THE MOLECULAR GEOMETRY OF d^8 FIVE-COORDINATION: AN ANALYSIS OF STATIC DEFORMATIONS

A thesis submitted to the
UNIVERSITY OF CAPE TOWN
in fulfilment of the requirement for the degree of
DOCTOR OF PHILOSOPHY

by
Thomas Paul Edwin Auf der Heyde, M.Sc. (Cape Town)

Department of Physical Chemistry
University of Cape Town
Rondebosch
7700
South Africa

March 1988

Modern physics thus pictures matter not at all as passive and inert but as being in a continuous dancing and vibrating motion whose rythmic patterns are determined by the molecular, atomic, and nuclear configurations. We have come to realize that there are no static structures in nature. There is stability, but this stability is one of dynamic balance, and the further we penetrate into matter the more we need to understand its dynamic nature to understand its patterns.

Fritjof Capra*

* Capra, F., The Turning Point: Science, Society and the Rising Culture, p.79, Fontana, London 1985.

FOREWORD

This dissertation is presented in five chapters:

Chapter 1: Introduction

Chapter 2: An Outline of the Relevant Statistical Techniques

Chapter 3: Data Search, Description of Conformation and Data Preparation

Chapter 4: Data Analysis and Discussion

Chapter 5: Summary and Conclusion

At the beginning of each chapter a detailed table of contents is presented. References cited are listed at the end of each chapter, and relevant appendices are also attached.

ACKNOWLEDGEMENTS

I wish to thank Professors L.R. Nassimbeni and H.B. Bürgi for the unrestricted use of their laboratory facilities, their patience and guidance through this study, and, above all, for the ready help and friendship offered to me. Sincere thanks also go to Leif Norskov and Gerhard Klebe, two other members of the *Bern Team*, for helpful discussions and for their interest and concern for my work, and the conditions under which it was performed. GK also kindly allowed me the use of one of his programs (see Chapter 3).

The University of the Western Cape gratiously granted me study leave and financial assistance. Through a travel grant Johannesburg Consolidated Investment Company Ltd. enabled my visit to Berne, where part of this work was done.

ABSTRACT

The geometries of the ML_5 fragments in 196 five-coordinate metal complexes (M = Ni(II), Pd(II), Pt(II), Rh(I), Ir(I); L = coordinated ligand atom) have been studied using multi-variate statistical techniques. For each molecular fragment the geometry has been precisely described by two sets of twelve non-redundant symmetry coordinates. These sets correspond to the two most common idealised five-coordinate conformations — the trigonal bipyramid (TBP) and the square based (or rectangular) pyramid (SQP). Each observed ML_5 fragment is considered to be represented by a point in a twelve-dimensional space spanned by the symmetry coordinates. There are two such spaces, depending on whether the observed structures are related to a TBP (T-space) or to a SQP (S-space). Cluster analysis reveals that the points aggregate around a D_{3h} TBP and a C_{2v} distorted SQP in T-space, and around a C_{2v} distorted TBP, a C_{4v} "elevated" SQP (eSQP) and a C_{4v} "flattened" SQP (fSQP) in S-space. It is shown that the SQP cluster in T-space is divided into the fSQP and the eSQP cluster in S-space and that, consequently, the results from the two data spaces are identical. Structural characteristics for the three archetypal conformations are presented and analysed. Factor analysis reveals the TBP to manifest static deformations mirroring distortions along an S_N2 coordinate, a Berry intramolecular exchange coordinate and one which indicates the preservation of a constant amount of bond order at the metal (the "glue" coordinate). The eSQP and fSQP distort along the glue coordinate and along a coordinate delineating ~~the pyramidalisation of the pyramid~~ deli pyramidalization. The sum of these is akin to an addition/elimination coordinate for the addition of a fifth ligand to square planar ML_4. Furthermore, the eSQP also distorts according to the Berry coordinate. Parallels are outlined between the static deformation and the solution dynamics of five-coordinate species, and these are discussed in the light of the structure correlation hypothesis.

CHAPTER 1
INTRODUCTION

1.1 Reaction Pathways and Static Deformations

Investigations into the dynamics of reacting chemical systems are always plagued by the lack of a suitable technique for the direct observation of the reacting species; the precise geometrical changes taking place along the reaction pathway are usually elusive. In general, the static molecular conformation of stable product and educt molecules, and in some cases that of isolable intermediates, may be elucidated with the aid of tools such as X-ray crystallography. However, descriptions of the mechanistic pathways connecting these molecules with each other are usually based either on *indirect* evidence or on the results of theoretical calculations.

The Born-Oppenheimer approximation to the Schrödinger equations assumes the separation of nuclear and electronic motions, and consequently enables the calculation of the potential energy of a polyatomic system as a function of the static arrangement of the atomic nuclei. If this calculation is repeated for a whole range of stationary nuclear arrangements a potential energy surface results. This surface is "an analytic function of the internal coordinates of a system that gives the potential energy as a function of geometry." [1] Consequently, where the system is defined by more than two coordinates the potential energy surface becomes defined in multi-dimensional space, i.e. it becomes a *hyper*surface. Stable arrangements of atoms, corresponding to product, educt or intermediate molecules, will tend to aggregate in "dips" or "wells" on the surface, while curves connecting such minima along energy "valleys" or over "passes" will represent minimum energy paths, or reaction pathways (reaction coordinates). It is these curves which detail the synchronous changes in bond distances and angles as educt molecules are transformed to product molecules.

Truhlar, Steckler and Gordon, from whom we have borrowed the above definition, have recently reviewed the use of the potential energy surface in the study of polyatomic reaction dynamics. [1] Their compilation, dealing with multi-atom systems of up to ten atoms, shows clearly, however, that the *ab initio* calculation of potential energy surfaces of such systems is computationally prohibitive. In the case of systems containing atoms bigger than ~~flourine~~ fluorine satisfactory calculations have yet to be performed; in the near term it seems extremely unlikely that potential energy surfaces can be used for the routine examination of reaction pathways for

real chemical systems.

Although direct *observation* of the reaction pathway does not seem feasible, its *visualisation* at least does. According to the structure correlation hypotheses, [2] the *gradual* distortion, or *static deformation*, which the 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.

The rationale behind the hypothesis is enticingly simple and insightful. Crystalline structures represent stable atomic arrangements — their representative points on the potential energy surface will consequently lie in, or close to, a local potential minimum, either a "well", a "dip" or a point in a "valley". Crystal structures containing closely related molecular fragments will be represented by points variously displaced from the potential minima, *along pathways of minimum energy*. Arranged in the correct sequence, then, these fragments can be assumed to map such pathways, their static deformation mirroring that expected along the coordinate. This idea has found expression in the structure correlation hypothesis: *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.* [2c]

1.2 The Structure Correlation Hypothesis in Action

Since it was originally outlined by Bürgi [2a] the structure correlation method has been used to map reaction paths for a variety of different reactions, including substitution and addition reactions, dissociations, conformational interconversions and isomerisations.* For example, an investigation of MX_4 and YMX_3 fragments of T_d or C_{3v} symmetry with M = S,P, Al,Sn and Y,X = O, Cl, Br, Ph *etc.*, has revealed deformations analogous to those expected for the S_N1 reaction, i.e. a heterolytic weakening and ultimate fission of the axial bond to yield planar MX_3 species. [2c] Figure 1.2.1 shows how the fragments were analysed, and gives a

* We will not give a full review of these here. For this, the reader is referred to references 2(b), 3 and 4.

scatterplot of the axial (r_2) and the equatorial (r_1) bond lengths, corresponding to the bond undergoing rupture and that within the MX_3 species, respectively, versus the average angle (θ) between the axial and the equatorial bonds. It shows that a lengthening of r_2 is accompanied by a decrease in θ (flattening of the YMX_3 or MX_4 tetrahedron), and a shortening of r_1 , indicative of an increase in bonding electron density in this bond at the expense of r_2 .

More recently, Liebau has proposed atomic scale mechanisms for condensation and decondensation reactions of silicates in aqeous solution, based on correlations obtained from a stereochemical analysis of the crystal structures of 15 closely related penta-coordinate silicon componds. [5] In his study Liebau closely follows the methods used by Bürgi in his analysis of penta-coordinate cadmium, [2a] and those used by Britton and Dunitz for tin compounds, [6] in mapping what essentially is an S_N2 pathway for tetra-coordinate silicon.

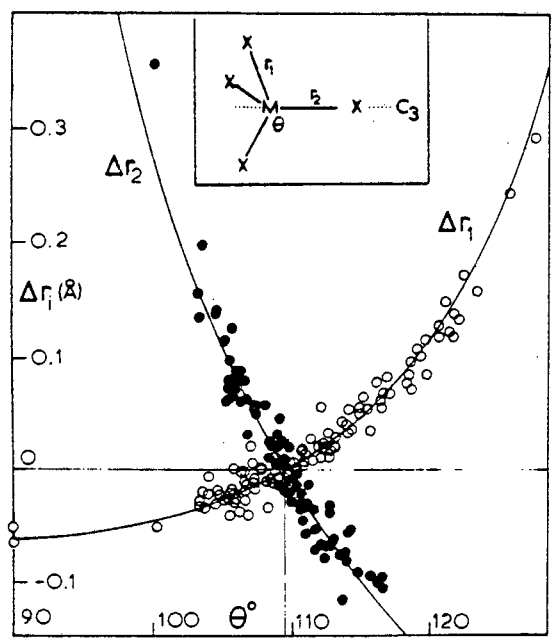


Figure 1.2.1 Diagram showing correlations between equatorial (r_1) and axial (r_2) bonds and average angle between them (θ) , for T_d or C_{3v} molecular fragments. Taken from reference 2(c).

