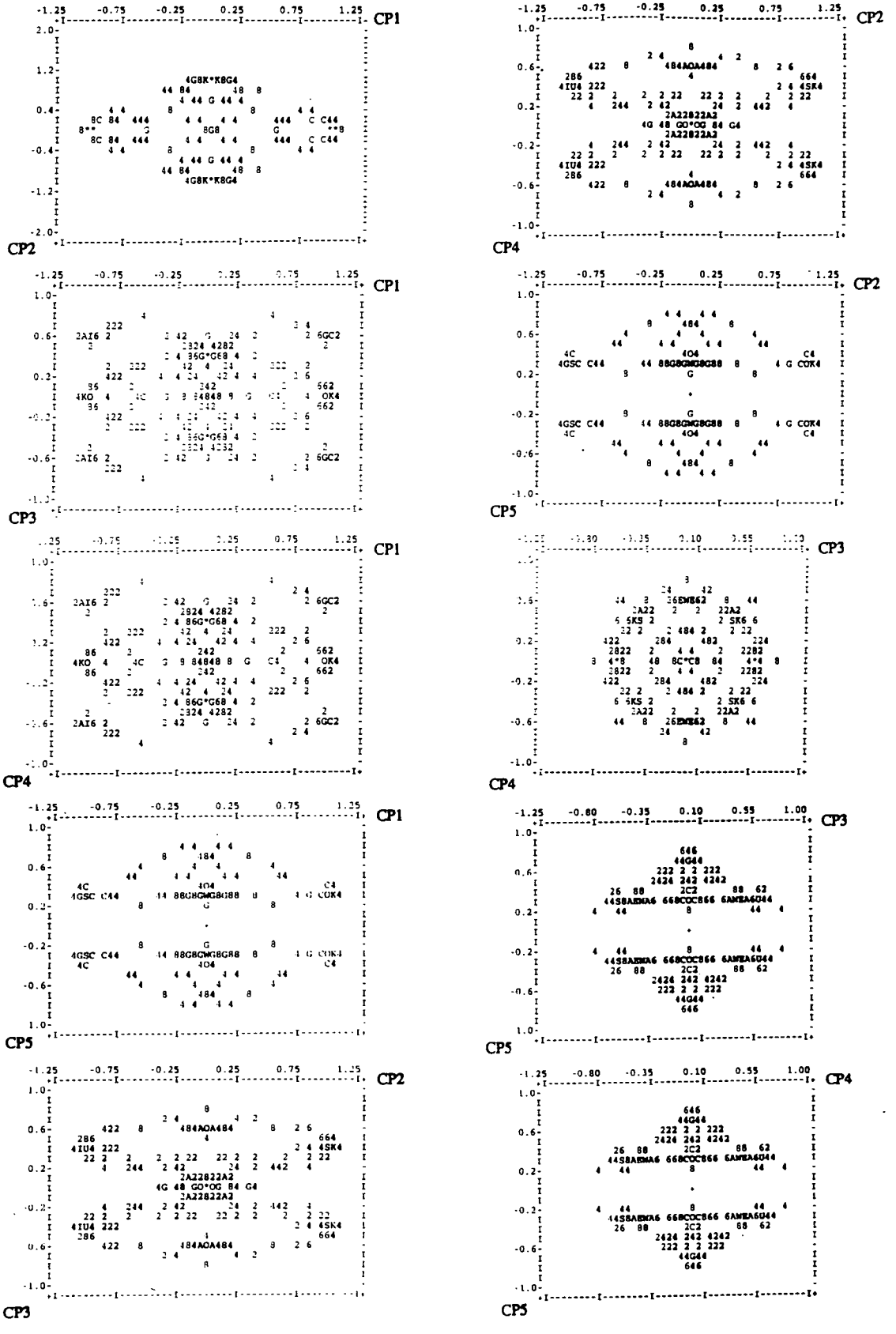


Figure 3.19

CP scatterplots for the symmetry-expanded 8C1 dataset

in the case of eight-membered ring systems to the unique crown conformer of D_{4d} symmetry. These CP coordinates for 5-dimensional space can be re-expressed in a Cartesian form using the following transformations, and visualized using Figure 2.8: CP1 = $q_2 \cos \phi_2$, CP2 = $q_2 \sin \phi_2$ map the B / TB conformers at $\theta = 90°$ (the equator) ; CP3 = $q_3 \cos \phi_3$, CP4 = $q_3 \sin \phi_3$ map the C / TC pathway along the meridians ; CP5 = q_4 - mapping the crowns at the poles of the sphere, $\theta = 0°$ and 180°.

The ten possible 2D scatterplots obtained from the Cartesian values for the 1024 fragments of the symmetry-expanded dataset 8C1 are shown in Figure 3.19. The true 5-dimensionality of the conformational space is apparent from the orthogonality of the five CP axes. Visualising the orthogonal views proves difficult unless the dataset is restricted to only the BC / TBCs, using a SELect $\theta = 70° - 80°$ command in GSTAT, and a torus is constructed (see Section 2.1.3 for details of the construction a pseudorotation torus for the 7-membered systems) for the 8N-pseudorotation itinerary. Orthogonal views seen with the toroidal model of the pseudorotational pathway are immediately identifiable with the patterns established with the 2D scatterplots, so demonstrating the nature of these CP-axes. This selected range of θ offers the simplest possible representation of the pseudorotation itinerary of the eight-membered rings. When θ is allowed to run over the entire 0°-180° range, and the inverse conformations are also taken into account, the CP plots of Figure 3.19 result.

In order to fully appreciate the axial relationships in CP space, one BC fragment (the MRF of cluster 1 from 8C1) was isolated, and the CP coordinates generated by GSTAT were input into the plotting progam MacMoMo (Dobler, 1990) on the Apple Macintosh computer. The coordinates are entered and viewed in sets of three axes, and pseudomolecules are constructed by joining adjacent equivalents for the BC conformer, such that the rotations of the pseudorotation pathways could be observed. Two sets of "snapshots" are shown. Figure 3.20(a) shows the patterns observed for the CP coordinate sets when θ is restricted to 0°-90°, and demonstrates the inter-relationships thus established by the appropriate axial rotations. Confining θ to this range enables us to consider only one of the two BC tori that exist for this fragment, and hence offers the most simplistic representation. When θ is allowed to run over the range 0°-180° the

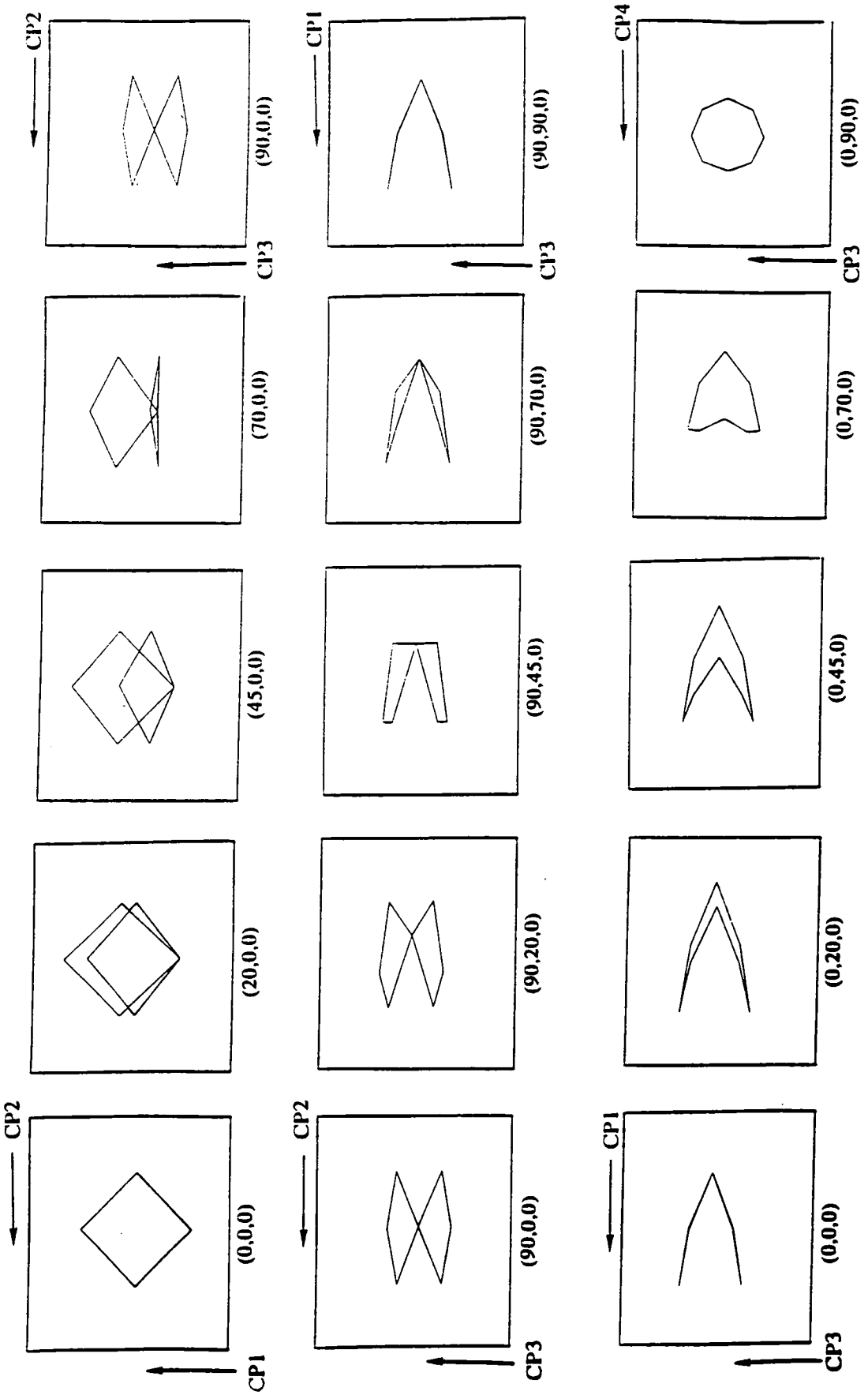


Figure 3.20(a)

Axial rotations of the CP axes for a BC conformation: θ restricted to 0°-90°

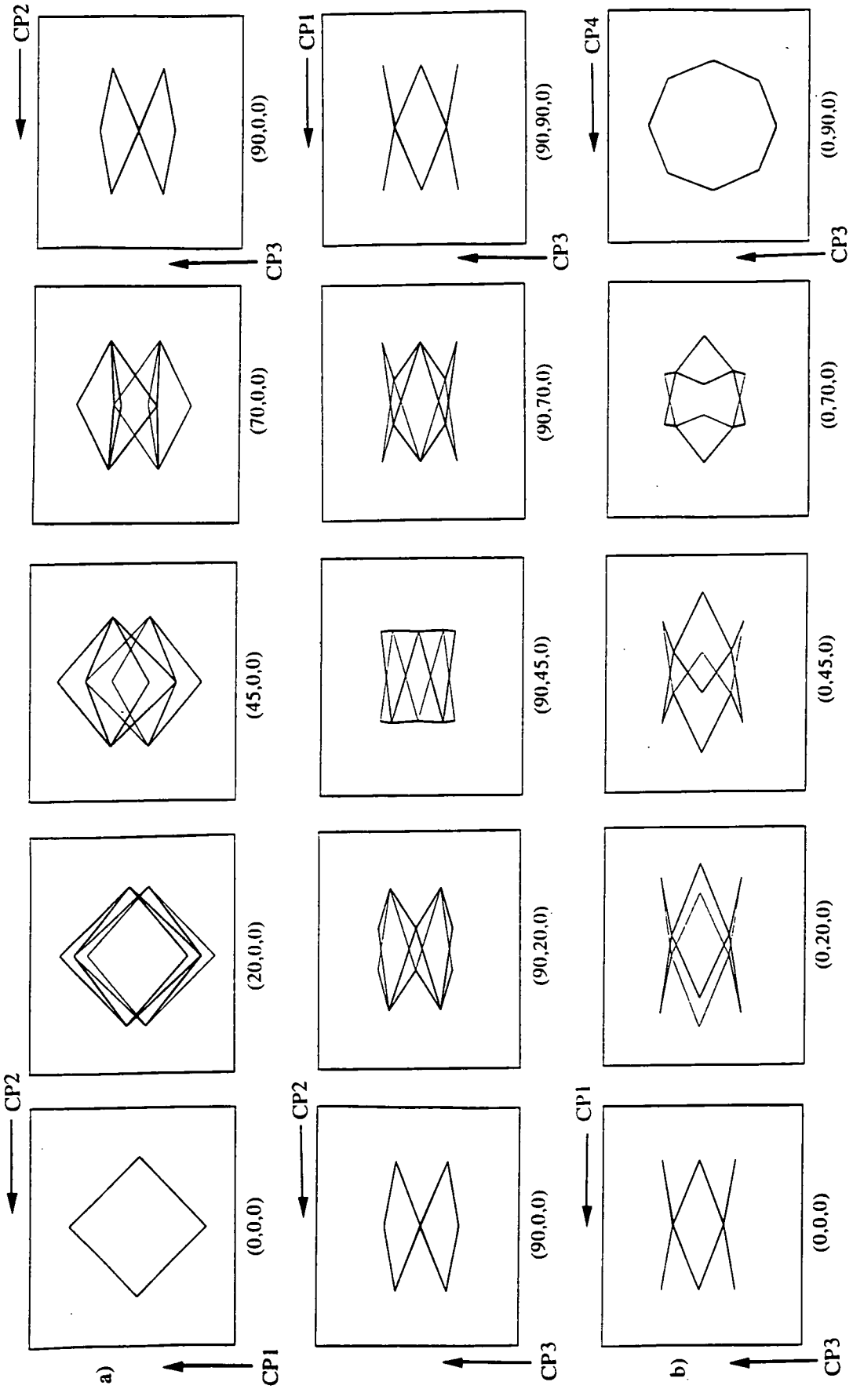


Figure 3.20(b)

Axial rotations of the CP axes for a BC conformation: θ in range 0°-180°

second torus superimposes itself onto the previously observed plots, as can be seen in Figure 3.20(b). It now becomes obvious that the second torus effectively runs in the opposite sense to the first, a point that can best be illustrated by comparing the CP1-CP3 plots obtained for both cases. With these simple rotational representations in mind, analysis of the 2D CP scatterplots generated by GSTAT can now be performed, and the orthogonal relationships of the axes can be appreciated.

The individual views of the 2D scatterplots can now be analysed. The CP1-CP2 plot, is a view of the tori from an orthogonal axis running from pole to pole of the sphere of Figure 2.8. Figure 3.17(b) shows that, for instance, the BC conformers exist at only 4 specific values of ϕ_2 (*e.g.* four conformers exist with $\phi_2 = 180°$). This fact is mirrored in the CP plot where the outer clumps of data density can be envisaged as being the 4 distinct groups of conformations on the BC / TBC torus, the density at the origin as representing the crown conformers, and the radial density emanating from the origin as representing the C / TC conformers which run as meridians in Figure 2.8. The CP1-CP3 and CP1-CP4 plots are equivalent views down two different orthogonal axes, and illustrate one "side on" view of the pseudorotational pathways present on the tori. The patterns observed can be correlated directly with those patterns seen previously (Figure 3.20). It must be remembered that the inverse conformations are also being included in these plots, together with the data for conformations which have no real toroidal dimensions, such as the CC / TCC torus of q_3=0, which form the line of data points through the middle of both plots. The CP2-CP3 and CP2-CP4 scatterplots are again equivalent views from two different orthogonal axes, but this time they represent the side on view of the torus rotated through 90° from those obtained in the CP1-CP3 and CP1-CP4 plots. The CP3-CP4 plot is the one obtained by looking down the axis that runs effectively through the middle of the toroidal rings. Here the number of possible ϕ_3 values for a certain conformation is eight, and this is shown by the plot. There are data point rings for three different conformations represented in the plot. The outer ring of data points is that seen for the BC / TBC torus and shows the eight very distinct and highly populated possible positions. The inner ring corresponds to the TCC / CC examples, and the data centred at the origin, to the crown conformations. All

those plots that show a CP5 dependency are effectively oblique views of the overall conformational space as illustrated in Figure 2.8, *i.e.*. a view of the sphere as a whole. The plot of CP2-CP5 for instance shows the highly populated bands of the BC / TBC running parallel to each other, with the crown conformations visible at the poles of the sphere upon which the data can be projected, and the meridians of the C / TC conformations emanating away from the polar positions. All of the other CP5 plots can be described similarly.

Symmetry-modified Principal Component Analysis (PCA)

PCA was performed on the eight torsion angles of each fragment of the symmetry-expanded dataset 8C1, the results being shown in Table 3.13. The top five PCs account for 99.9% of the total variance of the set, with a major factor and two degenerate pairs, as expected from the group-theoretical analysis of Bocian *et al.* (1975). Hence PC1 accounts for 37% of the total variance, PC2 and PC3 for 42.9% together, and PC4 and PC5 for 19.9% of the total variance. The PC loadings for 8C1 (Table 3.13) show that PC1 has the C_4 symmetry associated with the crown

Table 3.13

Results of the symmetry-modified Principal Component Analysis for Dataset 8C1.

(a) Variance analysis

Variance (%)

No.frags	PC1	PC2	PC3	PC4	PC5	Total
1024	37.268	21.450	21.450	9.889	9.889	99.946

(b) PC loadings

	PC1	PC2	PC3	PC4	PC5
$\%\tau_1$	44.868	27.609	39.435	-31.461	-8.865
$\%\tau_2$	-44.868	8.362	-47.407	8.865	-31.461
$\%\tau_3$	44.868	-39.435	27.609	31.461	8.865
$\%\tau_4$	-44.868	47.407	8.362	-8.865	31.461
$\%\tau_5$	44.868	-27.609	-39.435	-31.461	-8.865
$\%\tau_6$	-44.868	-8.362	47.407	8.865	-31.461
$\%\tau_7$	44.868	39.435	-27.609	31.461	8.865
$\%\tau_8$	-44.868	-47.407	-8.362	-8.865	31.461
Sym.	Crown	C	C	B	B

conformers, PC's 2 and 3 possess the mirror plane symmetry of the chair conformers, and PC's 4 and 5 have the boat type C_2 symmetry of loadings, as expected.

The CP plots and the PC plots have essentially the same form, but it must be noted that the two do not correlate directly. Hence the variances and loadings observed for the PC's can be explained in terms of the boat and chair characteristics of the BC conformations of which the dataset is largely composed.

3.5.4 Classification of conformations

The cyclooctane fragments were analysed in terms of their torsion angles using the symmetry-modified Jarvis-Patrick clustering algorithm (Jarvis & Patrick, 1972) in a similar procedure to the one described earlier (*e.g.* Section 3.2). Again, the number and composition of clusters formed by the algorithm was found to be highly dependent upon the K_{JP} / K_{NN} ratio. Values close to 0.5 generally yield chemically sensible clusters of fragments. After much testing with the 8C1 dataset the optimum results were achieved using $K_{NN} = 8$, $K_{JP} = 5$, $n = 2$ and $D_{lim} = 0.10$. Mean torsion angles and statistical descriptors for the three major clusters are given in Table 3.14.

Of the 32 fragments present in 8C1, 12 can be classed as pure BC conformers (cluster 1), 6 are classified as distorted TBCs with a geometry very similar to that of the [BC / TBC] conformation described by Anet & Krane (1973) (cluster 2), and 4 more are of a distorted BC / TBC type conformation (cluster 3). The centroids of cluster 1 are sufficiently close to a two-fold axis and a mirror plane, respectively, for coalescence with the symmetry-related counterparts to occur (Allen and Taylor, 1991). The results in Table 3.14 therefore relate to the 24 fragments in the coalesced cluster 1. Ten fragments remain unclustered and these singletons were analysed manually and seen to consist of one TCC and two crown conformations, with the rest of the outliers coming from severely distorted rings.

Runs performed at higher K_{NN} coalesce the three clusters into one, without managing to cluster any of the other remaining fragments from the optimum search. As this procedure does not provide any better clues as to the overall structure of the

Table 3.14

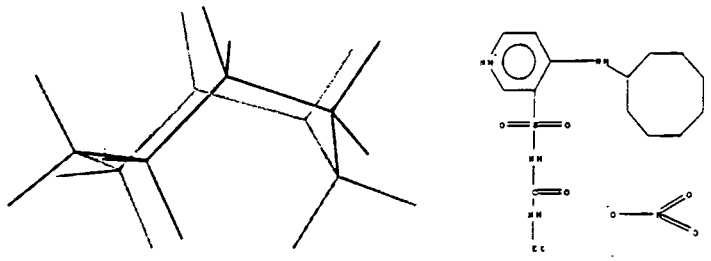
*Mean torsion angles and symmetry descriptors for the clustered data is marked *. Representative geometries for the other major conformations are also given*

dataset	refcode	symm.	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8
8C1	BAGPII*	BC	66.5	40.9	-99.2	66.4	-66.5	103.8	-45.3	-64.2
	COVLUU*	BC/TBC	78.1	24.2	-91.6	67.9	-64.7	104.8	-53.9	-57.0
	SPTZBN*	BC/TBC	52.8	53.8	-90.5	-0.6	79.4	-75.2	75.3	-105.4
	DEZPUT	Crown	70.4	-83.2	92.3	-73.3	63.8	-82.3	96.5	-82.0
	EOCNON10	TCC	47.7	-84.7	134.4	-85.3	48.7	-82.4	124.9	-80.7
8C2	IPRINC*	cis	0.3	87.1	-48.0	-53.8	104.0	-74.5	74.9	-79.7
	COCNBZ01	trans	138.1	-87.9	52.1	-80.0	112.5	-79.7	49.1	-86.9
8C3	GEDVAM*	C2	0.0	60.9	2.0	-96.8	80.9	-47.8	84.0	-96.8
	BUBXUR	C1	0.1	-70.5	8.4	83.5	-32.6	-72.3	54.8	28.8
8C4	TAFZOP	BC	0.4	-74.0	110.0	-72.8	-0.4	74.0	-110.0	72.8
	DBCODE	BB	-0.9	11.1	65.1	-81.7	-0.9	11.1	65.1	-81.7
8C5	AOCINB	BC(3)	71.7	32.4	-100.6	70.5	-62.7	103.3	-54.0	-59.4
	DEJJUX	BC(2)	63.4	47.8	-96.8	60.7	-70.3	103.1	-36.4	-69.8
8A1	LAURCN10	BC/TBC	69.0	47.1	-109.0	59.9	-57.2	100.5	-53.1	-57.5
	KEVBES	CC/TCC	77.0	-110.6	101.9	-73.6	86.4	-98.0	84.2	-71.6
8A2	CARYAV	BC	59.6	53.2	-109.6	64.6	-68.4	106.3	-46.9	-63.5
	BPHZOC10	Crown	94.5	-73.0	80.7	-101.3	96.6	-77.4	82.3	-103.1
8A3	TETIOC10	BC	79.4	51.1	-109.0	80.5	-81.9	109.9	-47.9	-81.8
	TOXOCN	Crown	83.7	-83.6	93.6	-94.0	83.7	-83.6	93.6	-94.0
	OMSIOC	BB	60.7	53.0	-57.1	-56.3	60.7	53.0	-57.1	-56.3

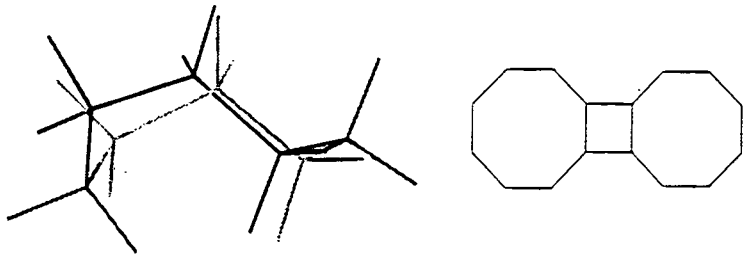
dataset, the parameters detailed above were retained as the best descriptors. The cluster analysis gives "most representative fragments" for each cluster located, and these MRF geometries are presented in Table 3.14 and shown in Figures 3.21, 3.22 and 3.23.

3.5.5 Conformational variety and chemical environment

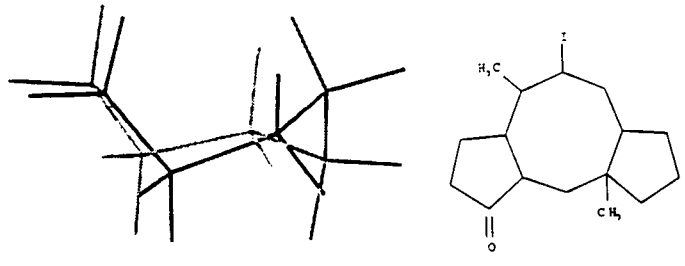
Having clustered the dataset 8C1 it is now essential to analyse the results and to try to rationalize the variety of conformations observed. Table 2.6 gives a list of



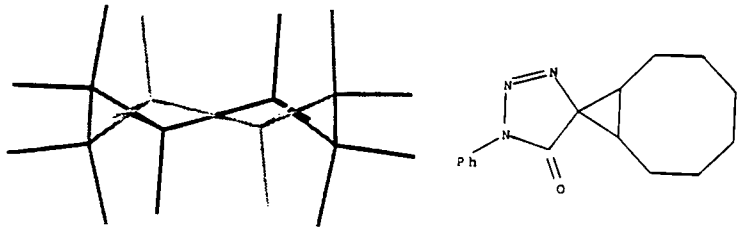
(i) Cluster 1 - BC
mrf -- BAGPII



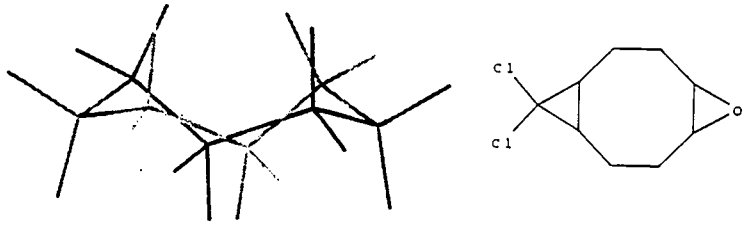
(ii) Cluster 2 - BC/TBC
mrf -- COVLUU



(iii) Cluster 3 - BC/TBC
mrf -- SPTZBN



(iv) Crown -- DEZPUT



(v) TCC -- EOCNON10

Figure 3.21

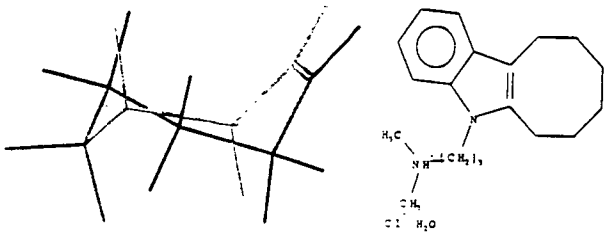
Most representative fragment geometries and environments for dataset 8C1

calculated energies for all of the conformations of the eight-membered rings. The commonly found minimum energy structure belongs to the BC conformer. Cluster 1 consists of 12 such BC examples, all of which exist as a free eight-membered ring with no adjoining ring systems, and few other external steric factors to contribute towards the resulting conformation. In this respect it is fair to say that the BC conformation is found to be the preferred form for the 8C1 dataset. Cluster 2 has six higher energy fragments. Analysis of these fragments shows that for all but one the eight-membered ring has an adjoining four- or five-membered ring. The inherent flexibility of the adjoining ring system determines the degree of distortion of the eight-membered ring away from its preferred BC conformation. Cluster 3 contains four higher energy examples of more distorted BC conformations. For these fragments the eight-membered rings are seriously affected by either adjoining cyclopropyl or epoxy groups, or by a seriously sterically-crowded external structure.

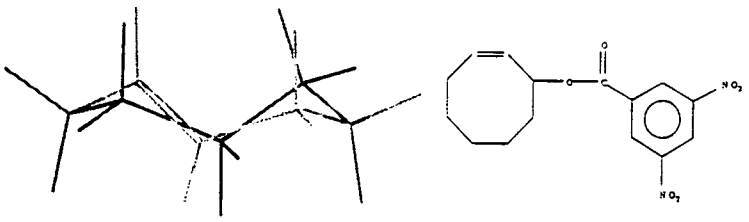
For the other unclustered fragments, the size of the adjoining rings or sub-groups appears to be a significant factor. The crown and TCC conformers are apparently favoured by the presence of two or more adjoining rings. For the crown, these neighbours are five-membered, with an element of flexibility inherent in these 5-rings, whereas the TCC conformers appear to be formed when the more sterically rigid 3- or 4-membered rings are involved. Hence, the chemical environment of the fragments can be seen to have an effect on the overall conformation adopted by the eight-membered ring: BC forms being the favoured, lowest energy conformer, and distortions away from this minimum being attributable to the size and flexibility of adjoining ring systems, together with the sterical attributes of any substituent groups.

3.5.6 Unsaturated eight-membered rings

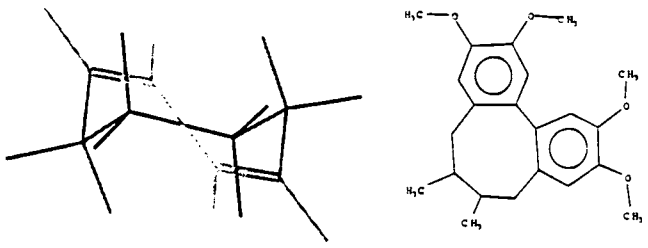
When analysing the entries for 8C2 and the other datasets of the unsaturated rings it is necessary in all GSTAT runs to specify which bond is actually the unsaturated one. This is achieved by using a TEST DISTance command in the fragment definition record. A distance test of between 1.30Å and 1.40Å was found to



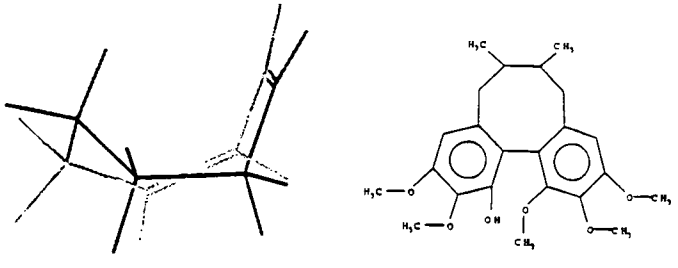
(i) Cluster 1 - cis-isomers
mrf -- IPRINC



(ii) trans-isomers
eg -- COCNBZ01



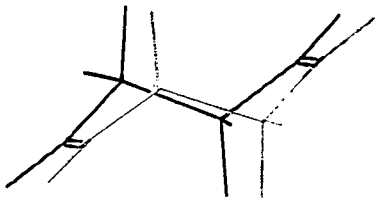
(iii) Cluster 1 - TB (C_2)
mrf -- GEDVAM

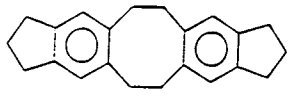


(iv) C_1-conformers
eg -- BUBXUR

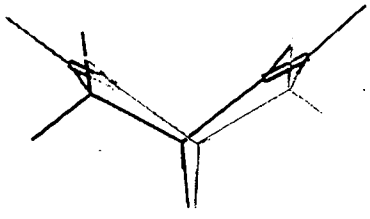
Figure 3.22

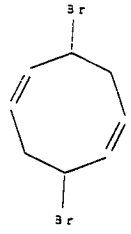
Fragment geometries and environments for dataset 8C2 [(i) and (ii)], and dataset 8C3 [(iii) and (iv)]



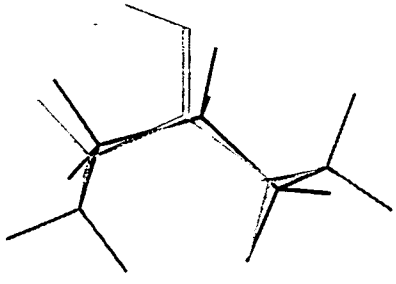


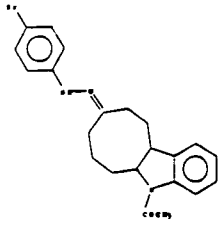
(v) di-benzo C-conformer
eg -- TAFZOP



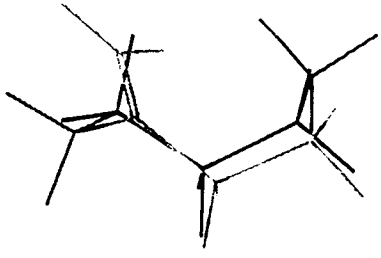


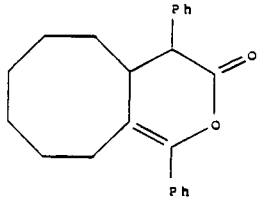
(vi) BB-conformer
eg -- DBCODE





(vii) BC-conformer with
exo-ene group at
position 3 of ring
eg -- AOCINB





(viii) BC-conformer with
exo-ene group at
position 2 of ring
eg.-- DEJJUX

Figure 3.22 (cont)

Fragment geometries and environments for dataset 8C4 [(v) and (vi)], and dataset 8C5 [(vii) and (viii)]

be appropriate in this case.

Dataset 8C2 contains all the entries found for the cyclooctene fragment. The *cis*- and *trans*-conformers have been well studied (Allinger and Sprague, 1972; Ermer and Lifson, 1973; Favini *et al.*, 1968) and are of interest here. Analysis of dataset 8C2 shows that of the 24 hits obtained, 9 are of the *cis*-isomer and 6 of the *trans*-isomer (using K_{NN}=8, K_{JP}=4, n=2, D_{lim}=0.08). Other conformations, including 2 BC's, are also observed but these appear to result largely from the steric influence of adjoining cyclopropyl or epoxy-groups. The six *trans*-isomers are easily identifiable by virtue of the expected 140° torsion angle (Traetteberg, 1975) about the double bond, compared with the 0° torsion angle found for the *cis*-isomers. The *cis*-isomer is quite a flexible ring and does not adopt one of the regular chair or boat forms. The preferred orientation, as can be seen from the torsional sequence belonging to the MRF (IPRINC: Table 3.14), involves an unsymmetrical ring which is related to the cyclooctane BC form. This fact is further emphasised by the values of θ calculated for the *cis*-isomers, which lie in the range 70° - 80°.

The *trans*-isomers observed correspond to the more symmetrical CC / TCC forms of cyclooctane, with θ lying in the range 30° - 40°. *trans*-cyclooctene is a strained ring structure capable of existing in two distinct conformations; the C_2-chair and the C_2-twist forms. The θ values calculated for the *trans*-isomer examples, lie in the range corresponding to the CC / TCC examples of the parent cyclooctanes. The twist form has been calculated to be the lowest lying in energy of the two, some 13.14 $kJmol^{-1}$ (Traetteberg, 1975) below the chair form, and is indeed found to be the only conformation observed here. All six hits exhibit the C_2-twist symmetry, with no chairs conformers present.

Dataset 8C3 contains the 21 examples of the *cis-cis*-1,3-cyclooctadiene fragments, of which 16 fragments exist as the symmetrical twist-boat (C_2) conformation, and 5 fragments as the unsymmetrical C_1 conformation (Figure 3.22). The dominance of the C_2 conformation over the C_1 form in the data, is a little unexpected since C_2 has been calculated to be only 1.76 $kJmol^{-1}$ (Allinger *et al.*, 1976) lower in energy than the C_1. A clustering run using K_{NN}=8, K_{JP}=4, n=2, D_{lim}=0.10

successfully groups 16 of the symmetrical fragments, with the remaining unclustered 5 fragments all being of the C_1 geometry.

The only other unsaturated eight-membered ring dataset of note is that belonging to the *cis-cis*-1, 5-cyclooctadienes, with 11 hits found. A complication of this data lies in the multiple occurrence of the diammine-{ε 4-cyclooctadiene} rhodium (I) fragment in crown ether complexes, which distorts the attempted analysis. These apart, the remaining compounds are all examples of the di-benzo-1, 5-cyclooctadiene fragment. This is an important observation since the normal modes of interconversion between the minimum energy twist-boat diene forms (White & Bovill, 1977) are not seen. The H/H eclipsing at the $C_{sp}2$ carbons of the 1, 5-cyclooctadiene can be relieved by twisting to give the TB forms. However at the $C_{sp}2$'s of the di-benzo analogues, the eclipsing is between H and another $C_{sp}2$, a situation which cannot be resolved by twisting *i.e.* there are the same torsional interactions in all geometries, and the transition state for ring inversion is therefore much higher for the benzo derivatives than for the ordinary dienes. The steric bulk of the benzo groups and the 1, 6 and 2, 5 steric interactions that would result from the formation of the boat conformers, pushes the ring towards the chair conformations observed (Figure 3.22). The only free fragment example, with two ordinary alkene type bonds, exists as the boat-boat type fragment shown, and confirms this argument. When the Jarvis-Patrick algorithm is applied to this dataset using the parameters K_{NN}=6, K_{JP}=3, n=1, D_{lim}=0.10, 8 of the 11 fragments in this group are successfully clustered.

3.5.7 Exo-unsaturated cyclooctanes

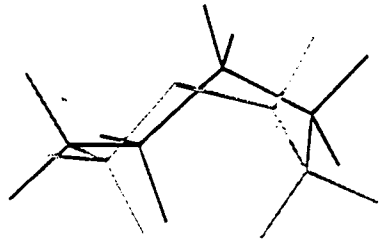
Dataset 8C5 contains the 12 fragments of the eight-membered C-rings possessing one external double bond to O, C or N. Analysis of this group as a whole shows that once again the BC conformation is highly favoured: all but one of the available fragments exist in this conformation. The only outlier adopts a TCC conformation in order to satisfy the steric constraints placed upon the ring by the presence of adjoining cyclobutyl and cyclopropyl rings. The fusion of larger, more

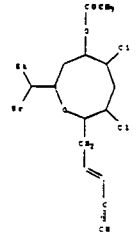
flexible rings does not inhibit the adoption of the BC conformation. Jarvis-Patrick clustering routines can be used to highlight the differences between conformations. K_{NN}=6, K_{JP}=3, n=1, D_{lim}=0.10 are the parameters found to give the most chemically sensible groupings. Typical coordinates(for the MRF's) are shown in Appendix 2. Most of the BC fragments exist with external double bond subtended from the 3-position of the ring as shown (Figure 3.22) with only one example found where the bond comes from the 2-position.

3.5.8 Eight-membered heterocyclic rings

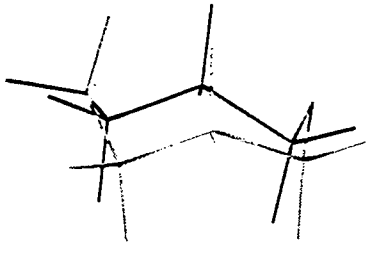
The only remaining significant datasets for the eight-membered rings are those which feature hetero-atoms in their ring structure. The analysis of these datasets (8A1, 8A2, 8A3) leads to some interesting conformational observations. For dataset 8A1, where only one ring C_{sp3} has been replaced by a hetero-atom, the 10 entries are found to be evenly split between the TBC / BC pseudorotational conformations and those of the TCC / CC. Conformations of the latter type appear as a result of the eight-membered rings having a fused ring, as well as another substituent group at one of the C_{sp3} sites. The resulting steric crowding forces the conformation of the ring away from the accepted minimum energy form (BC), and into the slightly higher energy conformation of the TCC / CC. All of the BC examples exist as "free" eight-membered rings with smaller steric interaction at the carbon centres. Clustering runs performed using K_{NN}=5, K_{JP}=3, n=1, D_{iim}=0.10 successfully cluster the 5 TCC examples, with the remaining 5 fragments all being BC conformers. Typical examples of each are shown in Figure 3.23.

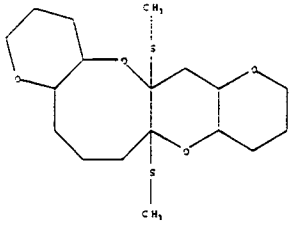
When analysing dataset 8A2 it is interesting to note that the number of different conformations has increased. This time the 12 fragments can be classified as 6 TBC / BC's, 2 TCC / CC, 1 crown and 3 TC / C. The examples of TBC / BCs are found for the rings with a sulphur atom in at least one of the two hetero-occupied positions. The presence of this large and diffuse atom obviously allows the ring the flexibility required to attain the low-energy BC conformation. When the nitrogen atom is introduced into



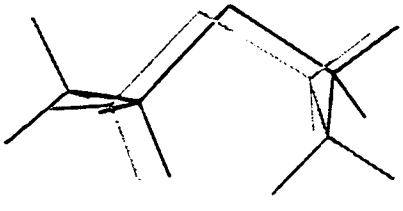


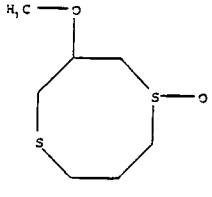
(i) TBC/BC-conformer
eg -- LAURCN10



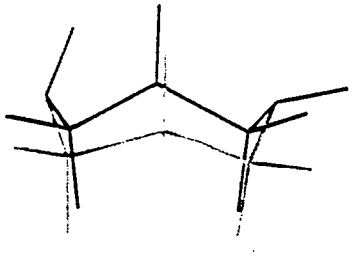


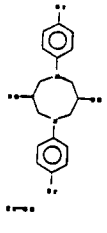
(ii) TCC/CC-conformer
eg -- KEVBES





(iii) BC-conformer
containing sulphur
eg -- CARYAV

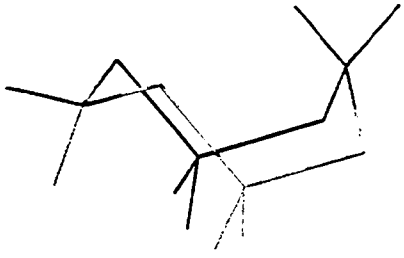


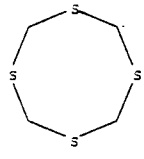


(iv) crown conformer
eg -- BPHZOC10

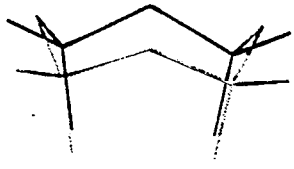
Figure 3.23

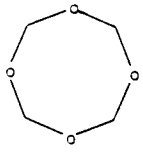
Fragment geometries and environments for dataset 8A1 [(i) and (ii)] and dataset 8A2 [(iii) and (iv)]



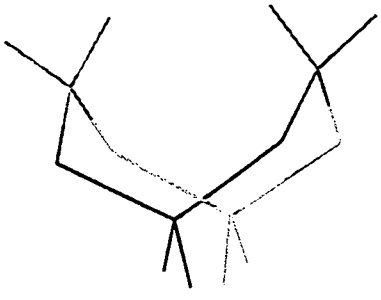


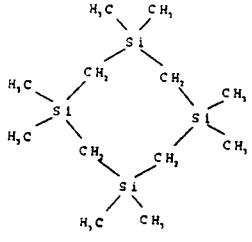
(v) BC-conformer for tetrathiocane -- TETIOC10





(vi) crown-conformer for tetroxocane -- TOXOCN





(vii) BB-conformer eg -- OMSIOC

Figure 3.23 (cont)

Fragment geometries and environments for dataset 8A3 [(v), (vi) and (vii)]

the ring more chair-like conformations result, and the only example of the crown conformation is observed when the ring has nitrogen atoms in both the 1 and 5 positions and with phenyl groups attached to both, as shown in Figure 3.23. For all of the examples found in dataset 8A2 none have fused ring systems, so the steric arguments for conformational variety expressed earlier are not important here. Instead the major factor behind the conformational diversity is the nature of the hetero-atoms present in the ring, with larger atoms such as sulphur allowing the eight-membered rings to adopt its minimum energy form, and smaller atoms, with their subsequently more restricted orbital structure and spatial requirements, forcing the eight-membered rings into a higher energy conformation.

The final dataset examined, 8A3, also exhibits the trends observed in the first two datasets. Here, of the 20 hits found in the CSD, 5 are of the TBC / BC type, 7 TCC / CCs, 5 crowns, 6 TC / C pseudorotation conformations, and finally 1 BB type conformation. Jarvis-Patrick analysis performed at K_{NN}=6, K_{JP}=3, n=2, D_{lim}=0.10, clusters the crowns and the CC's together but separates the TC cluster. The conformational groupings can be rationalised. The TBC / BC examples are found when the hetero-atoms in the ring are all of the large and diffuse type such as sulphur (in tetrathiocane) and selenium (in tetraselenocane), thus emphasising the need for non-rigidity in the ring if the lowest energy conformations are to be adopted. The BB example is the only true outlier here and is formed when the 4 hetero-atoms in the ring are all silicon. All of the other fragments have a degree of chair character to them which can be attributed to the hetero-atoms present in their rings. Examples which include two or four oxygen or nitrogen atoms in the ring being restricted to the chair family of conformations by their size and the more "rigid" nature of their bonding orbitals.