In many cases the reaction coordinates obtained by the structure correlation method closely parallel those from theoretical calculations, both *ab initio* and semi-empirical. Nucleophilic addition of an amine to a carbonyl group is a case in point. An investigation of a series of alkaloid and related structures, all containing a tertiary amino and a carbonyl group, revealed a correlation between the $R_3N.....CR_2O$ distance (d) and the pyramidalisation ($\triangle$) of the carbonyl group. [7] This is illustrated in Figure 1.2.2 together with the results of an *ab initio* calculation of the reaction coordinate for the simple model system

$$CH_2 = O + H^- \rightarrow CH_3O^-$$

corresponding to the (analogous) addition of hydride anion to formaldehyde. [8] The similarity is striking, with what difference there is being ascribable to the differing microenvironments in the two cases.

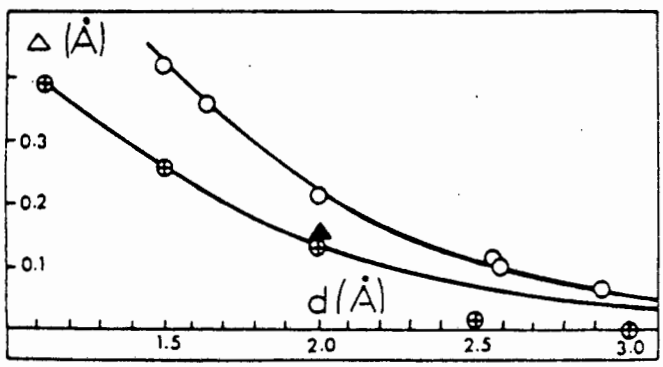


Figure 1.2.2 Plot of the pyramidalisation ($\triangle$) versus the nucleophile-carbonyl distance (d) for experimental data (top curve) and theoretical results (bottom curve). The black triangle is *calculated* for $NH_3 + H_2CO$. Taken from reference 8.

We have mapped the intramolecular ligand exchange pathway for pentacoordinate zinc from a stereo-chemical analysis of 33 different crystal structures. [9] A scatterplot of the data onto a plane defined by two of the internal coordinates describing the coordination sphere geometry around the zinc atom, agrees remarkably with a projection of the calculated potential energy surface onto the analogous

plane. Figure 1.2.3 illustrates the scatter plot, as well as the potential energy surface calculated by Favas and Kepert [10] for the system $[M(\text{unidentate})_5]$.

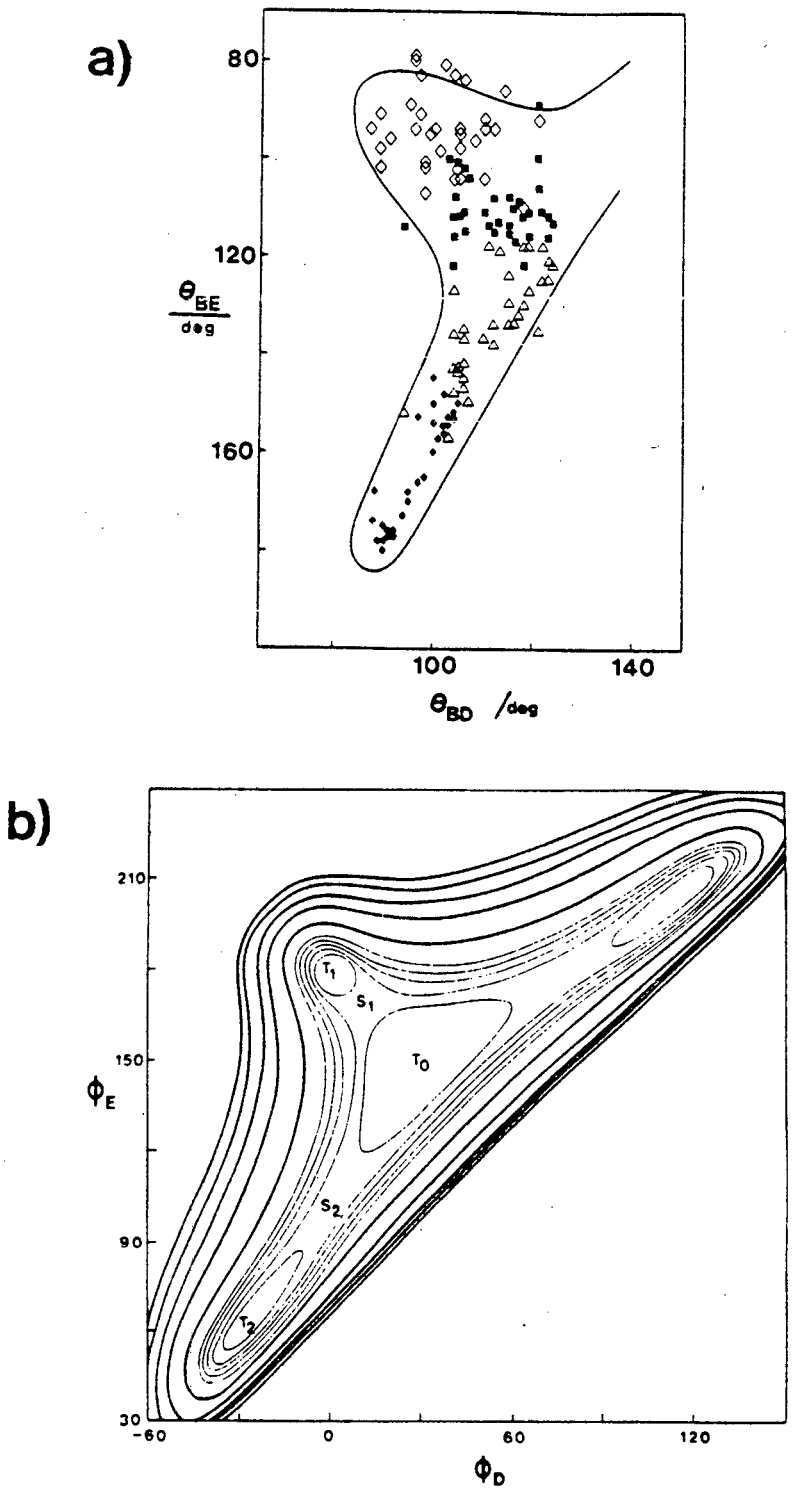


Figure 1.2.3(a) Scatterplot of data for 33 penta-coordinate zinc complexes. Taken from reference 9. (b) Potential energy surface calculated for a point-charge model of $[M(\text{unidentate})_5]$. The surface is symmetrical about the T_1, S_1, T_0 axis. Taken from reference 10.

A similar parallel between the distribution of data points in parameter space and the topography of the potential energy surface has been demonstrated recently by Klebe for intramolecular ligand exchange at penta-coordinate phosphorus. [11] This is graphically illustrated in Figure 1.2.4(a) and (b). Gilli and co-workers have mapped *cis-trans* isomerisation via rotation about the C–N bond in amides, anilines and related compounds. [12] They have expressed the coincidence of the crystal structures with low lying regions on the energy surface in a rather elegant graph, shown in Figure 1.2.4(c).

1.3 The Need for Multi-Variate Analysis

One factor which all the earlier studies employing the structure correlation principle had in common, was their reliance on two-dimensional scatterplots. Although these can often reveal highly informative trends in the data, a great danger of oversimplification of the problem is associated with them. All of the molecular fragments studied require considerably more than just two geometric coordinates for their complete definition.* An overreliance on scatterplots would consequently have the effect of collapsing a multi-dimensional problem onto two dimensions, with the loss of all information not associated with these two. In many cases this may be an appropriate simplification, although it cannot be so in general. Figure 1.3.1 shows very simply that a correlation in two dimensions does not necessarily imply the existence of one in three (or more) dimensions.

In the light of this more sophisticated analytical techniques are required which must be capable of revealing correlations in multi-dimensional space *simultaneously*, and not simply in the form of a number of bi-variate correlations. Moreover, such techniques should also enable the ready visualisation of the data distribution in parameter space. Multi-variate statistical techniques have been applied extensively in the social sciences, and to a lesser extent in analytical chemistry, and their mathematics are firmly established. Factor analysis extracts linear combinations of the original variables which describe, in turn, the greatest amount of sample variance, the second greatest, and so on, from the correlation or covariance matrix of the data. Cluster analysis groups together similar points in the data space, thereby

* The number of coordinates necessary is, of course, (3N-6), where N is the number of atoms constituting the fragment of interest.