3.6 Concluding remarks

A basic feature of crystallographic research is to obtain knowledge of the size and shape of three-dimensional molecules and chemical fragments from raw data in the

form of external coordinates referred to unit-cell axes. If enough data exists, this knowledge may be generalised and expressed in terms of typical dimensions or archetypal conformations. Essentially, this is a form of knowledge acquisition through pattern recognition and classification. In this Chapter we had four objectives: (i) to apply a variety of novel classification techniques to conformational descriptions of seven- and eight-membered rings, (ii) to interpret the results so as to provide a functional description of the relevant conformational hypersurface and to locate highly populated regions on that hypersurface, (iii) to study, modify or improve the available methodologies for data analysis, and (iv) to relate the conformational minima detected by these methods to the chemical environments of the fragments under investigation.

In Chapter 3.2 it was shown that the results of symmetry-modified principal component and cluster analyses are clearly interpretable in terms of *a priori* knowledge of the conformational space for cycloheptane. The work has shown that symmetry expansion of the datasets is an essential pre-requisite to a PCA for a symmetrical fragment. Indeed, even without the *a priori* knowledge the analysis would have indicated the four-dimensional nature of this space, and the two-fold degeneracy of the axial pairs. The analysis has, however, pointed to the importance of isolating a single asymmetric unit of conformational space (from the fully symmetrized dataset) as an adjunct to classification.

Chapter 3.3 mapped and classified the exo-unsaturated and heterocyclic seven-membered rings, with the results being in broad agreement with the features of the potential energy hypersurface, as calculated by force-field techniques within a novel algorithm, that does not involve ring cleavage. Given that these energy calculations were performed on the unsubstituted parent rings (V)-(XI), and that most of the observations of these rings in the crystal structures are heavily substituted and fused in very different ways, the qualitative agreement between theory and experiment is, perhaps, better than might have been expected. There is evidence here that, despite some relatively low energy barriers along the C/TC pseudorotation pathway, the individual rings attempt to adopt lower energy conformations even as sub-units much more complex structures. Whilst it is formally invalid to attempt to quantify

conformational energy differences on the basis of population ratios observed in crystal structures (Bürgi & Dunitz, 1988), surveys of this type add to a general pool of evidence that crystal conformations do tend to map the low-energy features of the relevant potential energy hypersurface.

One of the aims of this Chapter is to confirm that N_p-ordered classifications of fragment conformations derived from crystallographic observations are a reliable, albeit qualitative, guide to the relative energies of the different conformers. This was certainly shown to be true for the seven-membered ring fragments detailed in Chapter 3.2 and 3.3, which also exhibited a wide variety of substitution patterns, *i.e.* they were drawn from random chemical environments. However, in Chapter 3.4 the pitfalls of this description are starkly observed. Analysis of the dataset composed of 1, 3-cycloheptadiene fragments showed the chemical environment to be far from random, with the result that a higher-energy conformation appears more favourable in terms of N_p, over the calculated lowest-energy conformation. Once the specific environmental constraints have been accounted for, the basic thesis of the Chapter remains intact.

For the eight-membered rings of Chapter 3.5, the features of the five-dimensional conformational hypersurface are again well described by the relevant BPRS and Cremer-Pople coordinate systems. The major conformational feature of these symmetry-expanded systems is the BC/TBC pseudorotation itinerary, which dominates the low energy regions of the potential surface. Other high-energy conformers are observed when specific environmental factors are prevalent.

The Jarvis-Patrick clustering experiments have provided valuable insights concerning typical criteria to be employed in processing the nearest-neighbour table which underpins this algorithm. These insights indicate that a greater degree of automation of the pattern recognition process may be possible through (a) systematic computerised scanning of results obtained with small variations of the K_{JP} / K_{NN} ratio in the area of ≈ 0.5, and (b) further assessment of the statistical descriptors generated for each cluster. Nevertheless, results from the current semi-automated methods were obtained rapidly and provide a satisfactory numerical classification of conformations.

3.7 References

Allen, F.H., *Acta Cryst.*, **1990**, A46, C-139.

Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C.F., Mitchell, G.F., Smith, M.J. & Watson, D.G., *J. Chem. Inf. Comput. Sci.*, **1991**, 31, 187-204.

Allen, F.H., Doyle, M.J. & Auf der Heyde, T.P.E., *Acta Cryst.*, **1991**, B47, 412-424.

Allen, F.H., Doyle, M.J. & Taylor, R., *Acta Cryst.*, **1991b**, B47, 41-49.

Allen, F.H., Doyle, M.J. & Taylor, R., *Acta Cryst.*, **1991c**, B47, 50-61.

Allen, F.H., Garner, S.E., Howard, J.A.K. & Pitchford, N.A., *Acta Cryst.*, **1994**, B50, 395-404.

Allen, F.H., Howard, J.A.K. & Pitchford, N.A., *Acta Cryst.*, **1993**, B49, 910-928.

Allen, F.H., Howard, J.A.K., Pitchford, N.A. & Vinter, J.G., *Acta Cryst.*, **1994**, B50, 382-395.

Allen, F.H., Kennard, O., Watson, D.G., Orpen, A.G., Brammer, L. & Taylor, R., *J. Chem. Soc. Perkin Trans. 2*, **1987**, S1-S19.

Allen, F.H. & Johnson, O., *Acta Cryst.*, **1991**, B47, 62-67.

Allen, F.H. & Taylor, R., *Acta Cryst.*, **1991**, B47, 404-412.

Allinger, N.L. & Sprague, J.T., *J. Am. Chem. Soc.*, **1972**, 94, 5734-5747.

Allinger, N.L., Tribble, M.T. & Miller, M.A., *Tetrahedron*, **1972**, 28, 1173-1190.

Allinger, N.L., Viskosil, J.F.Jnr, Burkert, U. & Yuh, Y., *Tetrahedron*, **1976**, 32, 33-35.

Anet, F.A.L. & Krane, J., *Tetrahedron Letters*, **1973**, 50, 5029-5032.

Avirah, T.K., Molloy, T.B. & Cook, R.L., *J. Chem. Phys*, **1979**, 71, 2194-2201.

Biosym Technologies, *INSIGHT*, **1993**, Biosym Technologies, 9685 Scranton Road, San Diego, California, USA.

Bocian, D.F., Pickett, H.M., Rounds, T.C. & Strauss, H.L., *J. Am. Chem. Soc.*, **1975**, 97, 687-695.

Bocian, D.F. & Strauss, H.L., *J. Am. Chem. Soc.*, **1977a**, 99, 2876-2882.

Bocian, D.F. & Strauss, H.L., *J. Am. Phys.*, **1977b**, 67, 1071-1081.

Bocian, D.F. & Strauss, H.L., *J. Am. Chem. Soc.*, **1977c**, 99, 2866-2876.

Bürgi, H.B. & Dunitz, J.D., *Acta Cryst.*, **1988**, B44, 445-448.

Burkert, R. & Allinger, N.L., *Molecular Mechanics*, **1982**, ACS Monograph, No. 148, Washington D.C.: American Chemical Society.

Chiang, J.F. & Bauer, S.H., *J. Am. Chem. Soc.*, **1966**, 88, 420-425.

Courtois, A., Protas, J., Fixari, B. & Brunet, J.J., *Acta Cryst.*, **1975,** B31, 2064-2069.

Dillen, J. & Geise, H.J., *J. Chem. Phys.*, **1979**, 70, 425-429.

Dobler. M. *MacMoMo*. Molecular display and modelling program for Apple Macintosh computers. **1990**, ETH, Zurich.

Ermer, O. & Lifson, S., *J. Am. Chem. Soc.*, **1973**, 95, 4121-4132.

Favini, G., Maggi, A. & Todeschini, R., *J. Mol. Struct.*, **1983**, 105, 17-29.

Hagen, K. & Traetteberg, M., *Acta Chem. Scand.*, **1972**, 26, 3643-3648.

Hendrickson, J.B., *J. Am. Chem. Soc.*, **1967**, 89, 7047-7054.

Hashmi, S., Polborn, K. & Sziemies, G., *Chem. Ber.*, **1989**, 122, 2399-2401.

Irngartinger, H., Nixdorf, M., Riegler, N.H., Krebs, A., Kimling, H., Pocklington, J., Maier, G., Malsch, K.-D. & Schneider, K.-A., *Chem. Ber.*, **1988**, 121, 673-677.

Jarvis, R.A. & Patrick, E.A., *IEEE Trans. Comput.*, **1973**, 22, 1025-1034.

Jendralla, H., *Chem. Ber.*, **1980**, 113, 3557-3569.

Kirby, A.J., *The Anomeric Effect and Related Stereo-electronic Effects at Oxygen*, **1993**, Berlin, Springer-Verlag.

Kraus, W., Kypke, K., Grimminger, W., Sawitzki, G.R. & Schwinder, G., *Liebigs Ann. Chem.*, **1982**, 87-93.

Murray-Rust, P., Bürgi, H.B.& Dunitz, J.D., *Acta Cryst.*, **1978a**, B34, 1787-1793.

Murray-Rust, P., Bürgi, H.B.& Dunitz, J.D., *Acta Cryst.*, **1979**, A35, 703-713.

Noe, E.A. & Roberts, J.D., *J. Am. Chem. Soc.*, **1971**, 93, 7261-7288.

Norskov-Lauritsen, L., Bürgi, H.B., Hofmann, P. & Schmidt, H.R., *Helv. Chim. Acta*, **1985**, 68, 76-82.

Oberhammer, H. & Bauer, S., *J. Am. Chem. Soc.*, **1969**, 91, 10-16.

Saebo, S. & Boggs, J.E., *J. Mol. Struct.*, **1982**, 87, 365-373.

Tektronix Inc., *CaChe Scientific*, **1992**, PO Box 500, Beaverton, Oregon, USA.

Tochtermann, W., Luttman, K., Wolff, C., Peters, K., Peters, E.-M. & Von Schnering, H.G., *Chem. Ber.*, **1989**, 122, 1653-1660.

Traetteberg, M., *J. Am. Chem. Soc.*, **1964**, 86, 4265-4270.

Traetteberg, M., *Acta Chem. Scand.*, **1968**, 22, 2305-2312.

Traetteberg, M., *Acta Chem. Scand.*, **1975**, Sect. B, 29, 29-36.

Vinter, J.G., Davis, A. & Saunders, M.R., *J. Comput. Aided Mol. Des.*, **1987**, 1, 31-50.

White, D.N.J. & Bovill, M. , *J. Chem. Soc. Perkin Trans.*, **1977**, 2, 1610-1623.

Chapter Four

A Systematic Study of the Coordination Geometry of Three-Coordinated Transition Metals

4.1 Introduction

This Chapter describes a systematic study of the coordination sphere geometry of the formally three-coordinated transition metal complexes, as observed in the available crystal structure analyses located in the CSD. The archetypal geometry of this species is trigonal planar, with D_{3h} symmetry, but distortions towards planar T-shaped and Y-shaped geometries of C_{2v} symmetry, as well as towards the trigonal pyramidal coordination geometry of C_{3v} symmetry, are well known. The major objectives of this study are: (a) to map the observed distortions so as to facilitate the observation of configurational interconversions, and (b) to correlate these maps with both chemical and structural features and thereby seek to explain these observed distortions.

The principles of structure correlation have been stated and described earlier in Chapter 1, but can be summarized in two different ways, both of which are particularly relevant to the studies presented in this Chapter. The first (Auf der Heyde & Bürgi, 1989) states that "although direct observation of a molecule along the reaction pathway does not seem feasible, its visualisation at least does. According to the structure correlation hypothesis, the gradual distortion or static deformation that a molecular fragment of interest manifests collectively over a large variety of crystalline frameworks may be assumed to mirror the distortion which that fragment would undergo along a given reaction coordinate. The various crystal or molecular structures are considered to constitute a series of 'frozen-in' points, or snapshots, taken along the reaction pathway, which, when viewed in the correct order, yield a cinematic film of the reaction". In addition, the relationship between structural (geometrical) parameters and features of the relevant energy surface has been summarised (Murray-Rust, Bürgi & Dunitz, 1975) as: "If a correlation can be found between two or more independent parameters describing the structure of a given structural fragment in a variety of

environments, then the correlation function maps a minimum energy path in the corresponding parameter space". To illustrate this point it is informative to briefly review some of the more relevant studies that have been performed within this field, and which have particular significance for the experimental work described later in this Chapter.

The current work concentrates on the description of the molecular coordination geometry and on the processes involved in changing this geometry and, indeed, in changing the coordination number itself. In general there are two classes of reaction that can be attributed to changing the coordination at the reaction centre: addition/elimination and substitution. These processes have slightly different meanings in organic and inorganic chemistry, but can be summarised in the following way.

Addition and elimination are different aspects of the same process, which can be seen to increase or decrease the coordination number by one. A transition state exists in which the reactants form a partially bonded species. Substitution, however, involves no change in the coordination number and can take one of two different courses. The first involves an intermediate, and therefore requires two transition states; the second requires only one transition state. In the nomenclature of the organic chemist these are known as S_N1 and S_N2 reactions. Structure correlation has been applied to examples of both processes with notable success, and a few of the more relevant studies will now be briefly described.

As an example of the addition/elimination process one area of interest has been concerned with the deviations away from tetrahedral geometry that are observed for a variety of chemical systems (Dunitz, 1979; Bürgi, 1975; Bürgi & Dunitz, 1983). Deformations of the Y-MX_3 system that preserve the C_{3v} symmetry group can be described in the same terms for a wide variety of structures including Y-PO_3, Y-SO_3 and Y-$SnCl_3$. These systems have proved to be good models for the cleavage of the Y-M bond in a dissociative process. Observation of the systematic relationship between any of the geometrical parameters that "define" the reaction coordinate can be found from the crystallographic data at hand. So, as the Y-M bond is seen to lengthen under the particular forces that are prevalent in a certain structure, the rest of the tetrahedral

structure (for example the YMX bond angle, or the M-X bond length) responds in a predictable and correlated fashion.

Under the particular relationships of the chosen system this correlation must represent the most energetically favourable pathway for the structural change involved, which in this case is the elongation of the Y-M bond. This is by definition the same process that is involved in the cleavage of the Y-M bond when the system reacts and hence leads to the conclusion arrived at by Bürgi & Dunitz (1983) that the minimum-energy pathway being observed in the crystal is related to the minimum-energy pathway which defines the reaction coordinate for Y-M under the same conditions.

Of those studies that describe the S_N2 process, one of the earliest (Bürgi, 1973) introduced novel ideas on how the structural information obtained by a study of the $XCdS_3Y$ fragment could be used to effectively map the S_N2 type reaction at the metal centre. The geometrical parameters chosen there map the observed total deformation in terms of the change in bond distances for the two X and Y ligands (X and Y in this case occupying the apical positions of a trigonal bipyramid), the deviation of the cadmium atom out of the plane of the three S ligands, and also in terms of the observed distortion of the angles from those of the idealised, regular trigonal bipyramid.

The resulting data show that the ideal trigonal bipyramid (where the displacement out of the plane of the sulphur ligands of the cadmium atom equals 0Å) is found when the axial distances to both the X and the Y are approximately 0.3Å larger than the sum of the respective covalent radii, or about equal to the sum of the ionic radii. The data also show that a decrease in the length of a bond, say to the Y ligand, coincides with an observed increase in the bond length to the X ligand, together with a distortion of the cadmium atom out of the sulphur plane towards the Y atom, and an increase in the Y-Cd-S bond angles.

Logarithmic functions of the type first proposed by Pauling (1947, 1960), which link differences in interatomic distances to the logarithm of the bond order, were applied to the data and were observed to produce a very close fit. In addition, all of the points on the correlation curve obey the invariance condition of $n_X+n_Y = 1$ (where n_X and n_Y are bond numbers), which can be considered to mean that the total bond number

is not altered throughout the course of the "reaction". It is apparent therefore that the S_N2 reaction for this chemical system can be observed from the crystallographically obtained structural data, and that correlations with theoretical postulates is high.

The above analysis was based on only four crystal structures. However similar studies have followed which map S_N2 reactions for a much greater number of fragments (Britton & Dunitz, 1981), as determined by selective systematic searches on the Cambridge Structural Database. In a variety of similar systems the correlation between the geometric parameters found from crystallographic data and the theoretical laws is also very good. These studies are summarised by Bürgi & Shklover (1994).

An alternative approach to analysing the systematic geometrical data that can now be found in the crystallographic databases is to use a technique that has been rigorously developed by Murray-Rust, Bürgi and Dunitz (1978a, b; 1979). They seek to describe the observed fragment geometry as being a distorted version of a higher symmetry reference structure. Symmetry coordinates (the term "symmetry deformation coordinates", SDC's, is used throughout this thesis but an equivalent alternative used by other authors is "symmetry-adapted deformation parameters", SADP's) can be derived and used to partition the overall total distortion into various components, each of which preserves some of the symmetry elements of the parent: they are symmetry-adapted linear combinations of internal coordinates that transform as irreducible representations of the symmetry group of the reference structure. The symmetry coordinates are, by definition, all orthogonal to each other so that each component displacement from the parent symmetry is independent.

When these symmetry deformation coordinates are used to analyse a set of systematic data the results can be most enlightening. In the case above ($YCdS_3X$), a coordinate system could be designed to show the correlation between the movement of the cadmium atom out of the plane of the three equatorial sulphur atoms, and the bond length displacement for Cd-X. When independent parameters are seen to correlate in this fashion it can be regarded as proof of the concerted actions in one overall distortion, and hence illustrate the power of this technique.

These SDC's have been shown to give a very good representation of the mechanisms at work in a variety of systems and will be used in the study presented below to map the deformations observed for the 3-coordinated fragments found in this work. The derivation of pertinent coordinate systems is also discussed in depth.

It is apparent that the study of reaction type is possible with the methods developed for structure correlation, but other studies have also been performed which look at the changes associated with the geometry of the fragment, but which do not involve any changes in the coordination number at the metal centre. The structure correlation principle is therefore particularly pertinent to the study of pseudorotation mechanisms in a manner which is analogous to work described in Chapters Two and Three of this Thesis.

Most recently the mechanism for the Berry pseudorotation, observed for 5-coordinated metal fragments, has been investigated (Auf der Heyde & Bürgi, 1989a, b, c) using a general systematic search of the Cambridge Structural Database for five-coordinate d^8 complexes. The geometry of the ML_5 fragment is described by two sets of 12 non-redundant symmetry coordinates representing the two reference structures possible for this system: the trigonal bipyramid (D_{3h}), and the square-based pyramid (C_{4v}). When these coordinate systems are applied to the data it is possible to locate the individual motions that distort one configuration into the other. When the displacements are collected together the complete mechanism becomes evident, and the pseudorotation itinerary that takes the trigonal bipyramid into the square planar, and *vice-versa*, can be mapped. This study also makes extensive use of the numerical analysis tools that are available as part of the Cambridge Structural Database System. These applications have already been seen to be of great advantage in the handling of large quantities of data for the conformational analyses of the seven- and eight-membered rings (Chapter 3), and will be further exploited here. The conformational studies were seen to be crucially dependent on the symmetry-expansion of the dataset and this is also true in coordination studies. Auf der Heyde & Bürgi (1989) had to expand their dataset to account for the 12-dimensional nature of the

hyperspace spanned by the five-coordinated species, and similar expansions are vital for the three-coordinated species analysed below.

Studies of the coordination geometry of those four-coordinated fragments found in the database are quite numerous. Muetterties and Guggenberger (1974) were first to point out that the various conformations exhibited by different crystal structures of the four-coordinated fragment could be used to map the T_d to D_{4h} conversion that can occur via either of two mechanisms: the first a tetrahedral compression along an S_4 axis, and the second a digonal twist about this axis where the angles subtended at the metal are not constrained. The average bond length was found to increase sigmoidally as the fragment distorts away from the T_d configuration and towards the D_{4h}, this being due to the possibility of greater steric strain in the square-planar conformer. The most recent study that follows the principles of the structure correlation method has been performed by Klebe & Weber (1994), in which symmetry coordinates are used to analyse the coordination geometries of certain transition-metal complexes. Using a large selection of metal-organic 4-coordinated compounds, as found in the Cambridge Structural Database, interconversion pathways were plotted, and the mechanism of change was shown to be similar to that first proposed by Muetterties & Guggenberger (1974).

Other studies of the four-coordinated fragment include those of the gearing motions of PR_3 ligands that were investigated by Chandrasekhar & Bürgi (1983), and Norskov-Lauritsen & Bürgi (1984), and the studies performed by Luef, Keese & Bürgi (1987) and Luef & Keese (1992) that classify the distortions of the quaternary carbon in strained spiro compounds in terms of the symmetry coordinates for a regular tetrahedron.

Analyses of three-coordinated fragments are less common. One early discussion (Barrow *et al.*, 1976) considered the possibility of interpreting the changes in bond lengths and angles in terms of a low-energy pathway from the bent three-coordinate to linear two-coordinate complexes, and the reverse. Bürgi *et al.* (1989) later used the symmetry-expanded data from 19 crystal structures to investigate conformational preferences and changes. Neither study relates the 3-coordinated

fragments to either the 2-coordinate, or 4-coordinate fragments directly, but does imply that there may be a link.

In this study 3-coordinated transition metal compounds were located using a general search criterion on the CSD. Analysis of this dataset relies on the displacements of the observed angles away from those of the reference D_{3h} structural symmetry. Displacements involving bond lengths are somewhat more difficult to assess due to the nature of the dataset and the available analytical tools within the CSD system. The angular data could be symmetry-expanded and analysed using the PCA modules available in CSD, and the results so found can be compared and correlated with those derived using symmetry deformation coordinates.

With the advent of non-bonded interaction searches through the latest version of the Cambridge Structural Database (April 1994), it has now become possible to routinely study the influence of intermolecular contacts on the coordination geometry of the transition metal centre in question. It is therefore possible to study the mechanisms for transition in which a three-coordinated fragment may be the stable intermediate, *i.e.* it should now be possible to map the successive additions to a linear two-coordinated complex of first one ligand and then a second to carry a two-coordinate structure through to the three-coordinate and then the four-coordinated ideal geometries. In the work that follows this possibility is explored, and evidence for the expansion of the coordination sphere from 2 to 3 to 4 is given. The complete geometry of the addition processes are also investigated together with potential correlations of these processes with ligand type and metal oxidation states.

4.2 Geometrical descriptors and SDC's

The essential first reference point for all of the following analyses is the D_{3h} trigonal planar structure (Figure 4.1). This point group is the one with the highest symmetry possible for the MX_3 structures studied, and, as such, provides us with the reference geometry from which all of the observed deformations are calculated. As will be seen later, the D_{3h} configuration around the central trigonal transition metal atom is

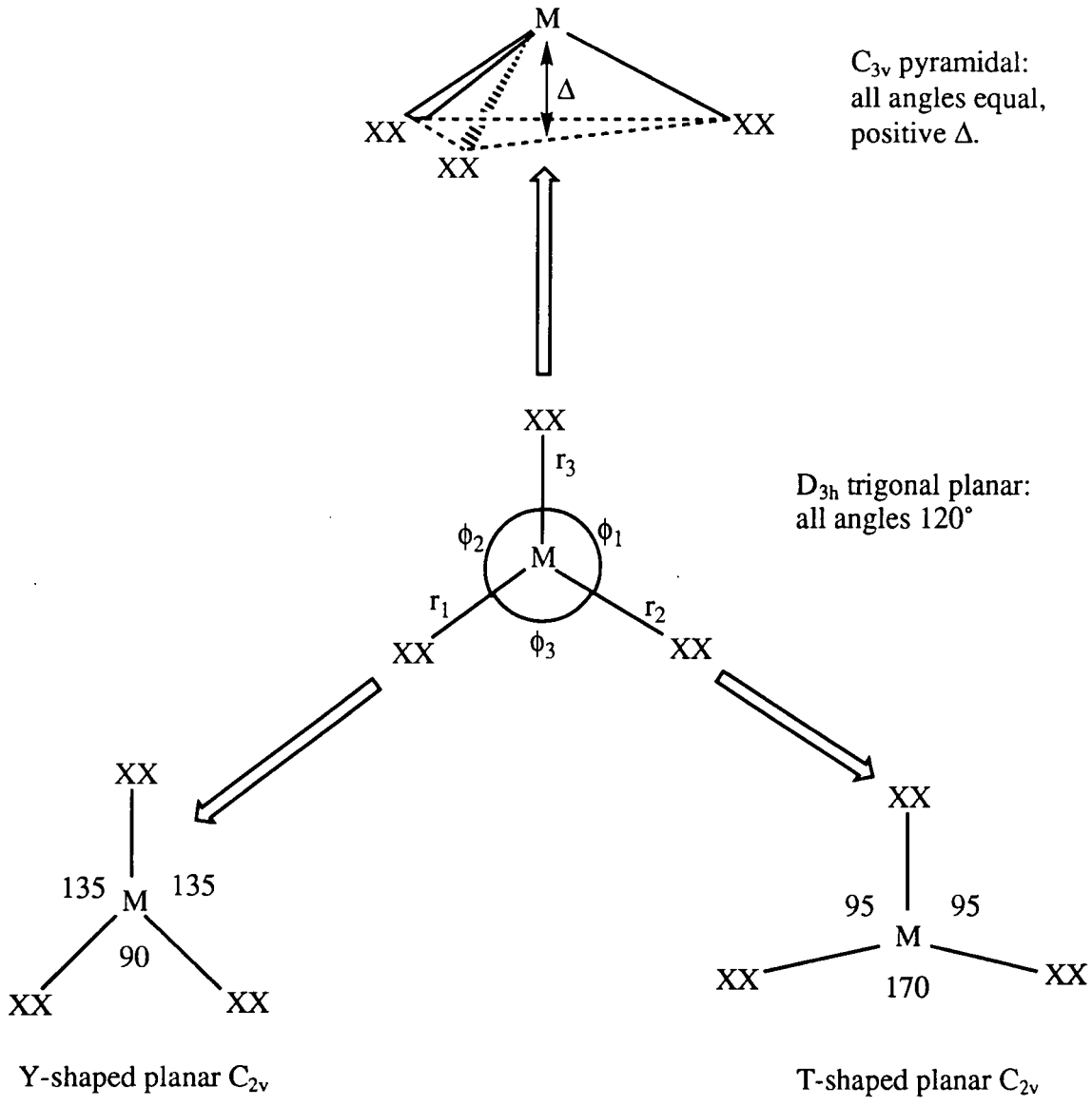


Figure 4.1

The three possible modes of distortion away from the D_{3h} reference geometry of the 3-coordinated metal fragment are shown, together with the resulting geometries.

the most prevalent for all of those "free molecules" found in the CSD, *i.e.* entries which show minimal steric or chemical constraints will generally be found in the D_{3h} configuration.

Deviations away from this reference geometry are analysed in terms of the angles formed by the three bonded ligands at the trigonal metal centre. These parameters are the "natural" descriptors of the shape of the fragment, as were the torsion angles for describing the conformations of the rings in Chapter 3. Bond lengths can be used in symmetry-deformation studies, but only when the two bonded atoms are known and their respective covalent radii can be accounted for during the analysis. The

datasets compiled in this study were deliberately constructed to be as general as possible, with the result that bond length deformations could not be automatically studied. The distortions detailed in this Chapter for exactly three-coordinate fragments, will therefore be reported exclusively in terms the deformations of the valence angles away from those of the reference structure.

Deformations of the D_{3h} structure that maintain co-planarity of the metal atom with the three bonded ligands follow two distinct paths: (i) the concerted increase of two of the basal angles from 120°, and the consequent reduction of the third angle, to form a Y-shaped structure, and (ii) the concerted decrease of two angles and subsequent increase in the third which leads to a T-shaped form.

For molecules exhibiting D_{3h} symmetry the ideal set of three basal angles would be 120°, 120°, 120° (with zero displacement of the metal atom out of the plane of the three ligands). For the "Y-shaped" molecules, experience shows that this set of angles will approach 135°, 135°, 90° and for the "T-shaped" examples the ideal situation would involve values of 180°, 90°, 90°. However molecules found as hits in the CSD have far from ideal basal angles and so the following cut off points were established for describing the general configuration at the metal centre and are used throughout this study:

D_{3h} :	130-120°,	130-110°,	120-110°
"T-shaped" :	>130°,	<120°,	<110°
"Y-shaped" :	>130°,	>120°,	<110°

4.2.1 SDC descriptions of molecular deformation

In the following analyses we want to be able to treat the observed structure of each molecular fragment as a distorted version of the symmetrical reference structure, *i.e.* the D_{3h} configuration. We can do this using symmetry deformation coordinates, following precedents established by Murray-Rust, Bürgi & Dunitz (1978a,b & 1979).

The essential feature of the symmetry-coordinate description is that the overall observed distortion of the molecular fragment away from the reference structure

geometry can be described as a sum of several components, each of which preserves some of the symmetry elements of the reference structure. Any displacement along a given SDC, which can be transformed as a particular irreducible representation (IR), preserves at least a certain symmetry, known as the kernel symmetry, of the IR in question. These kernel symmetries are derived below and can easily be found from group tables. A displacement along a symmetry coordinate transforming as a degenerate IR may preserve a higher symmetry than the kernel symmetry. In such a case the displacement may be transformed into itself by symmetry operations not belonging to the kernel of the IR. This higher symmetry is then known as the co-kernel (Melvin, 1956; McDowell, 1965) or epikernel (Ascher, 1977) symmetry.

In the outline given below the SDC's are derived from first principles for the reference structure D_{3h}, following established procedures summarised in Cotton (1990). Derivation of the kernel and co-kernel symmetries is also presented for this point group but these can also be found by reference to Wilson, Decius & Cross (1955) for all of the common point groups.

4.2.2 Derivation of SDC's for a D_{3h} reference structure

The vibrationally active normal modes of the reference molecule, can be directly compared to the processes involved in the distortion of the reference molecule towards the observed configuration, and so derivation of these modes is the crucial first step in the overall derivation of the SDC's. There are of course 3N-6 active modes to be

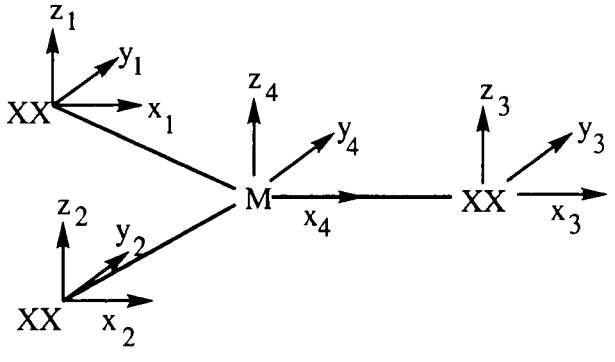


Figure 4.2

The 3N=12 Cartesian displacement vectors for the D_{3h}, trigonal planar reference structure

found, and for the four atom fragment shown in the diagram below this amounts to six normal modes of vibration.

By applying the D_{3h} point group operations to the 3N=12 Cartesian displacement vectors as represented below the following reducible representation can be found. A C_3 operation on this set of vectors will only affect those vectors at the central atom, transforming x, y into -1/2x, -1/2y while leaving z unchanged.

$D3_h$	E	$2C_3$	$3C_2$	σ_h	$2S_3$	$3\sigma_v$
τ_1	12	0	-2	4	-2	2

τ_1 can be re-expressed as the reducible representation :

$$\tau_1 = A_1' + A_2' + 3E' + 2A_2'' + E'$$

$$\text{rotations} = A_2' + E'$$

$$\text{translations} = E' + A_2''$$

$$\text{vibrations} = A_1' + 2E' + A_2'' \qquad \textbf{[1]}$$

We now have irreducible representations for the 3N-6 normal modes of vibration for a molecule exhibiting D_{3h} symmetry. To use these to find suitable symmetry coordinates we have first to change the perspective from which we are describing the molecular fragment, and recognise (see Figure 4.1) that it is the combination of the six physically accessible quantities of bond angles and bond lengths that will ultimately form the basis set for description of the total deformation.

D_{3h}	E	C_3	C_3^2	$C_2(a)$	$C_2(b)$	$C_2(c)$
$\hat{P}$ $A_1'(r_1)$	r_1	r_2	r_3	r_1	r_3	r_2
$\hat{P}$ $A_2''(r_1)$	$+1(r_1)$	$+1(r_2)$	$+1(r_3)$	$-1(r_1)$	$-1(r_3)$	$-1(r_2)$
$\hat{P}$ $E'(r_1)$	$+2(r_1)$	$-1(r_2)$	$-1(r_3)$	0	0	0

D_{3h}	σ_h	S_3	S_3^2	$\sigma_v(a)$	$\sigma_v(b)$	$\sigma_v(c)$
$\hat{P}$ $A1'(r_1)$	r_1	r_2	r_3	r_1	r_2	r_3
$\hat{P}$ $A_2''(r_1)$	$-1(r_1)$	$-1(r_2)$	$-1(r_3)$	$+1(r_1)$	$+1(r_2)$	$+1(r_3)$
$\hat{P}$ $E'(r_1)$	$+2(r_1)$	$-1(r_2)$	$-1(r_3)$	0	0	0

Using the bond angles and lengths as our descriptors the appropriate linear combinations are found by applying the projection operators of the IR's above, to just one of the members of the basis set, *e.g.* r_1. The results are shown above, and can be summarised as:

$\hat{P}\ A_1'(r_1) = 4\,(r_1 + r_2 + r_3)$ approximating to $(r_1 + r_2 + r_3)$

$\hat{P}\ A_2''(r_1) = 0$

$\hat{P}\ E'(r_1) = 2\,(2r_1 - r_2 - r_3)$ approximating to $(2r_1 - r_2 - r_3)$

As the E' representation is two-dimensional we need to find a second expression which will jointly be responsible for the overall representation. To arrive at this second function it is necessary to state that any member of a set of functions forming the basis for a representation, must be affected by the symmetry operations of the group in one of two ways :

1. It will go into ± 1 times itself.
2. It will go into another member of the set or a combination of members of the set.

If we chose an operation that does not convert the function into +1 times itself, for instance C_3, we find that under the operation :

$\hat{C}_3\,(2r_1 + r_2 + r_3) = (2r_2 - r_3 - r_1)$

which clearly does not go into itself ± 1 times. However this function is also not orthogonal to the first, as the partner must be. It follows therefore, that this function must be a linear combination of the first function and its partner, and so we can deduce the partner function by subtracting an appropriate number of the first function from the second, leaving the remainder as the partner.

$(2r_2 - r_3 - r_1) - (-1/2)(2r_1 - r_2 - r_3) = 3/2\ r_2 - 3/2\ r_3$ approximating to $(r_2 - r_3)$

This function is orthogonal to the first function and so the two together form the basis for the E' representation. The same calculations can be performed for the bond angles, the projection operators being applied to just one of these angles, *e.g.* θ_1. The

resulting set of normalised orthogonal functions form the basis of our symmetry deformation coordinates:

$$\frac{1}{\sqrt{3}}(r_1 + r_2 + r_3) \qquad \frac{1}{\sqrt{3}}(\theta_1 + \theta_2 + \theta_3)$$

$$\frac{1}{\sqrt{6}}(2r_1 - r_2 - r_3) \qquad \frac{1}{\sqrt{6}}(2\theta_1 - \theta_2 - \theta_3)$$

$$\frac{1}{\sqrt{2}}(r_2 - r_3) \qquad \frac{1}{\sqrt{2}}(\theta_2 - \theta_3)$$

It should be noted at this stage that the use of only the bond lengths and the bond angles has resulted in the derivation of a set of six basis functions, which in theory is all that we need to describe the displacements of the observed distorted geometries from that of the reference structure. However, these six functions come from only the A_1' and the E' irreducible representations, since the projection of the bond length or bond angle in the A_2'' representation always comes to zero. To find a function that describes this representation we have to recognise that the current set of functions describe only movement in the plane of a planar entity and take no direct account of the possible movement of the central atom out of the plane. If we ascribe this movement to the term Δ and take account of the sense of the displacement and then apply the projection operators as above we find that:

$\hat{P}\ A_1'(\Delta) = 0$

$\hat{P}\ A_2''(\Delta) = 12\,\Delta$ approximating to Δ

$\hat{P}\ E'(\Delta) = 0$

In presenting Δ as one of the six necessary functions for description of this system it is obvious that one of the functions already derived must become redundant. This function is

$$S_5 = \frac{1}{\sqrt{3}}(\theta_1 + \theta_2 + \theta_3)$$

which does describe the out of plane displacement of the central atom but gives no account of sense. A displacement along S_5 can however be related to Δ by the approximation (where d is the mean bond length):

$$S_2^2 = \Delta^2 \cong S_1^2\left(2\pi\sqrt{3} - S_5\right)/9 \cong d^2\left(2\pi/\sqrt{3} - S_5\right)/3$$

The complete set of symmetry coordinates can now be given for the D_{3h} point group as:

$$
\begin{aligned}
S_1(A_1') &= \frac{1}{\sqrt{3}}(r_1 + r_2 + r_3) \\
S_2(A_2'') &= \Delta \equiv S_5 \\
S_{3a}(E') &= \frac{1}{\sqrt{6}}(2r_1 - r_2 - r_3) \\
S_{3b}(E') &= \frac{1}{\sqrt{2}}(r_2 - r_3) \\
S_{4a}(E') &= \frac{1}{\sqrt{6}}(2\theta_1 - \theta_2 - \theta_3) \\
S_{4b}(E') &= \frac{1}{\sqrt{2}}(\theta_2 - \theta_3)
\end{aligned}
\qquad [2]
$$

The derivation of the relevant kernel and co-kernel symmetries can be found from the character tables for the point group. Here the only representations that need to be looked at are the A_1', A_2'' and E' **[1]**. The characters for these are shown below.

D_{3h}	E	$2C_3$	$3C_2$	σ_h	$2S_3$	$3\sigma_v$
A_1'	1	1	1	1	1	1
A_2''	1	1	-1	-1	-1	1
E'	2	-1	0	2	-1	0

The symmetry elements preserved in the kernel group are those whose characters equal the character of the identity element for each irreducible representation. It follows therefore that the S_2 coordinate, from the A_2'' irreducible representation, preserves the symmetry elements (E, $2C_3$ and $3\sigma_v$), *i.e.* those forming the C_{3v} point group, while the kernel symmetry of the S_3 and S_4 coordinates is that of the C_s point group. The kernel of the A_1' representation is similarly the D_{3h} group.

The S_{3a} and S_{4a} coordinates are chosen so that both of them are transformed into themselves by a mirror plane passing through r_1 and bisecting θ_3 in the reference structure. Displacements along either of these coordinates will therefore show co-kernel symmetry, CoK (E', σ_v) = C_{2v}.

When using the SDC's calculated above in the actual analysis of the datasets, it is important to recognise that the terms in the functions are displacements away from the ideal reference geometry. This means that instead of using actual bond angles, *e.g.*

ϕ, it is necessary to replace these in the derived functions by ($\theta = \phi$ - 120)°, since 120° is the bond angle appropriate to the reference frame of D_{3h} symmetry.

4.2.3 CSD bonding conventions and standard covalent radii

The CSD User Manuals (Appendix 10, 1992), report average covalent radii for the atoms present in all of crystal structures available in the database. The radii for those elements encountered in this study are given in Table 4.1 below. For metal complexes, the CSD staff assign a bond between two atoms, M and X, if the interatomic distance is less than the sum of their respective radii plus a bonding tolerance value of 0.4Å, *i.e.* when the following criterion is satisfied:

$$D(M\text{-}X) < [\ \text{radius}(M) + \text{radius}(X) + \text{tolerance}\ (0.4\text{Å})\ ] \quad [3]$$

Table 4.1

Covalent radii of the elements used in this Chapter. The radii are those described in Appendix 10 of the CSD User Manuals (1992), and used by the CSD System.

element	radius (Å)	element	radius (Å)
Ag	1.59	As	1.21
Au	1.50	B	0.83
Cd	1.69	Br	1.21
Co	1.33	C	0.68
Cr	1.35	Cl	0.99
Cu	1.52	F	0.64
Fe	1.34	I	1.40
Hg	1.70	N	0.68
Ir	1.32	O	0.68
Mn	1.35	P	1.05
Ni	1.50	S	1.02
Pd	1.50	Se	1.22
Pt	1.50	Si	1.20
Rh	1.45	Te	1.47
Zn	1.45		

However, in some publications it is difficult to decide whether a bond exists between a metal atom and a potential ligand atom, especially when the interatomic distance is unusually large. This "bond : no bond" problem has to be borne in mind when searching the CSD for transition elements with a specified coordination number. If doubt does exist as to whether or not a bond should be assigned, the author's comments are usually considered, particularly when the value for D (M-X) is close to the limit. In a few unusual instances this results in a bond being defined in the CSD, in compliance with the author's views, even if the D (M-X) just exceeds the limit, but a more common discovery follows the corollary of this argument: often a bond is not defined, because the author specifically does not describe it as such, even though the distance D (M-X) may be well within the tolerance limits allowed for bonding.

This situation will become apparent later in this Chapter, when non-bonded interactions to a three-coordinated centre are described. The non-bonded distances calculated by QUEST3D can be "normalised" by subtracting from the distance the sum of covalent radii for the atoms involved, *i.e.*

$$D_{norm} = [\ M \rightarrow X \text{ non-bonded distance }] - \text{radius (M)} - \text{radius (X)} \quad \textbf{[4]}$$

Analysis of these normalised distances show situations to exist where, in numerical terms, there should be a bond between the M and X atoms (*i.e.* where the normalised distance is less than 0.4Å), but where the author has not declared a bond. The normalisation of these non-bonded distances is imperative if the all of the different ligand types are to be expressed on the same graph, *i.e.* we must look at parameters generated on an equal footing, and not those that are dependent on external factors, such as ligand type.

4.3 Methodology

Version 5.0 (1st April 1994) of the Cambridge Structural Database was used throughout for substructure search, three-dimensional analysis of the extended crystal lattices, and data analysis using the programs QUEST3D and VISTA. Local software

amendments, not yet available in the distributed system, are summarised below together with precise definitions of search fragments. Where necessary, correlations between PC's and SDC's are performed using the statistical routines of GSTAT.

4.3.1 Database search and retrieval mechanisms

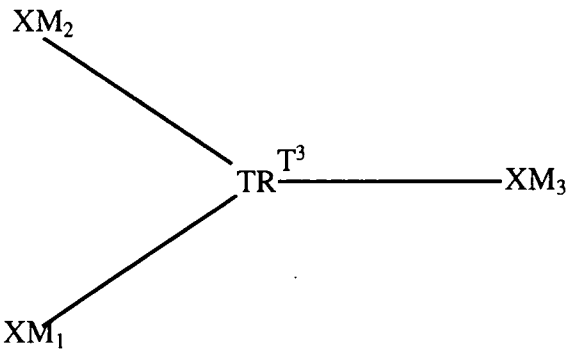
The chemical substructure examined here consisted of a general transition metal atom (TR) connected by single bonds to exactly three non-metallic elements (XM), using the ELDEF instruction to specify XM=B, C, Si, N, P, As, O, S, Se, Te, F, Cl, Br, I or At. Exact three coordination is provided by use of the 2D CONSTRAIN sub-menu command, TOTAL COORDINATION NO., which is specified to be 3 for the metal centre.

The search was further constrained by use of the CSD bit-screen search mechanisms (CSD User Manual, 1992) to locate entries with: (a) atomic coordinates available, (b) no residual numerical errors following CSD check procedures, (c) no reported disorder in the structure, (d) a crystallographic R-factor < 0.10, and which were (e) 'organometallic' compounds according to CSD chemical class assignments. The fragment to be located, together with the search script used by QUEST3D, is given in Figure 4.3 below.

The initial searches were performed on the entire database, but once the nature of the problem was established it became possible to restrict searches to specific subsets of the CSD, by utilising the "virtual database" facilities now operational in Version 5.0. There are four distinct subsets: (I) copper3 - containing the 161 refcodes (244 viable fragments) all with three-coordinate copper only, (II) mercury3 - 51 refcodes (63 fragments) all of 3-coordinate mercury, (III) silver3 - 50 refcodes (65 fragments), and (iv) others - comprising the non-copper, mercury, silver hits of all the other three-coordinate transition metals, 99 refcodes (134 fragments).