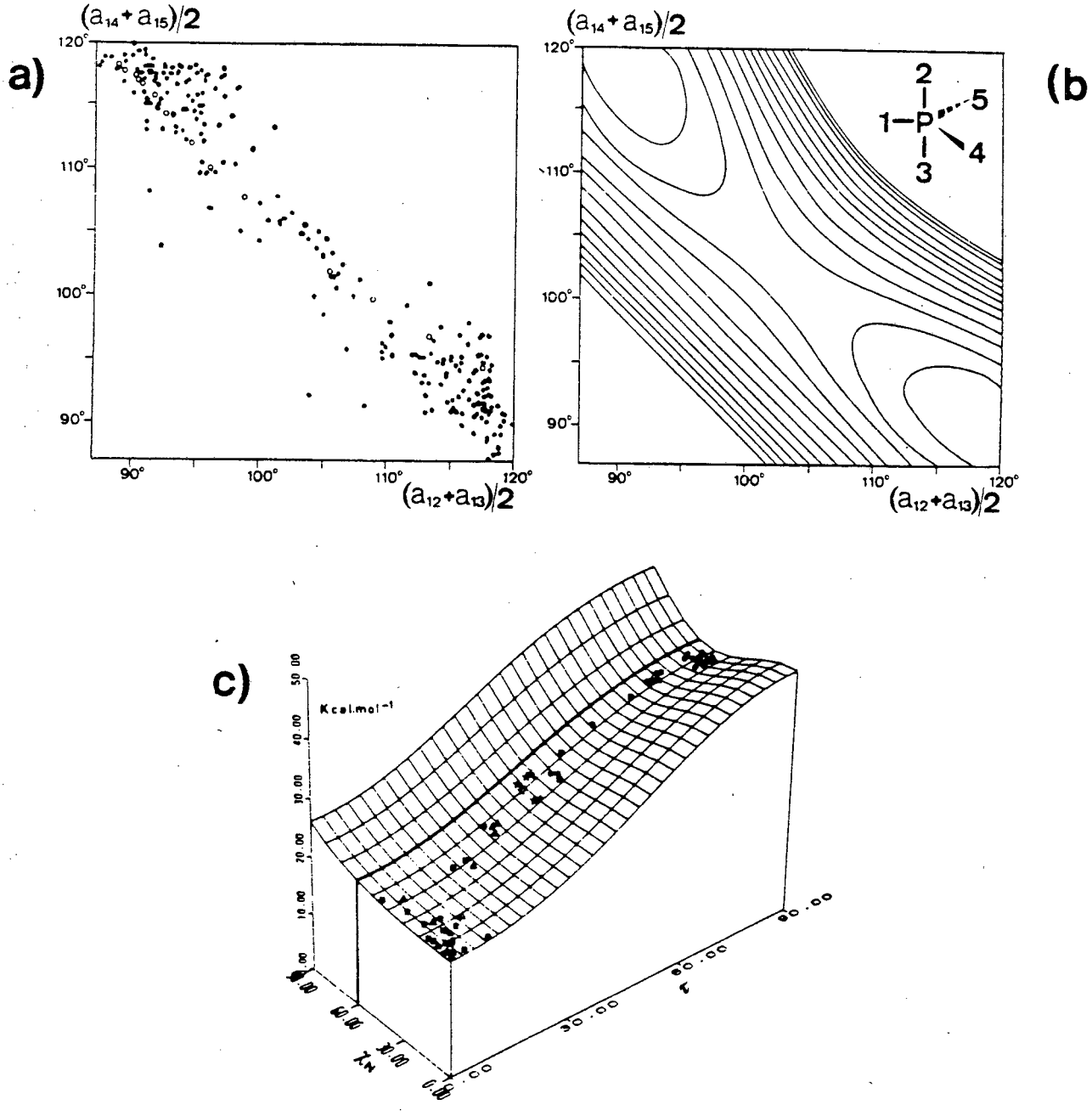


Figure 1.2.4 (a) Scatterplot of experimental data for penta-coordinate phosphorus. (b) Potential energy surface computed by MNDO for PF_5 ; contour interval is 1.kcal. (a) and (b) taken from reference 11. (c) Three dimensional representation of potential energy surface for amides and related structures, with experimental data projected onto it. The surface was calculated by molecular mechanics methods. Taken from reference 12.

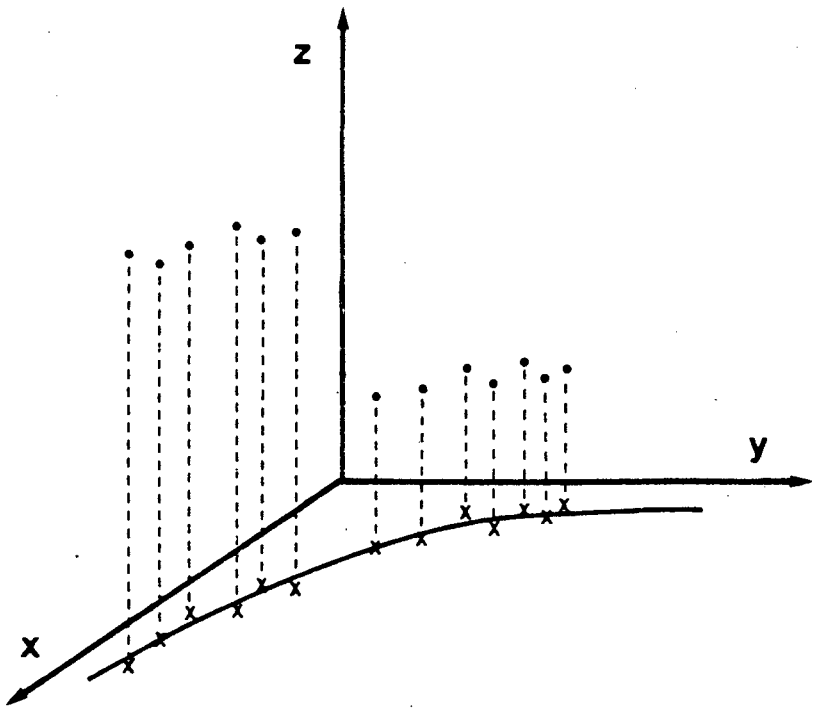


Figure 1.3.1 Diagram showing that a correlation in two dimensions (in the $x-y$ plane) does not necessarily imply one in more dimensions ($x-y-z$ space).

---

yielding clusters or clouds of data points which are often useful in classifying the data. Both techniques also simplify the visual interpretation of trends in the data.

Murray-Rust initially introduced the idea of factor extraction to the comparative analysis of molecular geometries in 1978, [13] and he first applied it to the geometry of the $\beta-1'-$aminofuranoside fragment. [14] Cluster analysis, although also indicated at the time, was first applied to an analysis of the type described here in 1985, when Nørskov-Lauritsen and Bürgi employed it for a conformational analysis of the $M(PPh_3)_2$ fragment. [15] Their study revealed a possible pathway for the conformational interconversion of this moeity, involving a "gearing motion of the two PC_3 fragments alternating with stepwise inversions of the helicities of the PPh_3 – propellers ." The data, consisting of 62 $XYM(PPh_3)_2$ fragment structures, were shown to aggregate into three essential clusters, each somewhat displaced along the reaction pathway. An analysis of the cluster *centrotypes* — the average conformation of the fragments classified within a given cluster — revealed detailed changes in torsion angles of the fragment, delineating the conformational inter-conversion.

1.4 The Aim and Scope of this Study

We will examine the molecular geometries of five-coordinate complexes containing the metals nickel, palladium, platinum, rhodium and iridium, all with a d^8 electron configuration. Cluster analysis will be employed to determine the average conformations which the complexes adopt, such as a square based- or rectangular pyramid, or a trigonal bipyramid. Their static deformations will be probed by factor analysis in order to establish correlations mapping distortion coordinates for these cluster averages.

The cluster centrotypes and their characteristic distortions will then be investigated in the light of the structure correlation hypothesis to establish the relationship between the static deformations mapped by the five-coordinate molecular fragments, and reaction coordinates of five-coordinate complexes. During the analysis comparisons will be made to the results of our earlier study of the dynamic stereochemistry of nickel, [4] which was based largely on the interpretation of scatterplots.

REFERENCES

1. Truhlar, D.G., Steckler, R., Gordon, M.S. *Chem Rev.* 1987, **87**, 217-236.

2.(a) Bürgi, H.B. *Inorg. Chem.* 1973, **13**, 2321-2325.

2.(b) Bürgi, H.B., Dunitz, J.D. *Acc. Chem. Res.* 1983, **16**, 153-161.

2.(c) Murray-Rust, P., Bürgi, H.B., Dunitz, J.D. *J. Am. Chem. Soc.* 1975, **97**, 921-922.

3. Auf der Heyde, T.P.E. *M.Sc. Dissertation*, Univ. of Cape Town, 1984.

4. Auf der Heyde, T.P.E., Nassimbeni, L.R. *Inorg. Chem.* 1984, **23**, 4525-4532.

5. Liebau, F. *Inorg. Chim. Acta* 1984, **89**, 1-7.

6. Britton, D., Dunitz, J.D. *J. Am. Chem. Soc.* 1981, **103**, 2971-2979.

7. Bürgi, H.B., Dunitz, J.D., Schefter, E. *J. Am. Chem. Soc.* 1973, **95**, 5065-5067.

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

9. Auf der Heyde, T.P.E., Nassimbeni, L.R. *Acta Cryst.* 1984, **B40**, 582-590.

10. Favas, M.C., Kepert, D.L. *Prog. Inorg. Chem.* 1980, **27**, 325-364.

11. Klebe, G. *J. Organomet. Chem.* 1987, **332**, 35-46.

12. Gilli, G., Bertolasi, V., Bellucci, F., Ferretti, V. *J. Am. Chem. Soc.* 1986, **108**, 2420.

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

14. Murray-Rust, P., Motherwell, S. *Acta Cryst.* 1978, **B34**, 2534-2546.

15. Nørskov-Lauritsen, L., Bürgi, H.B. *J. Comput. Chem.* 1985, **6**, 216-228.

CHAPTER 2

AN OUTLINE OF THE RELEVANT STATISTICAL TECHNIQUES

2.1 The Data Matrix

To start with, the data consisting of n readings or measurements on m objects are considered to constitute an $m \times n$ data matrix $\mathbf{D}$, the elements of which are x_{ij} (where x_{ij} is the value of the j-th variable for the i th object), and the columns of which list the variables $(1, 2, \ldots, j, \ldots, n)$.

$$\mathbf{D} = \text{objects} \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ i \\ \vdots \\ m \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & \ldots & j & \ldots & n \end{array} \\ \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1j} & \ldots & x_{1n} \\ x_{21} & & & & & \\ \vdots & & & \vdots & & \vdots \\ x_{i1} & \ldots & \ldots & x_{ij} & & \\ \vdots & & & \vdots & & \vdots \\ x_{m1} & \ldots & \ldots & \ldots & \ldots & x_{mn} \end{pmatrix} \end{array}$$

Hence object i can be thought of as being represented by a row vector $\bar{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ which is called the *pattern vector* for object i. Similarly, variable j can be represented by a column vector

$$\bar{x}_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{mj} \end{pmatrix}$$

in this case called the pattern vector for variable j, since $\bar{x}_j$ traces out the pattern of the various values of j for all the m objects.