4.3.2. Data analysis

Analysis of the structures represented in each of these subsets was carried out



(a)

```
T1 *CONN
NFRAG  1
ELDEF XM= B C N P AS O S SE F CL BR I
AT1 TR 3 T3                    :XY  491  523
AT2 XM 1                       :XY  262  674
AT3 XM 1                       :XY  627  746
AT4 XM 1                       :XY  297  246
BO 1 2 1
BO 1 3 1
BO 1 4 1
GEOM
PERMUTE
DEFINE LAB1 1
SETUP P1 3 2 4
DEFINE ?DEL 1 P1
TRANSFORM DELTA =  ABS ?DEL
DEFINE A1 2 1 3
DEFINE A2 2 1 4
DEFINE A3 3 1 4
TRANSFORM ?D3 = A3 - 120.0
TRANSFORM ?D2 = A2 - 120.0
TRANSFORM ?D1 = A1 - 120.0
TRANSFORM ?2A3 = ?D3 * 2
TRANSFORM ?S4AA =  ?2A3 - ?D2 - ?D1
TRANSFORM S4A =  ?S4AA * 0.408
TRANSFORM ?S4BA =  ?D2 - ?D1
TRANSFORM S4B =  ?S4BA * 0.707
TRANSFORM ?SUMA =  ?D1 + ?D2 + ?D3
TRANSFORM S5 =  ?SUMA * 0.577
C Overlap of crystal fragments permitted
C Search for ALL crystal fragments
NFRAG -99
SYMCHK OFF
C Rejection of symmetry equivalent CRYSTAL fragments is OFF
ENANT INVERT
END
QUEST T1
```

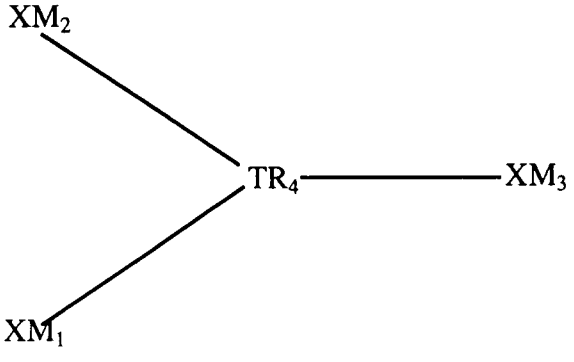
(b)

Figure 4.3

The QUEST3D alphanumeric instruction set (b) corresponding to the graphical search query shown (a), for exactly three-coordinated transition metals singly bonded to any of the atoms in the XM definition list, i.e. B, C, N, P, As, O, S, Se, F, Cl, Br or I.

using the QUEST3D routines available in the Version 5.0 3-D CONSTRAIN sub-menu. These allow for the calculation of all three of the XM-TR-XM angles present at the metal centre, as well as the calculation of the out of plane distance from the metal atom to the XM_1-XM_2-XM_3 plane. The TRANSFORM function was then used to calculate the symmetry deformation coordinates based on the observed valence angles and using the equations derived above [2].

The symmetry expansion facility, used earlier in Chapter 3 to expand a set of points over the entire hyper-dimensional parameter space, is also required for these studies. Here, the fragment location procedure of QUEST3D will ordinarily place each fragment in some arbitrary asymmetric unit of the 3D configurational space which is defined by the three valence angles (ϕ_1, ϕ_2, ϕ_3). Hence it is necessary to fill this space according to the D_{3h} topological symmetry of the planar MX_3 fragment, and with the inverse structures in three dimensions. For the MX_3 system then, the expansion must be 6-fold, and the appropriate atomic permutations required to completely fill the hyperspace are given in Figure 4.4.



(a)

XM enumerations			TR
1	2	3	4
2	3	1	4
3	2	1	4
1	3	2	4
2	1	3	4
3	1	2	4

(b)

Figure 4.4

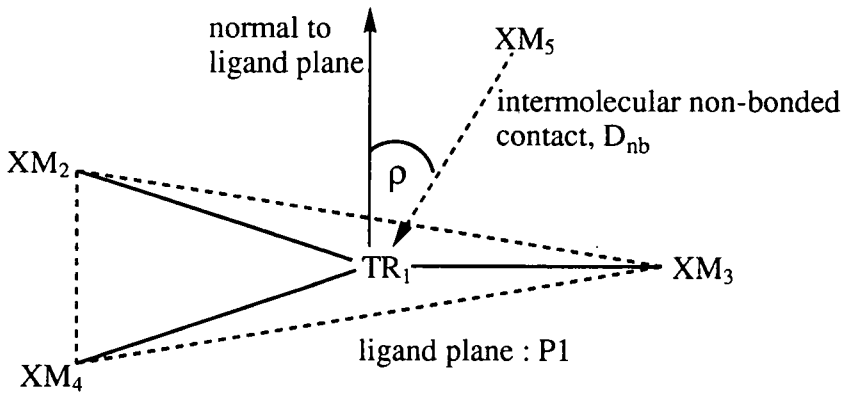
(a) one possible atomic labeling scheme, together with the necessary permutations of these labels (b) that are required to fully describe the configurational space.

A developmental modification (Johnson, 1994) to the substructure search algorithm in QUEST3D, allows the generation of the necessary permutational isomers for direct importation into VISTA. The normal atom-by-atom, bond-by-bond substructure search attempts to find a mapping of the query fragment (Figure 4.4a) to the atoms present in each "target" molecule in the CSD. The first result might, for example, map $XM_1 \rightarrow Cl$, $XM_2 \rightarrow N$, $XM_3 \rightarrow O$ and finally $TR_4 \rightarrow Cu$. The algorithm will also find five additional (and equivalent) mappings ($XM_1 \rightarrow Cl$, $XM_2 \rightarrow O$, $XM_3 \rightarrow N$, $TR_4 \rightarrow Cu$; $XM_1 \rightarrow N$, $XM_2 \rightarrow O$, $XM_3 \rightarrow Cl$; *etc.*). All six possible mappings, however, involve the same four atoms of the target molecule (Cu, N, O, Cl). Normally the last five mappings are rejected by the algorithm since the discovery of the first mapping satisfies the search question: "is fragment (a) present in molecule (b)?". However, the six mappings taken together for this D_{3h} symmetric fragment (a) represent the six permutational isomers required for the complete geometrical analysis.

Analysis in VISTA will show the symmetry-expanded nature of the dataset in question. This is an important modification as it allows a full Principal Component Analysis (PCA) on the dataset, and also the direct comparison of these generated PCA parameters with the symmetry deformation coordinates obtained from QUEST3D.

The symmetry deformation coordinates generated in QUEST3D are calculated using the three valence angle parameters. As was mentioned earlier, these angles give a "natural" description of the shape of the observed fragment, and can easily be compared with those of the reference molecule. Analysis of bond length displacements involves considerably more calculation, and ultimately provides no useful informational on the overall shape, and therefore, configuration of the fragment in question. For these reasons the bond length deformations are not used as descriptors of the configuration in this study.

In addition to the complete geometrical definitions of each dataset, QUEST3D also allows the user to search for non-bonded interactions. Within the query definition (Figure 4.5), searches were performed to assess the nature and frequency of such non-bonds within each of the subsets using the CONTACT function and by specifying the radii of each atom present in the search to be those found in Appendix 10 of the CSD



(a)

```
T1 *CONN
NFRAG  1
ELDEF XM= O F CL I S
AT1 Cu 3 T3                    :XY  445  530
AT2 XM 1                       :XY  243  693
AT3 XM 1                       :XY  650  726
AT4 XM 1                       :XY  430  266
AT5 XM 0                       :XY  765  485
BO 1 2 1
BO 1 3 1
BO 1 4 1
GEOM
VDW  CU 1.52 O 0.68 S 1.02 F 0.64 CL 0.99 I 1.40
VDW  TOL 0.75
CONTACT INTER 1 A 5 A
SETUP P1 2 3 4
DEFINE ?DEL 1 P1
TRANSFORM DELTA =  ABS ?DEL
DEFINE NBDIST 1 5
DEFINE LAB1 1
DEFINE LAB5 5
DEFINE A1 2 1 3
DEFINE A2 2 1 4
DEFINE A3 3 1 4
SETUP V1 5 1
DEFINE ?VECA7 V1 P1
TRANSFORM ?SVA7 =  SIN ?VECA7
TRANSFORM VECA =  ASIN ?SVA7
C Overlap of crystal fragments permitted
C Search for ALL crystal fragments
NFRAG -99
SYMCHK OFF
C Rejection of symmetry equivalent CRYSTAL fragments is OFF
ENANT NOIN
END
QUEST T1
```

(b)

Figure 4.5

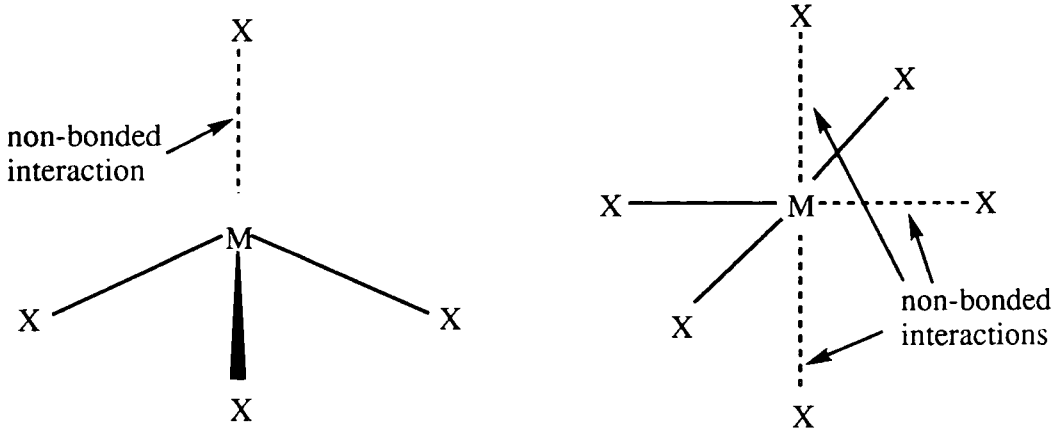
The QUEST3D alphanumeric script (b) for the non-bonded search query (a) locates all CSD entries with Cu to XM contact distance within a 0.75Å tolerance of eqn.[5]

User Manual (1992), *i.e.* the covalent radii. Contacts were found to non-metallic elements within a tolerance value of 0.75Å, *i.e.*

non-bonded distance (D_{nb}) < [sum of relevant covalent radii + tolerance (0.75Å)] **[5]**

In some cases, more than one contact is found within this non-bonded search criterion. Obviously it is necessary to be aware of this fact when studying the geometry of the MX_3 bonded species as there is likely to be a considerable difference between this geometry when only one non-bonded contact is present as opposed to the geometry found when, say, three non-bonded contacts are located (see Figure 4.6). For example, the presence of only one non-bonded ligand within the permitted contact-sphere, may result in the geometry of the three-coordinated fragment deforming away from the D_{3h} reference frame, and towards the trigonal pyramidal C_{3v} geometry (Figure 4.1), with the non-bonded ligand occupying a "pseudo-tetrahedral" position. However, if there are three non-bonded ligands present in the contact-sphere of the three-coordinated fragment, then the geometry of the 3-coordinated fragment may well distort to being more T-shaped (Figure 4.1), thereby allowing the three non-bonded ligands to occupy "pseudo-octahedral" positions.

It can be seen therefore, that in order to have a clear understanding of the geometry of the three-coordinated fragment, and of the factors that might influence that geometry, it is necessary to know the nature of all of the non-bonded contacts present



"pseudo" tetrahedral "pseudo" octahedral

Figure 4.6

Pseudo-geometries generated by non-bonded interactions at the 3-coordinated metal

within the pre-defined contact sphere. In the work described below the only non-bonded contacts studied are those for which the number of contacts is exclusively one: instances where there are a greater number of possibilities are removed prior to analysis, so as not to distort the interpretation of the data.

For instances where non-bonded contacts were observed, further testing was performed to establish the nature of this contact in terms of the geometry and direction of "approach" of the non-bonded ligand to the central metal atom. This was done in QUEST3D by defining a parameter to be the absolute angle between a vector, from the non-bonded ligand to the metal atom, and the normal to the plane of the three bonded ligands. Hence values for this angle in the region of 0° would show the direction of approach to be one "perpendicular" to the plane of the ligands, while a calculated angle of around 90° would illustrate a "co-planar" approach.

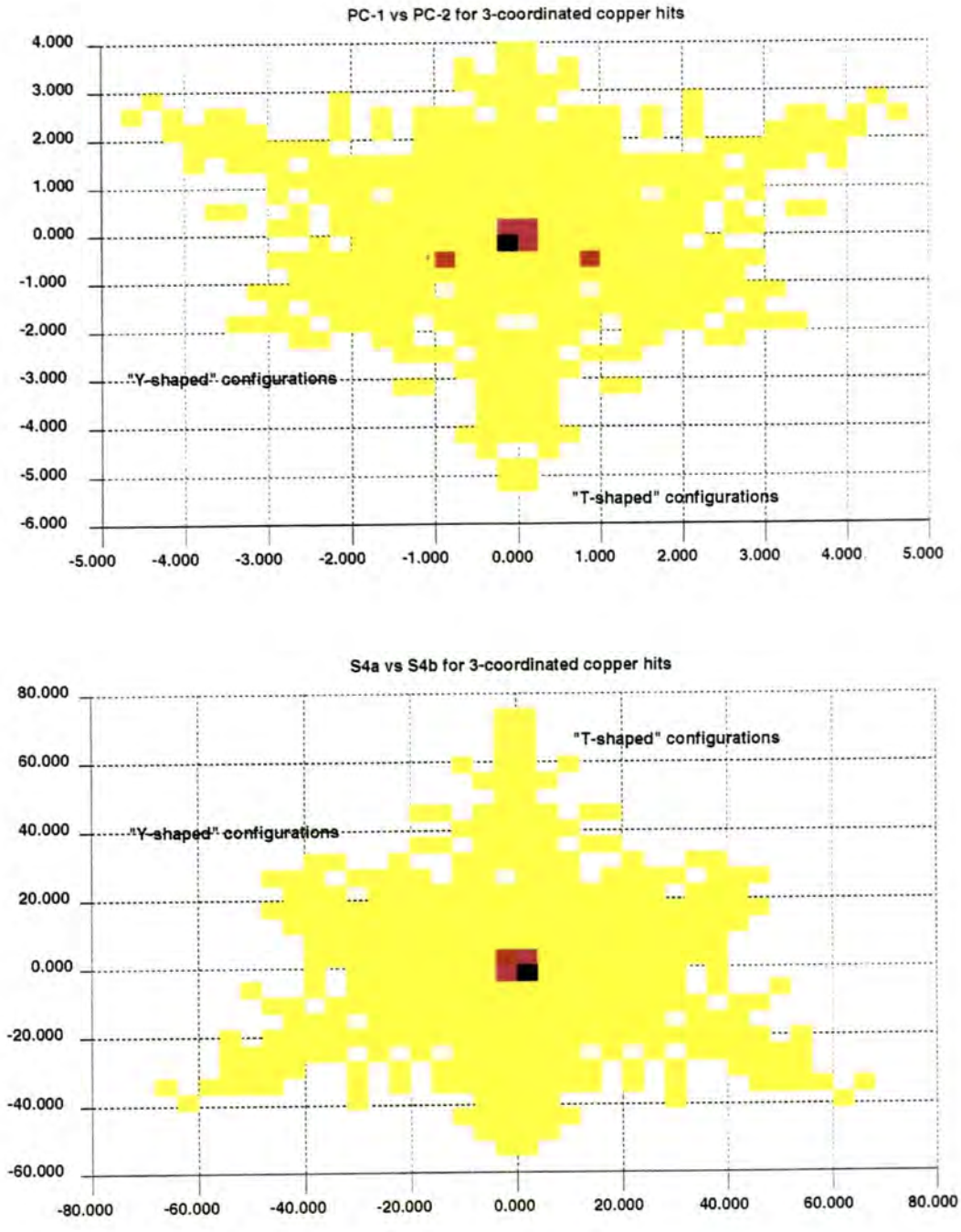
In order to analyse any possible effects that non-bonded ligand proximity might have on the geometry of the MX_3 structure, data can be extracted from the summary files of QUEST3D that shows the metal type, ligand type, and non-bonded distance. Simple programs can then be used to produce a list of specific metal covalent radius + ligand covalent radius figures which, when subtracted from the non-bonded contact distances calculated by QUEST3D, result in normalised values for this distance (D_{norm} **[4]**) that can then be correlated against any of the other geometrical parameters found in the earlier search, *e.g.* the distance that the metal atom has moved out of the plane of the three ligands, Δ.

4.4 Configurations at the 3-coordinate metal centre.

4.4.1. Copper

PCA and SDC's

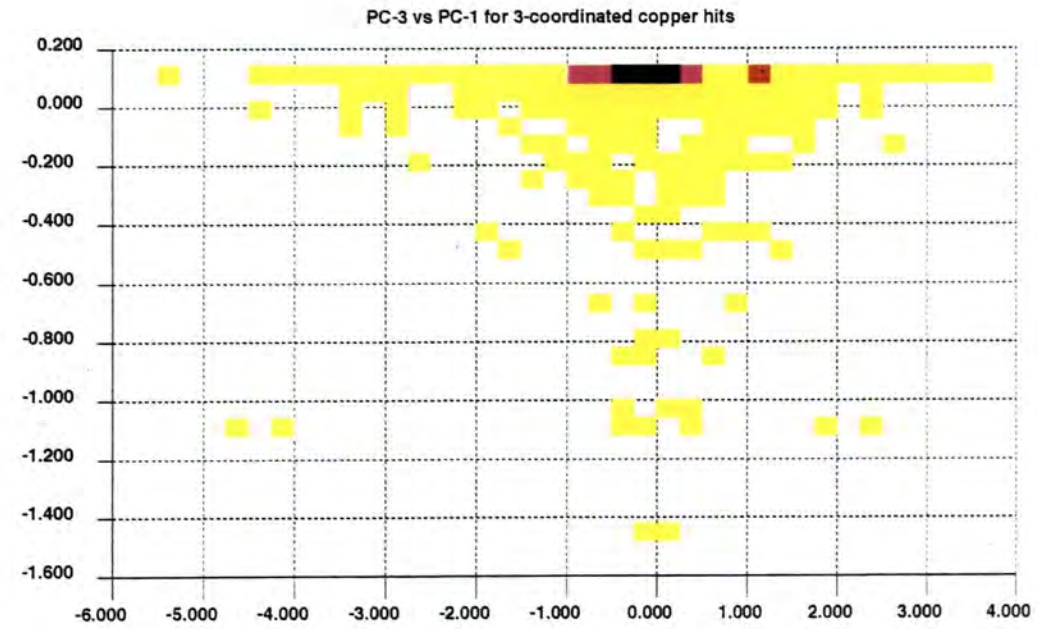
The dataset comprising those hits with 3-coordinated copper has 161 entries, generating 244 fragments of interest. The symmetry expanded scatterplots are shown in Figure 4.7a,b,c. These compare the principal component plots generated in VISTA with the symmetry deformation coordinate plots generated by QUEST. As can be seen the correlation is very high (neglecting the insignificant mirror image effect). As was

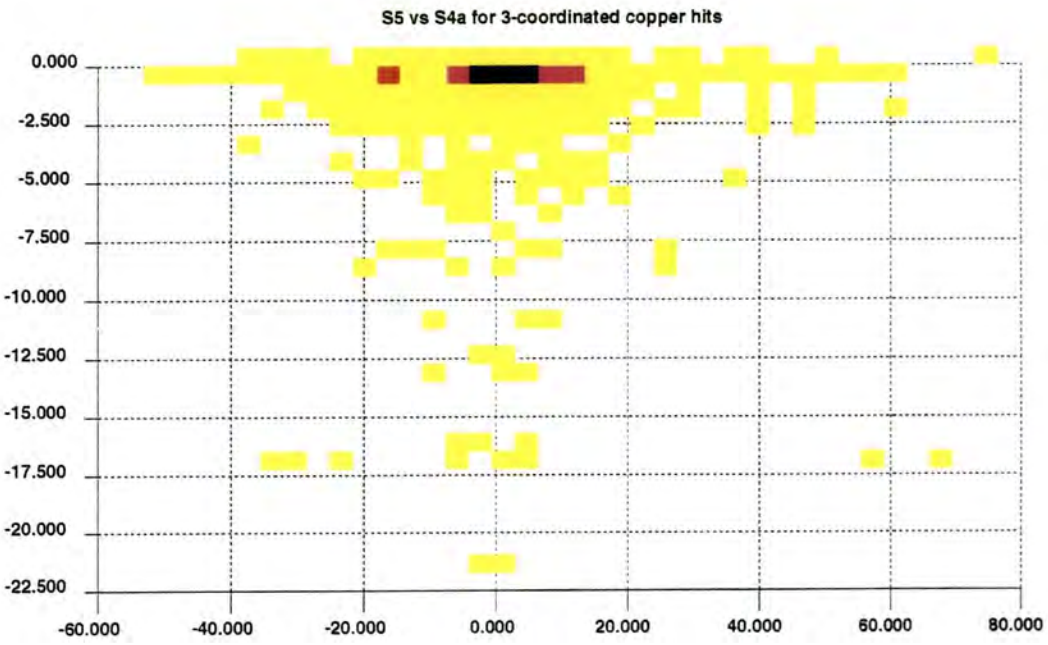


(a)

Figure 4.7

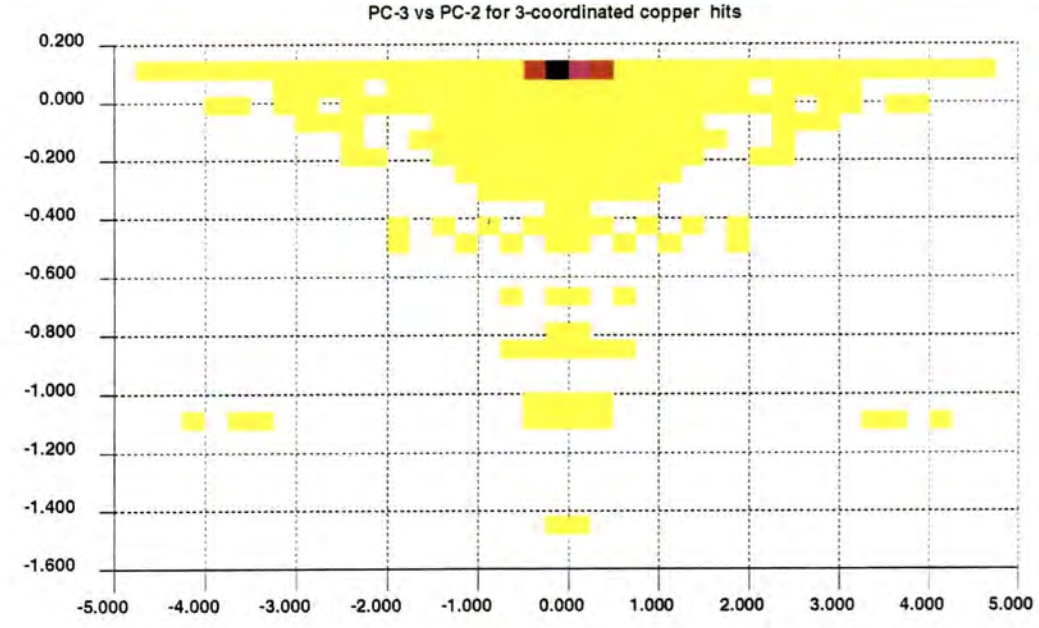
(a), (b) and (c): PC and related SDC scatterplots for 3-coordinated copper

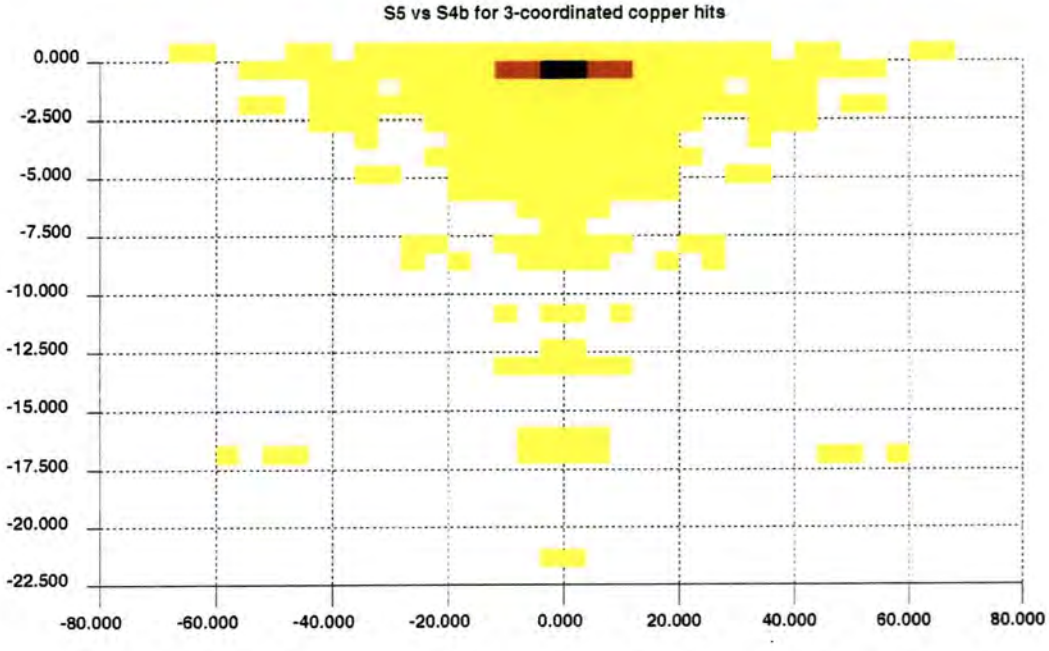




(b)

Figure 4.7 (cont.)





(c)

Figure 4.7 (cont.)

Table 4.2

PCA results and related SDCs for the copper dataset

	% of variance	equation	coordinate
PC-1	43.63	$\frac{1}{\sqrt{6}}(2\theta_1 - \theta_2 - \theta_3)$	$S_{4a}(E')$
PC-2	43.63	$\frac{1}{\sqrt{2}}(\theta_2 - \theta_3)$	$S_{4b}(E')$
PC-3	12.74	$\frac{1}{\sqrt{3}}(\theta_1 + \theta_2 + \theta_3)$	$S_5(A_2'')$

the case with the PCA / CP correlations generated in Chapter 3 for the medium ring systems, PCA can once again be shown to be a significant tool in the appraisal of general geometrical effects.

The PCA results (Table 4.2) show that the observed deformations are largely due to angular displacements following the degenerate E' coordinates, *i.e.* deforming away from the ideal D_{3h} and towards a C_{2v} (co-kernel) T- or Y-shaped structure. The small but significant variance in the PC-3 / S_5 direction maps the displacement of the copper atom out of the plane formed by the three bonded ligands; the "pyramidalisation" of the copper atom taking the structure from D_{3h} to the kernel C_{3v} description.

In broad terms, the first pair of scatterplots (4.7a) show the predominant D_{3h}-type structures as the core density at 0.0, 0.0 in both. These plots are effectively views from a point above the plane of the ligands looking down. The plots also show the 3-fold growth out of this core density of two distinct types of deformation: one mapping the T-shaped configurations, and the other mapping the Y-shaped configurations. The T-shaped ones show the greatest degree of overall distortion as they effectively map the transition from angles of 120°, 120°, 120° all the way to 90°, 90°, 180° whereas the total distortion to the Y-shaped configuration only involves an angular transition to 135°, 135°, 90°. In the second (4.7b) and third (4.7c) pairs of scatterplots the views are orthogonal to the one above. They may then be regarded as views from a point in the same plane as the ligands, *i.e.* in the D_{3h} σ_h plane. The PC-3 versus PC-2 scatterplot gives a view of the "trigonal" data from a head-on position, *i.e.* looking directly down one of the vertices, hence the symmetrical nature. The PC-3 versus PC-1 scatterplot

gives a view orthogonal to this one, *i.e.* a view from one side of the trigonal data, and therefore the data density is unsymmetrical.

The two pairs of scatterplots both show the trend for "pyramidalisation" well, with significant density being seen to evolve from the σ_h plane, and out along the S_5, or Δ, coordinate. With a few minor exceptions the displacement along the S_5 coordinate appears to maintain the equality of the basal angles formed by the three ligands to the metal atom. So even though the copper atom is effectively moving out of the plane of the three ligands the 3-fold symmetry of the resulting geometry is maintained.

When the symmetry-expanded dataset is analysed using the statistical package GSTAT, numerical figures can be attached to the loadings of the PC's, and direct correlations can be derived with the SDC's. The correlation results are shown in Table

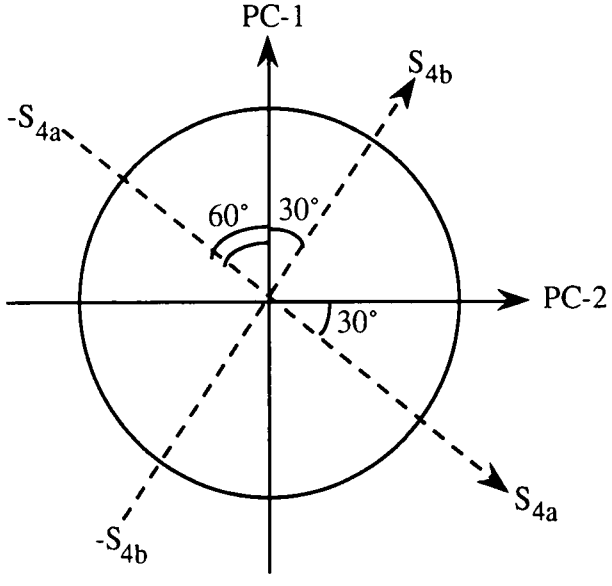
Table 4.3

(a) correlation coefficients for the SDC / PC relationship for 3-coordinated copper, together with the angles that they represent, and (b) a pictorial summary of the axial relationship between the SDC and PC representations.

(a)

	S_{4a}	S_{4b}	S_5
PC-1	-0.50 (120°)	0.87 (30°)	0.00 (90°)
PC-2	0.87 (30°)	0.50 (60°)	0.00 (90°)
PC-3	0.00 (90°)	0.00 (90°)	1.00 (0°)

(b)



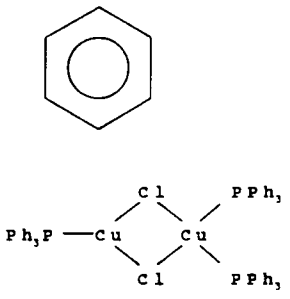
4.3(a) above, with a pictorial representation of the relationship between the SDC and PC axes shown in Table 4.3(b). It is obvious that the two descriptions are not directly correlated, but upon closer examination it is apparent that there is an axial rotation of the SDC's with respect to the PC axes.

Molecular Descriptors

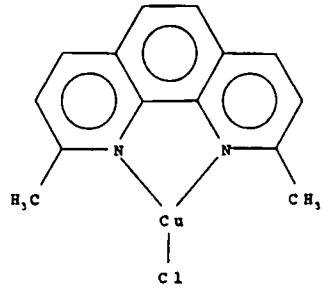
Of the 244 fragments that make up this dataset, 147 retain C_3 symmetry (within the angular limits established earlier), and are therefore D_{3h} or C_{3v} type structures. The remainder is comprised of 63 T-shaped, and 54 Y-shaped moieties, all with C_{2v} point group symmetry.

The structures that exhibit a Y-shaped configuration for the three-coordinated copper atom are almost all due to a specific steric effect. In nearly all of these cases the copper atom forms part of a 4- or 5-membered ring, with two of the three bonded ligands therefore also being members of the same ring. As the internal angles of these ring structures are ideally 90° and 108° respectively, this is all the constraint that appears to be necessary to force the molecule to adopt a Y-shape at the metal centre. Two typical examples of this effect are shown in Figure 4.8.

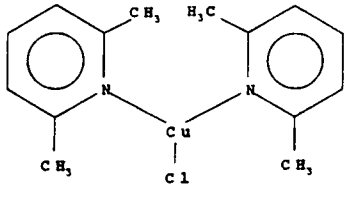
The molecules that have their three-coordinated copper centre in a T-shaped configuration are more interesting. Without exception the copper is in the +1 oxidation state. In aqueous solution this oxidation state is very unstable with respect to disproportionation (Greenwood & Earnshaw, 1986) but can be stabilised either in compounds of very low solubility or, alternatively, by complexing with ligands possessing π-acceptor character. Analysis of the subset of T-shaped moieties shows that the majority involve bonding to pyridine-type delocalised ring systems, *e.g.* DAMVOC (Dyason *et al.*, 1985; see Figure 4.8). More than one of these π-accepting ligands appears necessary for overall stability of the compound, but the result is that these bulky, electron-rich ligands repel each other quite strongly and move the configuration of the copper away from the D_{3h} ideal and towards the C_{2v} T-shaped configuration. Another reason for the formation of T-shaped molecules may be as a result of a 2+1 type addition reaction, where the resulting geometry is created by the



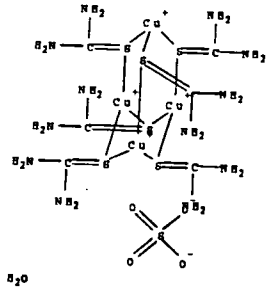
TPPCUC10



DIGWUL



DAMVOC



BILLOX

Figure 4.8

Examples of molecules which show deviation away from the ideal D_{3h} symmetry

attack of a third ligand on a linear 2-coordinated species. This possibility will be explored later, in Section 4.5.3.

It appears that molecules which are essentially unaffected by any of the above constraints will adopt a configuration that preserves ideal 3-fold symmetry. The variable influence of steric factors will of course determine just how close to that ideal situation the observed molecule lies.

Influence of oxidation states

Automatic classification according to oxidation state is only possible in the search routines of QUEST3D if the author has specified the oxidation number in the name of the compound when submitted for publication. Using this text query as a

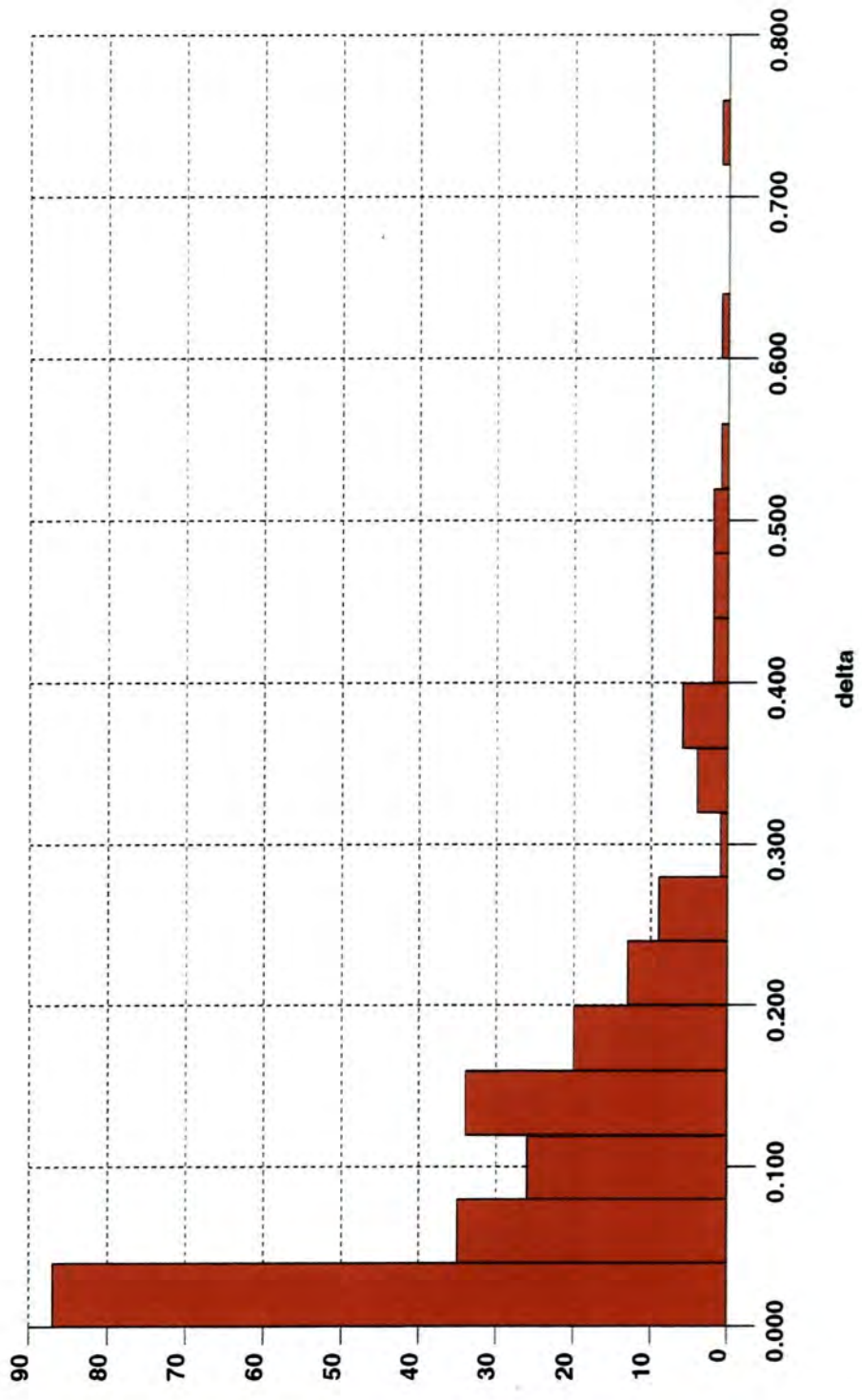


Figure 4.9

Histogram of the out of plane displacements (Å) for copper

supplementary search criterion it was found that the vast majority (74%) of the hits for the copper three-coordinated dataset was comprised of the metal in its +1 oxidation state. The occurrence of copper in the +2 state is much less common (4%), while those entries for which there is no specified oxidation state make up the remaining 22% of the total number of fragments in this dataset.

Influence of ligand type

The type of ligand bonding to the copper centre can be of importance to the resulting configuration. As mentioned above small rings force the adoption of Y-shaped configurations while the presence of pyridine-type ligands appears to be responsible for most of the instances of T-shaped configurations at the copper core. Beyond these very specific examples, however, the type of ligand to which the copper is bonded does not appear to have a dramatic effect on the overall structure. The variety of non-metallic ligand types bonded to copper in those structures that are essentially D_{3h} in character is large and encompasses virtually all of the electronegative elements to the right of the periodic table.

Out of plane displacement of the copper atom.

The histogram in Figure 4.9 shows that the copper atom does occur in a geometry which is not co-planar with the three bonded ligands. However, the vast majority (202/244) of these out of plane displacements occur in the region bounded by Δ in the range 0.0Å to 0.2Å. With reference to the scatterplots of Figure 4.7a,b,c it is also evident that these displacements are most significant for structures that tend to retain their 3-fold symmetry, *i.e.* we are observing the pyramidalisation of a D_{3h} structure along the S_5 or Δ coordinate and into the C_{3v} kernel configuration.

There are two possible explanations for this type of effect. Firstly that the environment of the copper is very constrained, *e.g.* a cage structure or a fused ring assembly. This type of environment is encountered in the cage-type structure of BILLOX (Van Meersche *et al.*, 1982; Figure 4.8), where the out of plane displacement, calculated to be 0.399Å, is obviously sterically induced. Secondly, the

pyramidalisation may be the result of a close contact between the copper atom and a fourth, non-bonded ligand. In this case the out of plane displacement of the copper atom would be significant evidence of an incipient addition reaction that converts the 3-coordinated system into a 4-coordinated system. This possibility will be explored in Section 4.5.1.

4.4.2 Silver

PCA and symmetry coordinates

The principal component scatterplots for the three-coordinated silver hits are shown in Figure 4.10a, and their corresponding SDC scatterplots are given in Figure 4.10b. These are quite interesting in that, although they retain some of the features that were observed for the copper PCA/SDC scatterplots, there are, nevertheless, some quite striking differences. It is immediately apparent that the two descriptions (PC or SDC) show the same gross features, and GSTAT analysis of the dataset provides the information in Table 4.4.

The PC loadings indicate that the PC's do in fact represent the original SDC's derived for the D_{3h} framework. Figure 4.11 indicates how the PC and SDC axes are related in 2D space. From this diagram, and by examination of the data in Table 4.4 above, it will be apparent that PC-1 is in fact 120° from the S_{4b} axis, and that PC-2 is 120° from the S_{4a} axis; the coordinates are therefore, directly correlated when the 3-fold

Table 4.4

Correlation coefficients relating the SDC and PC descriptions for 3-coordinated silver.

	% vaiance	S_{4a}	S_{4b}	S_5
PC-1	46.48	0.87 (30°)	0.49 (61°)	0.00 (90°)
PC-2	46.48	0.49 (61°)	-0.87 (150°)	0.00 (90°)
PC-3	7.03	0.00 (90°)	0.00 (90°)	1.00 (0°)

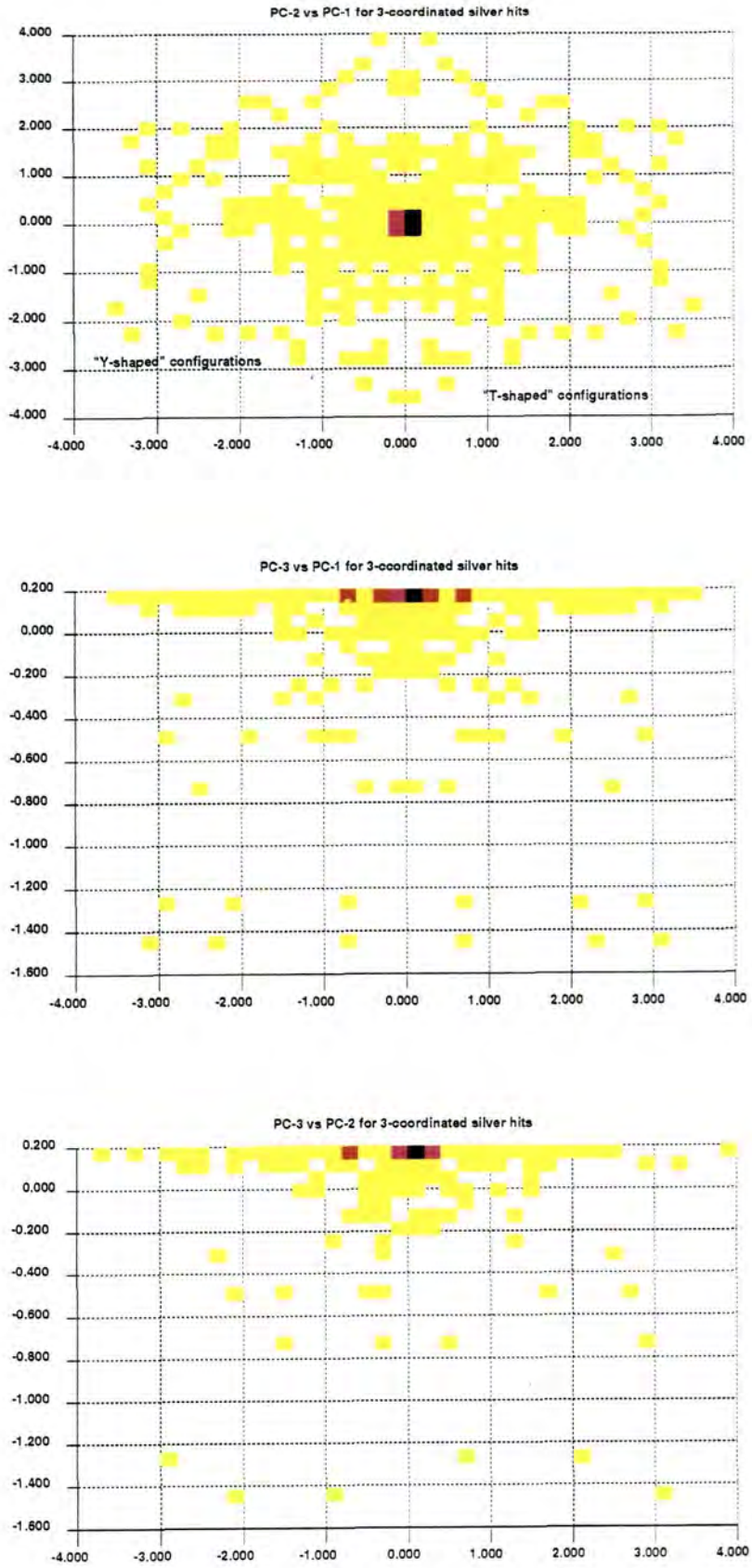


Figure 4.10(a)

PC scatterplots for 3-coordinated silver

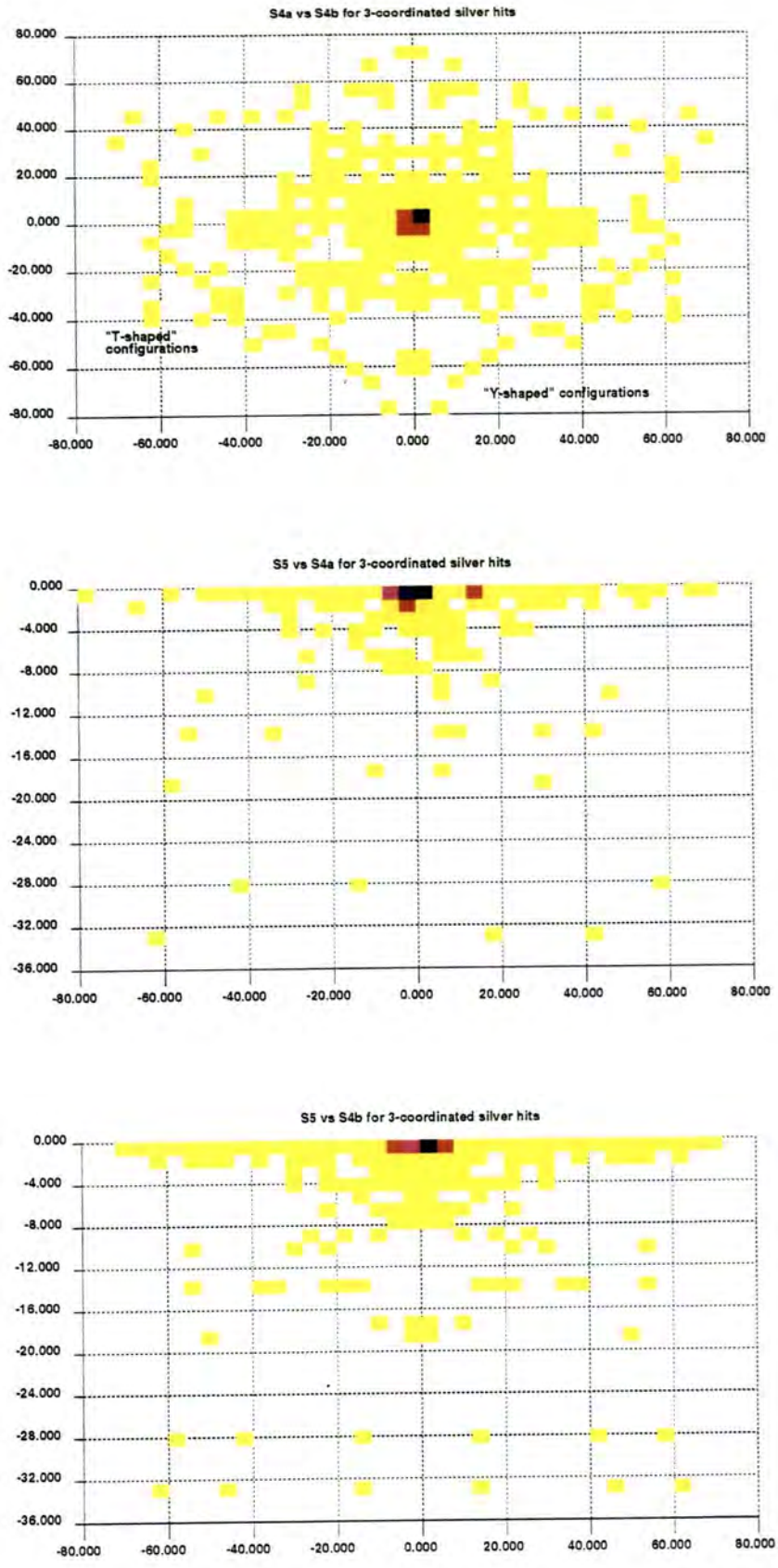


Figure 4.10(b)

SDC scatterplots for 3-coordinated silver

properties of the D_{3h} system are fully accounted for, *i.e.* the presence of the C_3 point group operation ensures that the S_{4a} and S_{4b} coordinate axes, each have two other equivalent representations at 120°, and 240° to the "original" axes. Although comparison the scatterplots for the PCA and SDC approaches would appear to support this view, it must be noted that S_{4a} and S_{4b}, and therefore also PC-1 and PC-2, are equivalent descriptions of a degenerate representation (E') and therefore the comparison of plots involving these axes is arbitrary.

In the PC-2 vs PC-1 scatterplot (Figure 4.10a) the overlapping 3-fold symmetries of the deviations towards both the Y- and T-shaped configurations from the central core of density which represents the D_{3h} configurations, can again be discerned. This time, though, there is an outer ring of density representing some heavily distorted versions of the T- and Y-shapes and then a concentration of the less severely distorted moieties closer to the inner core density.

The PCA plots that incorporate the PC-3 values, *i.e.* the plots that identify the movement of the silver atom out of the plane of the three bonded ligands along the S_5 or Δ coordinate, show the majority of the observed structures to be planar. The incidences which do show significant displacement along this coordinate are relatively few. This observation is reinforced by the weightings of the calculated principal components with the degenerate PC-1 and PC-2 each being responsible for 46.48% (Cu

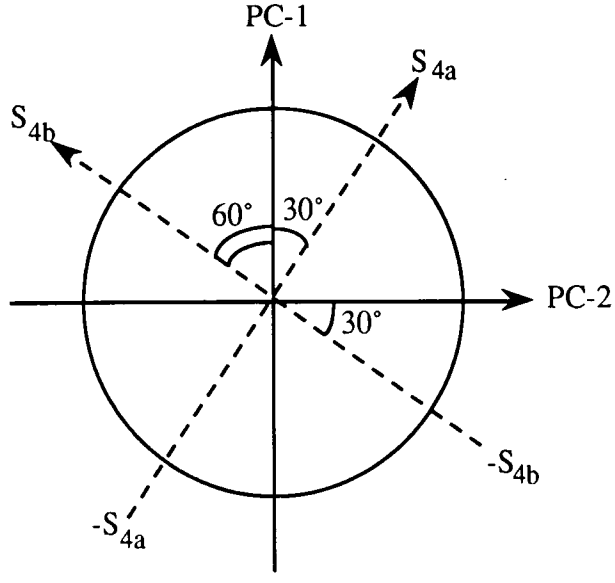


Figure 4.11

Pictorial representation of the relationship between the PC and SDC axes.

43.63%) of the total variance, and the PC-3 component only being responsible for 7.03% (Cu 12.7%) of the total variance. These figures show the in-plane angular distortions to be even more significant in this particular dataset than the displacement of the metal atom out of the ligand plane.

Molecular descriptors.

The 65 fragments that comprise this dataset can be classified as follows. 25 fragments retain the C_3 symmetry of the parent point group and so are either D_{3h} or C_{3v} type structures depending on the degree of displacement along the S_5 coordinate. 21 have T-shaped geometry, and 19 have Y-shaped geometry as defined earlier.

The Y-shaped configurations are again artefacts of a specific and severe geometrical restraint, all having the 3-coordinated silver atom in a small ring system, most commonly 4- or 5-membered. This restraint places two of the three bonded ligands in the same ring as well, and the resulting ring strain is sufficient to distort the angles created at the metal atom away from the ideal set of basal angles and towards those of the Y-shaped C_{2v} configuration.

The T-shaped moieties, for the most part, are again comprised of fragments which have a silver (I) atom at their core, and π-accepting ligands bonding to this metal. The resulting geometry appears to be related to the size of these ligands: the larger the ligands the greater the repulsion between the two and the more likely it is that a T-shaped compound will result. However, there is also a significant case for regarding the silver (I) compounds as being of the (2+1) coordination category (*i.e.* as examples of the addition of a third ligand to a linear AgL_2 framework), since the most commonly observed coordination of silver (I) compounds is the linear two-coordinated system. It would appear to be a relatively simple operation to expand this 2-coordinate system into the 2+1 and hence the 3-coordinated system with T-shaped geometry as is observed here. This possibility will be explored in Section 4.5.3.

As was found for copper, the molecules for which the above constraints do not apply will tend to form under the ideal geometry of the D_{3h} configuration. Distortions along the S_5 coordinate away from this ideal geometry will also be analysed later.

Oxidation states

For 3-coordinated silver the only "cited" (in the CSD) oxidation state is +1 (56% of the total number of entries), with situations where the oxidation state is undefined making up the remaining 44%. This is to be expected: silver (I) is the best-known oxidation state of silver, with a stable d^{10} electronic configuration, and instances of the +2 oxidation state are primarily observed as square-planar 4-coordinated complexes.

4.4.3 Mercury

Principal Component Analysis

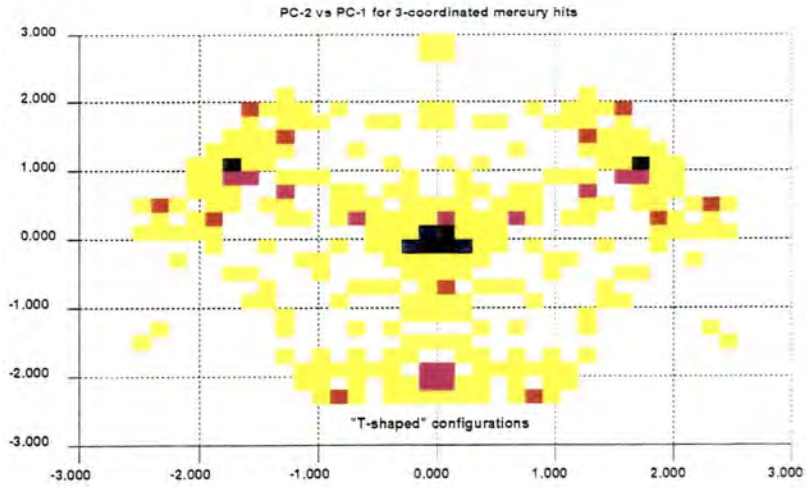
For mercury, the correlation between the calculated PC's and the derived SDC's is very similar to that of silver. Table 4.5 shows the correlation coefficients between the two descriptions. GSTAT finds two PC's are sufficient to fully describe this dataset (PC-1 and PC-2 each accounting for 48.76% of the total variance), immediate indication of the fact that out of plane displacement of the metal atom is a relatively minor effect in this dataset. Using similar arguments described for the observed correlations apparent in the silver dataset, PC-1 is directly correlated to S_{4a}, and PC-2 to S_{4b}, by a rotation in the inherent 3-fold axis of the D_{3h} description.

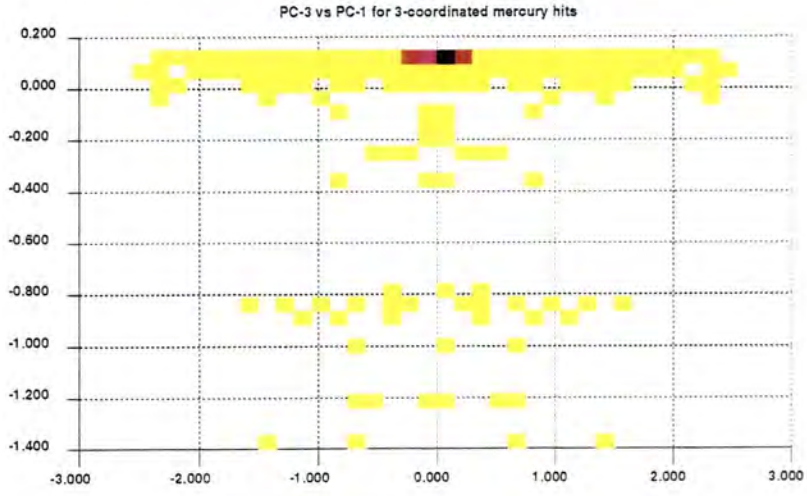
The PCA/SDP scatterplots for mercury again look similar in many respects to those found for copper, and yet there is a definite preference amongst the entries in this dataset to occur with a T-shaped configuration (Figures 4.12a and 4.12b). This is most

Table 4.5

Correlation coefficients for the relationship between the PC and SDC representations of the 3-coordinated mercury dataset

	% variance	S_{4a}	S_{4b}	S_5
PC-1	48.76	-0.50 (120°)	0.87 (30°)	0.00 (90°)
PC-2	48.76	0.87 (30°)	0.50 (60°)	0.00 (90°)





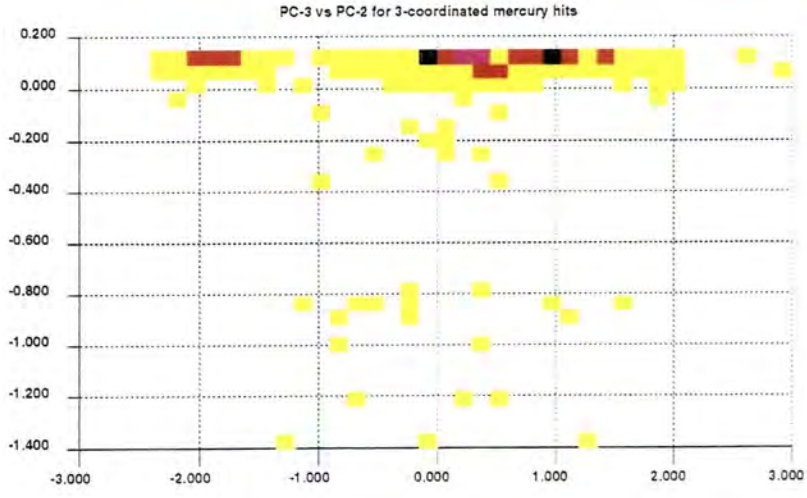


Figure 4.12(a)

PC scatterplots for 3-coordinated mercury

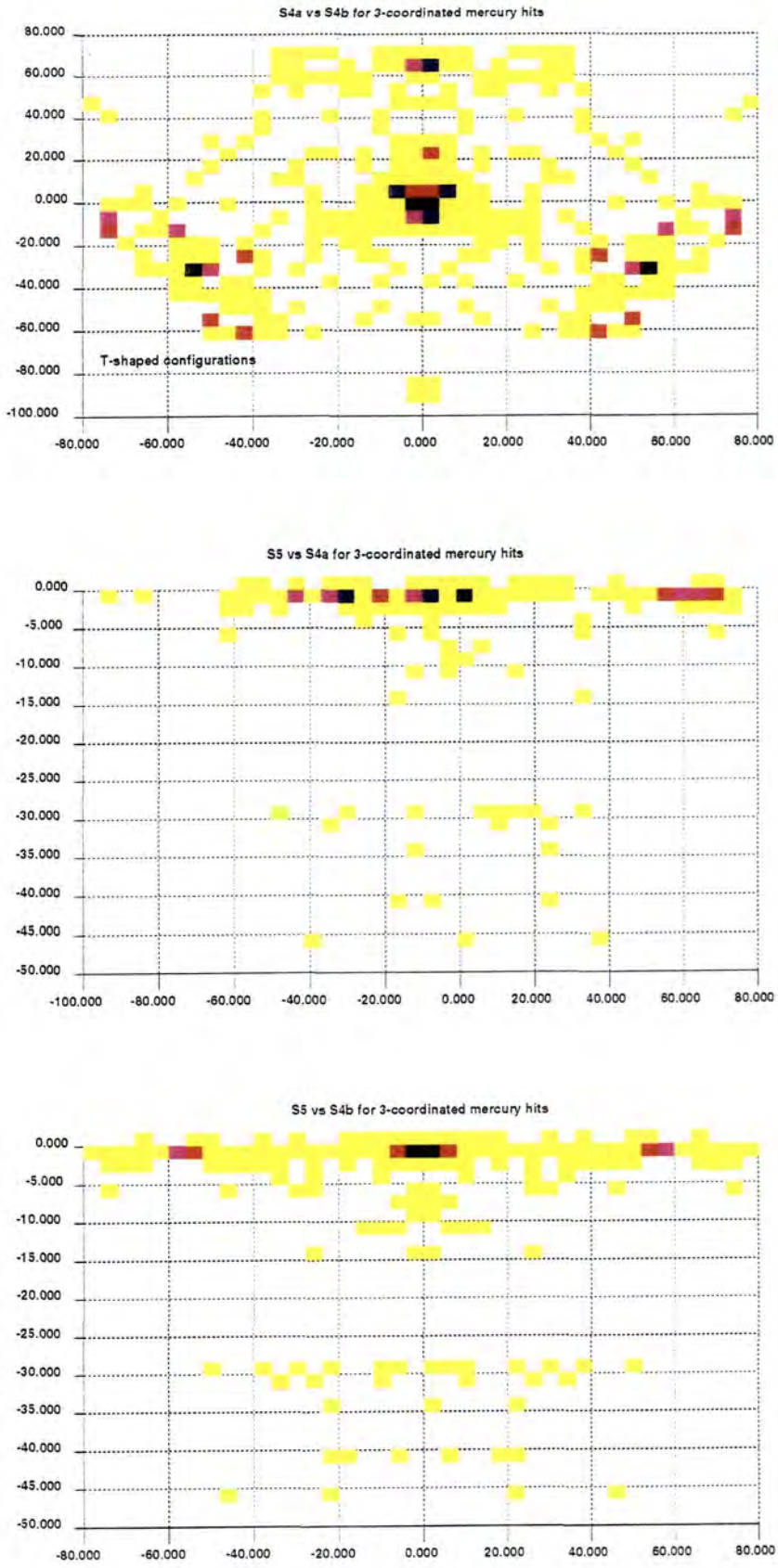


Figure 4.12(b)

SDC scatterplots for 3-coordinated mercury

evident in the PC-2 vs PC-1 principal component scatterplot. In this plot it is apparent that, apart from the central core of density, which represents the D_{3h} structures, there are also very noticeable centres of density at the points corresponding to the T-shaped entities. The lack of density linking the core D_{3h} structures to the prominent T related density indicates that very few structures in this dataset lie on a smooth distortion path, *i.e.* the distinctions between the two main configurations should be profound. Very few compounds appear to adopt the Y-shaped configuration, as will be obvious from the absence of points at the expected positions for these structures.

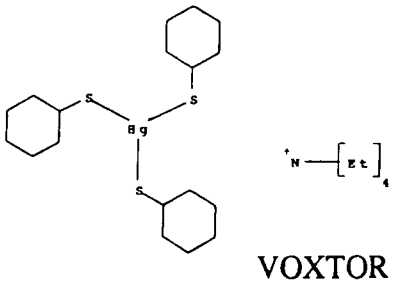
The distortion of the mercury atom out of the plane of the ligands is not a major factor here, as can be seen from the two plots involving PC-3. In fact the degenerate PC-1 and PC-2 components account for 48.76% of the total variance each, leaving only 2.47% of the total variance for the PC-3 component. Pyramidalisation of the central metal is not therefore a prominent feature in this dataset.

Molecular descriptors

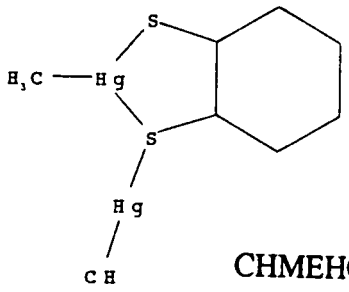
Of the 63 fragments which make up the 3-coordinated mercury dataset, by far the most dominant in terms of numbers is the T-shaped sub-set with 46 fragments. The D_{3h} fragments number 9 in comparison and the Y-shaped configurations only 8.

The Y-shaped configurations are again visible when the mercury is incorporated in a constrained ring system together with two of its bonded ligands. The typical ring sizes for this dataset however are five and three. There are also four instances where the geometry is best described as being Y-shaped but where the mercury is essentially free to adopt any configuration, *e.g.* VOXTOR (Alsina *et al.*, 1992; Figure 4.13). In four of these structures the mercury atom has a formal negative charge, and, in each case, all three of the bonded ligands are equivalent and bulky. With the exception of one fragment there are no significant non-bonded contacts to any of these mercury atoms.

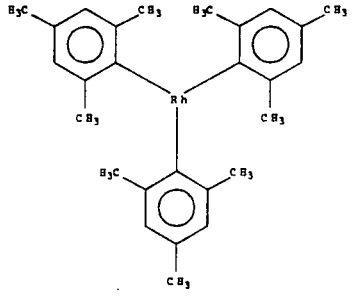
The trigonal D_{3h} configurations all arise in unconstrained systems in which the mercury atom appears free to adopt this configuration. Typically the ligand types involved are the common halogens or sulphur.



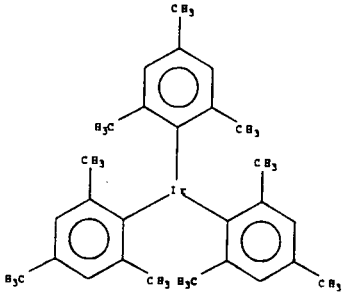
VOXTOR



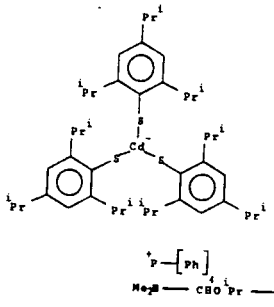
CHMEHG



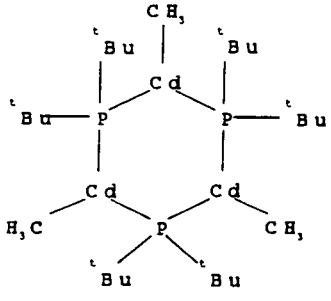
VEZNIX



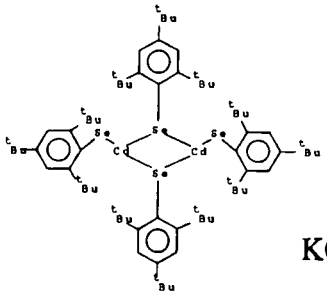
JURLUD



KELWAZ



KASCUC



KOCDUB

Figure 4.13

Characteristic molecules from the mercury and "others" datasets

There are no obvious chemical dissimilarity features that could account for the prevalence of the T-shaped entities in this dataset. Many of the hits are found to include the mercury in organic ring systems, *e.g.* CHMEHG (Alcock *et al.*, 1980; Figure 4.13), for which the angle subtended by the two sulphur atoms at the mercury is 84°, and the remaining angles are 108° and 168°, and this type of structure does appear with some regularity throughout the T-shaped subset. However this description is by no means an overwhelmingly common feature linking all of the observed hits.

Oxidation State

In terms of the oxidation states observed in this dataset the hits are split fairly evenly between mercury in the +2 state (54%), and those with no oxidation state cited in the CSD (46%). There are no hits for the +1 or +3 oxidation states.

There are precedents for the formation of the T-shaped entities. Grdenic (1965, 1981) has studied the complexes of mercury in detail and finds that the coordination number is not always a straightforward parameter, with the mercury exhibiting a propensity to increase its effective coordination. In such a fashion it would be reasonable to assume that the 3-coordinated compounds observed in crystal structures are artefacts of the addition of an extra ligand to a linear two-coordinated compound. Evidence for this comes from bond-length to bond-strength studies (*e.g.* Canty & Gatehouse, 1976; Hutton & Irving, 1979), which indicated that in most of the compounds of this kind, one bond is weak, and two bonds are strong, the two strong bonds being to the ligands that are closest to being diametrically opposite to each other. This is the most appropriate model for these T-shaped compounds

The 3-coordinated mercury dataset also contains a large number of entries for which additional non-bonded contacts fall within "reasonable" distances. These can have a large effect on the overall structural geometry adopted and are discussed in more detail in Section 4.5.1.

4.4.4 Others.

This dataset is made up of all of the 3-coordinated transition metal compounds with the exception of those belonging to the copper, silver and mercury datasets. The breakdown of the fragments in terms of the central metal under investigation is given in Table 4.6.

From Table 4.6 it can be seen that first row transition metals to the right of chromium all have 3-coordinated complexes to a lesser or greater extent. In addition coordination to only three ligands can also be observed for the heavier metals of Groups VIII, IB and IIB.

PCA scatterplots

The correlation coefficients for the PCA and SDC representation of this symmetry-expanded dataset are given in Table 4.7. In this instance the results are very easily interpreted, as the correlations are perfect, with PC-1 correlating directly to S_{4b}, PC-2 to S_{4a}, and finally PC-3 to S_5.

The PCA performed on the three basal angles of the members of this dataset produces interesting and descriptive plots, as can be seen in Figure 4.14a, and in

Table 4.6

Composition of the "others" dataset in terms of the number of entries retrieved from the CSD, and the number of fragments present in those entries

metal atom	no. fragments	no. entries
chromium	4	2
manganese	22	15
iron	11	6
cobalt	8	6
nickel	6	4
zinc	23	19
rhodium	2	2
palladium	6	4
cadmium	12	7
iridium	1	1
platinum	11	9
gold	28	24

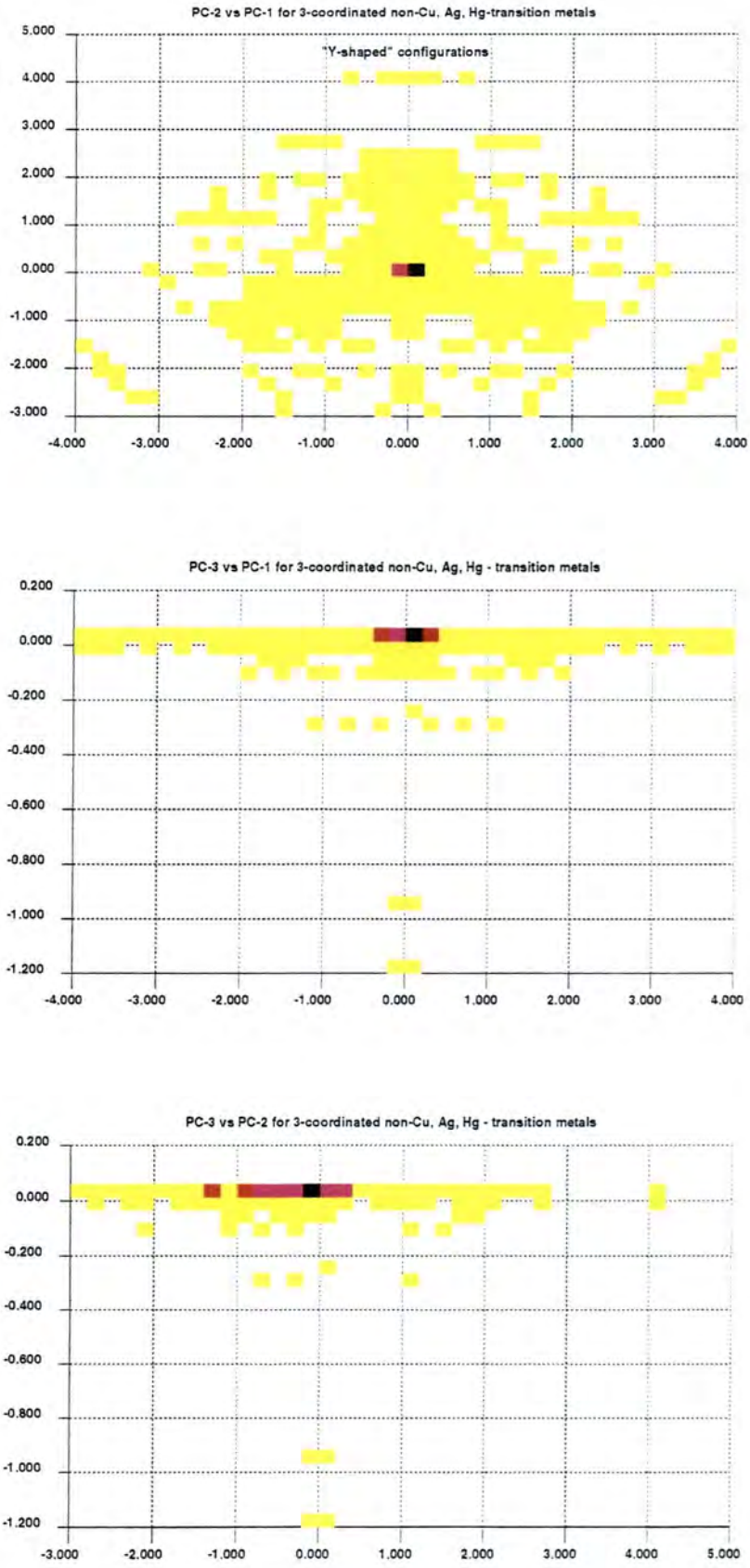
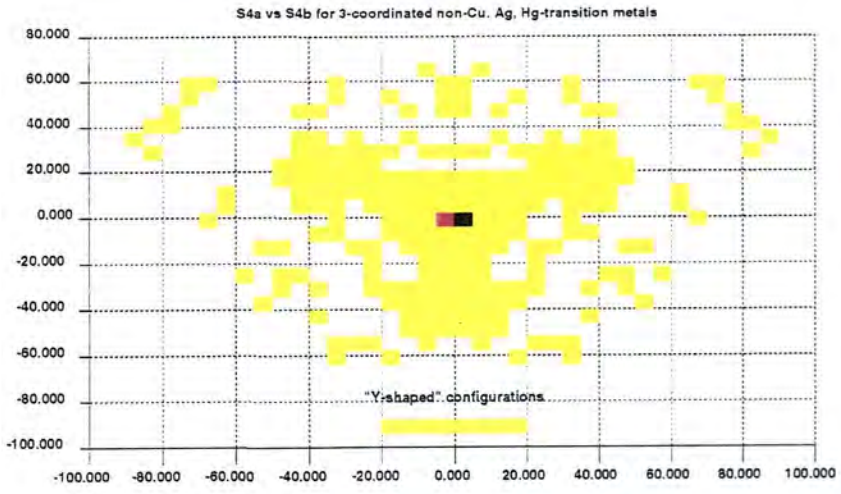
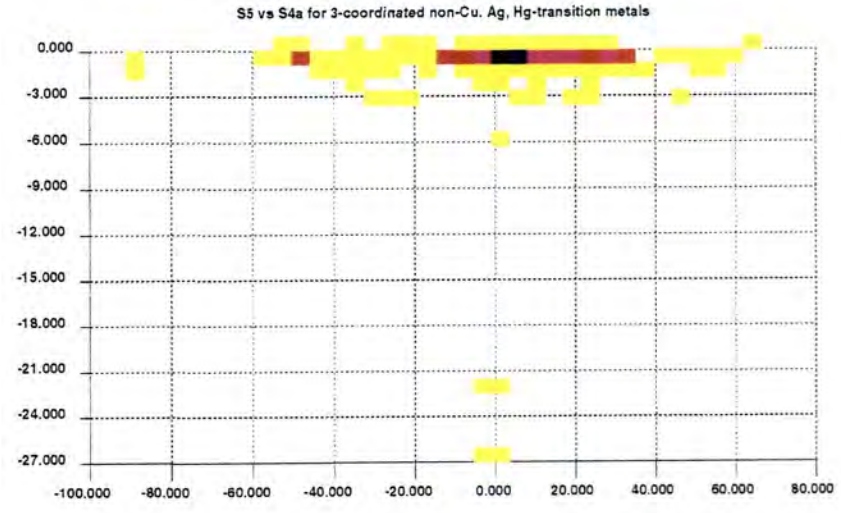


Figure 4.14(a)

PC scatterplots for 3-coordinated non-Ag, Hg, Cu transition metals





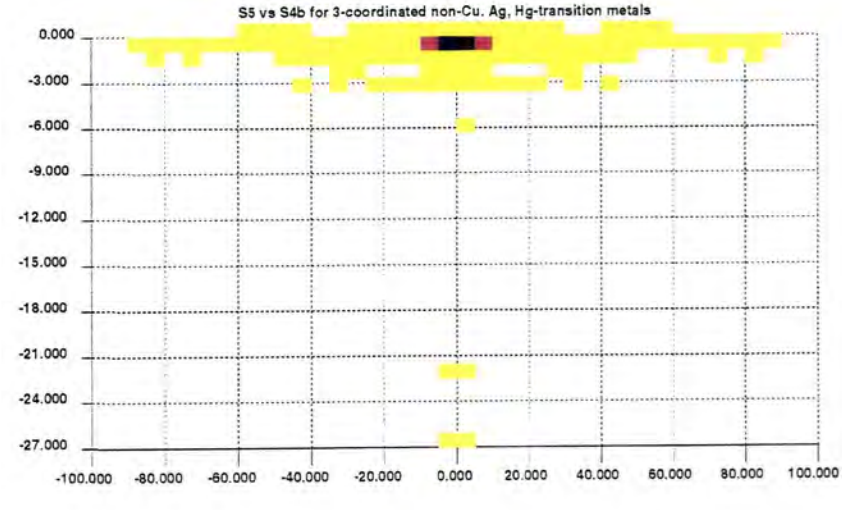


Figure 4.14(b)

SDC scatterplots for 3-coordinated non-Ag, Hg, Cu transition metals

conjunction with the SDC plots of Figure 4.14b. The plot of PC-2 against PC-1 indicates that the entries in this dataset tend to form in the D_{3h} or the Y-shaped configurations, and not the T-shaped configuration. The data point density is concentrated on D_{3h} but also spans the radial link to the concentrations at the areas representing the Y-shaped moieties. The virtual absence of points in the sectors that represent the T-shaped molecules, indicates that these configurations are not of great importance within this dataset.

With the exception of the points representing just one or two hits, the plots that include PC-3 show that almost all of the component molecules of this dataset have their central metal atom co-planar with the three bonded ligands. The weightings of the principal component axes are shown in Table 4.7. These data clearly show that the distortions along the S_5 coordinate that takes the structure from the D_{3h} configuration to the C_{3v} configuration are of minimal importance here.

Molecular Descriptors

From the histogram (Figure 4.15) mapping the out-of-plane displacement of the metal atom, it is evident that the majority of hits are essentially co-planar, with over 70% having a displacement of less than 0.1Å. Of the outliers, the most prominent occurs in the rhodium compound (VEZNIX: Hay-Motherwell *et al.*, 1990; Figure 4.13) of trimesityl-rhodium (III), which has an out of plane displacement of 0.798Å, and is pyramidal in shape. The other major outlier is again due to pyramidalisation at the metal centre, maintaining 3-fold symmetry, but driving the out-of-plane displacement to the value of 0.730Å; this instance being due to the trimesityl-iridium(IV) compound

Table 4.7

Correlation coefficients for the axial relationships between the SDC and PC representations for the "others" dataset.

	% variance	S_{4a}	S_{4b}	S_5
PC-1	49.7	0.00 (90°)	-1.00 (180°)	0.00 (90°)
PC-2	49.7	1.00 (0°)	0.00 (90°)	0.00 (90°)
PC-3	0.6	0.00 (90°)	0.00 (90°)	1.00 (0°)

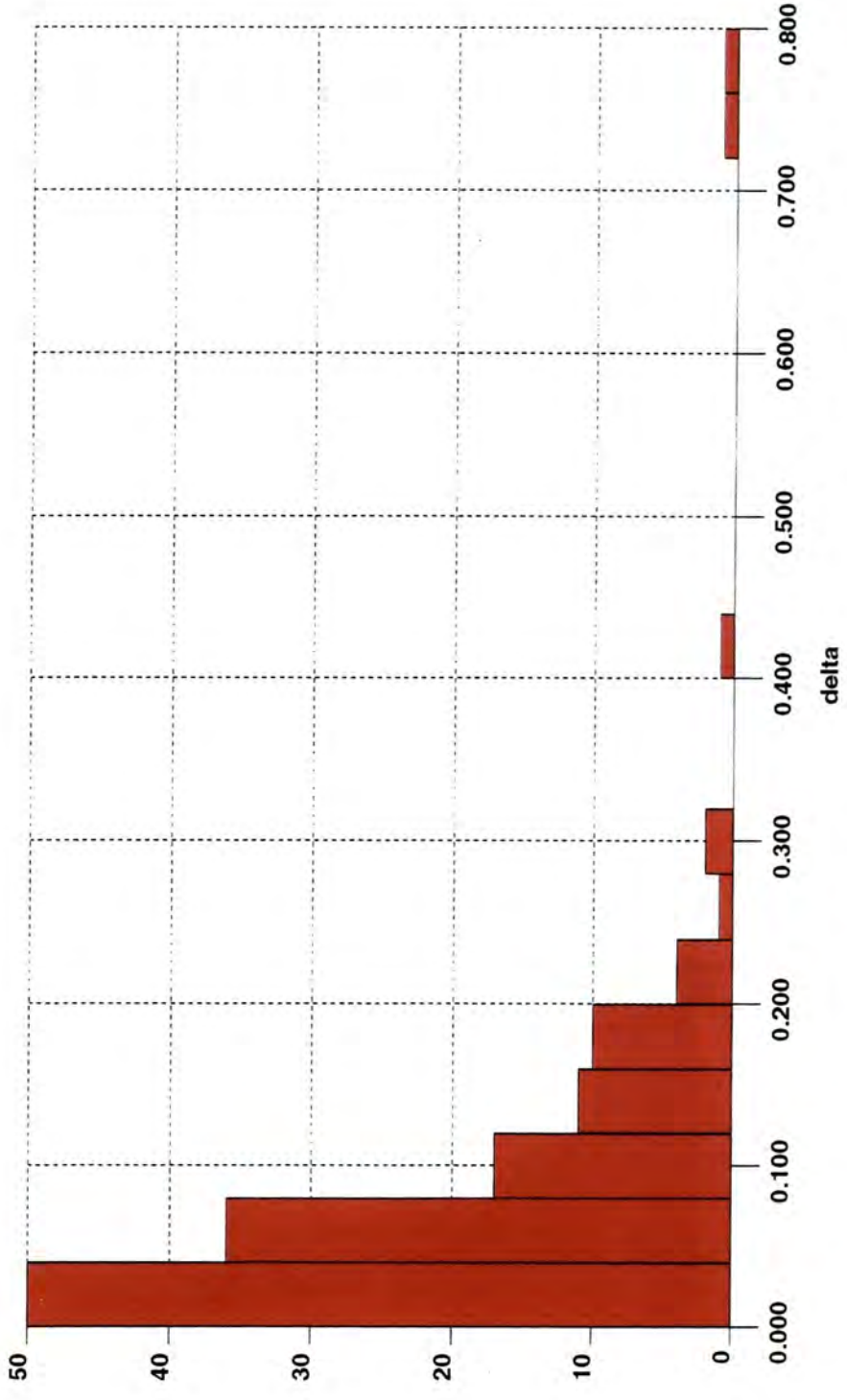


Figure 4.15

Histogram of the out-of-plane displacements (Å) of the non-Ag, Hg, Cu transition metal atoms

(JURLUD: Hay-Motherwell *et al.*, 1992; Figure 4.13). The classification of the remaining hits in terms of the geometrical reference structures discussed above, leads to the results in Table 4.8.

It is interesting to note that, while there appears to be a considerable number of fragments that can be classified as being essentially C_{2v} Y-shaped, and thus distorted away from the reference D_{3h} structure, these hits are largely due to the metal atom being in some form of constraining ring system. When molecules are analysed that allow the full degree of freedom of movement with respect to their central metal atom, then the overwhelming majority of structures show the expected ideal D_{3h} symmetry.

As an illustration of this point, we can consider the structures found for cadmium as the central metal atom. Of the total number of 12 hits, only 3 can be classified as maintaining their 3-fold nature. One of these is due to the "free" molecule KELWAZ (Gruff, 1990; Figure 4.13) which exhibits all of the geometrical properties of the D_{3h} system. The other 2 hits are due to the alternating Cd-P-Cd 6-membered ring system found for KASCUC (Benac *et al.*, 1989; Figure 4.13), where the internal angles are constrained to being approximately 120°, and so the overall representation is one of

Table 4.8

Classification of the observed geometries for members of the "others" dataset

metal	no. observed	D_{3h}	Y-shaped C_{2v}	T-shaped C_{2v}	pyramidal C_{3v}
chromium	4		4		
manganese	22	4	14	4	
iron	11	1	8	2	
cobalt	8	1	6	1	
nickel	6	3	3		
zinc	23	6	16	1	
rhodium	2	1			1
palladium	6	5			1
cadmium	12	4	8		
iridium	1				1
platinum	11	7		4	
gold	28	7	10	11	

maintaining the 3-fold nature. However, when the other hits for this subset are analysed it is found that they are all due to the 4-membered ring system Cd-X-Cd-X, which necessarily constrains the internal angles at the cadmium atoms, and thus forces the adoption of the Y-shaped C_{2v} structure.(*e.g.* KOCDUB: Bochmann *et al.*, 1991; Figure 4.13). This is an illustration of a general trend found throughout this dataset.

Of the few T-shaped molecules found in the dataset, the majority involve the heavier metals, notably platinum and gold. All of the gold T-shaped structures involve the metal in the +1 oxidation state, and the similarities to silver are drawn. In the +1 oxidation state, the metal readily complexes to two ligands in a linear fashion (Greenwood & Earnshaw, 1986), but may be susceptible to a weak bond coming in perpendicular to this strongly bonded linear unit and forming the observed T-shaped compound.

4.5 The Influence of Non-bonded Contacts

The main reason for analysing copper, silver and mercury separately from the other transition metals was because some members of each of these three datasets show evidence of close contacts from the metal in question to a "fourth" non-bonded ligand. All of the contacts studied are of the intermolecular type, with search criteria as described in Section 4.3.1, and with the tolerance in **[5]** set to 0.75Å. The results of the non-bonded contact searches are summarised in Table 4.9.

Table 4.9

Results of the search for intermolecular non-bonded contacts from the exactly 3-coordinated transition metals to a non-metallic atom (XM)

	number of contacts within tolerance of 0.75Å			
subset	zero	1	2	3
copper	222	25	0	0
silver	67	20	4	2
mercury	25	18	19	2
others	134	0	0	0

The majority of the contacts are single, *i.e.* only one "extra" ligand satisfies the contact criteria, and these will be analysed in depth below. It is interesting to note here, though, that the heavier the metal the more numerous are the contacts found. For the instances that show 3 + 2 coordination the general mode of "addition" of the two extra ligands is in such a way as to form a trigonal bipyramid with the non-bonded ligands approaching from opposite directions to the two faces of the plane of the bonded ligands. The geometry of this approach is rarely perfect but the overall result is clear.

Where three non-bonded contacts are involved the result is a complex with pseudo-octahedral geometry, as would be expected. The angles subtended at the metal centre by the 3 + 3 ligands all approximate to those angles found for a regular octahedron.

4.5.1. (3 + 1) coordination at the metal

The three datasets (copper, silver and mercury) that exhibit non-bonded intermolecular contacts, have already been analysed in terms of their inherent geometry. In general, there is a precedent for the metal atom to move out of the plane of the three bonded ligands and "pyramidalise". If this observed effect is related to the effective proximity of a fourth non-bonded ligand, then it should be possible to analyse the two effects together, and derive a correlation between them. To do this, it is necessary to know not only the out of plane displacement of the metal atom (Δ), but also a normalised distance (D_{norm}) between the two non-bonded atoms. The procedure and reasons for normalising this distance were discussed in Section 4.2.3.

Care was taken to exclude from the analysis all of the hits that have been found previously to have more than one possible non-bonded contact. In such a way the results presented below represent only the 3 + 1 coordination system.