The purpose of cluster analysis, in our case, is to *cluster* the m objects into groups according to their characteristic pattern vectors $\bar{x}_i$. It is also possible, however, to cluster together the variables according to their characteristic pattern.

2.2 Standardization and Scaling

Since the description of molecular geometry often involves a combination of interatomic distance and angle measurements, it follows that the data will sometimes need to be scaled or standardized in order that the effect of gross size or range of variation does not bias the analysis. Thus, for example, changes of the order of 1 Angström in interatomic distances might be relatively more important than changes of the order of 10 degrees in interatomic angles, though without appropriate scaling

their relative importance would be hidden. Similarily, two angle variables might have identical (or nearly so) means of 120^0 , say, while their ranges vary from 90^0 to 150^0 , and from 110^0 to 130^0 respectively. Obviously changes of 10^0 in the latter variable will be relatively more important than similar changes in the former variable, though again without appropriate standardization this effect would be lost on the analysis.

Two commonly used techniques for scaling and standardization are *ranging* or *range scaling* and the *z transformation* [1] . Ranging involves transforming the data matrix according to

$$y_{ij} = \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}}$$

so that, for all i and j , $0 \leq y_{ij} \leq 1$.

A statistically more sound procedure, however, is the z transform leading to a matrix $\mathbf{Z}$ composed of z scores

$$z_{ij} = \frac{x_{ij} - \mu_j}{s_j} \tag{2.2.1}$$

where $$\mu_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij} \tag{2.2.2}$$

and $$s_j^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_{ij} - \mu_j)^2 \tag{2.2.3}$$

Here μ_j is the *mean* of the sample while s_j^2 is the sample *variance*, or the square of the *standard deviation* (σ_j) of the sample. An important property (which will be illustrated later) of the z_{ij} values is that their covariance matrix is the same as the correlation matrix of the x_{ij} 's.

Standardization via the z transform exhibits some contradictory properties, though. On the one hand, its effect is to equalize the influence of variables with small variation and those with large variation, while on the other it excludes variables with no significant variation. This means that, starting from the most variable character, increasing weight is attached to the absolute degree of variation when the variability decreases, until suddenly no weight is attached at all to those characters whose variation is not high enough! This point will also be demonstrated later.

2.3 Measures of Similarity

In order to group together observations or molecular fragments some criterion of "similarity" will obviously need to be developed. Each object in the n dimensional space will need to be compared with every other object in order to group together into the same cluster those which are "similar", while assigning dissimilar ones to different clusters. Two such measures will be considered here.

(a) Covariance and Correlation

If the $m \times n$ data matrix $\mathbf{D}$ is pre-multiplied by its $n \times m$ transpose $\mathbf{D}^T$, after subtracting the mean of each variable, a $n \times n$ square matrix is obtained. After dividing its elements by the number of objects minus 1 it is called the *covariance* matrix $\mathbf{C}$.

An element of this matrix is given by

$$c_{k\ell} = \frac{1}{m-1} \sum_{i=1}^{m} (x_{ik} - \mu_k)(x_{i\ell} - \mu_\ell) \qquad (2.3.1)$$

$$\text{where} \qquad \mu_k = \frac{1}{m} \sum_{i=1}^{m} x_{ik}.$$

The matrix can be written as

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & & & \\ \vdots & & & \vdots \\ c_{1n} & \dots & \dots & c_{nn} \end{pmatrix}$$

and it should be noted, firstly, that the diagonal elements of this matrix are equal to the variances of the n variables and, secondly, that the matrix is symmetric about the diagonal. Moreover, the sum of the diagonal elements, or the *trace* of $\mathbf{C}$, is equal to the total variance in the data set.

$c_{k\ell}$ is large and positive when for most objects the values of variables k and ℓ deviate from the mean in the same direction. The covariance $c_{k\ell}$ of the two variables is therefore a measure of their association. This covariance or correlation between the two variables is often also expressed by the *correlation coefficient* $r_{k\ell}$ where

$$r_{k\ell} = \frac{c_{k\ell}}{s_k \cdot s_\ell} \qquad (2.3.2)$$

s_k and s_ℓ are the variances of variables k and ℓ respectively, and $r_{k\ell}$ is hence a standardized covariance which lies between -1 and +1. For each element $c_{k\ell}$ of the covariance matrix a correlation coefficient can be derived, and the covariance matrix $\mathbf{C}$ may consequently be transformed into a correlation matrix $\mathbf{R}$.

The covariance and/or correlation matrices almost invariably represent the basis of departure for the subsequent factor and cluster analyses.

(b) Distance measurements

In some cases, especially in cluster analysis, it may prove convenient to express the similarity of two observations or molecular fragments in terms of the *distance* between the two representative points in the n dimensional parameter space.

Thus, the *Euclidian distance* $d_{k\ell}$ between two points k and ℓ in n dimensional space is given by

$$d_{k\ell} = \left(\sum_{j=1}^{n} (x_{kj} - x_{\ell j})^2 \right)^{\frac{1}{2}} \qquad (2.3.3)$$

and can be seen to be easily derived from the two dimensional case.

Massart and Kaufman [1] have shown, however, that correlation between variables in the n dimensional space results in a distortion of the relation between the representative points, so that the Euclidean distance is an insufficient measure of similarity. The *Mahalanobis* distance, on the other hand, tends to compensate for this effect of correlation by incorporating the inverse of the covariance matrix $(\mathbf{C}^{-1})$ into the distance equation

$$d^2_{k\ell} = (\overline{x_k - x_\ell}) \cdot \mathbf{C}^{-1} \cdot (\overline{x_k - x_\ell})^T$$

where $(\overline{x_k - x_\ell})$ is the difference vector between the pattern vectors $\bar{x}_k$ and $\bar{x}_\ell$ for objects k and ℓ respectively, while $(\overline{x_k - x_\ell})^T$ is its transpose. Most cluster analysis computer packages offer the option of using this distance measure rather than the straightforward Euclidian distance.

2.4 Factor Analysis

(a) The Philosophical Basis

Essentially factor analysis involves the transformation of the n orthogonal axes (representing the variables) which span the parameter space into n new axes

(representing linear combinations of the variables), such that these new axes lie along the directions of maximum variance. This basic concept can easily be visualized with the help of a two dimensional example. Consider the distribution depicted in Figure 2.4.1(a).

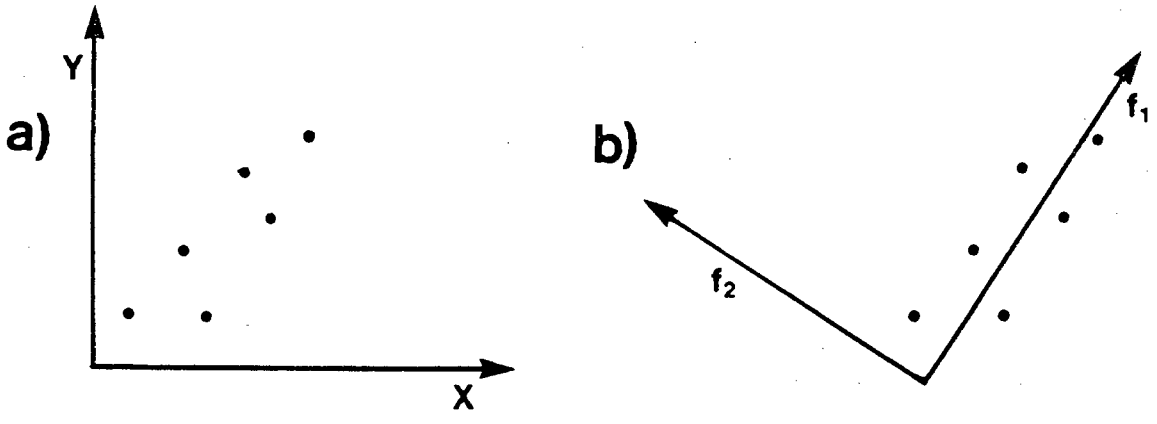


Figure 2.4.1 Diagram showing relation between data distribution and (a) variable axes x and y, and (b) factors f_1 and f_2.

It is obvious from 2.4.1(a) that the direction of maximum variance lies neither along the x axis nor along the y axis, but rather along some direction between them, i.e., along some combination of x and y. Similarly, the axis describing the direction of the second greatest amount of variation *away* from the principal direction of variance is coincident neither with x nor with y. Figure 2.4.1(b) depicts the identical distribution to that of 2.4.1(a), but referred to a new set of axes f_1 and f_2, such that f_1 represents the direction of greatest variance and f_2 that of the greatest variance *orthogonal* to f_1. Now, if the variation along f_2 is minimal compared to that along f_1, then it could justifiably be argued that the combination of x and y represented by f_1 is adequate in describing the distribution of the data points in the two dimensional space spanned by x and y. In other words, a reduction in the dimensionality of the data point distribution from two to one has been achieved.