Figure 4.16 shows the cumulative non-bonded contacts for all three of the metals being analysed. There is a general trend that can be identified: the closer the ligand, the greater is the out of plane distortion of the metal atom. However, the exact nature of this correlation is difficult to ascertain from the limited amount of data

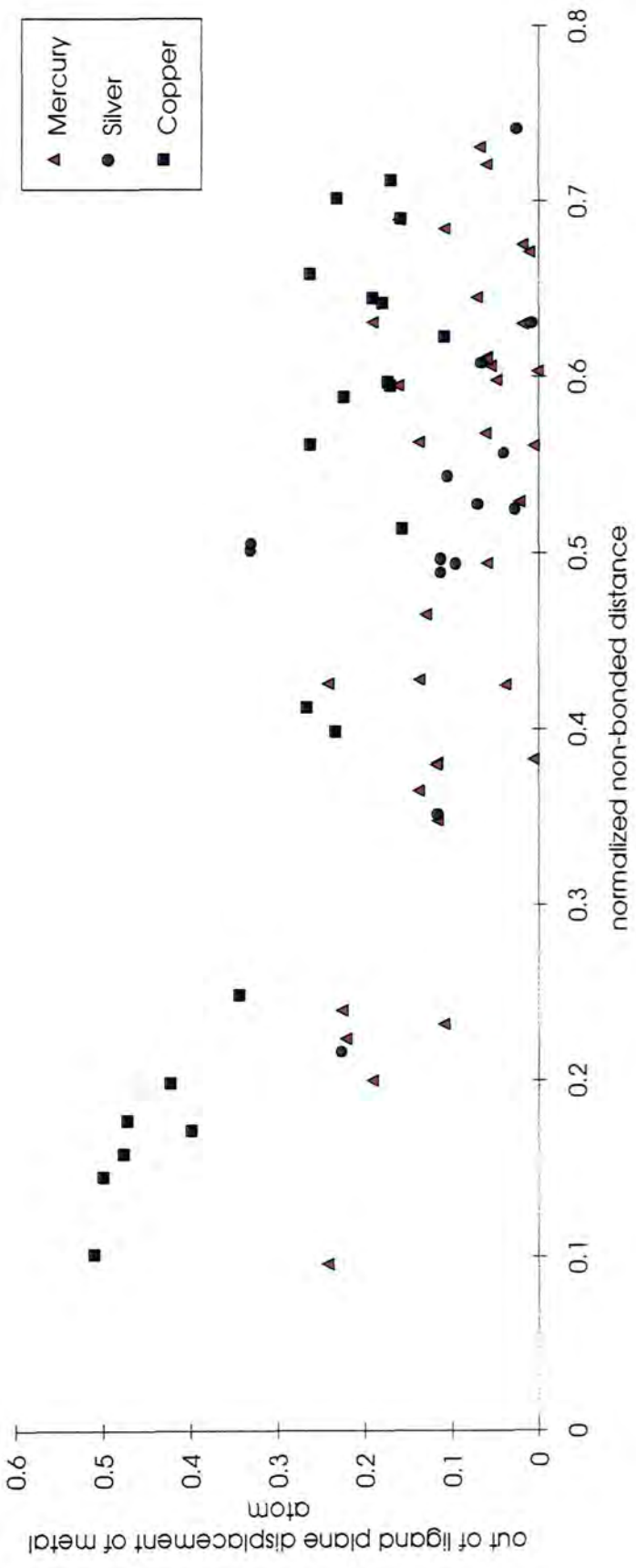


Figure 4.16

Plot of normalised non-bonded distance (D_{norm}) against out-of-plane displacement (Δ) for 3-coordinated copper, silver and mercury atoms

available. To the far right of the plot the non-bonded fourth ligand is too distant from the metal to interact significantly with it, and consequently the metal atom remains co-planar with its three bonded ligands. As the plot is traversed to the left, the non-bonded ligand is seen to get closer to the metal, and as it does so, the metal begins to elevate itself out of the plane formed by its bonded ligands, and towards the approaching fourth ligand.

The line of approach of these fourth ligands can be found by defining an angle (ρ) to be the acute angle subtended by the vector joining the non-bonded ligand to the metal, and the normal to the plane of the three bonded ligands (see Figure 4.5). The resulting histogram can be seen in Figure 4.17. It is evident that the majority of the contacts approach the metal centre along a vector almost perpendicular to the plane of the ligands, *i.e.* along the direction of the normal to the plane, and thus forming "tetrahedral" 3 + 1 coordinated compounds. There are a very few cases which show the fourth ligand to be entering the coordination sphere of the metal so as to create a "square planar" moiety, with the metal atom, the three bonded ligands, and the fourth non-bonded ligand, all essentially coplanar. These examples are all due to an $HgCl_3^-$

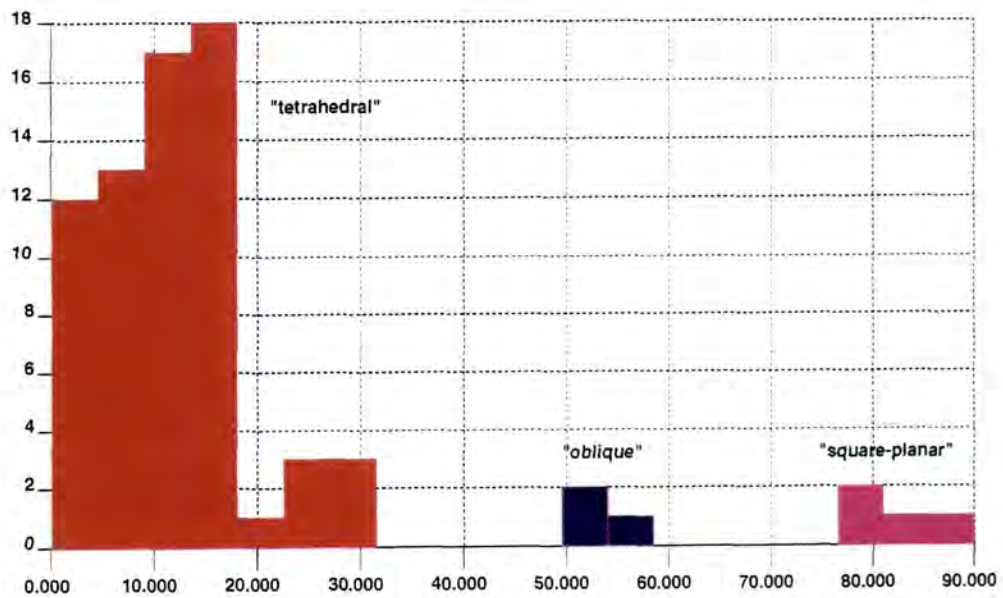


Figure 4.17

Histogram of the angle (ρ), mapping line of approach of the fourth non-bonded ligand

counter anion being present in the crystal structure, which is stabilised by the addition of a fourth ligand, usually chlorine, to complete a square planar configuration at the mercury atom.

4.5.2. (3 + 1)-coordination leading to 4-coordination

As the metal atom leaves the plane and "pyramidalises", the bond angles subtended at the metal by the three bonded ligands, show a concerted decrease. The greater the out of plane distortion, the smaller the angles become. Eventually therefore, a situation can be envisaged in which the 3 + 1 coordinated metal data will begin to coincide with the data that could be similarly derived for the metals that exist as exactly 4-coordinated species.

In order to test this hypothesis, exactly four-coordinate metal datasets were collected for M = Cu, Ag and Hg, where the metals were restricted to bonding only to non-metals (XM, Fig. 4.3, as before). Any entries for which non-bonded contacts were calculated to lie within the tolerance of 0.75Å were subsequently eliminated from the datasets, so as not to distort the analysis, *i.e.* to eliminate any possibilities of "pseudo" 5- or 6-coordination. Once the datasets were clearly defined the following geometrical parameters were extracted from them. For each hit, the bond length from one of the four ligands to the metal was calculated, and this distance subsequently normalised by subtracting the sum of the covalent radii of the relevant bonded atoms. The displacement of the metal atom out of the plane of the remaining three bonded ligands was also calculated. This procedure could then be repeated for each of the four bonded ligands in turn. By doing this, a complete analysis of the effect of the strength of the bond on the geometry of the species, can be formulated : the bond strength being related to the normalised bond length by the principles of Pauling's equation (1947, 1960) and as illustrated by Bürgi (1973) for S_N2 substitution in the $XCdS_3Y$ system. The equation:

$$\delta d = -a \log n \qquad [6]$$

relates the differences in interatomic distances (δd) to the logarithm of a bond number,

n. For the curves generated below the relevant expression would be:

$$\text{abs } \Delta = -a \log [(-D_{norm} + D_{norm\ max}) / 2\ D_{norm\ max}] + \text{constant} \qquad [7]$$

All of the values used in this form of the equation can thus be derived from the experimental data (abs Δ = absolute value for out of plane displacement, D_{norm} = normalised distance).

The number of hits found, particularly for 4-coordinated copper, and for mercury as well, is very large, and in the plots shown below this mass of data has been reduced by including only certain subsets of the total number of hits. It is possible however, to show all of the 4-coordinated hits for silver in a meaningful fashion, and so these are plotted (Figure 4.18) together with the 3 + 1 coordinate hits found earlier. A logarithmic fitting curve of the Pauling type described above is also shown. The curve, and the data that it represents, effectively shows that the non-bonded contacts to the 3-coordinated species are in fact the initial stages of an incipient addition reaction which takes the 3-coordinated trigonal planar silver through to a 4-coordinated tetrahedral configuration.

It should be noted from the plot that some of the bonds that have been stipulated to exist between the silver atom and one of its ligands, in fact fall outside the standard bonding tolerance of 0.4Å used in the CSD.

Of the other plots shown (Figures 4.19 and 4.20) the transition from 3+1 coordination towards complete 4-coordination is also apparent. Three-coordinated mercury is primarily approached by a non-bonded chlorine ligand (Figure 4.20), which causes it to begin to pyramidalise slowly. Once the fourth bond has been formed then it becomes apparent that the closer the chlorine gets to the mercury atom, the stronger the bond that is formed, and the greater the extent of the pyramidalisation at the metal, until the geometry finally becomes tetrahedral.

Oxidation states

Both the silver and mercury plots only extrapolate the 3 + 1 coordination sphere into the tetrahedral 4-coordination sphere. All of the evidence suggests that these two

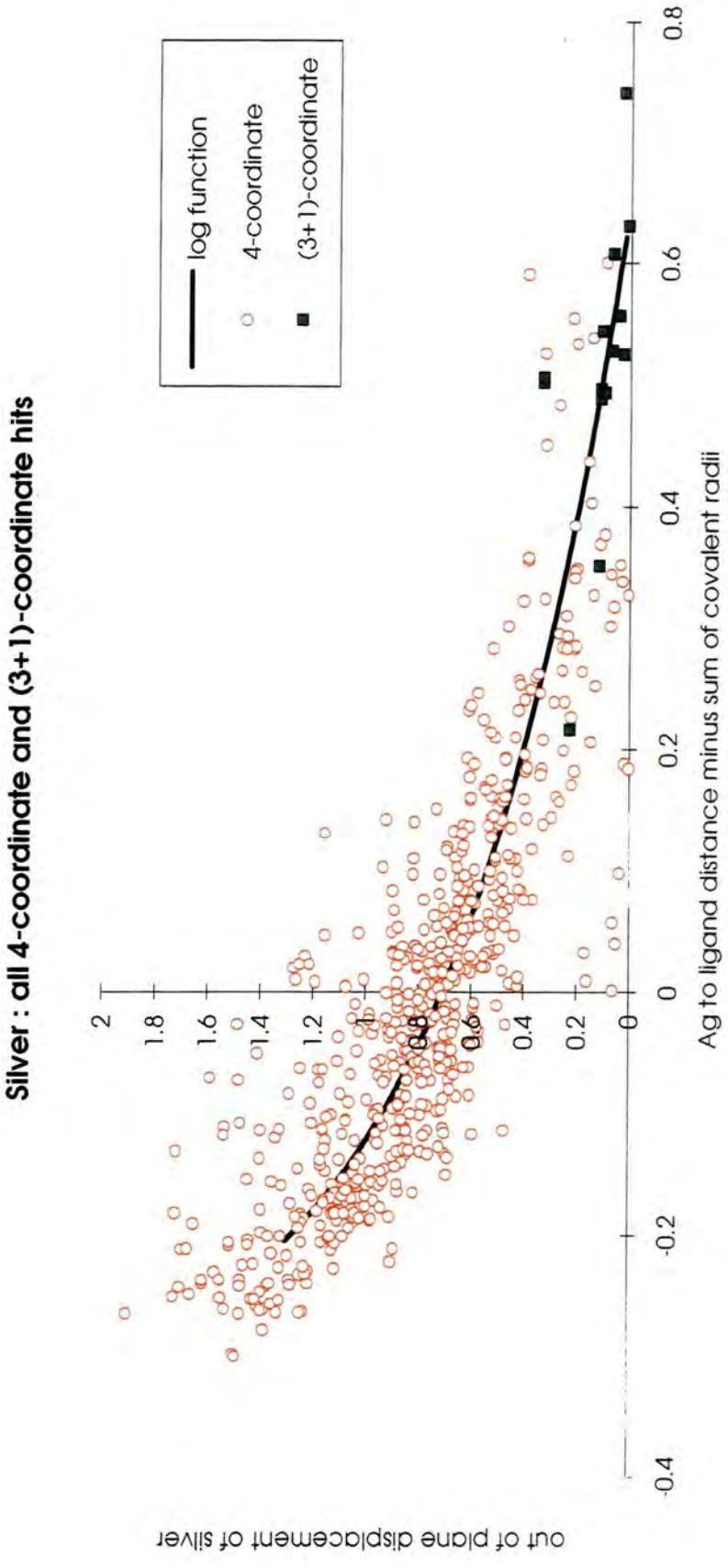


Figure 4.18

Normalised Ag -> L distances plotted against out-of-plane displacement for (3+1)- and 4-coordinated silver compounds.

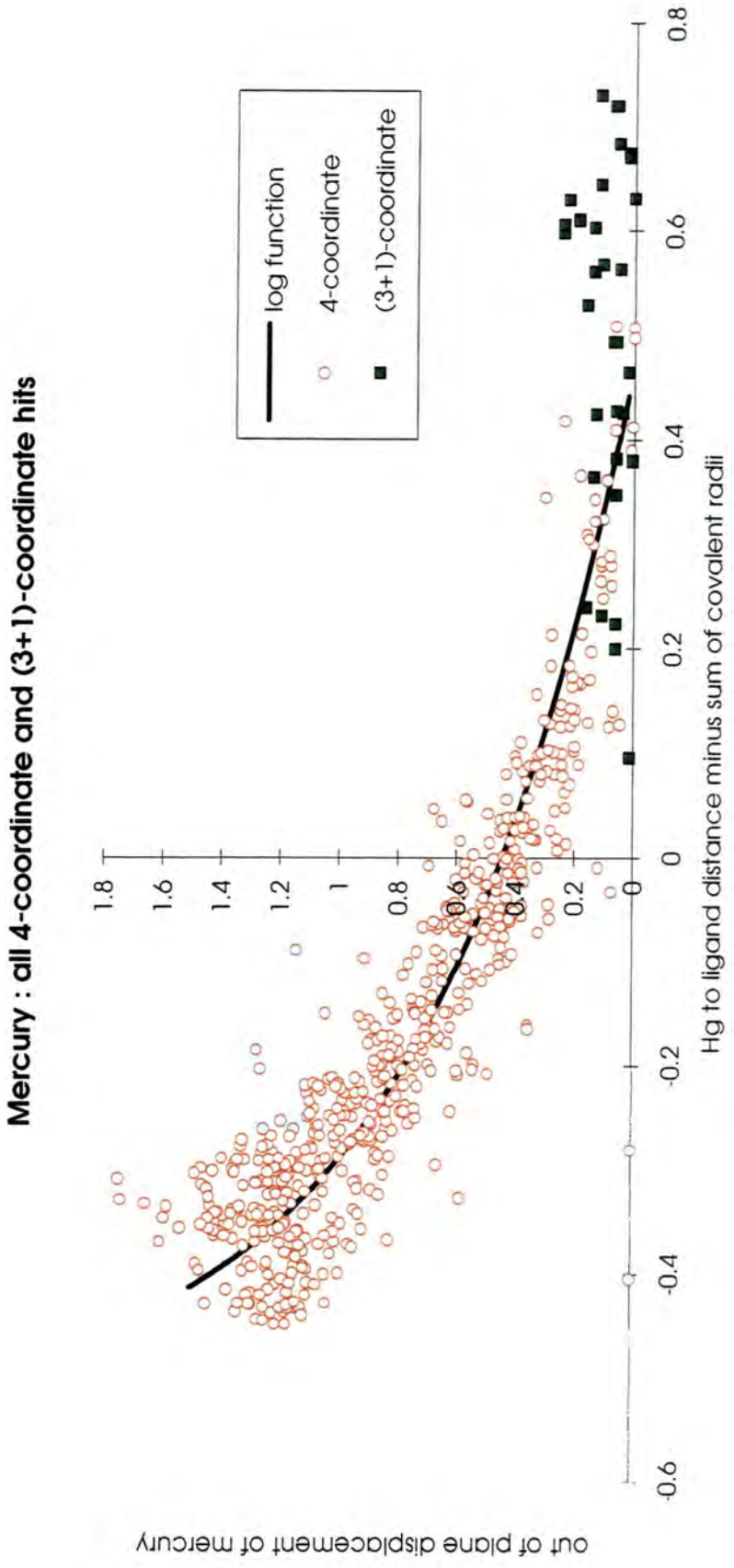


Figure 4.19

Normalised Hg -> L distances plotted against out-of-plane displacement for (3+1)- and 4-coordinated mercury compounds

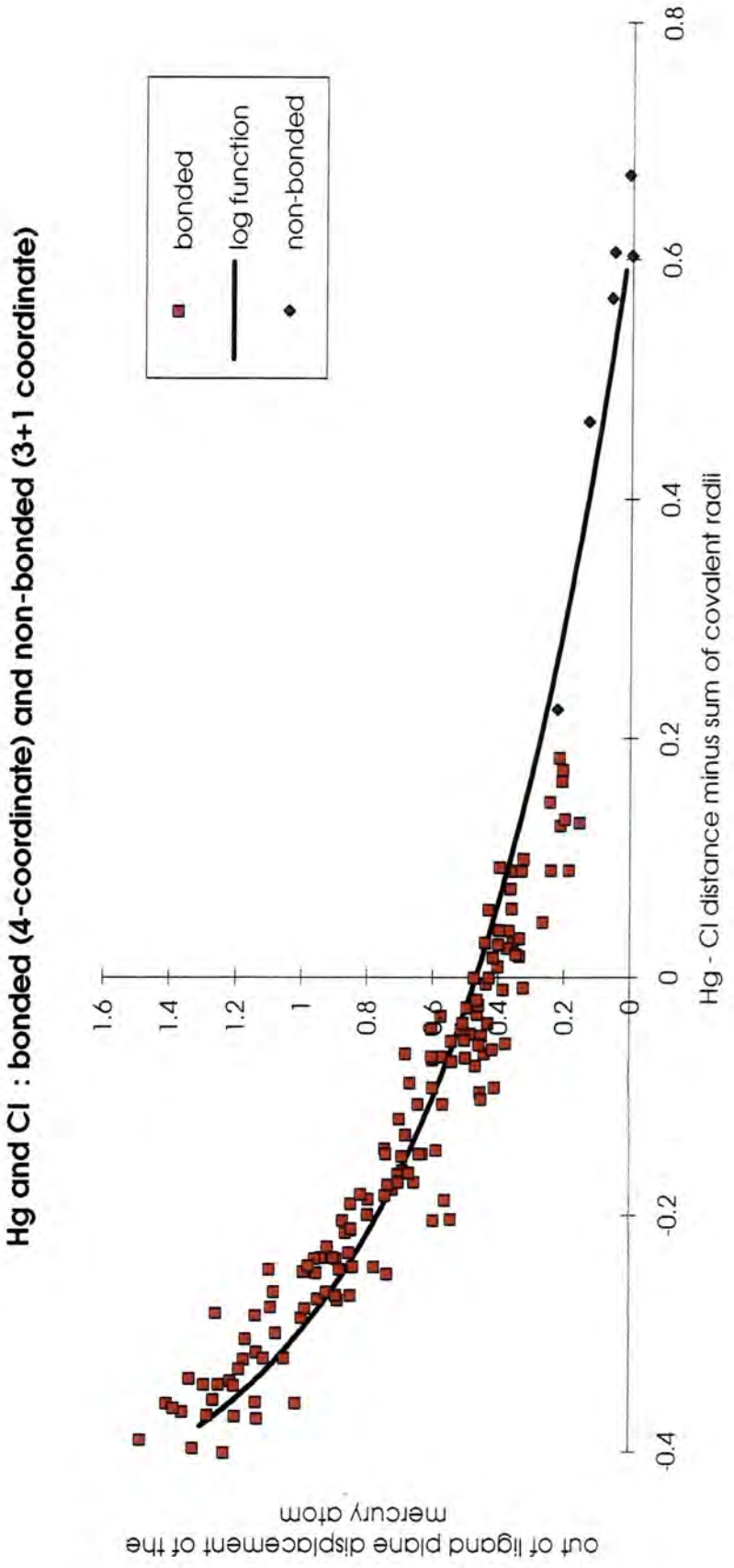


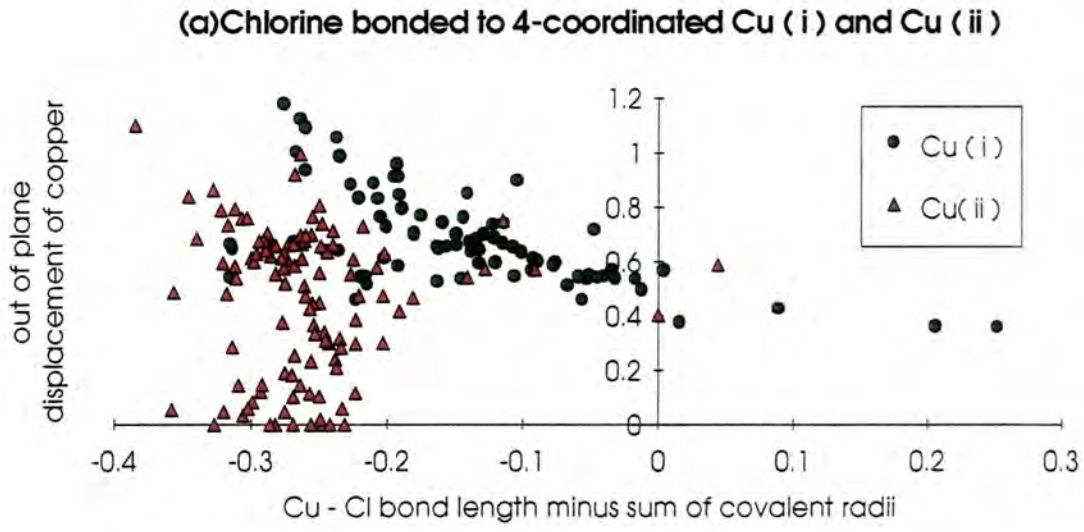
Figure 4.20

Normalised Hg -> Cl distances plotted against out-of-plane displacement for non-bonded [(3+1)-coordinated] and bonded [4-coordinated] mercury / chlorine interactions

metals will exhibit only tetrahedral geometries in the 4-coordinated state. For copper the situation is slightly different, and square planar geometries are well known. It is impressive then to look at Figure 4.21a which plots the displacement of the copper against the normalised bond length, for bonded chlorine ligands. The values have been categorised into those which are for Cu (I) and those which are Cu (II). There is a clear distinction between the two. All of the entries for copper (I) are shown to be tetrahedral in geometry, a result that is consistent with the d^{10} electronic configuration present. The copper (II) state however, has structures that are clearly square planar (those with zero displacement of the copper atom out of the plane of the ligands), yet also shows evidence of possessing tetrahedral moieties.

The fact that copper (II) has a d^9 electronic configuration means that the metal in this state is prone to Jahn-Teller distortions, and can be either tetrahedral, square planar, or any one of the distorted entities that link these two geometries, depending on the immediate environment surrounding the copper. The points that link the two distinct types of structure on the plot, can therefore be assumed to represent the geometrical interconversion states that lie on the potential energy hypersurface, describing the transition from tetrahedral to square planar geometry. Studies on the mechanism that surround this interconversion include that of Muetterties & Guggenberger (1974).

Another important point arising from this plot is that the mechanism of addition of the fourth ligand to 3-coordinated compounds only supports the formation of a tetrahedral geometry for 4-coordination (via a displacement towards tetrahedral in the 3+1 coordinated geometry). There is no evidence of a mechanism that would take the D_{3h} trigonal planar species directly into the D_{4h} square planar species, by addition of a fourth ligand in the plane of the existing three ligands. This plot therefore, also shows that there is no evidence for the co-planar addition of a fourth ligand to a 3-coordinated T-shaped species, to form a square-planar complex For this to occur we would expect to observe structures that directly link the two geometries along the x-axis of the plot. This phenomenon is at no time evident in any of the studies performed.



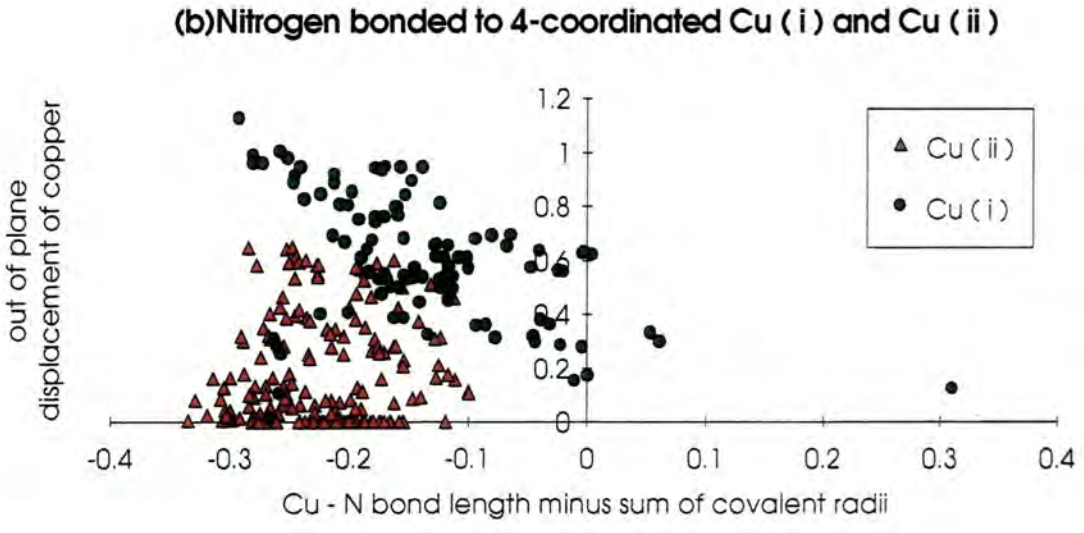


Figure 4.21

Normalised Cu -> L bond distances plotted against out-of-plane displacement for 4-coordinated copper in oxidation states +1 and +2: (a) copper bonded to chlorine, (b) copper bonded to nitrogen

Further evidence of the distinction between the two oxidation states of copper is shown in the plot given in Figure 4.21b, this time with respect to the bonds formed to nitrogen ligating species. The vast majority of the data once again supports the existence of tetrahedral copper (I), and of a mixture of geometries for copper (II) from square planar through to tetrahedral.

Finally in this description of the 3+1 coordination leading to 4-coordinated species, the differences that are apparent between the type of bonding shown by one ligand to the three different metals, are summarized. Oxygen is the only ligating species that is common to both the 3+1 coordination dataset and to the 4-coordination datasets, in sufficient number to make possible a comparative study of the way that it bonds to all three of the metals: copper, silver and mercury.

In the case of copper, the plot (Figure 4.22a) shows a marked tendency for the 4-coordinated species to exist as a square planar entity. From the previous analyses it is obvious, then, that these points show oxygen to be one of the main protagonists in driving the formation of a square planar complex from the tetrahedral geometry, for Cu (II). The few 3+1 coordinations (with oxygen as the fourth ligand) show quite a considerable effect on the displacement of the copper atom out of the ligand plane, and the resulting pyramidalisation into a pseudo-tetrahedral configuration follows exactly the predicted path from 3- to 4-coordination.

The case for mercury (Figure 4.22b) is less clear due to the lack of bonds between mercury and oxygen. This is symptomatic of the marked reluctance of mercury (II) to form covalent bonds to oxygen, instead preferring bonds with ionic character (Greenwood & Earnshaw, 1986, p.1413). This fact is again borne out in the strength of those bonds which have formed to the mercury in the 4-coordinated complexes shown: most having a bond length in excess of the sum of the covalent radii of the Hg and O atoms, and therefore relatively weak. The poor interaction between Hg and O atoms is evident here with the 3 + 1 coordinated examples showing only marginal displacements of the mercury atom out of the plane of its three bonded ligands.

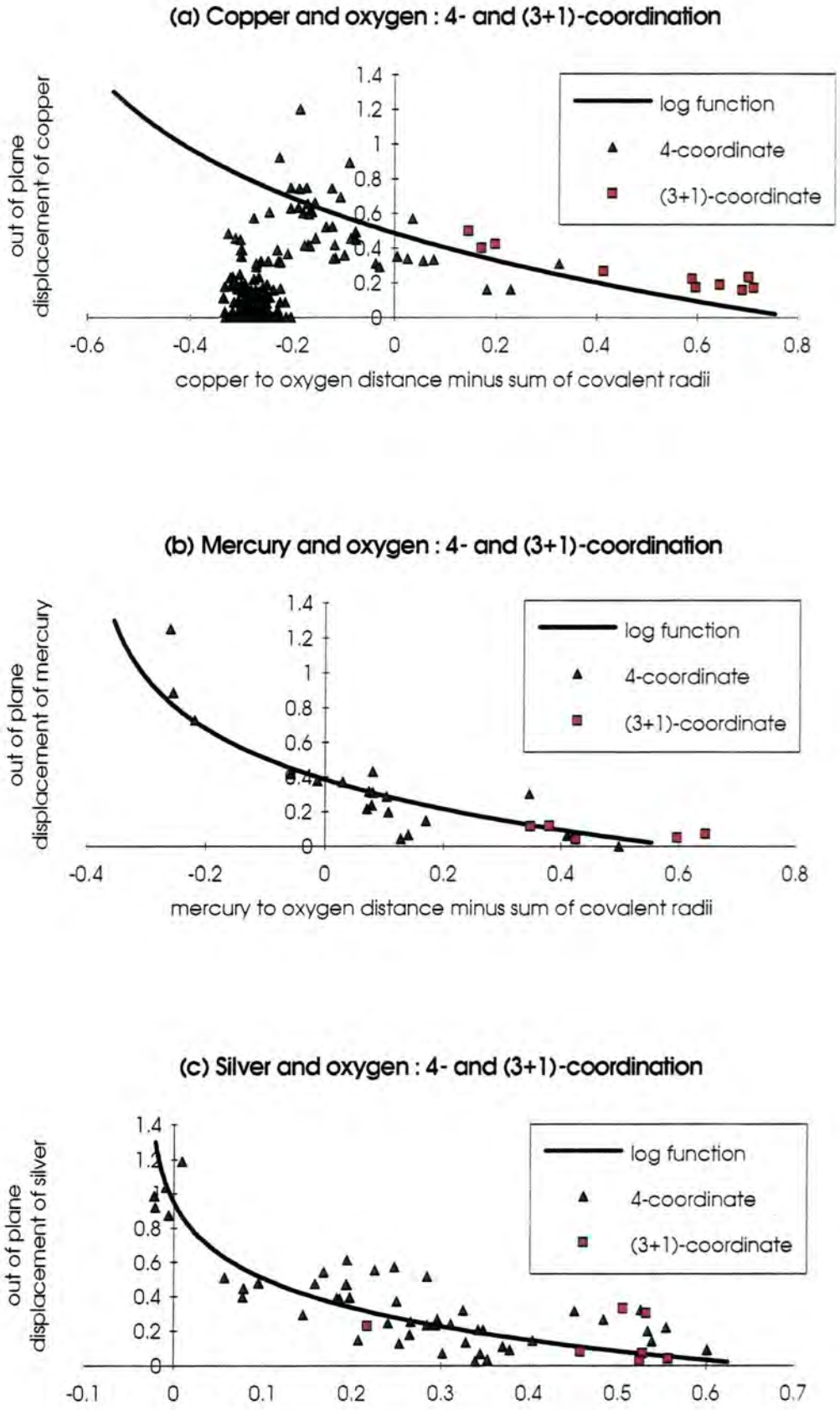


Figure 4.22

Normalised metal -> oxygen distances plotted against out-of-plane displacement for bonded [4-coordinate] and non-bonded [(3+1)-coordinate] Cu (a), Hg (b) and Ag (c)

For the silver with oxygen system (Figure 4.22c) the results are markedly different to the two plots found above. Silver has a great reluctance to coordinate to oxygen, particularly silver (I), and so the weak bonding observed in this plot is as expected. The general tendency is once again towards the formation of the tetrahedral 3+1 and later 4-coordinated entities by pyramidalisation at the central silver atom.

4.5.3. (2 + 1)-coordination leading to 3-coordination.

It has already been shown that the 3-coordinated geometry found for transition metal compounds can be classified into essentially three separate sub-groups. Firstly, those that maintain the three-fold symmetry of the parent point group, *i.e.* the D_{3h} and C_{3v} configurations. Secondly, the entries that have the metal atom bonded into a restrictive ring system, and thus usually configure themselves into the Y-shaped, C_{2v} geometry. And finally those fragments that adopt a T-shaped, C_{2v} configuration. Many of the literature references to this unusual T-shaped geometry (Bürgi *et al*, 1989; Grdenic, 1981; Canty *et al*, 1976) regard it as being an artefact of the addition of one ligand to the more common linear 2-coordinated complexes that are found for most transition metals.

In such an event the third ligand would be expected to bond relatively weakly to the metal centre, and at an angle roughly 90° to the other two ligands. As the strength of this bond increases therefore, we should be able to notice a concerted decrease in the size of the largest angle in the 3-coordinated compounds from 180° (essentially linear) down to *ca.* 120°, *i.e.* we should be able to follow the addition of a third ligand to a linear 2-coordinated complex all the way to the formation of a symmetrical 3-coordinated D_{3h} configuration. A plot of normalised metal to ligand distance against the size of the angle opposite to the approach of the third ligand, should show whether this situation occurs, and to what degree.

A dataset was extracted from the CSD containing all of the 2 + 1 coordinated entries for copper, silver and mercury, being careful to allow only one non-bonded contact within a tolerance of 0.70Å. When the normalised non-bonded distances are

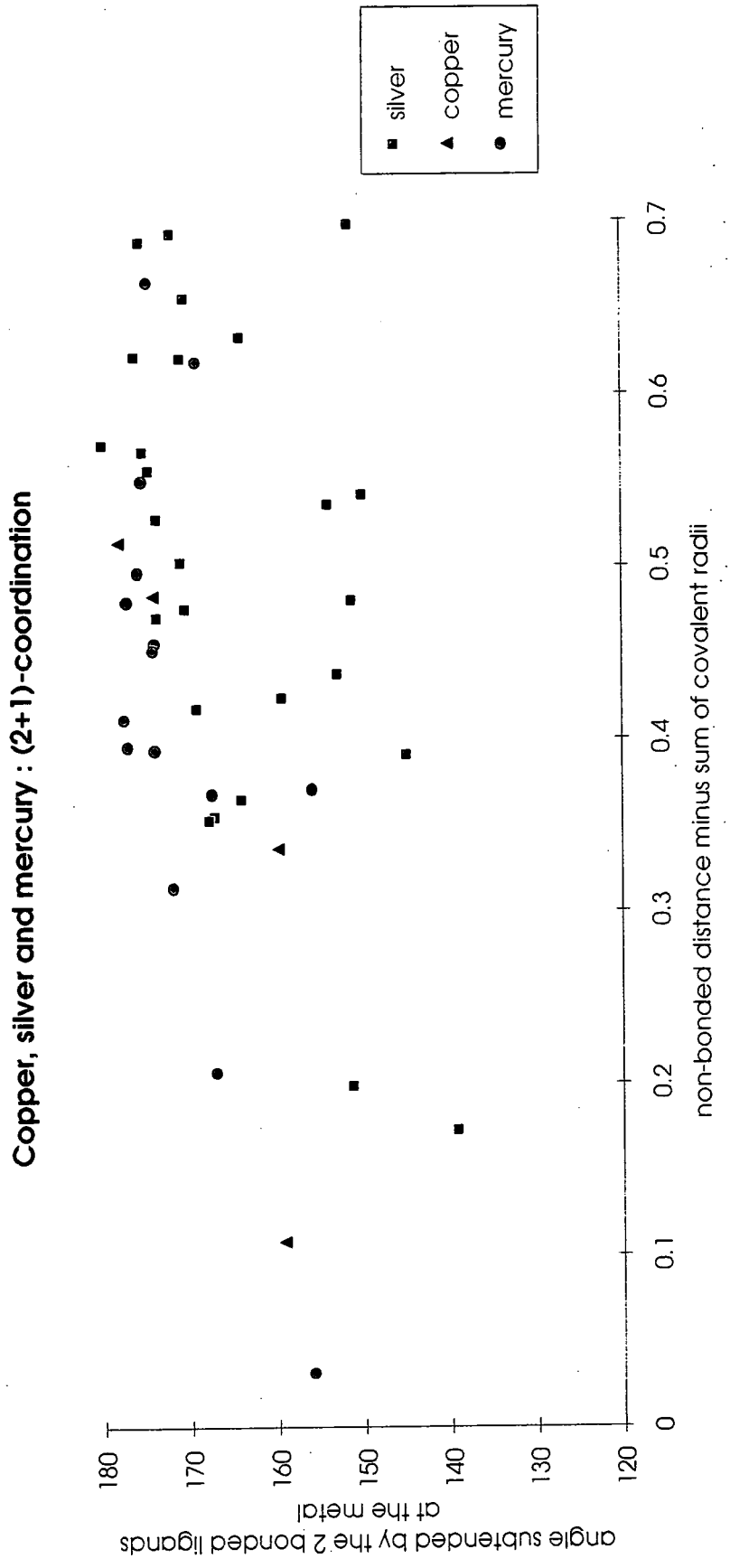


Figure 4.23

Normalised non-bonded distance plotted against the 2-coordinated ML_2 bond angle for copper, silver and mercury

plotted against the angle subtended by the two bonded ligands at the metal centre (Figure 4.23), the results seem to indicate that the further away the third ligand lies, the greater the degree of linearity that is maintained by the 2-coordinated system. However, as this third ligand interacts more strongly, the angle is driven down from 170-180° towards the 140-160° that is found for the T-shaped configurations.

It is possible to extrapolate this data into the region of the exactly 3-coordinated species by plotting the normalised bond length to each ligand from the metal, against the angle subtended at that metal by the two other ligands (a similar approach to that carried out for the 3+1 going to 4-coordination problem, but using bond angles instead of out of plane displacements). The results of these plots are shown in Figures 4.24-4.26.

From the copper dataset it is apparent that there are very few entries that adopt the T-shaped configuration for the 3-coordinate fragment, in comparison with the wealth of data that exists in the region that is defined by the D_{3h} parent symmetry. There are only four non-bonded interactions to the 2-coordinated copper species that fall within the tolerance limit of 0.70Å. However, these hits can be seen to fall on a line described by a logarithmic function similar to those of Pauling, that links the 2-coordinate linear moieties to the bulk of the 3-coordinate data.

The strength of the bond that is forming between the third ligand and the copper can be seen to increase as the plot is traversed right to left, and the angle subtended at the metal by the other two ligands simultaneously decreases. In doing so we go from a linear 2-coordinate system in which a third ligand is approaching the metal atom, through a series of 3-coordinate T-shaped molecules in which the third ligand is only "loosely" bound, and on to the 3-coordinate D_{3h} system in which the third ligand is now fully bonded and the geometry is as expected.

For mercury the situation is a little different, as many examples of the 3-coordinated T-shaped configuration are evident (Figure 4.12a, b). The linear 2-coordinate species is very stable, and it has been assumed (Bürgi *et al.*, 1989; Grdenic, 1965, 1981) that additions to this coordination sphere would try to preserve this linearity. Once again, the points on the graph can be seen to lie roughly on the line

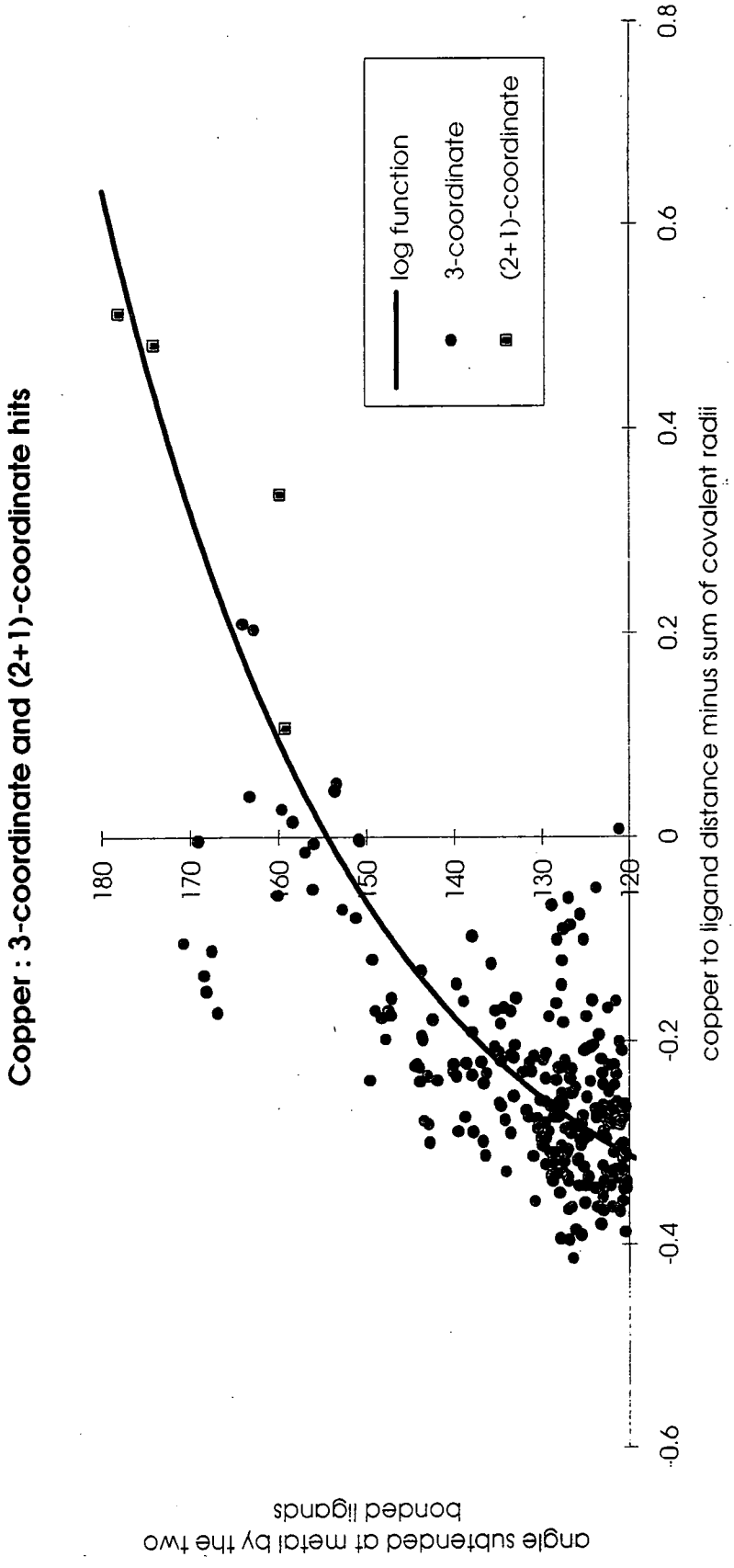


Figure 4.24

Normalised M -> L distance plotted against the 2-coordinated ML_2 bond angle for bonded [3-coordinated] and non-bonded [(2+1)-coordinated] copper

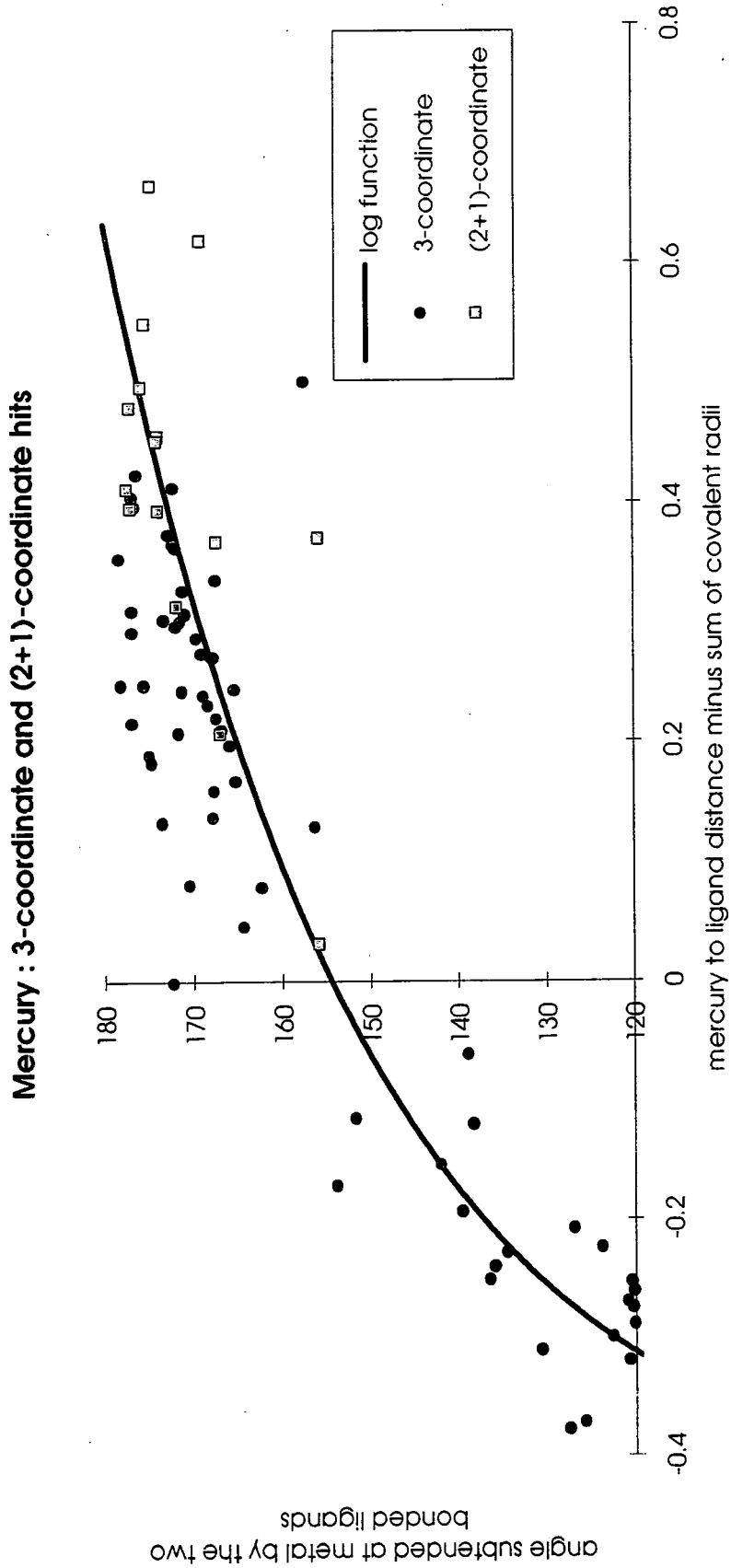


Figure 4.25

Normalised M -> L distance plotted against the 2-coordinated ML_2 bond angle for bonded [3-coordinated] and non-bonded [(2+1)-coordinated] mercury

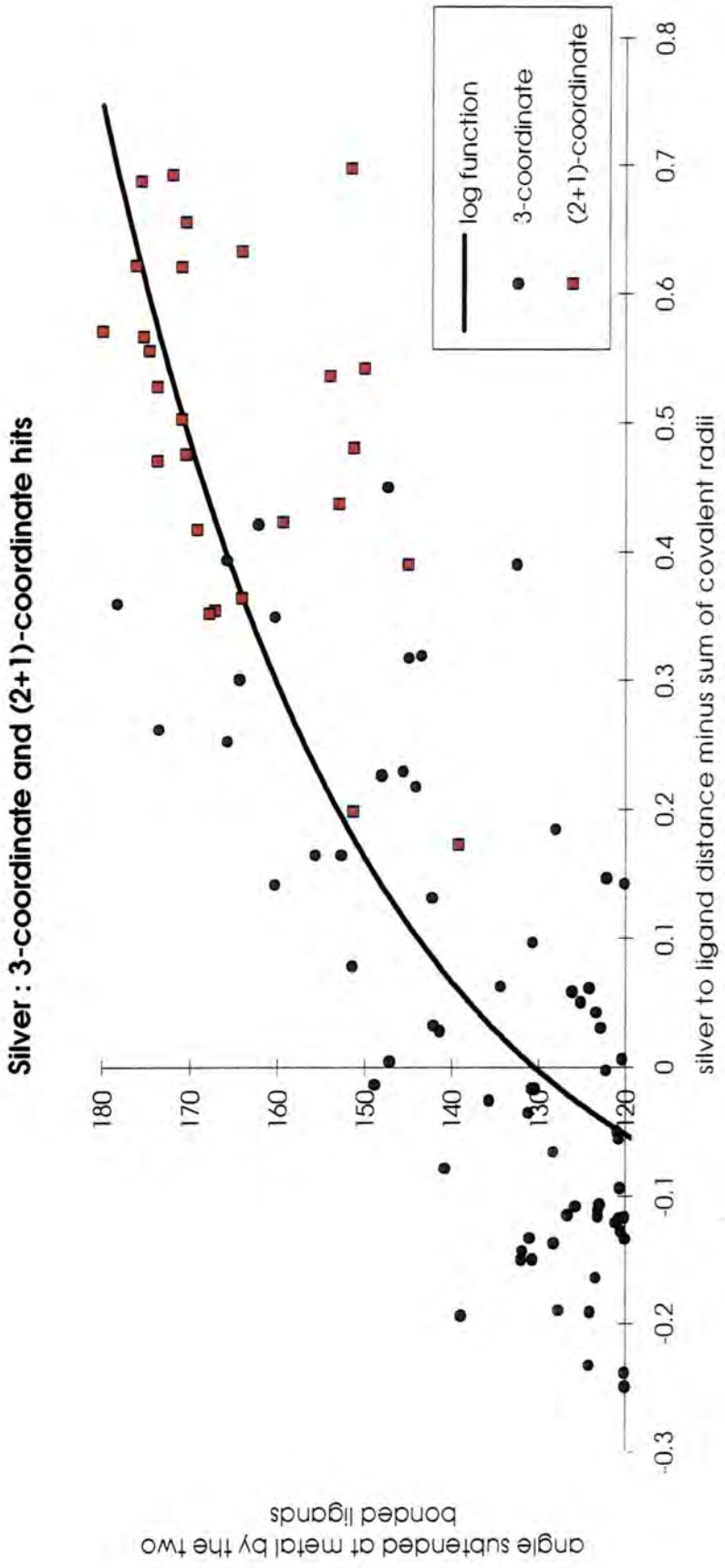


Figure 4.26

Normalised M -> L distance plotted against the 2-coordinated ML_2 bond angle for bonded [3-coordinated] and non-bonded [(2+1)-coordinated] silver

described by a Pauling-type function for the addition of one bond. The T-shaped 3-coordinate examples all have a relatively weak bond to the mercury, with the bond length between the metal and the third ligand being greater than the sum of the respective covalent radii. This is direct evidence of the mechanism of addition to the linear 2-coordinated mercury compounds, which manages to preserve the linearity of the parent system for as long as possible before the strength of the third additive bond overcomes the forces that maintain linearity and the molecule transforms into the 3-coordinated D_{3h} type system.

With silver the situation is far less obvious, with a great deal of scatter existing around the function that describes the transition from a linear 2-coordinate system to a trigonal planar 3-coordinate system. However, the underlying trend is one that follows the addition of one ligand to the linear 2-coordinated species, through a weak non-bonded interaction that seeks to preserve the linearity for as long as possible, until the attractive forces between the silver and the non-bonded ligand become so strong that the 3-coordinated molecule results. As such, there are no instances of a T-shaped molecule existing in which the third bond is a strong one, *i.e.* has a bond length less than the sum of the respective covalent radii.

4.6. Conclusions

The present results go a long way towards a complete description of all of the geometries, and of the trends associated with these geometries, for the 3-coordination sphere of the transition metals. The parent symmetry of this fragment is that which obeys the operations of the D_{3h} point group, and so the majority of the examples found in the Cambridge Structural Database are of trigonal planar geometry. Deviations away from this ideal geometry can be attributed to a variety of independent factors. Visualisation and characterisation of these effects can be successfully accomplished with symmetry deformation coordinates and principal component analysis.

One mode of deformation leads to a Y-shaped C_{2v} configuration. This geometry has been shown to be prevalent only in molecules whose structures place the

3-coordinated fragment as an integral constituent of a constraining ring system. The ring size is normally 4- or 5-membered, with the constraints of ring geometry, therefore, being the major factor in determining the shape established at the metal core.

The second mode of deformation is one resulting in a T-shaped C_{2v} configuration at the metal centre. These moieties have been shown to lie on the transition pathway between the ideal geometries of the 3- and 2-coordinate species, *i.e.* between trigonal planar and linear respectively. They feature one weakly bonded ligand and two strongly bonded ligands, with the largest angle (between 140° and 180°) being between the two strongly bonded ligands. The fragments can be considered to be dominated by the linearity of these two strong bonds, with the third ligand loosely attached to the metal centre. Extrapolation of this effect into the region belonging to the 2-coordinated species, with one non-bonded non-metallic ligand attacking shows excellent correlation, with the non-bonded examples adopting similar geometries and relative bond strengths as are found for the 3-coordinated T-shaped hits.

The third, and final mode of deformation that has been observed, is one which involves the displacement of the metal atom out of the plane of the three ligands to which it is bonded, and into a trigonal pyramidal configuration showing C_{3v} symmetry. This pyramidalisation effect has been shown to be linked to the proximity of a fourth non-bonded non-metallic ligand : the closer the ligand in terms of a normalised distance, the larger the effect. When data from 4-coordinated entries are introduced, it can be seen that the pyramidalisation and the relative bond strength can be described as obeying a Pauling type logarithmic function for the addition of one more ligand, and subsequent increase of the coordination number from 3, through (3+1), and up to 4. The mapping, therefore, is one that takes the trigonal planar arrangement into the tetrahedral assembly exclusively. No corresponding mechanism is observed that directly takes the trigonal planar configuration into the square planar 4-coordinated species. Formation of this species relies on the formation of a tetrahedral intermediate first, which can then rearrange to give the square planar product.

4.7 References

Alcock, N.W., Lampe, P.A. & Moore, P., *J. Chem. Soc., Dalton Trans.*, **1980**, 1471-1474.

Alsina, T., Clegg, W., Fraser, K.A. & Sola, J., *J. Chem. Soc., Dalton Trans.*, **1992**, 1393-1399.

Ascher, E., *J. Phys. C*, **1977**, 10, 1365-1377.

Auf der Heyde, T.P.E., Bürgi, H.B., *Inorg. Chem.*, **1989a**, 28, 3960-3969.

Auf der Heyde, T.P.E., Bürgi, H.B., *Inorg. Chem.*, **1989b**, 28, 3970-3981.

Auf der Heyde, T.P.E., Bürgi, H.B., *Inorg. Chem.*, **1989c**, 28, 3982-3989.

Barrow, M., Bürgi, H.B., Johnson, D.K. & Venanzi, L.M., *J. Am. Chem. Soc.*, **1976**, 98, 2356-2357.

Benac, B.L., Cowley, A.H., Jones, R.A., Nunn, C.M. & Wright, T.C., *J. Am. Chem. Soc.*, **1989**, 111, 4986-4988.

Bochmann, M., Webb, K.J., Hursthouse, M.B. & Mazid, M., *J. Chem. Soc., Dalton Trans.*, **1991**, 2317-2323.

Britton, D. & Dunitz, J.D., *J. Am. Chem. Soc.*, **1981**, 103, 2971-2979.

Bürgi, H.B., *Inorg. Chem.*, **1973**, 12, 2321-2325.

Bürgi, H.B., *Angew. Chem., Int. Ed. Engl.*, **1975**, 14, 460-473.

Bürgi, H.B. & Dunitz, J.D., *Acc. Chem. Res.*, **1983**, 16, 153-161.

Bürgi, H.B., Murray-Rust, J., Camalli, M., Caruso, F. & Venanzi, L.M., *Helv. Chim. Acta*, **1989**, 72, 1293-1300.

Bürgi, H.B. & Shklover, V., *Structure Correlation*, **1994**, 303-335, VCH, Weinheim.

Cambridge Structural Database, CSD User Manuals, 1992

Canty, A.J.& Gatehouse, B.M., *J. Chem. Soc., Dalton Trans.*, **1976**, 2018-20.

Chandrasekhar, K. & Bürgi, H.B., *J. Am. Chem. Soc.*, **1983**, 105, 7081-7093.

Cotton, F.A., *Chemical Applications of Group Theory*, **1990**, Wiley-Interscience, N.Y.

Dunitz, J.D., *X-ray Analysis and the Structure of Organic Molecules*, **1979**, Cornell University Press, Ithaca, New York.

Dyason, J.C., Healy, P.C., Pakawatchai, C., Patrick, V.A. & White, A.H., *Inorg. Chem.*, **1985**, 24, 1957-1961.

Grdenic, D., *Quart. Rev.*, **1965**, 19, 303-328.

Grdenic, D., *Structural studies on molecules of biological interest*, **1981**, OUP, New York, 207-221.

Greenwood, N.N. & Earnshaw, A., *Chemistry of the Elements*, **1986**, Pergamon, Oxford.

Gruff, E.S. & Kock, S.A., *J. Am. Chem. Soc.*, **1990**, 112, 1245-1247.

Hay-Motherwell, R.S., Hussain-Bates, B., Hursthouse, M.B. & Wilkinson, G., *J. Chem. Soc., Chem. Comm.*, **1990**, 1242-1243.

Hay-Motherwell, R.S., Wilkinson, G., Hussain-Bates, B. & Hursthouse, M.B., *J. Chem. Soc., Dalton Trans.*, **1992**, 3477-3482.

Hutton, A.T.& Irving, H.M.N.H., *J. Chem. Soc. Chem. Comm.*, **1979**, 1113-1114.

Johnson, O., *Personal Communication*, **1994**.

Klebe, G. & Weber, F., *Acta Cryst.*, **1994**, B50, 50-59.

Luef, W. & Keese, R., *J. Mol. Struct. Theochem.*, **1992**, 257, 353-368.

Luef, W., Keese, R. & Bürgi, H.B., *Helv. Chim. Acta*, **1987**, 70, 534-542.

McDowell, R.S., *J. Mol. Spectrosc.*, **1965**, 17, 365-367.

Melvin, M.A., *Rev. Mod. Phys.*, **1956**, 28, 18-44.

Muetterties, E.L. & Guggenberger, L.J., *J. Am. Chem. Soc.*, **1974**, 96, 1748-1756.

Murray-Rust, P., Bürgi, H.B.& Dunitz, J.D., *J. Am. Chem. Soc.*, **1975**, 96, 921-922

Murray-Rust, P., Bürgi, H.B.& Dunitz, J.D., *Acta Cryst.*, **1978a**, B34, 1787-1793.

Murray-Rust, P., Bürgi, H.B.& Dunitz, J.D., *Acta Cryst.*, **1978b**, B34, 1793-1803.

Murray-Rust, P., Bürgi, H.B.& Dunitz, J.D., *Acta Cryst.*, **1979**, A35, 703-713.

Norskov-Lauritsen, L. & Bürgi, H.B., *J. Comp. Chem.*, **1984**, 6, 216-228.

Pauling, L., *J. Am. Chem. Soc.*, **1947**, 69, 542-553.

Pauling, L., *The Nature of the Chemical Bond*, **1960**, Cornell University Press, Ithaca, N.Y.

Van Meersche, M., Kamara, R., Germain, G. & Declercq, J.P., *Bull. Soc. Chim. Belg.*, **1982**, 91, 553.

Wilson, E.B., Decius, J.C. & Cross, P.C., *Molecular Vibrations*, **1955**, McGraw-Hill, New York.

Chapter Five

Crystal Structure Determination by X-ray Diffraction Methods

5.1 Introduction

Crystal structure elucidation is one of the principal techniques for obtaining accurate information about the nature of a molecule. It has been shown in the previous Chapters, that without the reliable 3D atomic coordinate data provided, primarily by X-ray analyses, the methods of structure correlation would be worthless. Over the last few decades the technology implemented in the field of crystallography has developed dramatically, resulting in faster and easier structure solutions. These technological advances range from the ever-increasing power of computers that are now readily available, together with their integrated software packages, to the new forms of diffractometer hardware, such as area detectors, that can reduce the time required for standard data collections from days down to hours. The advances being made in this field can be judged by the increasing number of crystal structures being published, with the CSD (Allen, Davies *et al.*, 1991; Allen & Kennard, 1993) increasing in size from 107,000 to over 126,000 in the years 1993-94.

With this technique having gained such importance this century, there is an associated wealth of primary literature available, and many of the particularly mathematical techniques highlighted in this Chapter, will be referenced and not described in depth here. General introductions to crystallography can be found in: Glusker & Trueblood, 1985; Giacovazzo *et al.*, 1992; Stout & Jensen, 1989; Cheetham & Day, 1991. This Chapter aims to give a clear summary of some of the more important stages, and precedents, involved in performing an accurate crystal structure determination. Following this, details about some of the crystal structure determinations carried out during these studies will be given, and any particular points of interest concerning either the methods used to solve the structure or the geometry of the molecule itself, will be described.

The crystallography reported in this Chapter was performed using either a Rigaku AFC6S 4-circle diffractometer (with low temperature data obtained using an Oxford Cryostreams device), or a Siemens P4 four-circle diffractometer.

5.2 Procedures prior to data collection

5.2.1 Crystal selection and mounting.

Crystal selection is of great importance to the experiment as an enormous amount of information is obtained from what is a tiny piece of matter. The quality of sample crystals can be assessed optically, particularly by use of a polarising microscope. Viewed under such conditions, a suitable crystal will have sharply defined extinction boundaries upon rotation. This is a good preliminary indicator of whether the crystal is single or twinned. The real test of the quality of the crystals, however, is how they respond to X-rays, as many perfectly good diffracting crystals have a poor external appearance.

The size of the crystal is important: the scattering power of the crystal is proportional to the crystal volume, and to the contents of the unit cell, but a practical ceiling is defined by the dimensions of the X-ray beam, with many beams being homogeneous only over the central fraction of the collimator aperture. The crystal is generally chosen to be small enough to be totally bathed in the incident X-ray beam. Absorption increases exponentially with the size of the crystal, and a rough guide for the required size of the crystal is $2/\mu$ (where μ is the linear absorption coefficient for the X-ray source being used; Stout & Jensen, 1989). Where strong absorption is present, extreme crystal shapes (*e.g.* long needles or thin plates) should be avoided if at all possible, or modelled accordingly.

Crystals can be mounted in a number of ways, the most common of which are either by adhesion to a glass fibre, or in a low-density glass capillary. The latter method is favoured when the crystal is known to be sensitive to air or water, the Lindemann capillary providing an air-tight, inert (usually N_2) environment for the

duration of the data collection. For open glass fibre mountings modern polymer glues (*e.g.* epoxy resins) are commonly used as the adhesive. These adhesives give a rigid fixing without pulling during the short setting time.

Once the crystal has been securely mounted on the goniometer head, it is then transferred to the diffractometer, and optically centred with respect to the machine's axes. When viewed using the instrument's optical alignment system, a well centred crystal will not precess while it is rotated through all the available angles in both phi and chi.

5.2.2 Collection conditions

The most commonly used radiation types are those obtained from copper (wavelength Cu-Kα = 1.54178Å) or molybdenum (wavelength Mo-Kα = 0.71069Å) target materials. Copper gives a higher X-ray intensity for the same power rating, a factor of some importance for poorly diffracting crystals, but molybdenum radiation is less strongly absorbed by most materials. Also, higher Bragg angles are found with copper radiation, which may be of use if the unit cell is large, as the reflections are then better resolved. However, mechanical restrictions can make the measurement of the high-angle copper data impossible, whereas the relevant molybdenum data are more accessible (for the experiments detailed below, Mo-Kα radiation is used exclusively).

The data collection temperature is also important. At lower temperatures thermal motion is reduced and the higher Bragg angle diffraction intensities are improved. Low temperatures will also be necessary for crystals that are unstable at room temperature. For very low temperature work it is possible for the crystal to be mounted onto the fibre using a film of inert oil. This film freezes to form a rigid non-crystalline solid on cooling which affords both mechanical stability and protection from the atmosphere for the crystal.

5.2.3 Unit cell and orientation matrix

Once the crystal is centred, the next steps involve finding reflections, assigning

indices to these and determining the unit cell and orientation matrix, refining and checking the unit cell, and then determining the Laue symmetry. These procedures are largely automated in modern software for data set-up and data collection.

On the Rigaku diffractometer the initial reflections are found using a blind zig-zag search which maximises the amount of reciprocal space scanned. These reflections are then accurately centred so that the diffracted beam enters the detector at it's mid-point, this being the Bragg reflecting position. An omega scan moves the crystal through the Bragg position, keeping the other axes fixed. This leads to a sharp reflection profile, the breadth of which depends on the crystal mosaic spread as well as on the diffractometer. The Siemens P4 collects the initial peaks using a hemi-spherical search mechanism. This routine finds one reflection, then searches for reflections which are close to that reflection (within a hemisphere in reciprocal space), automatically indexing the reflections and determining the unit cell parameters as it finds each one. The search finishes when 10 reflections are found for which no significant changes in the unit cell parameters are observed.

The Texsan software for the AFC6S determines the orientation matrix and unit cell from the unindexed reflections by operating in reciprocal space. The aim is to find the simplest cell and matrix which allow integral indices to be assigned to all of the reflections. The observed vectors x must correspond to the reciprocal lattice points, and to the list of initial (search) vectors are added all the vectors $x_i \pm x_j$, which also represent lattice points. From the complete list, the three shortest non-coplanar vectors are chosen as the reciprocal cell axes, and the nine xyz coordinates become the nine elements of the orientation matrix. The indices for all of the reflections can now be found. If any indices are found to be simple fractions then one or more of the reciprocal axes is too long and the necessary corrections must be made. If indices are found to differ significantly from integer values then twinning may be present, or spurious reflections may be observed, both of which require investigation before proceeding.

The reflections chosen for unit cell refinement should be well separated in reciprocal space, leading to better determined cell parameters. The higher the Bragg

angles used, the greater the precision, but this has to be weighed against the lower intensities of these reflections and a compromise reached. The unit cell's Laue symmetry can be checked by comparing the intensities of supposedly equivalent reflections. Absorption may cause the intensities of equivalents to differ dramatically, and in such cases of uncertainty it is best to assume a lower symmetry for the data collection, which is done automatically by both machines used.

5.3 Data collection and processing

5.3.1 Collection of intensity data

The aim of the experiment is to measure intensity data to the greatest possible accuracy and precision, in the minimum time, and with the avoidance of errors. Poisson statistics suggest that the precision of a measurement is proportional to the square of the time spent counting it, and in order to double the precision therefore, a four-fold increase in the measuring time is required.

The type of scan to be used is left to the operator's discretion. In general, unusually low mosaic spread requires an omega/2theta scan, whereas for large unit cells an omega scan is used to reduce the overlap of neighbouring reflections. The scan type can also be correlated with the optimum detector aperture, with too large an aperture increasing background and hence reducing precision, but too small an aperture truncating reflections and introducing systematic errors. Similarly, the incident beam size should be chosen to ensure that the crystal sample is completely and evenly irradiated, but if the collimator is too large the background counts will increase disproportionately.

The time needed for data collection depends to a large extent on the maximum Bragg angle required. Compounds containing heavy atoms can diffract strongly to high angles, but most of this high angle scatter will be due to the diffracting power of the heavy atoms and does not aid the accurate location of the lighter atoms in the molecule. In practice, good diffracting crystals were collected out to 2θ (Mo-Kα) = 50-55°,

ensuring a large amount of high angle data for structure solution and refinement purposes, yet not requiring excessive diffractometer time for collection.

5.3.2 Correction of intensity data for systematic errors

The collected intensity data has to be reduced to either structure factor amplitudes ($|F_o|$), or their squares (F_o^2). Correction factors are applied in the data reduction process, which compensate for instrumental factors and systematic errors.

The Lorentz and polarisation corrections arise from the movement of the crystal through the reflecting position during the measurement of each reflection. The geometry of the machine gives the Lorentz factor, which for 4-circle diffractometers is simply sin 2θ, *i.e.* each intensity must be multiplied by this factor. Upon reflection the X-ray beam becomes partially polarised, the degree of which depends on the Bragg angle:

$$p = \left(1 + \cos^2 2\theta\right) \Big/ 2 \qquad [1]$$

The monochromator also partially polarises the incident beam and so the form of this correction must also depend on the orientation of the monochromator crystal relative to the equatorial plane of the diffractometer (for both the P4 and the Rigaku this is parallel geometry). In addition, the degree of polarisation by the monochromator depends on the physical characteristics of its crystal and is, therefore, not necessarily ideally behaved.

The effects of X-ray absorption can be corrected for, both empirically and semi-empirically. However, in all of the experiments to be detailed below, the only absorption method employed was that of psi-scans, which is purely empirical. Here, extra measurements are made of the variation of intensity for chosen reflections at different azimuthal angles. The variation in intensity is assumed to be solely due to absorption effects, and is modelled by a Fourier series of the diffractometer angles. Once the Fourier coefficients have been calculated, they are used to correct each reflection in the entire dataset for absorption.

Extinction reduces the observed intensity of strong reflections particularly at low Bragg angles and is marked for crystals with low mosaic spread. The wide variety of corrections available for extinction are generally used during structure refinement and not in data reduction. Irradiation of the crystal sometimes increases the mosaic spread and hence reduces extinction effects, making very strong reflections increase in intensity, while other reflections decrease. For this reason, very strong low-angle reflections should not be used as standard reflections for the data collection.

Multiple reflections lead to a reduction in the intensity of the strong reflections together with an enhancement of generally weak reflections. This commonly results in the observation of significant intensity when a systematic absence is expected. This problem is not easily corrected for during data reduction. It can be recognised more readily when the merging of equivalent reflections reveals discrepancies. Multiple reflections can be treated by re-measuring the reflection intensity at different azimuthal angles: any inconsistent intensity values then being rejected from the collected data.

For non-centrosymmetric structures, Friedel's law does not strictly hold and the extent to which reflections and their opposites (for every reflection *hkl*, the Friedel opposite is the reflection $\bar{h}\bar{k}\bar{l}$) differ in intensity depends on the presence of atoms with significant anomalous dispersion effects. Treatment of anomalous dispersion is best applied during structure refinement (Cruickshank & McDonald, 1967; Rogers, 1981;Bernardinelli & Flack, 1985), but where the effects are expected to be significant it is important to measure the Friedel pairs, and essential that the equivalents are not merged during the data reduction. The Friedel pairs must also be collected for any sample that requires an absolute configuration to be determined, *i.e.* where chirality is present.

Thermal diffuse scattering (TDS) is another systematic error that is present in collected intensity data. TDS is inelastic scattering and is due to the cooperative thermal motion of the atoms in the crystal lattice (lattice vibrations or "phonons"), resulting in a diffuse peak. These peaks occur at Bragg reflecting position, so measured intensities contain a small TDS contribution which artificially enhances them.

Corrections for TDS are not normally made except for very accurate work for which empirical methods have been developed (Blessing, 1987).

5.4 Structure solution and refinement

5.4.1 Solution

There are several alternative methods available for structure solution, all of which rely on the mathematical analysis of the observed intensity data from the experiment. It would be impossible to describe these methods fully within the context of this Chapter and so brief overviews will be given together with suitable leading references (*e.g.* Glusker & Trueblood, 1985; Giacovazzo *et al.*, 1992; Stout & Jensen, 1989; Cheetham & Day, 1991).

For any particular Bragg reflection, a summation over all contributions from the different atoms in the unit cell gives the Structure Factor, F_{hkl}:

$$F_{hkl} = \sum_j f_j \exp\left(B_j \sin^2\theta / \lambda^2\right) \exp\left(2\pi i\{hx_j + ky_j + lz_j\}\right) \qquad [2]$$

where f_j is the scattering amplitude of the jth atom (form factor), (x_j, y_j, z_j) are fractional coordinates with respect to the unit-cell axes, and B_j is a measure of the amplitude of thermal vibration. Equation **[2]** relies on the prior knowledge of the crystal structure which we obviously do not have at this stage. However, if the unit-cell contents are described in terms of electron density, ρ(x, y, z), at any point (x, y, z) in the cell (with no reference to atoms), equation **[3]** can be derived:

$$\rho(x,y,z) = \frac{1}{V}\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} F_{hkl} \exp\left(2\pi i\{hx + ky + lz\}\right) \qquad [3]$$

This would imply that if all F_{hkl} are measured, and enough values of ρ are calculated to draw up a map of the electron density in the unit cell, then this map will show all of the atomic positions, and the structure will be solved. However, there is a fundamental pitfall, and this is known as the "phase problem". F_{hkl} corresponds to a

sum of oscillatory wave motions, which have both amplitudes and phases. X-ray detectors measure cumulative intensities over time, but the phases themselves cannot be detected experimentally; the only quantity that can be directly found being the modulus $|F_{hkl}|$, given that:

$$|F_{hkl}|^2 \propto |I_{hkl}| \qquad [4]$$

The phase problem then, is how to extract phases from the set of observed moduli, and there are several ways of doing this.

Patterson maps

If F_{hkl} in **[3]** is replaced by F^2_{hkl}, then the phase information is lost and the resulting vector maps represent the relationships between all of the different atoms in the unit cell. The Patterson map shows other atoms surrounding the first atom (the origin), so if two atoms exist in real space at (x_1, y_1, z_1) and (x_2, y_2, z_2), the map produces a peak at the point $(x_1\text{-}x_2, y_1\text{-}y_2, z_1\text{-}z_2)$ [$\equiv (u_1, v_1, w_1)$ in vector space]. As the number of atoms in the structure increases, the Patterson map gets more complex, as there should be $n(n\text{-}1)$ peaks for n atoms, together with a large origin peak (since all atoms are at a zero distance from themselves). Deconvolution of the map is a problem and can be impossible for structures containing large numbers of atoms with relatively even scattering powers, *e.g.* large organic molecules. However, symmetry elements in the unit cell lead to some simplifications in the form of Harker lines and sections. Further, the form of the Patterson relationship itself causes peak heights to be proportional to the product of the scattering powers for the atoms involved: since scattering power is proportional to the atomic number, interactions between heavy atoms stand out dramatically in the map, allowing their positions to be defined and deconvolution of the entire map to begin.

Direct methods

Theoretical studies (Harker & Kasper, 1948; Karle & Hauptman, 1950; Woolfson, 1961) showed that certain probability relationships exist between the signs or phases of reflections having certain combinations of indices, and that these

relationships have practical applications to structure solution. Direct methods is a statistical method for predicting the phases (Sayre, 1952) and assumes that: (i) all of the atoms in the molecule have the same X-ray scattering power, (ii) that the distribution of these atoms within the unit cell is random, and (iii) that the electron density at an atom is positive but is zero everywhere else.

The first applications were to centrosymmetric structures, as the phases here are either 0 or π radians, corresponding to plus or minus signs. Nowadays, non-centrosymmetric structures are also solved using direct methods, or statistical phasing. Organic compounds usually contain no strong scatterer and so are particularly reliant on direct methods to produce a solution, as the even distribution of scattering power throughout the unit cell is one of the basic assumptions in the theory of direct methods.

In all but one of the experiments detailed below direct methods were used to generate the structure solutions, using the program XS in the SHELXTL PLUS software system (Sheldrick, 1991).

Fourier Methods

The methods illustrated above provide only a partial solution to the structure: Patterson maps usually only revealing the heavy atoms, while direct methods rely on phase information for the strongest reflections in the dataset. Extension of the phasing can, however, be carried out by using the known atomic position and atomic scattering amplitudes in a structure factor calculation, using equation **[2]**. Phases calculated from this are applied to the observed amplitudes, which are used in **[3]** to produce an electron density map, also known as a Fourier map. This map shows peaks of positive sign at the positions of greatest electron density, the heights of which are a measure of the atomic numbers of the atoms located there. The map should show the previously located atoms, together with further undiscovered atoms.

In direct methods we start with an E-map (electron density map: where the normalised structure factors, E, replace F as coefficients in the Fourier series), and as atoms are located they are given chemical sensible identities, while peaks in obviously wrong positions are ignored. The identified atoms are then re-located into the structure-

factor calculation, and the whole process is repeated iteratively until all of the atoms in the molecule have been located and identified.

5.4.2 Refinement and precision indicators

Once all of the atoms have been located, approximately positioned, and assigned initial amplitudes for thermal vibration (B_j in **[2]**), the next stage involves refinement of the model to the point where the discrepancies between the observed and calculated structure factor amplitudes, |F(obs)| and |F(calc)|, are minimized. One way of doing this is with a "difference Fourier map", which is based on |F| = |F(obs)| - |F(calc)| in **[3]**, and shows positive peaks where electron density exists but has not been specified (as atomic sites) in the current solution, and troughs where unwarranted scattering power has been built into the model. Thus, a slightly misplaced atom produces a peak on one side and a trough on the other. The coordinates and temperature factors can then be adjusted until the difference map has as few features as possible remaining in it.

The more powerful alternative is the purely mathematical, "least-squares method", in which a direct minimization of discrepancies in the structure factors amplitudes is obtained by altering the structural parameters. This procedure relies on the fact that a large amount of good data is collected during single-crystal experiments, and consequently approximately 10-20 quantitative observations are made for each parameter used in the model. It is for this reason that such detailed and precise information on the molecular dimensions of the sample in question can be found by crystal structure determination.

In all of the experiments detailed below, least-squares refinement has been employed to finalise the structure solution. These refinements were initially carried out using the XLS program in the SHELXTL software package, which was then superseded by the more powerful SHELXL-93 (Sheldrick, 1993) program used for the later structure determinations.

In structure refinement by least-squares a number of parameters are determined from a larger number of data, and the quantity minimised is:

$$\sum_{i=1}^{N} w_i \Delta_i^2 \qquad [4]$$

where Δ_i is usually either $|F_{obs}|_i$ - $|F_{calc}|_i$, or $|F_{obs}|_i^2$ - $|F_{calc}|_i^2$ and each of the N reflections has a weight w_i. The esd values depend on (a) the minimised function, (b) the number of data and parameters, and (c) the diagonal elements of the inverse least-squares matrix $\mathbf{A}^{-1}$, such that:

$$\sigma(p_j) = \left\{ \left(\mathbf{A}^{-1}\right)_{jj} \cdot \frac{\sum_{i=1}^{N} w_i \Delta_i^2}{N - P} \right\}^{\frac{1}{2}} \qquad [5]$$

where p_j is the jth of the P parameters. Low esd values (high precision) are seen to occur when a combination of good agreement between observed and calculated data (small numerator) and a large excess of data over parameters (large denominator) is used.

In general, refinement is continued until the solution is optimised, *i.e.* until the refinement has reached convergence. In order to gauge this, certain common indicators are invoked: "R-factors", "weighted R-factors" and "goodness of fit" parameters are the most widely used.

$$R = \sum_i |\Delta|_i \Big/ \sum_i |F_{obs}|_i \qquad [6]$$

$$wR = \left(\sum_i w_i \Delta_i^2 \Big/ \sum_i w_i F_{obs,i}^2 \right)^{\frac{1}{2}} \qquad [7]$$

$$S = \left[\sum_i \left\{ \frac{\Delta_i}{\sigma(F_{obs})_i} \right\}^2 \Big/ (N - P) \right]^{\frac{1}{2}} \qquad [8]$$

All three residuals describe the degree of correlation between the experimental data and the model: the lower the values of R, or wR, the better the fit. S is the goodness of fit and should equal unity if the model is a true representation of the X-ray scattering power of the structure. In SHELXL-93 the refinement procedure is based on F^2 values and similar residuals are generated by replacement of F values in **[6]**-**[8]** with the corresponding F^2 values.

Temperature factors can also be good indicators of the quality of the model. These parameters provide information about both the time-averaged temperature-dependent movement of the atoms about the mean equilibrium positions (dynamic disorder), and their random distribution over different sets of equilibrium positions from one unit cell to another (static disorder) (Johnson, 1980; Dunitz, Maverick & Trueblood, 1988). For isotropic refinement a single parameter (B or U) is generated:

$$f'(\theta) = f(\theta)\exp\left(-B\sin^2\theta/\lambda^2\right)$$
$$f'(\theta) = f(\theta)\exp\left(-8\pi^2 U\sin^2\theta/\lambda^2\right) \qquad \textbf{[9]}$$

where $f(\theta)$ is the scattering factor for a stationary atom and $f'(\theta)$ is the scattering factor for the vibrating atom. B and U both have units $Å^2$, and U represents a mean-square amplitude of vibration. For an anisotropic model six parameters are used, and the exponent in **[9]** becomes variously:

$$-\left(\beta_{11}h^2+\beta_{22}k^2+\beta_{33}l^2+2\beta_{23}kl+2\beta_{13}hl+2\beta_{12}hk\right) \qquad \textbf{[10a]}$$

or

$$-\frac{1}{4}\left(B_{11}h^2a^{*2}+B_{22}k^2b^{*2}+B_{33}l^2c^{*2}+2B_{23}klb^*c^*+2B_{13}hla^*c^*+2B_{12}hka^*b^*\right) \qquad \textbf{[10b]}$$

or

$$-2\pi^2\left(U_{11}h^2a^{*2}+U_{22}k^2b^{*2}+U_{33}l^2c^{*2}+2U_{23}klb^*c^*+2U_{13}hla^*c^*+2U_{12}hka^*b^*\right) \qquad \textbf{[10c]}$$

The first form **[10a]** is the most compact, however the six β terms are not directly comparable; the second form **[10b]** is equivalent to the isotropic B, and the third form **[10c]** is equivalent to the U expression.

These parameters can be represented graphically by "thermal ellipsoids", but only if certain inequality relationships among the 6 parameters are satisfied. Otherwise

they become "non-positive definite" and the corresponding ellipsoid does not have three real principal axes. This could indicate a significant problem in the model (*e.g.* a disordered atom), or just be due to imprecise (high esd's) U parameters, in which case the anisotropic model is probably not justified.

At the end of the refinement process a list of fractional positional coordinates for each of the atoms in the structure, together with an estimate of their reliability (given in the form of an esd, or standard uncertainty), is produced. From this list the normal geometrical descriptors of the structure can be derived, *e.g.* bond lengths and angles, together with a measure of their reliability. It is also from this list of coordinates that the more specific structural parameters are generated which were used for the structure correlation studies reported in the earlier Chapters of this Thesis.

5.5 Experimental details

The X-ray crystal structure determinations for the eight molecules shown in Figure 5.1 are detailed below. The atomic coordinates for each structure are given in Appendix 3.

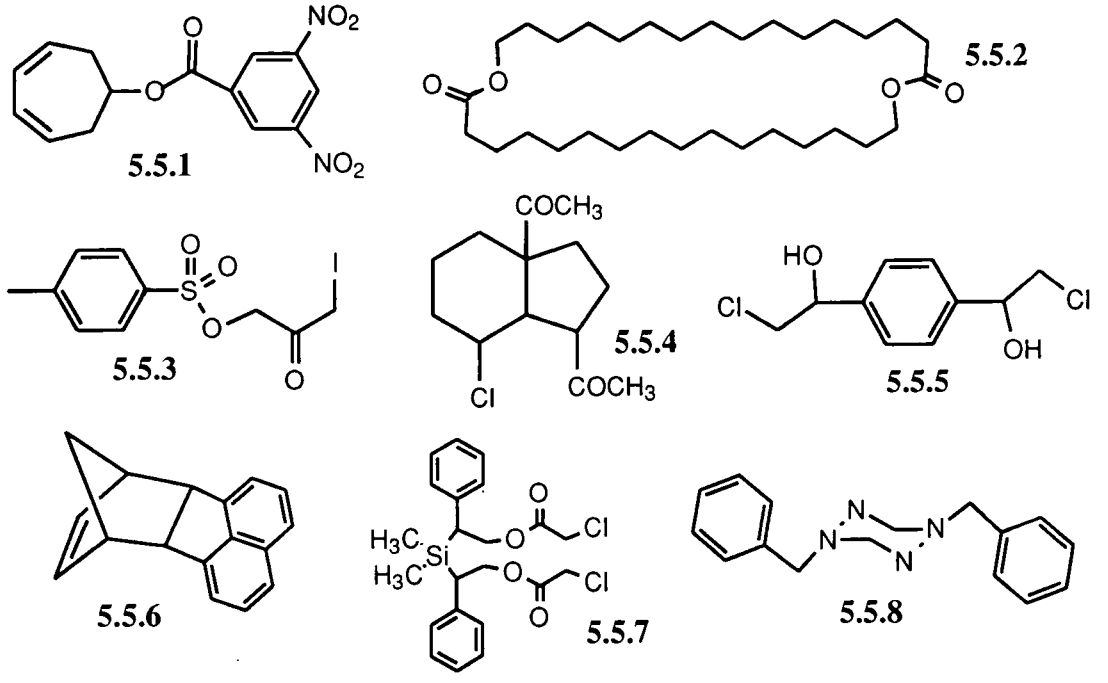


Figure 5.1

Chemical diagrams for the structure solutions reported in this Chapter

5.5.1 3, 5 cycloheptadienyl - 3, 5 dinitrobenzoate ($C_{14}H_{12}O_6N_2$) [Figure 5.2]

Empirical formula	C14 H12 N2 O6
Formula weight	304.26
Diffractometer	Rigaku AFC6S
Temperature	150(2) K
Wavelength	0.71073 Å
Crystal system	Monoclinic
Space group	$P2_1/n$
Unit cell dimensions	a = 6.7210(10) Å alpha = 90° b = 9.002(2) Å beta = 91.46(3)° c = 22.914(5) Å gamma = 90°
Volume	1385.9(5) $Å^3$
Z	4
Density (calculated)	1.458 Mg/m^3
Absorption coefficient	0.116 mm^{-1}
F(000)	632
Crystal size	0.50 x 0.30 x 0.30 mm
Theta range for data collection	2.88 to 25.05°
Index ranges	0<=h<=8, 0<=k<=10, -27<=l<=27
Reflections collected	2661
Independent reflections	2438 [R(int) = 0.0409]
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	2320 / 0 / 199
Goodness-of-fit on F^2	1.063
Final R indices [I>2sigma(I)]	R1 = 0.0518, wR2 = 0.1431
R indices (all data)	R1 = 0.1115, wR2 = 0.2078
Largest diff. peak and hole	0.269 and -0.338 $e.Å^{-3}$

This experiment was carried out on the crystal of the molecule shown below (Figure 5.2). The major feature of interest here was the conformation of the seven-membered ring, as this is the first example of a 1,3-cycloheptadiene that is unconstrained by benzo fusion, *i.e.* has 2 purely ethylenic bonds. Structure correlation studies performed on similar fragments located on the CSD (Section 3.4.3) showed a preponderance of the C_2 (1,3-T) conformation of the ring. This was contrary to the results of the energy calculations which predict that the C_s (1,3-E) envelope conformation should be the most energetically favourable by values in the range of 3.4-6.7 $kJmol^{-1}$ (calculation package dependent). The majority of the hits used in the structure correlation study displayed benzo-fusion, and the minimisation of H-H interactions forced the adoption of the C_2 conformation.

For this molecule the endo-unsaturated bonds are not involved in any benzo-fusion and the C_s conformation for the 1,3-diene is observed. The corresponding intra-annular torsion angles are given in Table 5.1. The absence of any serious chemical constraints allows the 7-membered ring to adopt the minimum energy conformation, the C_s (1,3-E) form, and serves to highlight one of the potential pitfalls

Table 5.1

Torsion angles of the seven-membered ring

C7 - C1 - C2 - C3	77.62 (0.32)
C1 - C2 - C3 - C4	-36.57 (0.45)
C2 - C3 - C4 - C5	-3.91 (0.55)
C3 - C4 - C5 - C6	7.19 (0.59)
C4 - C5 - C6 - C7	4.61 (0.57)
C5 - C6 - C7 - C1	18.86 (0.47)
C2 - C1 - C7 - C6	-66.24 (0.34)

of structure correlation studies of small datasets: that when the molecules aggregated in a dataset have very similar chemical environments, then a distorted view of the conformational diversity can be obtained. In this case the preponderance of the C_2 form in the dataset would suggest that this is the minimum energy conformation for the 1,3-cycloheptadienes, however, remove the H-H interactions present when benzo-fusion occurs, and the true energy minimum for the free ring is the C_s form observed here.