In the case of an n dimensional problem what factor analysis therefore yields are up to n orthogonal factors (linear combinations of the original variables) lying along, respectively, the axis of largest variance, the axis of second largest variance, of third largest variance, and so on. Often the number of factors needed to describe, say, ninety percent of the sample variance is less than n, so that factor analysis essentially affords one a technique whereby the dimensionality of the parameter space can be reduced, i.e., it is a dimension reduction method.

However, factor analysis offers a second important tool for multidimensional analysis which derives, in fact, from its original application in the social sciences and from which it got its name. Consider, for example, a hypothetical survey of lung cancer sufferers. These might be asked to complete questionnaires in which, amongst many other items, they are asked to indicate whether they are male or female, what the colour of their hair is, how many cigarettes they smoked daily, what their incomes are, and so on. When the results of such a survey are subjected to factor analysis, what would very conceivably arise is a situation whereby one factor would be seen to account for most of the variance in the sample population, with other factors adding very little additional information. If this principal factor were examined for the components of the original variables present in it, it is very likely that the number of cigarettes smoked would feature as one of the components, while sex, for example, would not. The conclusion then would be that smoking is one of the "factors" which contributes to lung cancer!

In other words, factor analysis can also reveal those underlying factors or combinations of the original variables which principally determine the structure of the data distribution, and which, not infrequently, are related to some *real* influencing factor in the sample population. The task of the chemist, in our case would then be to *interpret* in chemical terms those underlying factors extracted out of the data matrix by factor analysis.

(b) The Mathematical Basis

Various slightly different techniques have been developed, depending on whether the original data matrix, the covariance or the correlation matrix serve as the starting point for the factor analysis. Essentially, though, the mathematical basis of factor analysis rests on *eigenanalysis* of the covariance or correlation matrix [1,2].

The descriptive approach outlined by Murray-Rust in a series of papers on computer analysis of molecular geometry [3-9] will be used here to sketch in broad outline the mathematical treatment involved.

Since the covariance matrix $\mathbf{C}$ is symmetrical about its diagonal, it will have real and non-negative *eigenvalues* λ, and its corresponding *eigenvectors* $\overline{X}$ can hence be obtained. (A section describing very briefly how to obtain eigenvalues and eigenvectors from a 3×3 matrix is appended.) Thus, eigenanalysis of an $n \times n$ covariance matrix, say, will yield n eigenvalues λ and n eigenvectors $\overline{X}$. The n factors are then obtained from

$$\mathbf{F} = \mathbf{E} \cdot \Lambda^{\frac{1}{2}} \tag{4.1}$$

where $\mathbf{F}$ is the $n \times n$ matrix of the factors, $\mathbf{E}$ is the $n \times n$ matrix of the column eigenvectors $\overline{X}$ and $\Lambda^{\frac{1}{2}}$ is the $n \times n$ diagonal matrix composed of the square roots of the eigenvalues.

Ordinarily the matrix $\mathbf{E}$ is composed of the normalised eigenvectors $\overline{X}$, since if this is the case then

$$\mathbf{C} = \mathbf{F} \times \mathbf{F}'$$

i.e., there is a check offered of whether the factors extracted from $\mathbf{C}$ are the correct ones, in that multiplying the factor matrix $\mathbf{F}$ by its transpose $\mathbf{F}'$ ought to again yield the original covariance matrix $\mathbf{C}$. (This point is explained in more detail in the Appendix to this chapter.)

The factors appear as linear combinations of the original variables in the form

$$f = ax_1 + bx_2 + \cdots ix_i$$

where the x_i represent the original variables, while the coefficients $a, b, \ldots i$ give an indication of the relative importance of the corresponding variable in the factor. These coefficients are often called factor loadings.

A further important point to note is that the λ's represent the proportion of the variance which the corresponding eigenvectors (factors) explain. Thus the factor with the largest eigenvalue will be the most important, or principal factor, and will lie along the axis of maximum variance of the data.

The factors f_n obtained in this way from the factor matrix $\mathbf{F}$ are sometimes called *abstract* factors in that they often do not relate directly to any chemical information, but represent rather a composite mixture of the original variables. In order to obtain chemically meaningful factors the abstract factor matrix $\mathbf{F}$ needs to be transformed or *rotated* into chemically meaningful data. This is accomplished using a $n \times n$ rotation matrix $\mathbf{A}$ such that the new factors g are obtained from

$$\mathbf{G} = \mathbf{F} \cdot \mathbf{A} \tag{2.4.2}$$

Essentially such a rotation corresponds to a rotation of the axes representing the abstract factors in the factor space, until they become coincident with a set of chemically meaningful "chemical" factor axes in that space. There are two methods for doing this. The first, orthogonal rotation, preserves the orthogonal relation of the abstract factor axes on rotation, while the second, oblique rotation, does not preserve the angles between the axes.

However, since rotation may lead to a subjective interpretation of the results of factor analysis and because the analogy between the unrotated factors lying along the eigenvectors and molecular vibrations (as described later) are lost, rotation needs to be treated cautiously, as indeed has been pointed out repeatedly by Murray-Rust [5,6].

Finally, in order to analyse graphically the results of factor analysis it is necessary to convert the original data matrix $\mathbf{D}$ (or the matrix $\mathbf{Z}$ of z scores in the case of standardized data) into a matrix $\mathbf{S}$ of *factor scores*

$$\mathbf{S} = \mathbf{D} \cdot \mathbf{F} \tag{2.4.3}$$

The factor scores for a given observation simply represent the coordinates of its representative point in the $n-$dimensional space spanned by the $\mathbf{n}$ factors, in much the same way as the values of the variables represent the coordinates for the data point in the original data space. The representative point in the original data space is therefore simply transformed into a new one, as the original data space is transformed into the new factor space. Consequently, whatever symmetry there may be in the data set must stay preserved during the transformation. This implies that not only the symmetry evident in the spatial distribution of the data must

remain, but also that the factors themselves must reflect the symmetry relations between the original variables which constitute them. Thus two variables which are symmetrically distributed in the data set, must have equal absolute loadings in whatever factor they may appear.

Therefore if a data set exhibits symmetrical properties, then these must be reflected also in the transformation. In such a case, therefore, we are afforded another means of checking that the factor analysis is proceeding correctly, by simply searching for symmetry both in plots of the data distribution in the factor space, and in the factors themselves, that is in the loadings of symmetry related variables.

The philosophical relation between eigenanalysis and factor analysis is premised, in fact, on the definition of an eigenvector $\overline{X}$ of a matrix $\mathbf{M}$ as *a vector which is transformed into a multiple of itself by* $\mathbf{M}$, i.e.

$$\mathbf{M} \times \overline{X} = \lambda I \cdot \overline{X}$$

where λ is a scalar called the eigenvalue of $\mathbf{M}$, and I is the identity matrix.

Suppose a covariance matrix $\mathbf{C}$ can be obtained from a given data set $\mathbf{D}$, i.e., a matrix which describes the covariance between the variables describing $\mathbf{D}$. Suppose further that some given linear combination $\overline{X}$ of these variables describes the axis of maximum variance in $\mathbf{D}$. Now, if more data taken *from the same parent population* as $\mathbf{D}$ were added to $\mathbf{D}$, then this should not in any way influence the axis of maximum variance, since the axes of maximum variance of any subset $\mathbf{D}$ of the parent population should be identical. All that this additional data added to $\mathbf{D}$ should do, if indeed the axis found represented the vector of maximum variance, is to *reinforce* this vector. In other words, the *direction* of $\overline{X}$ should not change, but its length might. Consequently $\overline{X}$, in fact, represents an eigenvector of the covariance matrix $\mathbf{C}$, since it can only be transformed into a multiple of itself by $\mathbf{C}$, while its direction remains unchanged.

2.5 A Worked Example

In order to demonstrate some of the statistical techniques outlined above a simple, hypothetical three dimensional data set consisting of 12 observations will be subjected to a simple analysis.

Starting with the data matrix $\mathbf{D}$ with the 12 cases of measurements of 3 variables x, y and z.

$$\mathbf{D} = \begin{array}{c} \textit{Case} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{array} \begin{array}{c} x \\ \left(\begin{array}{r} -5 \\ -5 \\ -4 \\ -3 \\ -2 \\ -1 \\ \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 6 \end{array} \right. \end{array} \begin{array}{c} y \\ \begin{array}{r} -1 \\ -4 \\ -2 \\ -4 \\ 0 \\ -2 \\ \\ 2 \\ 3 \\ 0 \\ 2 \\ 3 \\ 1 \end{array} \end{array} \begin{array}{c} z \\ \left. \begin{array}{r} 2 \\ -1 \\ 3 \\ 4 \\ 1 \\ 0 \\ \\ -2 \\ -1 \\ -3 \\ -1 \\ -5 \\ -7 \end{array} \right) \end{array}$$

$$\mu = 0.50 - 0.17 - 0.83$$

$$s^2 = 4.30 \quad 2.48 \quad 3.19$$

$$\text{range} = 11 \quad 7 \quad 11$$

where the means μ and sample variances are obtained as shown in equations (2.2.2) and (2.2.3).