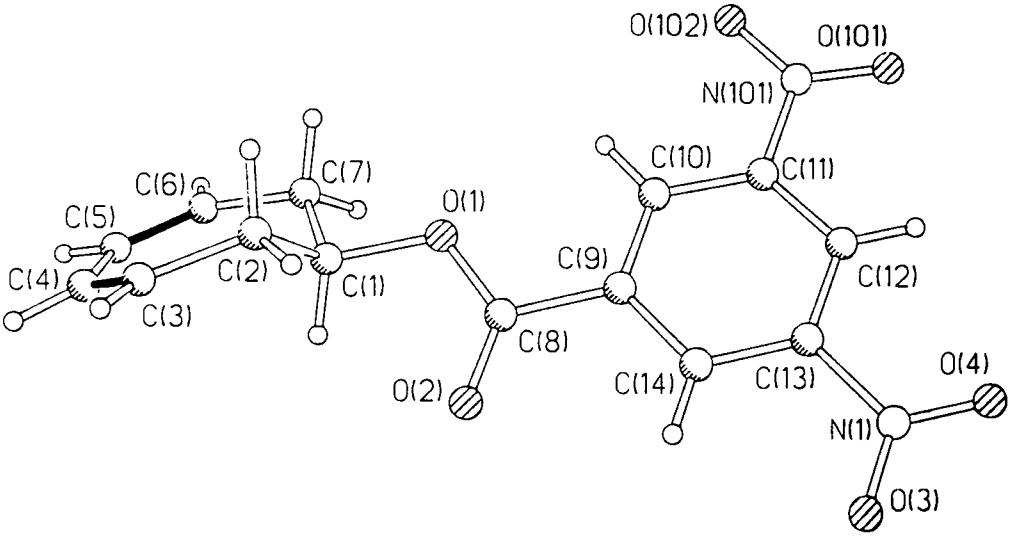


Figure 5.2

$C_{14}H_{12}O_6N_2$ molecule showing the C_s conformation of the 1,3-cycloheptadiene ring

5.5.2 A 34-membered diolide (C_{32} H_{60} O_4) [Figure 5.3]

Empirical formula	C32 H60 O4	
Formula weight	508.80	
Diffractometer	Siemens P4	
Temperature	293(2) K	
Wavelength	0.71073 Å	
Crystal system	triclinic	
Space group	P bar 1	
Unit cell dimensions	a = 5.330 Å	alpha = 90.81(1)°
	b = 8.647(1) Å	beta = 92.43(1)°
	c = 17.936(1) Å	gamma = 107.92(1)°
Volume	785.51(10) $Å^3$	
Z	1	
Density (calculated)	1.076 Mg/m^3	
Absorption coefficient	0.068 mm^{-1}	
F(000)	284	
Crystal size	0.4 x 0.15 x 0.1 mm	
Theta range for data collection	2.27 to 25.00°	
Index ranges	0<=h<=6, -10<=k<=9, -21<=l<=21	
Reflections collected	3117	
Independent reflections	2792 [R(int) = 0.0108]	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	2787 / 0 / 163	
Goodness-of-fit on F^2	0.958	
Final R indices [I>2sigma(I)]	R1 = 0.0385, wR2 = 0.0633	
R indices (all data)	R1 = 0.0846, wR2 = 0.0747	
Largest diff. peak and hole	0.120 and -0.107 $e.Å^{-3}$	

This structure shows a surprisingly rigid ring conformation. It was anticipated that the 34-membered nature of the open ring would prove to be very flexible and thereby render the structure difficult to solve, but instead the ring adopts a "cigar" shape with the length of the "cigar" being a straight chain of carbon atoms (analogous to the aliphatic straight chains of alkanes), and the turning points provided by the geometry of the lactone groups. The torsion angles along the chain are approximately 180°, and the minimum C-C distance across the ring (*e.g.* C10-C8A) is 4.164Å. The torsion angles at one end of the "cigar" are given in Table 5.2, and show the atoms of the lactone group (O1, C1, O2) to be co-planar with C16A. This results in an intramolecular non-bonded H...O contact distance of 2.412Å. The closest intermolecular C-H...O is 2.806Å.

Table 5.2

Torsion angles for the lactone group

C16A - O2 - C1 - O1	0.33 (0.20)
C16A - O2 - C1 - C2	-179.30 (0.11)
O1 - C1 - C2 - C3	-123.93 (0.16)
O2 - C1 - C2 - C3	55.69 (0.17)

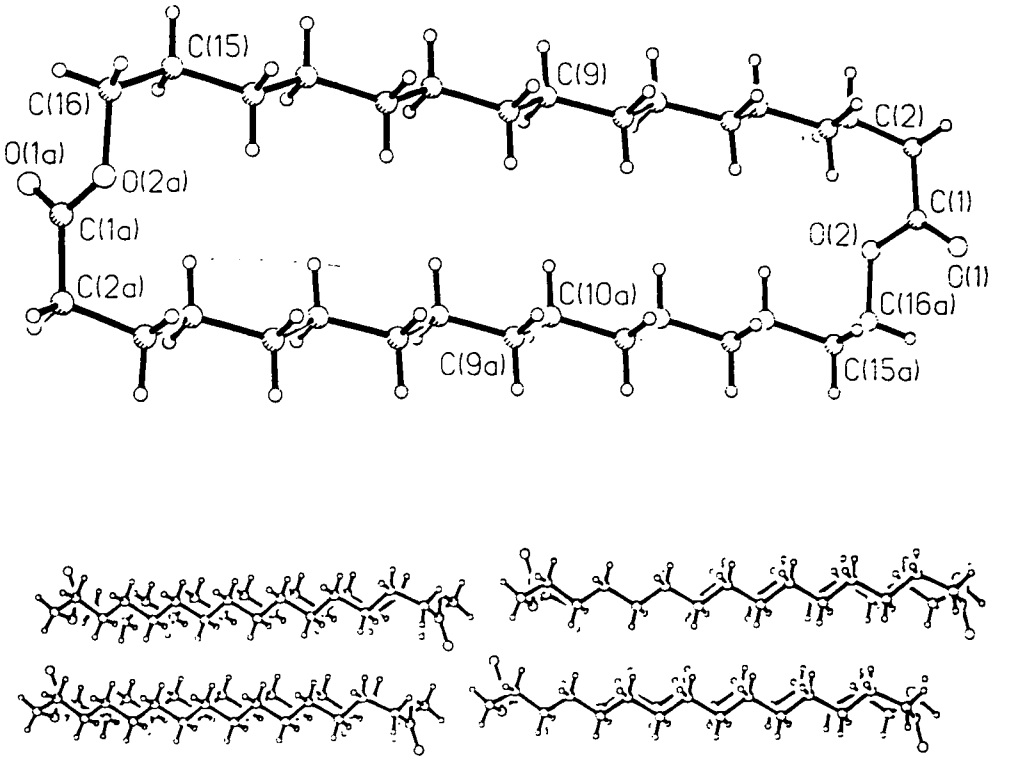


Figure 5.3

C_{32} H_{60} O_4 "cigar"-shaped molecule

5.5.3 1-Iodo-3-tosyloxy-propan-2-ane (C_{10} H_{11} O_4 I S) [Figure 5.4]

Empirical formula	C10 H11 I O4 S	
Formula weight	354.15	
Temperature	150(2) K	
Diffractometer	Rigaku AFC6S	
Wavelength	0.71073 Å	
Crystal system	triclinic	
Space group	P bar 1	
Unit cell dimensions	a = 4.9740(10) Å	alpha = 96.18(3)°
	b = 8.344(2) Å	beta = 94.85(3)°
	c = 14.630(3) Å	gamma = 90.54(3)°
Volume	601.4(2) Å^3	
Z	2	
Density (calculated)	1.956 Mg/m^3	
Absorption coefficient	2.831 mm^{-1}	
F(000)	344	
Crystal size	0.50 x 0.20 x 0.10 mm	
Theta range for data collection	2.69 to 27.49°	
Index ranges	-6<=h<=6, 0<=k<=10, -18<=l<=18	
Reflections collected	2963	
Independent reflections	2773 [R(int) = 0.0056]	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	2773 / 0 / 146	
Goodness-of-fit on F^2	1.089	
Final R indices [I>2sigma(I)]	R1 = 0.0156, wR2 = 0.0376	
R indices (all data)	R1 = 0.0176, wR2 = 0.0381	
Largest diff. peak and hole	0.367 and -0.520 e.Å^{-3}	

The reason for this structure solution was to confirm the product of the synthesis. The data for this structure were collected using omega scans with the Lehmann-Larsen profile fitting program. Initial scans showed the peaks to be very close to each other in reciprocal space, and the Lehmann-Larsen routine was chosen to minimise the effects of this, *i.e.* countered the inaccurate measurement of background intensity by not measuring it at all, and estimating it instead.

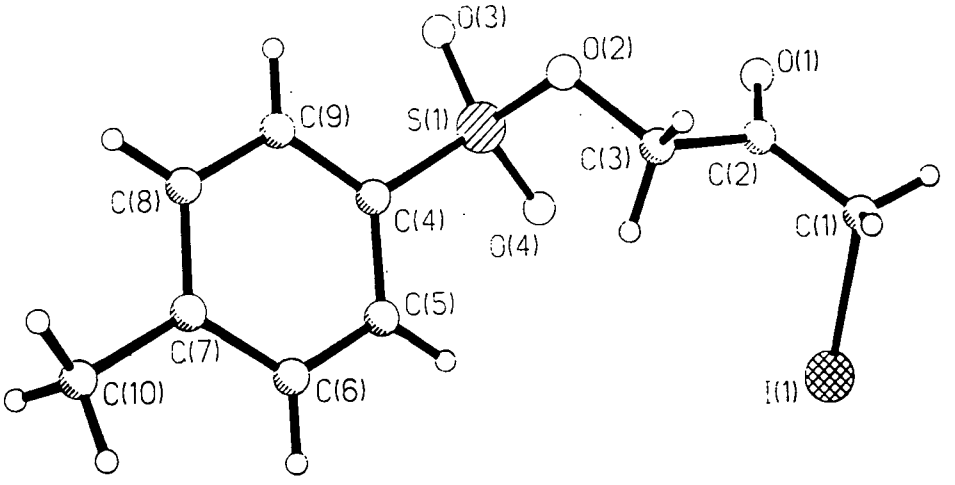


Figure 5.4

A plot of the $C_{10}H_{11}O_4IS$ molecule

5.5.4 1β, 9β-Diacetyl-7α-chloro-cis-hydrindane ($C_{13}H_{19}O_2Cl$) [Figure 5.5]

This molecule (Figure 5.5) was examined as part of a study into the Friedel Crafts acylation of hydrindane (Davison *et al.*, 1993). The X-ray crystal structure shows that the 6-membered ring exists as the symmetrical D_{3d} chair conformation (Hendrickson, 1967). The 5-membered ring conforms to the parameters first postulated by Pitzer & Donath (1959) describing the C_s "envelope" conformation. The tip of the envelope is at the ring juncture atom C(1), with the four other carbon atoms of the ring

Table 5.3

Selected torsion angles (°) for the ring carbon atoms

Five-ring		Six-ring	
C8 - C1 - C2 - C3	3.6 (0.2)	C9 - C4 - C5 - C6	58.7 (0.3)
C2 - C1 - C8 - C9	-26.6 (0.2)	C5 - C4 - C9 - C8	-53.5 (0.3)
C1 - C2 - C3 - C9	20.7 (0.3)	C4 - C5 - C6 - C7	-57.6 (0.3)
C2 - C3 - C9 - C8	-36.7 (0.2)	C6 - C7 - C8 - C9	-51.1 (0.3)
C1 - C8 - C9 - C3	39.1 (0.2)	C5 - C6 - C7 - C8	55.0 (0.3)
		C7 - C8 - C9 - C4	48.7 (0.3)

(C2-C7-C8-C9) being co-planar. Both rings are in their minimum energy conformations (Kolossvary & Guida, 1993) although the envelope is technically a saddle point on a zero-energy pseudorotation pathway between two twist conformations. Selected torsion angles for the rings are shown in Table 5.3.

Empirical formula	C13 H19 O2 Cl	
Formula weight	242.75	
Temperature	293 K	
Diffractometer	Rigaku AFC6S	
Wavelength	0.71073 Å	
Crystal system	monoclinic	
Space group	$P2_1/c$	
Unit cell dimensions	a = 10.867(1) Å	alpha = 90°
	b = 10.334(2) Å	beta = 102.65(1)°
	c = 11.869(1) Å	gamma = 90°
Volume	1300.6(3) $Å^3$	
Z	4	
Density (calculated)	1.326 Mg/m^3	
Absorption coefficient	2.78 mm^{-1}	
F(000)	520	
Crystal size	0.80 x 0.80 x 1.00 mm	
Theta range for data collection	2.0 to 30.00°	
Index ranges	0<=h<=15, -14<=k<=14, -16<=l<=16	
Reflections collected	3789	
Independent reflections	1894 [R(int) = 0.0549]	
Refinement method	Full-matrix least-squares on F	
Data / restraints / parameters	1894 / 0 / 203	
Goodness-of-fit on F	1.5182	
Final R indices [I>2sigma(I)]	R1 = 0.0475, wR = 0.0484	
R indices (all data)	R1 = 0.1064, wR = 0.0798	
Largest diff. peak and hole	0.28 and -0.20 $e.Å^{-3}$	

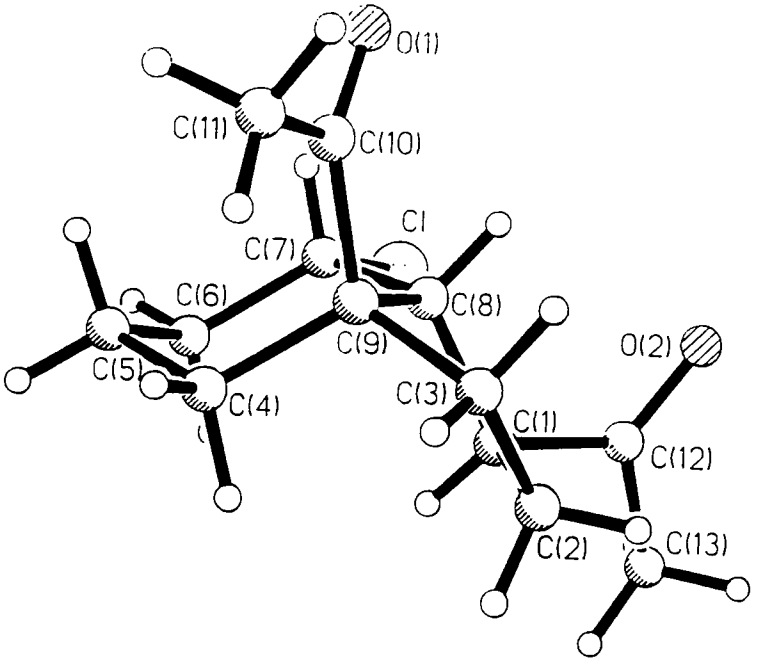


Figure 5.5

A plot of the 1β, 9β-Diacetyl-7α-chloro-cis-hydrindane ($C_{13}H_{19}O_2Cl$) molecule

5.5.5 (R,R) - 1,4 - bis (2'-chloro-1'-hydoxyethyl) benzene (C_{10} H_{12} O_2 Cl_2)

Empirical formula	C10 H12 Cl2 O2
Formula weight	235.10
Temperature	150(2) K
Diffractometer	Rigaku AFC6S
Wavelength	0.71073 Å
Crystal system	orthorhombic
Space group	$P2_1\,2_1\,2_1$
Unit cell dimensions	a = 4.8100(10) Å alpha = 90°
	b = 9.212(2) Å beta = 90°
	c = 24.028(5) Å gamma = 90°
Volume	1064.7(4) $Å^3$
Z	4
Density (calculated)	1.467 Mg/m^3
Absorption coefficient	0.580 mm^{-1}
F(000)	488
Crystal size	0.5 x 0.2 x 0.2 mm
Theta range for data collection	2.79 to 26.00°
Index ranges	-6<=h<=6, -12<=k<=12, -31<=l<=31
Reflections collected	3148
Independent reflections	1576 [R(int) = 0.0331]
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	1572 / 0 / 137
Goodness-of-fit on F^2	1.074
Final R indices [I>2sigma(I)]	R1 = 0.0351, wR2 = 0.0793
R indices (all data)	R1 = 0.0687, wR2 = 0.1123
Absolute structure parameter	0.17(14)
Largest diff. peak and hole	0.212 and -0.227 $e.Å^{-3}$

The reason for determining the structure of this compound was to find the absolute configuration of the molecule: two of the carbon atoms (C2, C9) being chiral. Friedel opposites were therefore collected at 150K for each reflection. The resulting solution refined to a Flack parameter of 0.1669 with an esd of 0.1363. The expected values are 0.0 (within 3 esd's) for the correct, and +1.0 for the inverted, absolute structure, and so it is possible to state that the R, R enantiomer determined here (Figure 5.6) is optically pure.

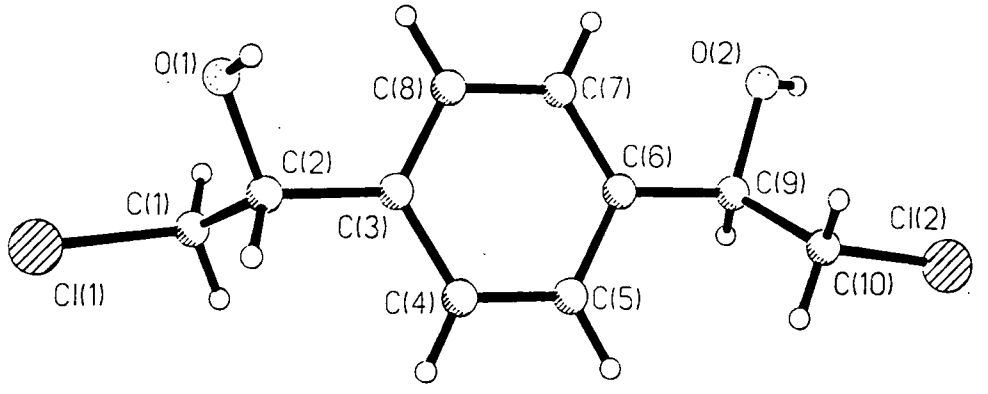


Figure 5.6

A plot of the R, R enantiomer determined for the molecule C_{10} H_{12} O_2 Cl_2

5.5.6 Monomer unit of $C_{17}H_{14}$ [Figure 5.7]

The molecule whose structure is shown in Figure 5.7 is a parent monomer for ring opening metathesis polymerisation (ROMP) reactions. The crystal structure determination showed that the *trans*-form is present, as expected, since this provides the maximum separation between the hydrogens on the apical carbon atom (C7) and the napthalene group.

Empirical formula	C17 H14	
Formula weight	218.3	
Temperature	150 K	
Diffractometer	Rigaku AFC6S	
Wavelength	0.71073 Å	
Crystal system	monoclinic	
Space group	$P2_1/c$	
Unit cell dimensions	a = 17.049(2) Å	alpha = 90°
	b = 9.687(1) Å	beta = 91.85(11)°
	c = 14.209(2) Å	gamma = 90°
Volume	2345.5(50) $Å^3$	
Z	8	
Density (calculated)	1.122 Mg/m^3	
Absorption coefficient	0..07 mm^{-1}	
F(000)	928	
Crystal size	0.6 x 0.4 x 0.35 mm	
Theta range for data collection	2.49 to 27.45°	
Index ranges	0<=h<=17, 0<=k<=12, -20<=l<=20	
Reflections collected	4123	
Independent reflections	2854 [R(int) = 0.0242]	
Refinement method	Full-matrix least-squares on F	
Data / restraints / parameters	2854 / 0 / 335	
Goodness-of-fit on F	1.5559	
Final R indices [I>2sigma(I)]	R1 = 0.0370, wR = 0.0397	
R indices (all data)	R1 = 0.0680, wR = 0.1098	
Largest diff. peak and hole	0.14 and -0.17 e.$Å^{-3}$	

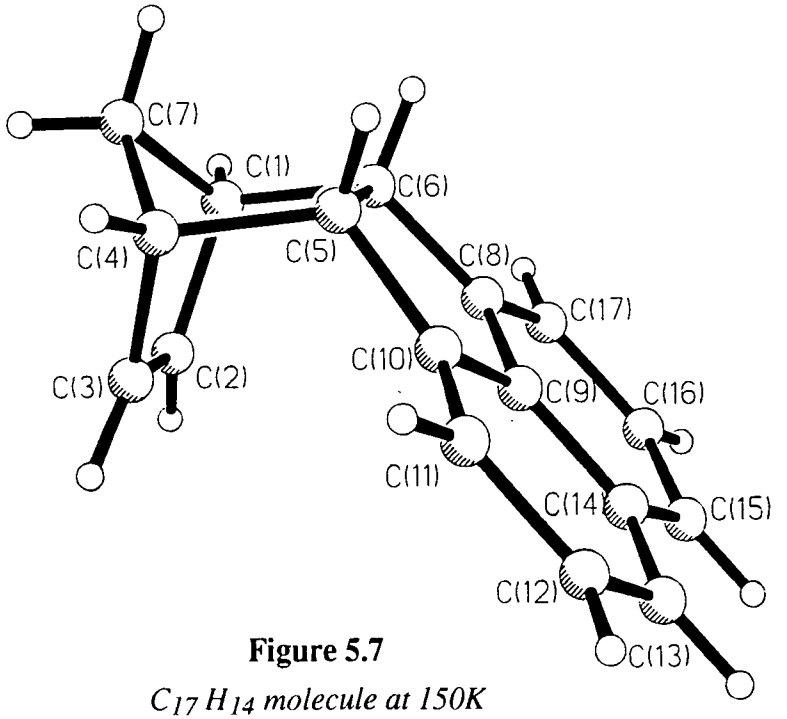


Figure 5.7

$C_{17}H_{14}$ molecule at 150K

5.5.7 1,5 - di (2'-chloroacetoxy) - 3,3 - dimethyl - 2,4 - diphenyl - 3 - silapentane ($C_{22}H_{26}O_4Cl_2Si$) [Figure 5.8]

Empirical formula	C22 H24 Cl2 O4 Si	
Formula weight	451.40	
Temperature	293(2) K	
Diffractometer	Siemens P4	
Wavelength	0.71073 Å	
Crystal system	triclinic	
Space group	P bar 1	
Unit cell dimensions	a = 10.959(1) Å	alpha = 69.05(1)°
	b = 11.003(1) Å	beta = 78.52(1)°
	c = 11.610(1) Å	gamma = 62.01(1)°
Volume	1153.7(2) $Å^3$	
Z	2	
Density (calculated)	1.299 Mg/m^3	
Absorption coefficient	0.358 mm^{-1}	
F(000)	472	
Crystal size	0.4 x 0.35 x 0.2 mm	
Theta range for data collection	2.77 to 27.50°	
Index ranges	-10<=h<=13, -12<=k<=13, -14<=l<=15	
Reflections collected	7377	
Independent reflections	5157 [R(int) = 0.0159]	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	5155 / 0 / 267	
Goodness-of-fit on F^2	1.112	
Final R indices [I>2sigma(I)]	R1 = 0.0694, wR2 = 0.1709	
R indices (all data)	R1 = 0.0969, wR2 = 0.2001	
Largest diff. peak and hole	0.744 and -0.420 $e.Å^{-3}$	

The quality of the data collected for this sample was very good, with well defined and intense peaks. However, refinement of the initial solution showed areas of disorder in the structure (Cl1, O3 and C22) which required further modelling in order to obtain the final solution. With all of the non-hydrogen atoms anisotropically refined it was found that the O3 and C22 disorder could be modelled as pairs of the corresponding half-occupied atoms (*i.e.* O3B and C22A = part 1, while O3A and C22B = part 2). At Cl1 though, the refinement of different numbers of anisotropic part chlorine atoms merely resulted in the generation of more possible positions, even with geometry restraints in place. The best model of the Cl1 movement around the C1 atom is provided by the isotropic refinement of 4 quarter-occupied Cl atoms, as can be seen from Figure 5.8. The four positions so generated are each chemically reasonable distances from the C1 atom, and angularly separated so as to define positions on the ring of electron density that represents the overall average position of the Cl1 atom.

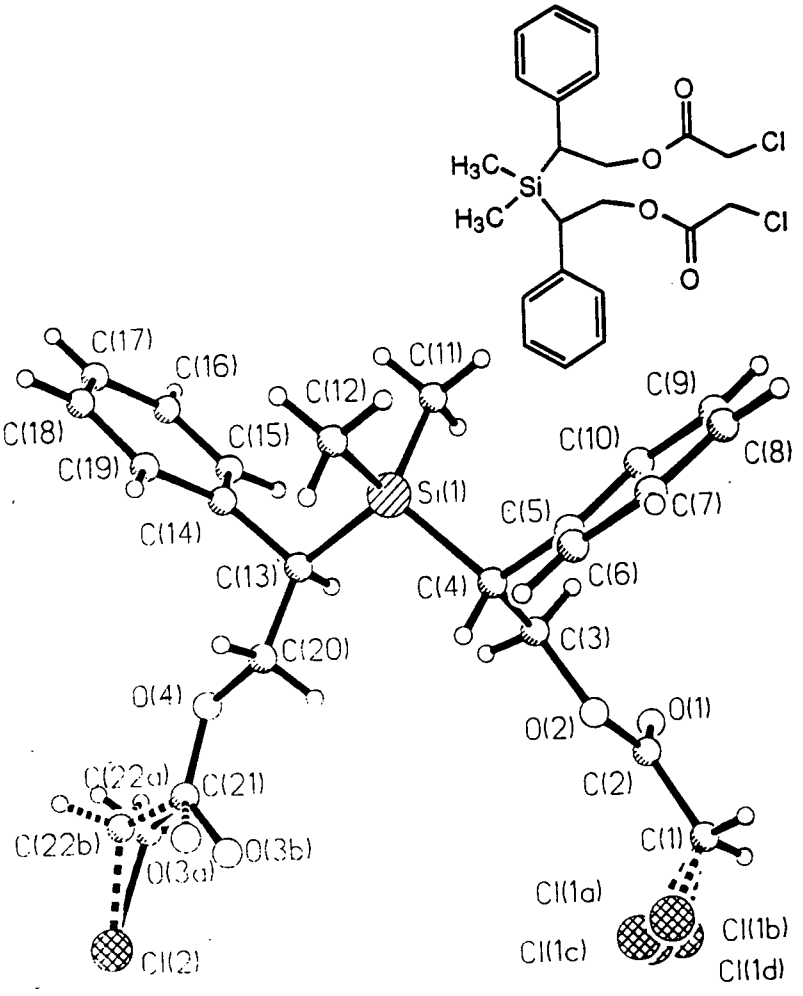


Figure 5.8 (a)

The disordered structure of the $C_{22}H_{26}O_4Cl_2Si$ molecule

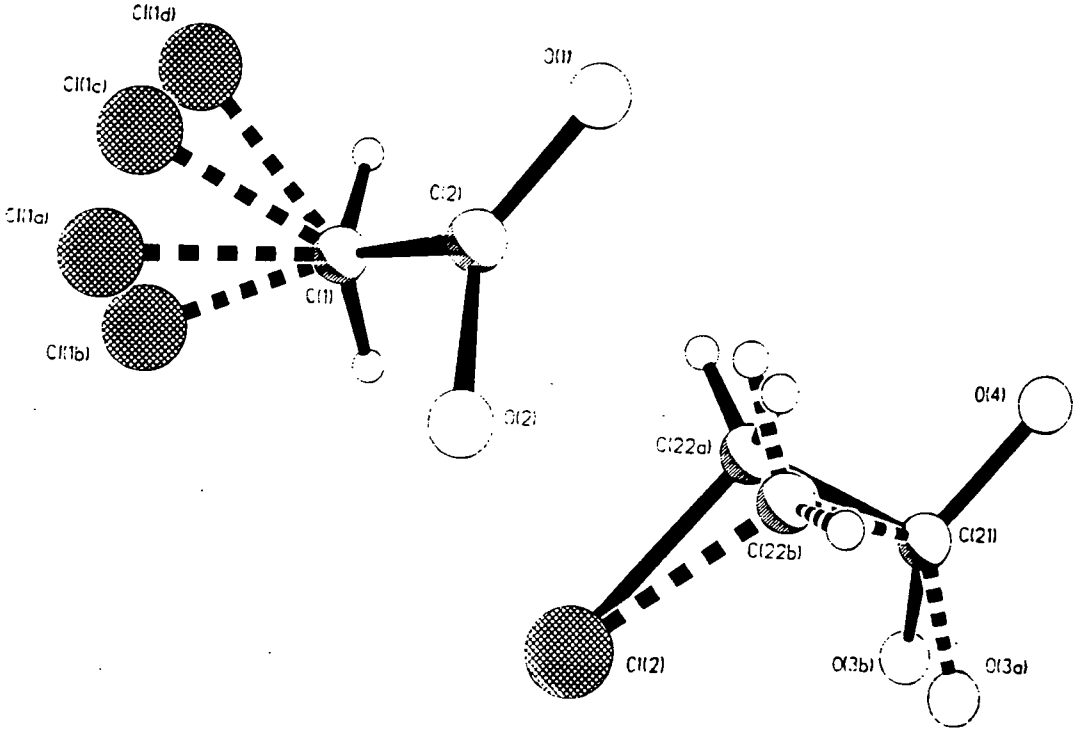


Figure 5.8 (b)

Models of the disorder for C22, O3 and Cl1

5.5.8 1,4 dibenzyl - 1,2,4,5 - tetraazacyclohexane ($C_{16}H_{20}N_4$) [Figure 5.9]

Empirical formula	C16 H20 N4	
Formula weight	268.36	
Temperature	293(2) K	
Diffractometer	Siemens P4	
Wavelength	0.71073 Å	
Crystal system	monoclinic	
Space group	$P2_1/c$	
Unit cell dimensions	a = 15.371(1) Å	alpha = 90°
	b = 5.050 Å	beta = 99.93(1)°
	c = 9.485(1) Å	gamma = 90°
Volume	725.23(9) $Å^3$	
Z	2	
Density (calculated)	1.229 Mg/m^3	
Absorption coefficient	0.076 mm^{-1}	
F(000)	288	
Crystal size	0.5 x 0.25 x 0.25 mm	
Theta range for data collection	2.69 to 28.00°	
Index ranges	-20<=h<=20, -6<=k<=0, -12<=l<=12	
Reflections collected	3514	
Independent reflections	1759 [R(int) = 0.0247]	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	1757 / 0 / 91	
Goodness-of-fit on F^2	1.046	
Final R indices [I>2sigma(I)]	R1 = 0.0616, wR2 = 0.1730	
R indices (all data)	R1 = 0.0849, wR2 = 0.1962	
Largest diff. peak and hole	0.560 and -0.464 $e.Å^{-3}$	

The proposed structure for this compound was that of a triazine, however it soon became apparent that the diazine (Figure 5.9) was in fact present. Data from the elemental analysis and the mass spectrometry support this structure. The diazine six-membered ring exists as a chair conformation with the bulky substituents both in equatorial positions.

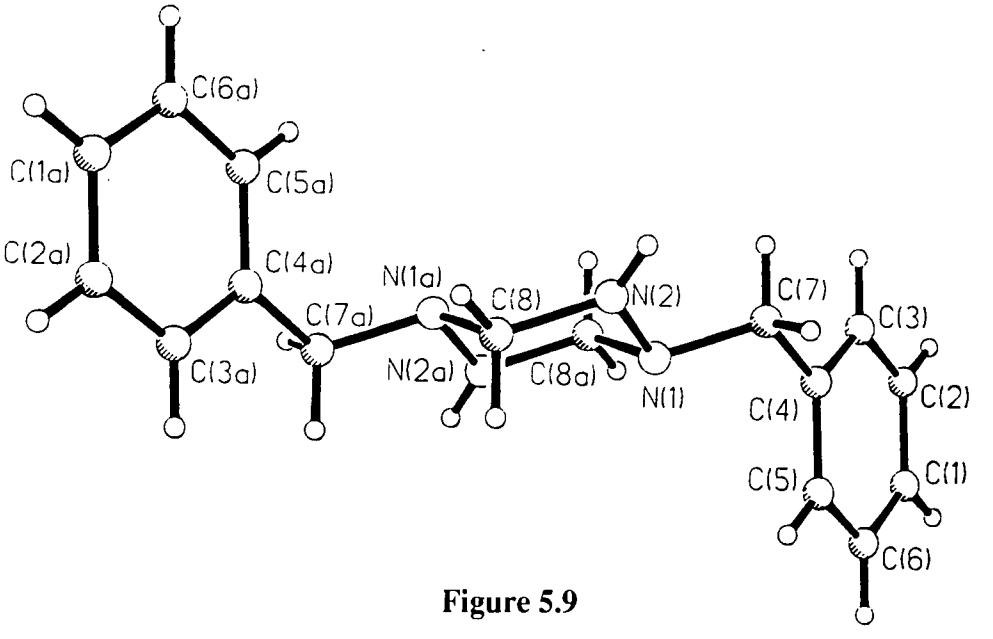


Figure 5.9

The diazine molecule of $C_{16}H_{20}N_4$

5.6 References

Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C.F., Mitchell, G.F., Smith, M.J. & Watson, D.G., *J. Chem. Inf. Comput. Sci.*, **1991**, 31, 187-204.

Allen, F.H. & Kennard, O., *Chem. Design Automation News*, **1993**, 8, 1.

Bernardinelli, G. & Flack, H.D., *Acta Cryst.*, **1985**, A41, 500-511.

Blessing, R.H., *Cryst. Rev.*, **1987**, 1, 3-58.

Cheetham, A.K. & Day, P., *Solid State Chemistry Techniques*, **1991**, OUP, New York.

Cruickshank, D.W.J. & McDonald, W.S., *Acta Cryst.*, **1967**, 23, 9-11.

Davison, G.R., Howard, J.A.K., Pitchford, N.A., Jones, A.M., Rasburn, J.W. & Simpson, A.J., *Tetrahedron*, **1993**, 49, 10123-10132.

Dunitz, J.D., Maverick, E.F. & Trueblood, K.N., *Angew. Chem. Internat. Edit. English*, **1988**, 27, 880-895.

Giacovazzo, S.C., Monaco, H.L., Viterbo, D., Scordari, F., Gilli, G., Zanotti, G. & Catti, M., *Fundamentals of Crystallography*, **1992**, OUP, Oxford.

Glusker, J.P. & Trueblood, K.N., *Crystal Structure Analysis. A Primer.* 2nd edition., **1985**, OUP, Oxford.

Harker, D. & Kasper, J.S., *Acta Cryst.*, **1948**, 1, 70.

Hendrickson J.B., *J. Am. Chem. Soc.*, **1967**, 89, 7036-7043.

Johnson, C.K., *Computing in Crystallography*, **1980**, eds. Diamond, R., Ramaseshan, S. & Venkatesan, K., Indian Academy of Sciences, Bangalore.

Karle, J. & Hauptman, H., *Acta Cryst.*, **1950**, 3, 181

Kolossvary I. & Guida W.C., *J. Am. Chem. Soc.*, **1993**, 115, 2107-2119.

Pitzer K.S. & Donath W.E., *J. Am. Chem. Soc.*, **1959**, 81, 3213.

Rogers, D., *Acta Cryst.*, **1981**, A37, 734-741.

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

Sheldrick, G.M., *SHELXTL PLUS*, **1991**, University of Göttingen, Germany.

Sheldrick, G.M., *SHELXL-93*, **1993**, University of Göttingen, Germany.

Stout, G.H. & Jensen, L.H., *X-ray Structure Determination. A Practical Guide.* 2nd edition., **1989**, Wiley, New York.

Woolfson, M.M., *Direct Methods in Crystallography*, **1961**, OUP, Oxford.

Chapter Six

Summary and suggestions for future work

6.1 Summary of structure correlation studies

This Thesis has demonstrated the use of the Cambridge Crystallographic Database for structure correlation experiments in two very different fields. The first part of the Thesis (Chapters 2 and 3) was concerned with the systematic conformational analysis of medium-sized rings and satisfied the objectives of this study by: (i) applying novel classification techniques to the conformational descriptions of both the seven- and eight-membered rings, (ii) interpreting the results in terms of the relevant conformational hypersurface by locating the highly populated regions of that hypersurface and mapping the interconversion pathways, (iii) studying, modifying and improving the available methodologies for data analysis, and (iv) relating the conformational minima found using these methods to both the chemical environments of the fragments under investigation, and to energetic features of the hypersurface obtained by computational methods.

The second major structure correlation study involved the analysis and description of 3-coordinated transition metal complexes in terms of both simple geometrical models and the group-theoretically based symmetry deformation coordinates. Deviations away from the ideal D_{3h} symmetry of this species were attributed to a variety of structural and environmental factors. Non-bonded interactions were seen to play a significant part in the geometry of the 3-coordinated fragment, and extrapolation of these results led to the rationalisation of an addition/elimination scheme linking 4- and 2-coordinated fragments through the intermediate 3-coordinated species.

6.2 Suggestions for further work

6.2.1 CSD software

The software developments integrated into the CSD System during the period of this research work have made possible the easy extraction and analysis of crystallographic data for structure correlation purposes. However, there are a number of identifiable areas where a further development of the search and analysis software would enable more complex structure correlation studies to be undertaken.

Deficiencies in the current software capabilities were most apparent for the experiments involving transition metals. Chief amongst these was the inability of QUEST3D to "normalise" distances between any two atoms by subtracting the sum of the respective covalent radii for those atoms from the calculated distance between the atoms (Section 4.2.3, eqn. [**4**]). This normalisation can only be done outside the CSD system, and subsequently the analysis of these normalised parameters must also be done using external software. Provision of an internal "look-up" table of covalent radii for each of the elements, and the "availability" of these values when designing a search query and generating a geometrical data matrix, would clearly make studies of the type carried out in Chapter 4 much more accessible. There is an enormous range of reaction types that can be analysed in terms of changes in "bond order", and the provision of this facility would be of immense advantage to research in both organic reaction pathways, and those of the organometallic and transition metal compounds.

It was also noted in Chapter 4 that analysis of the transition metals in terms of their oxidation states is currently reliant on the author's specification of oxidation states in the name of the compound. This situation is not ideal and many CSD entries carry no oxidation number information at all. If an electron-counting algorithm could be developed and incorporated into the search software, then structure correlation studies would be possible of the chemistry associated with specific electronic configurations and oxidation states. For instance, it would then be possible to study the effects of Jähn-Teller distortions over all 4-coordinated metals with d^9 electronic configurations,

and to correlate the observed configuration at the metal centre with the types of ligands to which it is bonded.

In order to provide a complete facility for structure correlation studies the CSD System also needs to incorporate more advanced tools for statistical analysis. The provision of automated symmetry-expansion facilities is therefore absolutely necessary, as the manual entering of permutational possibilities (Chapters 2, 3 and 4) is a cumbersome process which is prone to input errors, even for the simplest system. The continued development of the VISTA program to include options for more detailed statistical analysis (*e.g.* least-squares curve fitting, calculating correlation coefficients between parameters, cluster analysis, *etc.*) would also prove to be a valuable aid to the user: as a statistical tool VISTA still has considerable scope for development. Having said that, however, there do exist a wide variety of statistics packages for most computer platforms and close links between VISTA and some of these packages would also be desirable.

The non-bonded contact facility now present in the CSD System is an immensely useful tool, particularly for determining and analysing reaction coordinates and pathways present in certain chemical systems. However, there is currently no way of limiting the number of intermolecular contacts found for any one CSD entry, and instances arose during the study of non-bonded interactions to the 3-coordinated metals where a large number of contacts existed within the stipulated distance limits. This leads to an immense drain on computer resources (for one CSD entry in excess of 6 hours cpu-time were required to process that entry completely) as all of the contacts are located and tested against search requirements. An option to include a sensible cut-off point (*e.g.* maximum 2 contacts for any one entry) could therefore facilitate fast and efficient searches of a general nature.

From the descriptions of the clustering routines that are currently used in the CSD System (Chapters 2 and 3) it is apparent that extensive work is still needed to make these processes more "automatic". At present, satisfactory conclusions of cluster analyses are largely guided by the *a priori* chemical knowledge of the user, with parameters adjusted until chemically sensible clusters result. The problem with many

of the available clustering algorithms is that they do not "know" when to stop agglomerating data, *i.e.* programs may continue the process of clustering even after chemically sensible clusters have been identified. The potential for the use of cluster analysis in structure correlation studies is wide, and so the development of these algorithms should continue to the point where they become a powerful automated tool in pattern recognition. Such a tool could then be used to generate "fragment libraries", *i.e.* libraries of common conformations exhibited by a range of substructural fragments. Such libraries would, of course, find considerable use in molecular modelling and drug design applications.

6.2.2 Structure correlation studies of conformation and configuration

The conformational analyses of the medium sized carbocyclic rings has now been completed up to the 8-membered ring level. The scope for further analysis of larger rings is dependent upon the number of entries present in the CSD for the ring in question: the number of 9- and 10-membered carbocyclic rings is less than 10 of each, and so a complete conformational analysis is not practical. However, investigations into the conformational properties of the diolide crystal structure determined in Section 5.5.2 showed that similar molecules with much smaller ring structures (12-14 members) exist and display interesting conformational features. Chemical systems such as these do exist in the CSD and the opportunities are therefore present for conformational analyses of their collective structures.

The determination of specific spatial properties for ring systems, and the observation of how these properties are affected by chemical environment and different ring substituents, is of great interest in the field of rational drug design, as illustrated above and in Section 1.2.5. This is another area in which a great deal of work can be done, and would follow on directly from the analyses performed in Chapters 2 and 3. It is hoped that studies of environmental effects, similar to those carried out manually in Chapters 3 and 4, can become automated: the CSD screen records are bit-encoded

representations of 2D structure and automated comparisons of these records opens the possibility of machine learning in this area.

Possible structure correlation studies on the transition metal compounds present in the CSD are numerous. In Chapter 4 the configurations found for purely 3-coordinated transition metal complexes were analysed, and future work in this area may go on to study the configurational variety for all possible coordination spheres. One problem likely to be encountered in doing this, however, is that for the more common coordination numbers, *i.e.* 4, 5 and 6, the number of entries that will be located in the CSD far exceeds that found for the 3-coordinated metals, and file sizes and data handling will become less manageable. Also, as the coordination number increases then so too does the number of parameters required to describe the geometry, as well as the number of permutations required to account for the higher topological symmetry. When combined with the greater number of fragments, these factors will result in huge data matrices for these studies.

Using the intermolecular non-bonded contact search routines it will also be possible to study the addition/elimination type schemes through all possible coordination numbers. Again this raises the problem of data handling and computational time required, but with the continued development of faster and more powerful computers this type of problem may become less of a consideration. Future structure correlation studies may also concentrate on other aspects of coordination chemistry such as the influence of ligand type on the geometry of the coordinated metal, or the analysis of specific electronic configurations and oxidation states. The possibilities are apparently open-ended, but are currently limited by the availability of suitable software for search and analysis, as was illustrated above.

6.2.3 Crystal structure determinations designed to fill knowledge gaps

It was noted in Section 3.4.3 that the examples present in the CSD for the 1,3-cycloheptadiene fragment showed a preference for the 1,3-T conformation. This was largely unexpected as the 1,3-E conformation for this molecule is the energetically

preferred form. In order to investigate this further, a collaboration with a synthetic chemist who was aimed to produce a number of these types of 1,3-dienes with different functionality present around the ring so that the conformational properties of the ring could be studied. One such compound was synthesised and its structure determined (Section 5.5.1). The conformation of the ring was found to be that of the minimum energy 1,3-E form. Other attempts to produce synthetic analogues of this ring type failed, due to the marked preference for oil formation during synthesis instead of crystal formation,. However the concept remains sound: that if there are obvious "knowledge gaps" in the crystallographic literature, then it is possible to perform combined synthetic and crystallographic studies specifically designed to fill in these gaps. There are numerous possibilities for this field of study because for structure correlation purposes there can never be too much data.

Appendix One

CSD refcodes for all of the datasets of Chapters 2 and 3

Dataset 7C1

AUTUMN	BUHXAD	DASPAT	FIYJUS	HYMINB	PLENOL10	STRAMB10
BAHIAL	CAZJUI	DIPHEP10	FOSFAU	HYMXON	PULCHL10	TAGHUE
BAHMIG	CEBBUG	DIPHIT10	FUGFUI	HYSTER	RADIAT	TDCOBZ
BAKVAK	CEWSAY10	DITYEK	GEXXOW	IPTDCN	RUDMOL	VALGEU
BERDAD	CEWSEC10	DIYKEB	GEYNAZ	JATGUG	SAJPAU	VALGIY
BEXFUF	CHBRBZ	FABFET	GUANLD	JEHMEO	SAJPEY	VAWBAW
BEXGAM	CIFBIC	FEGWAP	HAZFNO	JIDBON	SCHOHX10	VEKGAT
BIGVOC	CILXEA	FEGWAP01	HBCDEO	KAYVIP	SEGFUF	VEMMOP
BIKNIS	CPMOXZ	FEYDAO	HBCUDO	KAZNOO	SEHBOW	
BITENL10	CSNLRB	FGUOLD	HDAPTX	KEJPEU	SEZLUE	
BOLKUI	CUFVII10	FIBSOY	HOXTHD	KEKXON	SILYIV	
BOLWUU	DACFOC	FIBSOY10	HXMXCT10	KIMKOG	SPDXLN	
BRMEXE10	DANNAH	FIWMIH	HYMINA	KRLAPY	SPICHB10	

Dataset 7C2

ABCOCX	BUXRER	CUTVIW	FAPKUC	GRAYAN	KIKZOT	SIFNOK
ACONTN10	BUXRER10	CUVTES	FAPLAJ	HFPCDN	KIKZOZ	SINXAO
ACONT010	BZOTRP	CUWHOR	FEMJUC	HMCYDL10	LAPACN10	SINXES
ADHKAU	BZTRMS	CUXTOE	FEMKAJ	HMPCD020	LEUCOA	STPHSN
AJMALN10	CAFCIV	CXMTUN	FEMKEN	HMSDCN10	LEUCOB	TCNONC
ANABSN10	CALKEF	DAFGAS	FEMKIR	HMTCDO	LYCOTD	TCTCDC
BABKEU	CALKEF10	DAFWAI	FEXKOI	IPCUND	MBOXPH	TRPHDT10
BAHCOC	CANHUU	DAKWUH	FEYSEH	JAJBEB	MBZAZL	TSBEYA
BAZFEN	CANLEI	DAPLAC10	FIDXUL	JAJVOF	MCUNDB	TZPHDO
BCUNDL	CANNIO	DATMOA	FIJYIG	JALFUX	MENTCN	VACZEE
BCYLON10	CAXPOG	DEBPAB	FIKXAY	JAPBIL	NHMENT	VAHWUW
BECKOJ	CDBOTN	DECZAM	FINLOO	JAPBOR	NNHXNT	VAHXAD
BEHCIA	CECKUQ	DELNUD10	FIVYEO	JAPLUH	NORCBR	VATJIJ10
BEHDUN	CEGZAP	DELPSC10	FIZPIN	JATWEG	OCTHEN	VATMAE
BEKGUT	CEKZIB	DELRUH	FIZVUF	JAYRUW	OHPHYL	VEDTIH
BEMOTD	CEXCIR	DEXWEI	FIZXUH	JAYZEO	OHRPRT	VEWKUD
BERLIT	CEXGER	DICKUV	FIZYAO	JENFAJ	PCCHLD	VEYPEU
BEXGUG	CHASBC10	DITCDC	FOLNAV	JEPMOG	PEIAJM10	VIGW'UD
BEYMIB10	CHTDMS10	DITDUF	FOVSUE	JEPMUM	PHALDR	VIKNEI
BHPCHD10	CIJSET	DIWKOJ	FOVTIT	JIDNAL	PHMRNT	VIKYAP
BHYOSM	CIKSAQ	DIWLOK	FOWZAS	JIFRAR	PHYBAC	VIKYET
BIBXEP	CINZEY	DIWLUQ	FOWZAS10	KAMBUV	PIERIG	VIKYIX
BIGSIT	CITHUI	DLPHFL10	FOYPIS	KAMCAC	PIMARB	VIKYOD
BIRGEO	CITJAQ	DOBHIL	FUDHAN	KASNAT	PTSANO	VIKYUJ
BLONGA10	CIZWEN	DOFBEF	GADSIN	KAVDOA	SAJWIJ	VILCOI
BLONGF	CNADMT	DOFHUB	GAHGAX	KAVDUG	SAKDOX	VINVAP
BNORS010	COCAIN10	DOKJUI	GAKPAJ	KAYSEI	SALTII	VINVET
BOBCIE	COCMEI	DONFER	GANCED	KECGOO	SALXOS	VUWZES
BOBCIE10	COCSES	DONHET	GANCIH	KEGJIP	SAMWAE	VUYYUJ
BODBEB	COCSES10	DOTHUP	GARCAD	KELYUV	SAPLOK	XMTDIB
BORRIJ	COLBAG	DUGKOF	GAWNOH	KERYIP	SAPMAX	COYFEB
BOTMOM	COLBEK	DUMRAE	GAYDIT	KETFIY	SARBIW	FATKAM
BOTNUT	COXNIM	DUMRAF	GAZSIJ	KETREG	SBLVES	FOJPID
BRLCTA	CPETLD	DUTNIP10	GEBLEE	KETRIK	SEGKOE	GRYNTX
8RLOFE10	CPRSBR	DUYREU	GERZIM	KETRUW	SEJPOM	SEMKOK
BRTRPN	CSNLRA	DUYRIY	GETCAJ	KETSAD	SEVFUU	SEMLAX
BTCUND	CTCDEC	DUZHAH	GIBJAC	KIDMIT	SEYZOL	
BUCHAI	CUCYII	ETCOGR10	GIGJEL	KIKYOS	SEZNUG	
BUHXEH	CUGLEV	EUPALB	GIHXAW	KIKZAF	SIFFIW	
BUMFEU	CUGLOF	FAJHAZ	GILYUV	KIKZIN	SIFFOC	

Methylenecycloheptanes (V)

ACHOLB	CELPIS	FAWLUK	SALREC
AIBFOR10	CELPOY	FEBXUF	MIKTIN
BETDOT	CENSIN01	FEJHOR	SANGOD
BGIGAC	CENSIN10	FEMYAX	SAYZOH
BISJUI	CEWRUR10	FIJZED	SERRAT10
BTMUNE	CHSPOX	FOYMEL	SOLSIN
BUMPII	CIDYOD	GAYFOB	TBCMUN
BUVPEN	CODSUJ10	GRLCTN	THIELN10
BUWSAN	CUNPIK	HGUDOL	VENPOT
BUWSAN01	CUPCIZ	HIMCHC	VEPXIX
CARGIL	DAJKOO	HNANDB	XANTHB
CAWGIQ	DANETN	HYPOCH	
CAWGIQ01	DAPTIX	ITNGRO	
CELPAK	DHYGUA	JIDKUC	
CELPEO	DIPXOP	KEJNUI	

Cycloheptanones (VI)

APORON	DINMUI10	FOLKUM	HONGUA
BISYUX	DOTJOL	FONFAP	JAMFOS
BOXXOB	FECHEA	FUFLAT	JIDHUZ
CAYZIL	FEGSOZ	FUSNEM	JILJIX
CIJMUD	FEYCER	FUSPAK	LASROL
CIJNAK	FEYREG	GEBRAG	SADYEB
CURFAW	FIKDOS	GETNAU	
DAFCOC	FIXMUU	HAESTD	

Azocycloheptanes (VII)

BOCBEA	BUVVAP	CACBEN	DOTJUR
BOCBIE	BUVVET	CENMIR	DOTKAY
BPAZBH	CABZUA	DORJUP	KANDOS
BPAZTS	CACBAJ	DORKAW	

Oxacycloheptanes (VIII)

ACGLSP	BUNSIM11	JAMVOI	KEVSEJ
AHIDIT	CIHDIG	JAMVOI10	MAGSEP10
AIPGLC	EIPGLU10	JAMVUI	MIPGSP
AIPMGH	FAJMOS	JAMVUI10	MPASEP
BATLAJ	FISHAQ	JAPGAI	SEVKIN
BERFAF	FISHAQ10	JAPGEM	VEZKAM
BUFFOX	GETKEV	KETVUA	
BUJRED	GETKEV10	KETXUC	
BUJXIN	GLUSPT10	KEVBAO	

Azacycloheptanes (IX)

AZBQUB	CABCUD01	JEBTEP	VUZDID
BACMEC10	CAVFOU	JIDBAZ	XIMBZA
BHAUST	CROOMI	KAVLIC	XIMBZB
BUCXAY	FECYIV	MECILN	XIMBZB01
BUFPEX	FOZKUA	STEMON10	
CABCUD	HEXAMC	VANRAD	

ε-Lactones (X)

AXHNOM	FAWLEU	JOCCUZ	VESVIY
BEBSIK	JOCCIN	JOCDIO	
BRASSA01	JOCCOT	PRIEUB10	

ε-Lactams (XI)

BAJZOB10	CAPRES	FEFSUE	VAXYOI
BILJOV	CAZCHI	FULMUU	VENTEN
CAPLAC	DIKVAU	FULNAB	VENTIR
CAPLAC01	DOKMUL	FULNEF	VIMXIY

Cycloheptenes (XII)

AESTME10	CUDTOK	GESJOD	ONYCHL
AHBCPB	CUDTUQ	GESJUJ	PARMOL10
ARTEGA	CUKJIB	GIHCEF	PLENAC
BAHHAT	DAZPID	GIPCAJ	PORTSH10
BEBPED	DEFSEM	INSERR10	PSVALB10
BERLAN	DEFSIQ	IVLBIN	SACSEU
BIHNOV	DIDHIH	JAHHIJ	SEBBUW
BRGAIL10	DOLTAC10	JEXJOL	TAKROM
CAPLAG	DORKIE	KAHDUS	TCTDOL
CAXJUG	DOZMOU	KAMBAB	VEBNUL
CAZXIK	ERGYZN	KEHFIM	VEGTUW
CHEBAR	FICCUP	LACTUC10	VEKBUI
CILXAW	FUCVII	MEOCHP	VEZPEV
CILXEA	FUPYAQ	MXSERR10	
CIWCAM10	GASCHA	NPHRBL10	
CIWCEQ10	GAYDUF	OCMETD01	
COVPAE	GAYFAN	OCMETD10	
CUDSUP	GEJYEZ	OHDLST	

Cyclohepta-1,3-dienes (XIII)

CAJCOF	EXPOCP	MELOCH	VEFCAK
CHMHAZ	GELWOJ	NACCOL	XMOHAZ
DOYHUU	HCMCHP	PBOPOS	
DUTGAA	KENLUK	SATGUP	

Cyclohepta-1,4-dienes (XIV)

BERWOK	CIKVIB	JASGIT	JIDCUU
CHSBHA	DPHTRP	JIDCII	SAHROI
CIKVEX	HBZCHP	JIDCOO	

Cyclohepta-1,3,5-trienes (XV)

BAYRIC	CMCHBZ	DMDPCH	HPHAZO
BAYROI	CPHHTZ	DOZSIU	KEXNIK
BEDWAI	DAJXAN	DUGWOR	MTBCHO
BIDTEN	DALXOD	FAPMOY	TUJBRP10
BSPDOD10	DELCUS	FEKTEU	VAMKAV
BUKTAC	DELDAZ	FUNCOG	
BUZTUL	DIGGEF	GIJXIG	
CIDCEX	DIGGIJ	HPHAZL	

Dataset 8C1 (XVI)

AMCOCA	BAGPII	BCOCTB	CLCOCT
COCOAC	COCOXA10	COVLUU	CURBIA
CUVZEY	CYOCDL	DEZPUT	ECOTDA
EOCNON10	GATRAU	GIVBAO	HOXTHD
HUMULB10	KESVIN	MSOCYO10	OCSHYD
PCDODO	SATKIH	SEJFIW	SPOCTC10
SPTZBN	VALGOE	VASWOB	

Dataset 8C2 (XVII)

BMEOPH10	BUNPRA	BUTGOM	CACKAS
CBCOEK	CEPHAL10	COCNBZ01	COCNBZ10
CRPLST10	CTSCOC	DEMRES	DINSIC10
DIWZOY	EOCTEC	FEKRUI	FEMGEJ
FETKEU	IPRINC	KAKHEJ	TCHPDO

Dataset 8C3 (XVIII)

BRISTG	BUBXUR	CETZUW	CICHOL
CICHOL01	CIFNAG	CILSOF	COLBOU
COLBUA	DENWOI	FETRAX	GEDVAM
HISPIC	ISTEGO	JECKIL	JECKOR
KAKHAF	KEHTOG		

Dataset 8C4 (XIX)

BINWIE	COCPAL	COCPAL10	COCPEP
COCPEP10	DBCODE	DSTYBZ	DUDTOL
FODZED	TAFZOP	VAWPAK	

Dataset 8C5 (XX)

AOCINB	BOVMUU	BURJED	CAWLAN
CYCOXA	CYCOXB	DEJJUX	DEMRIW
POCARA	POCARD	POITDL	

Dataset 8A1 (XXI)

FINZAD	HEPAMC	KETVUA	KEVBES
KEVBIW	LAURCN10	VEZJUF	

Dataset 8A2 (XXII)

ACTHCO10	BODVAR10	BPHZOC10	CARYAV
CARYEZ	CATJIQ	CATJIQ01	FAVMOE
MSAZOP01	MSAZOP10	THETAC	TSHOCZ

Dataset 8A3 (XXIII)

ANTZCO	BALRIP	BEJVIV	BUKDAM
COWBEV	DEDBUJ	KAKLIR	METALD
MOAOSP10	MTCHLR	OCHTET	OCHTET01
OCHTET03	OCHTET04	OCHTET12	OMSIOC
TACTAZ10	TETIOC10	TOXOCN	TSELOC

Appendix Two

Orthogonal coordinates (Å) referred to molecular axes for the most representative fragment of each of the datasets in Chapter 3 (where applicable)

Dataset 7C1

Cluster One - KIMKOG

0.00000	0.00000	0.00000
1.51487	0.00000	0.00000
2.18141	1.36601	0.00000
2.37870	1.94067	1.40075
1.10784	2.41097	2.08581
-0.01962	1.38328	2.17856
-0.68148	1.06375	0.83655

Cluster 3 - SAJPEY

0.00000	0.00000	0.00000
1.54766	0.00000	0.00000
2.03042	1.45320	0.00000
1.57473	2.37581	-1.17107
0.16042	2.29335	-1.73791
-0.26326	0.95919	-2.34454
-0.76263	-0.07631	-1.34485

Cluster 2 - CUFVII10

0.00000	0.00000	0.00000
1.60477	0.00000	0.00000
2.33494	1.33460	0.00000
2.08713	2.19579	-1.26956
0.79047	2.98718	-1.25613
-0.50318	2.17122	-1.26496
-0.74229	1.32784	0.00082

Cluster 4 - HYMINB

0.00000	0.00000	0.00000
1.55151	0.00000	0.00000
2.19854	1.40347	0.00000
1.28529	2.52860	-0.49351
0.06844	2.81084	0.42985
-0.12022	1.92323	1.63433
-0.67267	0.52646	1.27965

Dataset 7C2

Cluster 1 - SINXES

0.00000	0.00000	0.00000
1.56067	0.00000	0.00000
2.11660	1.42403	0.00000
1.38140	2.39355	0.96135
-0.13320	2.19235	0.83813
-0.63899	2.30382	-0.62580
-0.54260	0.84947	-1.18050

Cluster 6 - GAYDIT

0.00000	0.00000	0.00000
1.54278	0.00000	0.00000
2.19822	1.37129	0.00000
1.66385	2.23518	-1.15581
1.83993	1.46168	-2.47930
0.95315	0.23359	-2.20881
-0.24975	0.78077	-1.33405

Cluster 2 - VINVAP

0.00000	0.00000	0.00000
1.54735	0.00000	0.00000
1.94993	1.48414	0.00000
2.04872	1.88941	-1.49430
1.70404	0.61615	-2.29551
0.16942	0.67275	-2.48888
-0.35262	-0.37426	-1.46300

Cluster 7 - CXMTUN

0.00000	0.00000	0.00000
1.54218	0.00000	0.00000
2.16806	1.40616	0.00000
2.27461	2.04582	-1.39324
0.99440	2.34990	-2.17346
-0.35930	2.04657	-1.51937
-0.62557	0.55848	-1.28848

Cluster 3 - DLPHFL10

0.00000	0.00000	0.00000
1.54482	0.00000	0.00000
1.93295	1.49593	0.00000
2.34675	2.01601	1.39530
1.27127	1.63060	2.43221
-0.14321	1.77647	1.89005
-0.38751	1.45872	0.38680

Cluster 8 - VEWKUD

0.00000	0.00000	0.00000
1.53443	0.00000	0.00000
1.88149	1.49424	0.00000
1.89275	2.13475	1.39811
0.59086	2.00357	2.25086
0.17211	0.56481	2.58384
-0.65152	-0.08030	1.40968

Cluster 4 - JAPBIL

0.00000	0.00000	0.00000
1.54827	0.00000	0.00000
2.00729	1.46960	0.00000
2.98058	1.53847	1.16055
2.01083	0.93246	2.17780
1.55347	-0.36903	1.50983
0.00095	-0.36542	1.52592

Cluster 9 - JAJVOF

0.00000	0.00000	0.00000
1.54979	0.00000	0.00000
1.89723	1.51488	0.00000
1.91064	1.52060	1.55135
0.39870	1.86331	1.49115
0.50277	2.19587	-0.02515
-0.42762	1.24791	-0.74653

Cluster 5 - IPCUND			Cluster 10 - CUXTOE		
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.56350	0.00000	0.00000	1.51624	0.00000	0.00000
1.97812	1.46123	0.00000	1.83048	1.55204	0.00000
0.97525	1.82890	-1.07990	2.67980	1.74778	-1.32032
1.10143	0.72003	-2.17490	1.83928	1.62201	-2.61076
-0.36298	0.18153	-2.33644	1.03011	0.35142	-2.28970
-0.37688	-0.88200	-1.24121	-0.26262	0.67658	-1.42445

Dataset (V)

DHYGUA (TC^3)			SAYZOH (TC^2)		
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.50358	0.00000	0.00000	1.50772	0.00000	0.00000
2.17781	1.38807	0.00000	2.26336	1.36046	0.00000
1.49259	2.37847	-0.95133	1.63617	2.39534	0.92283
0.36465	3.18414	-0.28213	0.22297	2.82629	0.50687
-0.91252	2.35684	-0.01834	-0.90945	2.06808	1.18663
-0.68924	1.06924	0.75887	-0.71858	0.54840	1.20206
-0.70284	-0.89600	-0.60902	-0.65071	-0.52492	-1.02679

MIKTIN (C^1)			CELPEO (TC^4)		
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.52450	0.00000	0.00000	1.50165	0.00000	0.00000
2.18488	1.40539	0.00000	2.18852	1.36494	0.00000
2.19412	2.06571	1.38994	2.42940	1.96180	1.37973
0.82107	2.31810	1.94112	1.15560	2.37842	2.13195
0.00397	1.07563	2.29081	0.12659	1.28513	2.25029
-0.83876	0.56049	1.12395	-0.65802	0.99222	0.94871
-0.52682	-0.59448	-1.08300	-0.73447	-0.76002	-0.82935

Dataset (VI)

FUSNEM (TC^3)			JIDHUZ (C^1)		
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.51762	0.00000	0.00000	1.55622	0.00000	0.00000
2.06884	1.42898	0.00000	2.23573	1.40511	0.00000
1.96689	2.15876	1.32568	2.01909	2.19661	-1.30489
0.57419	2.32105	1.93329	0.61521	2.67560	-1.58754
-0.03168	0.96499	2.40565	-0.32108	1.58686	-2.03477
-0.72791	0.06980	1.35189	-0.97740	0.74304	-0.92738
-0.61308	-0.07906	-1.03782	-0.51641	-0.67905	0.88123

HAESTD (C^2)			FUFLAT (TC^2)		
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.51343	0.00000	0.00000	1.50442	0.00000	0.00000
2.09341	1.44500	0.00000	2.00222	1.46619	0.00000
1.75258	2.24283	-1.26976	1.19191	2.46682	-0.79464
0.23614	2.52169	-1.46132	0.66534	1.96387	-2.15243
-0.64802	1.42626	-2.09765	-0.66095	1.22887	-2.07337
-0.73458	0.03736	-1.37888	-0.64034	-0.11910	-1.37118
-0.59096	0.04601	1.01386	-0.66720	0.17902	0.98740

Dataset (VII)

BUVVAP (C^1)			CACBEN (TC^3)		
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.49541	0.00000	0.00000	1.50245	0.00000	0.00000
2.14162	1.41170	0.00000	2.02605	1.41140	0.00000
2.23837	1.95959	1.42863	1.84130	2.14993	-1.33630
0.89844	2.21199	2.14012	0.39338	2.22905	-1.80559
0.03181	0.98070	2.37973	-0.21095	0.91689	-2.38703
-0.78236	0.53703	1.17157	-0.70615	-0.11662	-1.34973
-0.53392	-0.54193	-1.03372	-0.72987	0.11417	1.02031

Dataset (VIII)

BUJRED (TC^3)			JAMVUI (TC^2)		
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.40522	0.00000	0.00000	1.43053	0.00000	0.00000
2.04349	1.39395	0.00000	2.00424	1.41098	0.00000
1.17212	2.43110	0.72810	1.55739	2.26954	1.16852
0.07480	3.03218	-0.16051	0.06867	2.63575	1.07566
-1.04931	2.08816	-0.58109	-0.89961	1.61508	1.63848
-0.58084	0.74220	-1.07548	-0.57961	0.15957	1.30190

GETKEV (C^2)			JAMVOI (C^3)		
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.40590	0.00000	0.00000	1.43224	0.00000	0.00000
2.09921	1.31010	0.00000	1.98341	1.42837	0.00000
1.28553	2.60102	-0.09094	1.58119	2.23441	-1.21621
0.29929	2.76885	-1.23763	0.14151	2.73513	-1.14879
-0.95594	1.93558	-1.14459	-1.00406	1.71858	-1.35997
-0.74651	0.42573	-1.15047	-0.58639	0.25333	-1.29961

Dataset (IX)

VANRAD (TC^3)			STEMON10 (C^1/TC^4)		
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.48828	0.00000	0.00000	1.55854	0.00000	0.00000
2.03658	1.40626	0.00000	2.21872	1.29511	0.00000
1.90804	2.15833	-1.30637	2.46882	1.78536	-1.47090
0.50004	2.23969	-1.86485	1.24348	2.11485	-2.21095
-0.04381	0.91608	-2.35368	0.24843	0.97435	-2.33141
-0.70283	0.04590	-1.32346	-0.64551	0.74635	-1.14548

Dataset (X)

JOCCOT (C^1)		
0.00000	0.00000	0.00000
1.32224	0.00000	0.00000
2.06360	1.24093	0.00000
1.95490	2.02429	-1.28780
0.68568	2.83993	-1.49781
-0.59408	2.00187	-1.46013
-0.76150	1.28091	-0.11813
-0.54021	-1.07369	0.10815

Dataset (XI)

CAPLAC01 (C^1)		
0.00000	0.00000	0.00000
1.32918	0.00000	0.00000
2.17969	1.20670	0.00000
2.07426	2.03174	1.25308
0.79808	2.87822	1.35857
-0.49710	2.07639	1.33635
-0.73967	1.31674	0.04476
-0.64028	-1.06621	-0.06508

Dataset (XII)

OCMETD01 T_b			SACSEU C_4		
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.52084	0.00000	0.00000	1.54759	0.00000	0.00000
2.25363	1.35251	0.00000	2.14699	1.42841	0.00000
2.05086	1.99874	1.34347	2.02893	2.08859	1.36029
1.15703	1.82133	2.32094	0.91667	2.30095	2.00808
-0.18406	1.12689	2.37713	-0.48893	1.95531	1.61368
-0.70728	0.96814	0.93829	-0.68913	0.46958	1.29907

Dataset (XIII)

HCMCHP 1,3-C_2(ii)		
0.00000	0.00000	0.00000
1.53814	0.00000	0.00000
2.13764	1.37701	0.00000
2.06811	2.19878	-1.05454
1.32642	1.85406	-2.28697
0.08807	1.29466	-2.19260
-0.61366	1.08447	-0.89657

Dataset (XIV)

CIKVEX 1,4-A		
0.00000	0.00000	0.00000
1.54819	0.00000	0.00000
2.35564	1.14628	0.00000
1.92899	2.60423	0.02838
0.51747	2.98543	-0.35987
-0.48175	2.45466	0.64287
-0.64526	1.08698	0.86750

Dataset (XV)

	BAYRIC C_S			CPHHTZ C_S	
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.51495	0.00000	0.00000	1.49304	0.00000	0.00000
2.22856	1.13437	0.00000	2.11364	1.20010	0.00000
1.69860	2.47892	0.15652	1.44514	2.43489	-0.31433
0.61721	2.81712	0.87958	0.38560	2.59681	-1.14237
-0.24666	1.91095	1.59783	-0.27326	1.54860	-1.86009
-0.49359	0.64322	1.26432	-0.38672	0.29023	-1.40597

	KEXNIK C_S	
0.00000	0.00000	0.00000
1.49886	0.00000	0.00000
2.19122	1.21716	0.00000
1.46623	2.50230	0.09809
0.34427	2.82188	-0.59893
-0.34458	1.93904	-1.55800
-0.50078	0.56560	-1.30264

Dataset 8C1 (XVI)

	BAGPII			COVLUU	
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.52762	0.00000	0.00000	1.53112	0.00000	0.00000
2.19721	1.37394	0.00000	2.24403	1.39734	0.00000
2.76966	1.79002	1.34996	1.54860	2.63663	0.53241
1.78416	2.04234	2.47819	1.54478	2.70862	2.05961
1.00786	0.82488	2.98225	0.91001	1.50903	2.76467
-0.41006	0.66109	2.44614	-0.51245	1.18377	2.29782
-0.66730	0.95578	0.96609	-0.65608	-0.03767	1.38347

	SPTZBN			DEZPUT	
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.54608	0.00000	0.00000	1.52156	0.00000	0.00000
2.18742	1.37521	0.00000	2.24255	1.37356	0.00000
2.15379	2.12182	1.30620	2.72291	1.96532	1.32290
0.96838	2.84315	1.79335	1.72478	2.71169	2.18437
-0.32055	2.88017	1.01942	0.51238	2.01606	2.79809
-1.24623	1.73302	1.40556	-0.68202	1.60693	1.94073
-0.63868	0.34613	1.31661	-0.75594	0.18392	1.31387

	EOCNON10	
0.00000	0.00000	0.00000
1.51951	0.00000	0.00000
2.28581	1.29594	0.00000
2.35940	1.88210	-1.42977
0.98689	1.83739	-2.00530
0.50037	0.64076	-2.64241
-0.87922	0.13252	-2.32894
-0.77703	-0.73842	-1.06039

Dataset 8C2 (XVII)

	IPRINC			COCNBZ01	
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.37865	0.00000	0.00000	1.34074	0.00000	0.00000
2.26675	1.19095	0.00000	2.05661	1.29314	0.00000
2.56463	1.70978	1.41016	2.22801	1.71433	-1.46797
1.37976	1.87189	2.30851	0.91577	1.70377	-2.27324
0.21360	2.70188	1.82478	0.39416	0.33614	-2.80812
-1.03510	1.89986	1.34532	-0.90975	-0.23945	-2.21401
-0.86434	1.22089	-0.00617	-0.76809	-0.94314	-0.84770

Dataset 8C3 (XVIII)

	GEDVAM			BUBXUR	
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.38645	0.00000	0.00000	1.37840	0.00000	0.00000
2.13573	1.28559	0.00000	2.24680	1.19222	0.00000
2.00189	2.20651	1.03886	2.37221	2.04060	-1.12565
1.12934	1.95024	2.25284	1.42041	1.81001	-2.31956
-0.28367	2.56084	2.20021	0.01631	2.49909	-2.15036
-1.30721	1.73987	1.38991	-0.44882	2.54506	-0.62435
-0.81583	1.27904	-0.00035	-0.87919	1.21933	-0.00234

Dataset 8C4 (XIX)

	TAFZOP			DBCODE	
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.40675	0.00000	0.00000	1.37198	0.00000	0.00000
2.19453	1.28650	0.00000	2.29553	1.16889	0.00000
2.21190	2.05483	-1.34928	1.73268	2.54388	0.25319
1.43589	3.34250	-1.35761	1.18041	2.70542	1.60192
0.02915	3.34250	-1.35761	-0.04295	2.34225	2.10575
-0.75864	2.05600	-1.35761	-1.17586	1.68495	1.39602
-0.77600	1.28767	0.01752	-0.94497	1.12111	0.01752

Dataset 8C5 (XX)

	AOCINB			DEJJUX	
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.50528	0.00000	0.00000	1.51450	0.00000	0.00000
2.14028	1.38241	0.00000	2.11297	1.43060	0.00000
1.86087	2.27159	-1.20865	1.36533	2.47854	-0.80754
0.41367	2.79418	-1.29403	1.40690	2.34143	-2.31445
-0.55036	2.06473	-2.27098	0.77304	1.09795	-2.92824
-0.33298	0.56539	-2.47596	-0.63120	0.75150	-2.41331
-0.72881	-0.30307	-1.28090	-0.70217	-0.30315	-1.29513

Dataset 8A1 (XXI)

	LAURCN10			KEVBES	
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.45528	0.00000	0.00000	1.46496	0.00000	0.00000
2.02358	1.40648	0.00000	1.95100	1.46452	0.00000
1.80871	2.17695	1.27361	1.79062	2.22226	-1.32660
0.39699	2.33546	1.69287	0.32515	2.89715	-1.46730
-0.50840	2.92753	0.58169	-0.68409	2.14615	-2.26481
-1.37088	1.98212	-0.20095	-1.47354	1.04929	-1.55542
-0.71665	0.85855	-0.92414	-0.79311	-0.24619	-1.17069

Dataset 8A2 (XXII)

	CARYAV			BPHZOC10	
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.82359	0.00000	0.00000	1.47484	0.00000	0.00000
2.49685	1.37064	0.00000	2.03780	1.41799	0.00000
2.33991	2.18940	-1.26936	1.89303	2.15362	-1.32613
0.61493	2.66838	-1.60776	0.61185	2.73869	-1.58320
0.18686	1.65599	-3.05400	-0.34934	2.02978	-2.50852
0.44554	0.16389	-2.87216	-1.35878	1.18349	-1.74187
-0.31231	-0.50432	-1.72012	-0.76844	-0.09896	-1.25955

Dataset 8A3 (XXIII)

	TETIOC10			TOXOCN	
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.81520	0.00000	0.00000	1.41102	0.00000	0.00000
2.71468	1.57189	0.00000	1.97232	1.28232	0.00000
2.68287	2.01727	-1.75799	2.30235	1.83877	1.24676
1.64216	3.42622	-2.22076	1.29986	2.64675	1.78883
0.08543	3.06904	-1.38174	0.28012	1.97606	2.49685
-1.07716	1.92352	-2.15247	-0.73504	1.47169	1.67559
-0.38788	0.29368	-1.74707	-0.61245	0.13941	1.24785

	OMSIOC	
0.00000	0.00000	0.00000
1.90323	0.00000	0.00000
2.84777	1.59542	0.00000
2.40536	2.73349	-1.48415
2.60238	2.02194	-3.24381
1.50426	0.49753	-3.54810
-0.31858	0.56510	-3.21614
-0.73758	1.05804	-1.40642

Appendix Three

Atomic coordinates (x 10^4) together with equivalent isotropic displacement parameters ($Å^2$ x 10^3) for the crystal structures presented in Chapter 5 of this Thesis. U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

3, 5 cycloheptadienyl - 3, 5 dinitrobenzoate ($C_{14}H_{12}O_6N_2$) [5.5.1]

	x	y	z	U(eq)
C(1)	1692(4)	1921(3)	9644(1)	30(1)
C(2)	2704(5)	3371(3)	9814(1)	37(1)
C(3)	2744(5)	3640(3)	10459(1)	42(1)
C(4)	1322(5)	3296(3)	10837(1)	41(1)
C(5)	-635(5)	2667(3)	10722(1)	42(1)
C(6)	-1470(5)	2138(3)	10230(1)	38(1)
C(7)	-553(4)	1941(3)	9645(1)	38(1)
C(8)	3833(4)	812(3)	8943(1)	30(1)
C(9)	3983(4)	323(3)	8319(1)	28(1)
C(10)	2440(4)	564(3)	7913(1)	27(1)
C(11)	2663(4)	69(3)	7345(1)	27(1)
C(12)	4354(4)	-660(3)	7164(1)	32(1)
C(13)	5837(4)	-897(3)	7583(1)	32(1)
C(14)	5700(4)	-418(3)	8155(1)	32(1)
O(1)	2146(3)	1551(2)	9032(1)	31(1)
O(2)	5107(3)	543(2)	9309(1)	41(1)
O(3)	8987(3)	-1872(2)	7761(1)	46(1)
O(4)	7605(3)	-2233(3)	6905(1)	53(1)
N(1)	7607(4)	-1728(3)	7403(1)	39(1)
N(101)	1036(4)	294(3)	6913(1)	34(1)
O(101)	1220(4)	-247(3)	6427(1)	57(1)
O(102)	-433(3)	991(2)	7064(1)	37(1)

A 34-membered diolide ($C_{32}H_{60}O_4$) [5.5.2]

	x	y	z	U(eq)
O(1)	2073(2)	2791(1)	4959(1)	67(1)
O(2)	-18(2)	3793(1)	4068(1)	52(1)
C(1)	170(3)	2666(2)	4550(1)	47(1)
C(2)	-2303(3)	1235(2)	4502(1)	49(1)
C(3)	-3021(3)	474(2)	3720(1)	49(1)
C(4)	-837(3)	-54(2)	3378(1)	48(1)
C(5)	-1433(3)	-581(2)	2561(1)	52(1)
C(6)	765(3)	-1019(2)	2189(1)	52(1)
C(7)	168(3)	-1473(2)	1364(1)	55(1)
C(8)	2354(3)	-1889(2)	975(1)	54(1)
C(9)	1719(3)	-2348(2)	153(1)	57(1)
C(10)	3881(3)	-2766(2)	-247(1)	55(1)
C(11)	3169(3)	-3262(2)	-1062(1)	56(1)
C(12)	5302(3)	-3667(2)	-1482(1)	53(1)
C(13)	4500(3)	-4190(2)	-2290(1)	54(1)
C(14)	6612(3)	-4531(2)	-2741(1)	51(1)
C(15)	5675(3)	-5009(2)	-3550(1)	52(1)
C(16)	7741(3)	-5252(2)	-4045(1)	55(1)

1-Iodo-3-tosyloxy-propan-2-ane (C_{10} H_{11} O_4 I S) [5.5.3]

	x	y	z	U(eq)
I(1)	2250(1)	7130(1)	4733(1)	27(1)
S(1)	59(1)	5730(1)	7924(1)	17(1)
C(1)	3157(4)	9054(2)	5826(1)	23(1)
C(2)	1776(3)	8646(2)	6647(1)	19(1)
C(3)	3429(3)	7638(2)	7285(1)	21(1)
C(4)	2223(3)	4181(2)	8223(1)	17(1)
C(5)	3518(4)	3264(2)	7544(1)	20(1)
C(6)	5351(4)	2117(2)	7809(1)	22(1)
C(7)	5885(3)	1864(2)	8734(1)	21(1)
C(8)	4535(4)	2793(2)	9397(1)	22(1)
C(9)	2724(4)	3956(2)	9149(1)	21(1)
C(10)	7881(4)	628(2)	9029(2)	31(1)
O(1)	-461(3)	9111(2)	6795(1)	26(1)
O(2)	1978(2)	7291(1)	8052(1)	21(1)
O(3)	-1786(3)	6062(2)	8613(1)	25(1)
O(4)	-901(3)	5425(2)	6969(1)	23(1)

1β, 9β-Diacetyl-7α-chloro-cis-hydrindane (C_{13} H_{19} O_2 Cl) [5.5.4]

NB. No scale factors used in this table

	x	y	z	U(eq)
Cl	0.0036(1)	0.2162(1)	0.6357(1)	0.062(1)
C(1)	-0.1597(2)	0.3335(2)	0.8047(2)	0.042(1)
C(2)	-0.2194(2)	0.4239(3)	0.8829(3)	0.059(1)
C(3)	-0.1712(2)	0.5598(3)	0.8692(2)	0.054(1)
C(4)	-0.2416(2)	0.5745(2)	0.6515(2)	0.052(1)
C(5)	-0.2099(2)	0.5402(3)	0.5367(2)	0.062(1)
C(6)	-0.1637(2)	0.4015(3)	0.5377(2)	0.055(1)
C(7)	-0.0500(2)	0.3833(2)	0.6358(2)	0.047(1)
C(8)	-0.0707(2)	0.4208(2)	0.7543(2)	0.038(1)
C(9)	-0.1272(2)	0.5581(2)	0.7540(2)	0.043(1)
C(10)	-0.0296(2)	0.6636(3)	0.7530(2)	0.056(1)
C(11)	-0.0695(3)	0.8018(3)	0.7604(3)	0.083(1)
C(12)	-0.0968(2)	0.2201(2)	0.8758(2)	0.048(1)
C(13)	-0.1772(3)	0.1041(3)	0.8798(3)	0.077(1)
O(1)	0.0787(2)	0.6379(2)	0.7501(2)	0.084(1)
O(2)	0.0109(2)	0.2257(2)	0.9303(2)	0.074(1)

(R,R) - 1,4 - bis (2'-chloro-1'-hydoxyethyl) benzene (C_{10} H_{12} O_2 Cl_2) [5.5.5]

	x	y	z	U(eq)
Cl(1)	14998(2)	-878(1)	-3471(1)	57(1)
Cl(2)	4963(2)	702(1)	444(1)	50(1)
O(1)	16220(5)	-2100(3)	-2333(1)	52(1)
O(2)	6350(5)	-2057(2)	-261(1)	41(1)
C(1)	12854(7)	-959(4)	-2875(1)	46(1)
C(2)	14472(6)	-876(3)	-2342(1)	36(1)
C(3)	12506(6)	-832(3)	-1850(1)	34(1)
C(4)	11301(7)	459(3)	-1694(1)	40(1)
C(5)	9399(7)	505(3)	-1265(1)	40(1)
C(6)	8692(6)	-735(3)	-975(1)	32(1)
C(7)	9916(7)	-2028(3)	-1130(1)	35(1)
C(8)	11799(7)	-2077(3)	-1565(1)	38(1)
C(9)	6605(6)	-670(3)	-503(1)	32(1)
C(10)	7560(8)	411(4)	-74(1)	48(1)

Monomer unit of $C_{17}H_{14}$ [5.5.6]

NB No scale factors used in this Table

	x	y	z	U(eq)
C(01)	0.0921(1)	-0.3432(2)	0.0781(1)	0.037(1)
C(02)	0.1786(1)	-0.3109(2)	0.0931(2)	0.041(1)
C(03)	0.2068(1)	-0.2793(2)	0.0099(2)	0.041(1)
C(04)	0.1405(1)	-0.2902(2)	-0.0627(1)	0.036(1)
C(05)	0.0837(1)	-0.1662(2)	-0.0447(1)	0.030(1)
C(06)	0.0499(1)	-0.2033(2)	0.0543(1)	0.030(1)
C(07)	0.0922(1)	-0.4078(2)	-0.0208(1)	0.041(1)
C(08)	0.0696(1)	-0.0798(2)	0.1158(1)	0.029(1)
C(09)	0.1097(1)	0.0184(2)	0.0618(1)	0.027(1)
C(010)	0.1204(1)	-0.0253(2)	-0.0317(1)	0.028(1)
C(011)	0.1589(1)	0.0605(2)	-0.0914(1)	0.034(1)
C(012)	0.1862(1)	0.1903(2)	-0.0575(1)	0.035(1)
C(013)	0.1752(1)	0.2335(2)	0.0333(1)	0.033(1)
C(014)	0.1357(1)	0.1472(2)	0.0970(1)	0.028(1)
C(015)	0.1191(1)	0.1756(2)	0.1925(1)	0.035(1)
C(016)	0.0807(1)	0.0794(2)	0.2455(1)	0.038(1)
C(017)	0.0552(1)	-0.0496(2)	0.2080(1)	0.035(1)
C(1)	0.3599(1)	-0.6017(2)	0.2998(1)	0.031(1)
C(2)	0.3793(1)	-0.4545(2)	0.2732(1)	0.031(1)
C(3)	0.3889(1)	-0.4511(2)	0.1807(1)	0.033(1)
C(4)	0.3770(1)	-0.5965(2)	0.1436(1)	0.033(1)
C(5)	0.4491(1)	-0.6837(2)	0.1800(1)	0.030(1)
C(6)	0.4361(1)	-0.6899(2)	0.2893(1)	0.028(1)
C(7)	0.3139(1)	-0.6475(2)	0.2103(1)	0.036(1)
C(8)	0.5111(1)	-0.6343(2)	0.3339(1)	0.026(1)
C(9)	0.5630(1)	-0.5988(2)	0.2627(1)	0.027(1)
C(10)	0.5302(1)	-0.6226(2)	0.1712(1)	0.028(1)
C(11)	0.5750(1)	-0.5922(2)	0.0952(1)	0.034(1)
C(12)	0.6525(1)	-0.5417(2)	0.1116(1)	0.036(1)
C(13)	0.6846(1)	-0.5220(2)	0.2005(1)	0.035(1)
C(14)	0.6394(1)	-0.5487(2)	0.2804(1)	0.030(1)
C(15)	0.6630(1)	-0.5333(2)	0.3766(1)	0.034(1)
C(16)	0.6121(1)	-0.5662(2)	0.4461(1)	0.034(1)
C(17)	0.5356(1)	-0.6178(2)	0.4264(1)	0.031(1)

1,4 dibenzyl - 1,2,4,5 - tetraazacyclohexane ($C_{16}H_{20}N_4$) [5.5.8]

	x	y	z	U(eq)
N(2)	5003(1)	1563(3)	3768(2)	39(1)
N(1)	5812(1)	908(3)	4730(2)	36(1)
C(1)	9098(2)	-246(7)	6373(3)	76(1)
C(2)	8815(2)	-1829(6)	5209(4)	79(1)
C(3)	7974(2)	-1496(5)	4406(3)	61(1)
C(5)	7715(1)	2010(5)	5957(3)	54(1)
C(6)	8547(2)	1669(6)	6756(3)	66(1)
C(4)	7418(1)	438(4)	4771(2)	43(1)
C(7)	6519(1)	892(5)	3873(2)	48(1)
C(8)	4296(1)	1663(4)	4601(2)	39(1)

1,5 - di (2'-chloroacetoxy) - 3,3 - dimethyl - 2,4 - diphenyl - 3 - silapentane ($C_{22}H_{26}O_4Cl_2Si$) [5.5.7]

	x	y	z	U(eq)
Cl(1A)	-458(5)	13899(5)	1570(6)	78(1)
Cl(1B)	-981(5)	13901(5)	2020(4)	77(1)
Cl(1C)	-602(6)	14228(6)	772(6)	82(1)
Cl(1D)	-1055(7)	14539(7)	296(6)	102(2)
Cl(2)	7390(2)	9275(2)	344(1)	105(1)
Si(1)	1712(1)	6499(1)	3478(1)	46(1)
O(1)	-1257(3)	11812(3)	271(3)	80(1)
O(2)	-490(2)	10939(2)	2181(2)	55(1)
O(4)	5109(2)	7241(3)	1737(2)	66(1)
C(1)	-1675(5)	13447(5)	1368(6)	100(2)
C(2)	-1107(3)	11984(4)	1180(4)	58(1)
C(3)	117(3)	9454(3)	2129(3)	50(1)
C(4)	510(3)	8454(3)	3412(3)	47(1)
C(5)	-720(3)	8542(3)	4310(3)	46(1)
C(6)	-749(3)	8711(4)	5449(3)	55(1)
C(7)	-1870(4)	8796(4)	6277(4)	68(1)
C(8)	-2977(4)	8719(4)	5987(4)	74(1)
C(9)	-2965(3)	8522(4)	4875(4)	68(1)
C(10)	-1844(3)	8434(3)	4036(3)	56(1)
C(11)	885(4)	5728(4)	2885(4)	70(1)
C(12)	2168(4)	5475(4)	5123(3)	71(1)
C(13)	3323(3)	6449(3)	2453(3)	50(1)
C(14)	4273(3)	4905(3)	2454(3)	46(1)
C(15)	4273(3)	4472(4)	1469(3)	59(1)
C(16)	5081(4)	3066(5)	1450(4)	72(1)
C(17)	5893(4)	2066(4)	2402(4)	73(1)
C(18)	5921(4)	2445(4)	3391(4)	76(1)
C(19)	5115(3)	3863(4)	3417(3)	63(1)
C(20)	4022(4)	7200(4)	2708(4)	65(1)
C(21)	5652(4)	8116(4)	1656(4)	67(1)
O(3B)	5081(6)	9148(7)	2070(6)	74(2)
C(22A)	6563(9)	8192(9)	432(8)	57(2)
O(3A)	5537(6)	8588(7)	2488(6)	74(2)
C(22B)	6873(9)	7872(10)	823(9)	67(2)

Appendix Four

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
May 1992	Modern Perspectives in Inorganic Crystal Chemistry, Erice, Sicily
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 17	Dr. R. W. Kemmitt, 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. Orsted 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