The corresponding covariance matrix $\mathbf{C}$ as obtained from equation 2.3.1 is

$$\mathbf{C} = \frac{1}{11}\mathbf{D}^T \cdot \mathbf{D} = \begin{pmatrix} 18.45 & & \\ 8.64 & 6.15 & \\ -11.09 & -4.79 & 10.15 \end{pmatrix}$$

and the corresponding correlation matrix $\mathbf{R}$ as obtained from equation (2.3.2) is

$$\mathbf{R} = \begin{pmatrix} 1.00 & & \\ 0.81 & 1.00 & \\ -0.81 & -0.61 & 1.00 \end{pmatrix}$$

The matrix $\mathbf{Z}$ of $z-$scores obtained from the data matrix $\mathbf{D}$ according to equation 2.2.1 is

$$\mathbf{Z} = \begin{pmatrix} -1.28 & -0.33 & 0.89 \\ -1.28 & -1.54 & -0.05 \\ -1.05 & -0.74 & 1.20 \\ -0.81 & -1.54 & 1.52 \\ -0,58 & 0.07 & 0.57 \\ -0.35 & -0.74 & 0.26 \\ & & \\ 0.35 & 0.87 & -0.37 \\ 0.58 & 1.28 & -0.05 \\ 0.81 & 0.07 & -0.68 \\ 1.05 & 0.87 & -0.05 \\ 1.28 & 1.28 & -1.31 \\ 1.28 & 0.47 & -1.94 \end{pmatrix}$$

$$\mu = 0.0 \quad 0.0 \quad 0.0$$

$$s^2 = 1.0 \quad 1.0 \quad 1.0$$

$$\text{range} = 2.56 \quad 2.82 \quad 3.46$$

The corresponding covariance matrix $\mathbf{C}_z$ is

$$\mathbf{C}_z = \frac{1}{11}\mathbf{Z}^T \cdot \mathbf{Z} = \begin{pmatrix} 1.00 & & \\ 0.81 & 1.00 & \\ -0.81 & -0.61 & 1.00 \end{pmatrix}$$

which is identical to the correlation matrix $\mathbf{R}$ of the unstandardized data matrix $\mathbf{D}$, as was pointed out in section 2.2 above. An important point to note is that the relative sizes of the ranges of the x and y variables reverse during the standardization. Thus, for the raw data the range for variable y was smaller than that of x, whereas it becomes greater after standardization.

The effect of standardization has therefore been to increase the importance of the y variable in the $\mathbf{Z}$ matrix relative to its importance in the $\mathbf{D}$ matrix. This observation and how it relates to the problem of standardization has already been discussed under section 2.2.

From the values of the correlation coefficients r in the matrix $\mathbf{R}$, it becomes obvious that there is a high degree of correlation between x, y and z and, furthermore, that any pair of variables can describe between 37% and 66% of the variance of the sample. This may be gleaned from the squares of the correlation coefficients which each represent the proportion of the variance that can be explained by the linear relatedness of the two parameters involved.

Since in this case both the scale and the range of the raw data values are almost identical for each variable, no scaling or standardization will be employed for the subsequent factor analysis and, moreover, the covariance matrix will be used as the point of departure.

In order to extract from the covariance matrix $\mathbf{C}$ the eigenvalues λ and the corresponding eigenvectors $\overline{X}$ (and hence the factors), it is necessary to solve the following equation for λ

$$|\mathbf{C} - \lambda I| = 0$$

where I is the identity matrix with diagonal elements equal to unity and the off-diagonal elements equal to zero, and the vertical lines indicate that the determinant of the difference matrix between the lines should equal zero.

The above equation can be readily solved using standard matrix algebra and formulae for the solution of cubic equations. In this case the roots extracted from $\mathbf{C}$ are $\lambda_1 = 30.17$, $\lambda_2 = 3.22$ and $\lambda_3 = 1.36$, i.e., these are the eigenvalues of the covariance matrix.

In order to obtain the corresponding eigenvectors $\overline{X}$ the following equation needs to be solved for the various λ 's

$$(\mathbf{C} - \lambda I) \cdot \overline{X} = 0$$

where $\overline{X}$ is the column vector of the three variables x, y and z . Hence, the product of the difference matrix in brackets with the three dimensional column vector $\overline{X}$ must equal zero.

On solution, the three eigenvalues yield eigenvectors

$$\overline{X}_1 = \begin{pmatrix} 2.03 \\ 1.00 \\ -1.36 \end{pmatrix} \qquad \overline{X}_2 = \begin{pmatrix} 0.39 \\ 1.00 \\ 1.31 \end{pmatrix} \qquad \overline{X}_3 = \begin{pmatrix} -0.84 \\ 1.00 \\ -0.51 \end{pmatrix}$$

for λ_1, λ_2 and λ_3 respectively.

The eigenvectors are ordinarily normalized, which in this case yields

$$\overline{X}_1 = \begin{pmatrix} 0.77 \\ 0.38 \\ -0.52 \end{pmatrix} \qquad \overline{X}_2 = \begin{pmatrix} 0.23 \\ 0.59 \\ 0.77 \end{pmatrix} \qquad \overline{X}_3 = \begin{pmatrix} -0.60 \\ 0.71 \\ -0.36 \end{pmatrix}$$

The matrix of eigenvectors $\mathbf{E}$ is hence

$$\mathbf{E} = \begin{pmatrix} 0.77 & 0.23 & -0.60 \\ 0.38 & 0.59 & 0.71 \\ -0.52 & 0.79 & -0.36 \end{pmatrix}$$

and that of the square roots of the eigenvalues is

$$\Lambda^{\frac{1}{2}} = \begin{pmatrix} 5.49 & 0 & 0 \\ 0 & 1.79 & 0 \\ 0 & 0 & 1.17 \end{pmatrix}$$

When these two matrices are combined as in equation (2.4.3), the factor matrix $\mathbf{F}$ emerges.

$$\mathbf{F} = \mathbf{E}\cdot\Lambda^{\frac{1}{2}} = \begin{pmatrix} 0.77 & 0.23 & -0.60 \\ 0.38 & 0.59 & 0.71 \\ -0.52 & 0.77 & -0.36 \end{pmatrix} \cdot \begin{pmatrix} 5.49 & 0 & 0 \\ 0 & 1.79 & 0 \\ 0 & 0 & 1.17 \end{pmatrix} = \begin{pmatrix} 4.23 & 0.41 & -0.70 \\ 2.09 & 1.06 & 0.83 \\ -2.86 & 1.38 & -0.42 \end{pmatrix}$$

Consequently the three factors constituting the factor matrix are

$$f_1 = \quad 4.23x + 2.09y - 2.86z$$

$$f_2 = \quad 0.41x + 1.06y + 1.38z$$

$$f_3 = -0.70x + 0.83y - 0.42z$$

A check on whether the correct eigenvalues have been found is afforded by a comparison of the sum of the λ's with the sum of the variances of the original variables. These should obviously be equal, since the variance in the sample should be the same both *before* factor analysis and *after*. In this case the variances of the variables add up to 34.75 (= 18.45 + 6.15 + 10.15) as indeed do those of the factors (30.17 + 3.22 +1.36) also! Furthermore, the proportion of the sample variance explained by each factor can be estimated from its eigenvalue. Thus, f_1 has $\lambda = 30.17$, which represents $(30.17/34.75) \times 100$ percent of the variance. Hence f_1, f_2 and f_3 describe respectively 86.8, 9.3 and 3.9 percent of the sample variance.

As pointed out in section 2.4 above multiplication of the factor matrix $\mathbf{F}$ by its transpose $\mathbf{F}'$ affords a means of checking whether the correct factors have been extracted from the covariance matrix $\mathbf{C}$, since

$$\mathbf{C} = \mathbf{F} \cdot \mathbf{F}'$$

if the eigenvectors making up $\mathbf{F}$ have been normalised.

In this case

$$\mathbf{F} \cdot \mathbf{F}' = \begin{pmatrix} 18.55 & & \\ 8.69 & 6.18 & \\ -11.25 & -4.88 & 10.21 \end{pmatrix}$$

which is very close to the original covariance matrix obtained. Hence in this case the factors obtained are the correct ones within the limits of accuracy of these calculations.

The first or principal factor therefore adequately describes the variance in the sample, and a reduction of dimensionality from three to one has consequently been

achieved. Obviously, these factors have absolutely no real significance, and rotation would thus be meaningless.

However, a graphical analysis of the results of the factor analysis might prove instructive. In order to accomplish this, it is necessary to transform the data matrix $\mathbf{D}$ into a factor score matrix $\mathbf{S}$ which represents the projection of each of the original observations onto the factor axes. This is done as in equation (2.4.3)

$$\mathbf{S} = \mathbf{D} \cdot \mathbf{F}$$

and yields

$$\mathbf{S} = \begin{array}{c} \textit{Case} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{array} \begin{array}{c} f_1 \\ \left(\begin{array}{r} -29.0 \\ -26.7 \\ -29.7 \\ -32.5 \\ -11.3 \\ -\ 8.4 \\ \\ 18.4 \\ 21.8 \\ 25.5 \\ 28.2 \\ 46.0 \\ 47.5 \end{array} \right. \end{array} \begin{array}{c} f_2 \\ \begin{array}{r} -0.4 \\ -7.7 \\ 0.4 \\ 0.1 \\ 0.6 \\ -2.5 \\ \\ 0.2 \\ 3.0 \\ -2.5 \\ 2.8 \\ -1.3 \\ -6.1 \end{array} \end{array} \begin{array}{c} f_3 \\ \left. \begin{array}{r} 1.8 \\ 0.6 \\ -0.1 \\ -2.9 \\ 1.0 \\ -1.0 \\ \\ 1.1 \\ 0.8 \\ -1.5 \\ -2.3 \\ 0.4 \\ -0.4 \end{array} \right) \end{array}$$

The correlation matrix corresponding to $\mathbf{S}$ is

$$\mathbf{R} = \begin{pmatrix} 1.00 & & \\ 0.01 & 1.00 & \\ -0.06 & 0.05 & 1.00 \end{pmatrix}$$

Thus, within the context of the crude eigenanalysis performed and the truncation of numbers, the three factors are, for all intents and purposes, orthogonal to each other and consequently independent and uncorrelated, whereas the original variables x, y and z were highly correlated. This emerges very clearly from an examination of Figures 2.5.1 and 2.5.2.

Figure 2.5.1 represents plots of the original variables against each other, and it reveals the correlations between them. It is also easy to see, furthermore, that the data actually represent two clusters. Further information, however, cannot be gleaned from these plots.

Figure 2.5.2 represents plots of the factor scores against one another. From the scatter of the data points it becomes immediately obvious that f_1, f_2 and f_3 are not correlated. Moreover, in 2.5.2(a) and 2.5.2(c) the two clusters of points are very nicely separated, and it is this feature which makes factor analysis such a useful tool in more general and real cases (as will be seen later). Additionally, 2.5.2(a) reveals two possible outliers, while 2.5.2(b) reinforces this observation very dramatically.

An examination of the data matrix reveals in fact, that cases 2 and 12 (and possibly 11) are slight outliers.

In summary then, factor analysis is a method for reducing the dimensionality of the problem being investigated, it is capable of revealing significant underlying chemical factors which explain the variance in the data, and it offers a useful graphical technique for the representation of both clusters and outliers in the sample.

Factor analysis of a more general n dimensional data matrix would involve eigenanalysis of the $n \times n$ covariance or correlation matrix, and hence necessitate the solving of an equation of the $n-$th power. The mathematical algorithms which have been developed for this purpose incorporate least squares methods whereby the eigenvectors are consecutively calculated so as to minimize the residual error in each step. Thus each successive eigenvector accounts for a maximum of the variation in the data.

The procedure involves essentially the following steps. One, the eigenvector associated with the largest eigenvalue is orientated in the factor space so as to account in a least squares sense for the greatest possible variance in the data. Two, the second eigenvector associated with the second largest eigenvalue is directed orthogonally away from the first, and in the direction of maximum variance. These steps are then repeated for up to the $n-2$ eigenvectors left, each step being subject to the conditions (i) that the eigenvector be orthogonal to each preceding one, and (ii) that it account for the maximum variance possible. In this way each eigenvector (or factor) which emerges from the iteration is orthogonal to all the previous ones, and is oriented in the direction that maximises the sum of squares of all projections onto that axis (factor).

Since each successive eigenvector accounts for a smaller fraction of the total

variance in the data, it often occurs that the first four factors, say, describe up to ninety percent of the variance, the last ten percent being accounted for by the other $n - 4$ factors. In order, therefore, to know how many factors are necessary some tests have been devised. For example, Kaiser's criterion, which has been incorporated into the BMDP factor analysis program used in this study, retains factors whose eigenvalues are greater than unity. An alternative criterion might be to retain all those factors which collectively account for, say, ninety percent of the variance and discard all others.

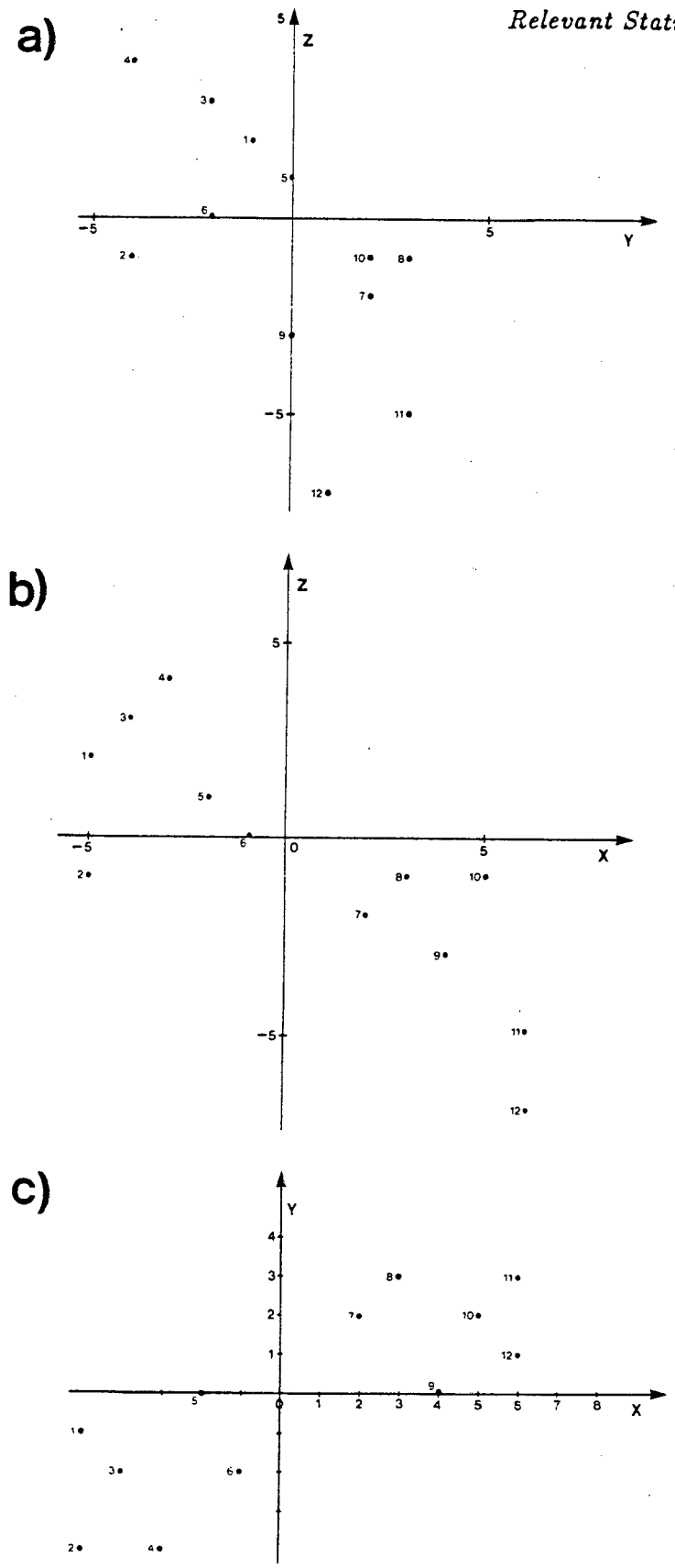


Figure 2.5.1 Scatterplots of data for worked example. Axes used are original variable axes.

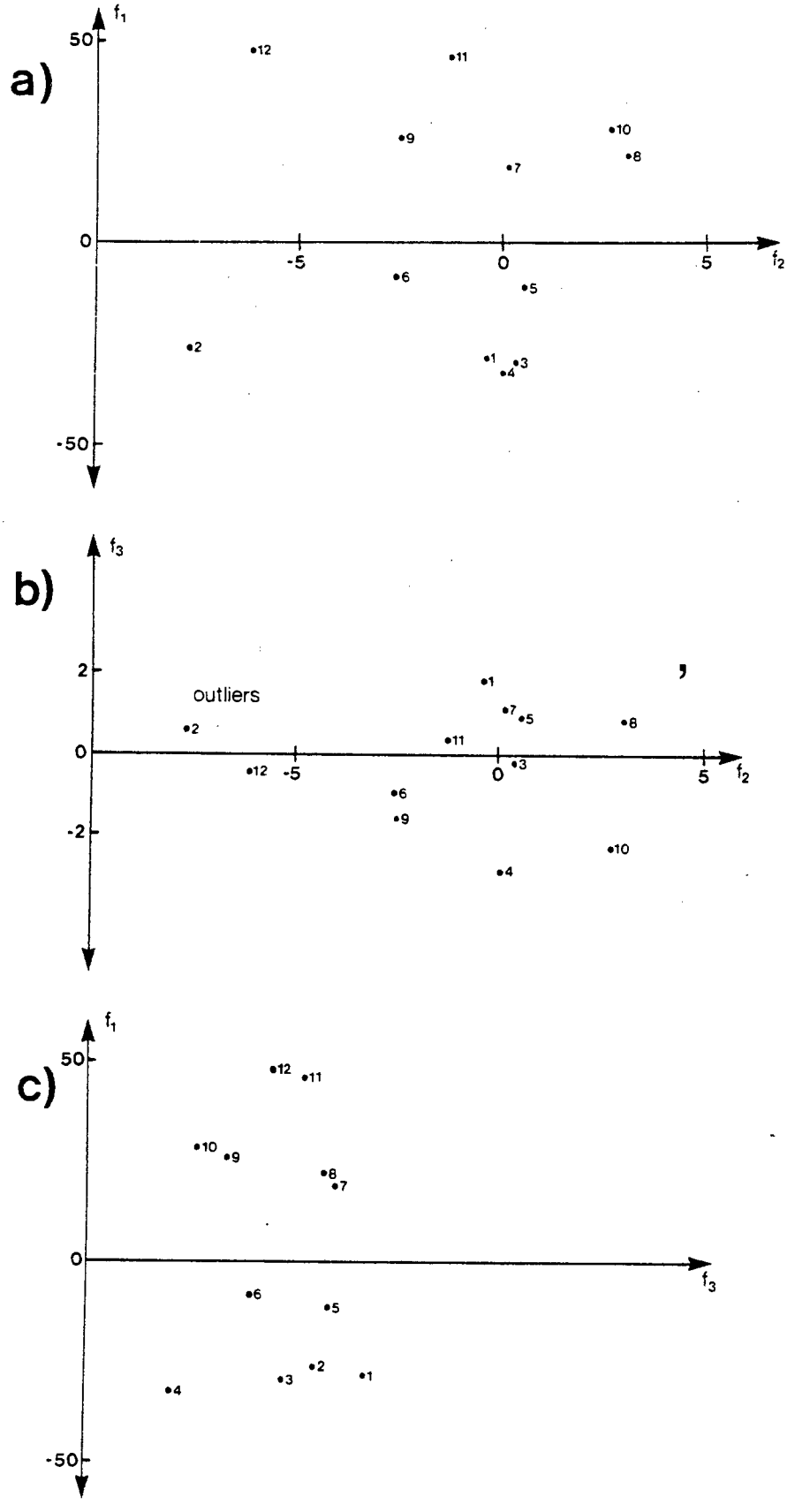


Figure 2.5.2 Scatterplots of factor scores for data of worked example.

If the data were free of experimental error or superfluous information, then factor analysis would yield only c eigenvectors, one for each of the controlling "chemical" factors, where c is less than n. The computer algorithms cannot decide, however, which of the n eigenvectors have physical meaning. All that they can do is reject some of the more insignificant factors according to some pre-set condition. It is the task of the chemist to judiciously examine the results of the computer iteration and to then interpret these according to his or her understanding of the chemical basis underlying the analysis.

2.6 Cluster Analysis

As pointed out earlier, factor analysis enables one to represent graphically the general *qualitative* characteristics of an n dimensional data point distribution and to investigate possible correlations between the elements of that data set *as a whole*. It does not,however, offer a means of examining the detailed *quantitative* relations in n dimensional space that might exist between individual clusters of data points *within* the data set, such as the number of clusters, the inter-cluster distances or the cluster characteristics.

Cluster analysis, on the other hand, enables one to "investigate the relationships that exist within a multivariate data set for which no *a priori* categorization is known" [10] . It therefore represents an extremely objective tool, since it does not presuppose any notions about the underlying characteristics of the data set, which, instead, emerge from the analysis.

What all clustering algorithms essentially do is to cluster together similar or neighbouring points into clusters in the n dimensional space. Their differences lie mainly in the criteria used for establishing similarity and in the rationale according to which clusters are fused together. Generally, two types of algorithms are distinguished, these being hierarchical and non-hierarchical or relocation clustering. Both methods require the calculation of a similarity matrix which contains a number indicating the "similarity" between each pair of observations of the original data set. This similarity, which is really a measure of the proximity of the pair of observations in the n dimensional space, is usually expressed in terms of either the Euclidian or the Mahalanobis distance between the two points, as outlined in section 2.3(b) above.

Once this similarity matrix has been established the various clustering techniques can be applied to it.

(a) Hierarchical Clustering Methods

There are two opposing approaches to hierarchical clustering, these being agglomerative and divisive procedures. In agglomerative clustering each observation in a data set is initially considered in a cluster on its own, and the hierarchical classification is built up by a series of linkages in which the most similar pairs of clusters are merged until all of the compounds are in a single cluster. Conversely, the divise algorithm begins by placing all the observations into one cluster which is then progressively subdivided into smaller ones until, finally, each observation is again in a cluster of its own. This approach may, consequently, be dubbed a "top down" technique, while the agglomerative algorithm represents a "bottom up" technique.

(b) Non-hierarchical (Relocation) Clustering

Relocation (or partitioning) methods attempt to partition a data set into some number of disjoint clusters such that related or similar compounds fall into the same cluster, with compounds unrelated to that cluster being distributed among the other, well-separated clusters in the set [1,11]. In general, the algorithm will generate a particular partition or clustering, determine the "goodness" of fit in some statistical sense and then relocate individual observations among the clusters until an optimum fit has been achieved. Since all the clusters are generated simultaneously, the resulting classification is non-hierarchical.

The number of clusters to be generated can either be specified in advance, or it may be optimised by the algorithm itself according to certain criteria. The use of some of these criteria is fraught with difficulties, however, and the latter type of algorithm must consequently be handled with care.

(c) Linkage Criteria

There are a large number of different criteria which have been developed to decide which individual elements and/or clusters should be merged together and in which way the similarity between a newly obtained cluster and other clusters or objects is defined. It is important to realise that the same algorithm may well give

different results for a given data set depending on what linkage and similarity criteria are used. It is therefore important to apply different techniques or to complement the clustering method with graphical techniques (such as factor analysis) wherever possible.

Single linkage is the oldest and simplest procedure, and in it the distance between objects and/or clusters is simply considered to be equal to the shortest distance between two individual elements, one from each cluster.

Complete linkage is the opposite of single linkage, in that the distance between two clusters is now considered to be equal to the largest distance between two individual elements, one from each cluster.

Average linkage defines the intercluster distance as the average distance between all pairs formed by elements from each of the two clusters respectively.

Centroid linkage focuses on the distance between the centroids of two clusters, or between the centroid of a cluster and an object outside of it.

Figure 2.6.1 shows graphically three of the linkage criteria outlined above.

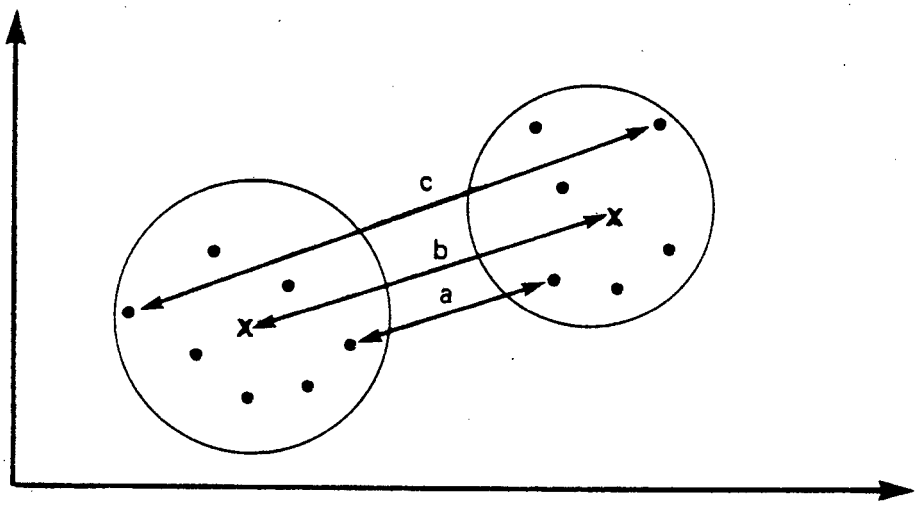


Figure 2.6.1 Diagram showing (a) single linkage, (b) centroid linkage and (c) complete linkage.

The four clustering methods described so far rely on the use of an *a priori* joining criterion, such as a minimal or maximal intercluster distance. Because of

this they are unsuitable for analysis of a data set about whose n dimensional distribution no *a priori* assumptions can be made.

In contrast to the above, the *error sum of squares* (or Ward's) method applies *a posteriori* criteria. It involves defining the "heterogeneity" or scatter of a cluster in terms of the sum of the squared distances of each element in that cluster from its centroid, i.e., the scatter E_i of a cluster i containing k objects, each a distance d_{ik} away from the centroid, is

$$E_i = \Sigma_k d_{ik}^2$$

The algorithm will then consider every possible cluster that might result from a fusion between cluster i and any other cluster (or object) j in the data space and estimate the scatter E_{i+j} of the resulting cluster. Eventually it will merge those two clusters i and j for whom, on fusion, the increase in the scatter or heterogeneity will be a minimum, i.e., the algorithm will tend to minimize

$$E_{i+j} - E_i - E_j$$

The *K-Means* clustering method has been devised for use with relocation algorithms exclusively, in contrast to the previous five linkage methods. Essentially this technique involves locating K centroids within the data space, such that the sum of the distances from the data points to each nearest centroid is minimized. Obviously this will need to be done via an iterative procedure, since the first, usually randomly chosen distribution of the centroids is unlikely to correspond to that of the true centroids of the clusters in the data space. Depending on the algorithm used, however, the original K centroids can be specified or the number of clusters to be determined may be specified if there is some *a priori* notion of what the distribution is likely to be. The advantage of this method lies, of course, in lower computation times, although it would seem to be unfeasible for large data sets, unless these are highly ordered and the number of clusters is reasonably low.

It has been suggested [1] that for hierarchical clustering Ward's method and the average linkage method are to be preferred, since these determine similarity from entire clusters, rather than from just two points, as in single and complete linkage. Similarily, it appears [1,4] that for relocation clustering Ward's method, the K-Means method and the centroid linkage method offer the best results. It must be

emphasized, though, that different algorithms can yield different clusterings, and that the results of a cluster analysis are therefore not necessarily unambiguous and, consequently, need to be supported by at least one other technique.

Finally it must be pointed out that different clustering algorithms handle symmetry relations within data sets in quite different, largely haphazard ways, since none of the software has been written so as to explicitly handle symmetric sample distributions. In other words, the algorithms do not necessarily follow up one particular clustering step with another, symmetry related one, and thus often destroy the symmetry which would otherwise have been evident. In this regard Bürgi and Norskov-Lauritsen [13] have shown how single linkage agglomerative clustering is able to maintain the symmetry in a hypothetical sample, while Ward's method fails to do so.

However, if it is implicitly accepted that a symmetric data set ought to yield a symmetric clustering pattern, then this assumption can be used in deciding whether a given result is feasible or not. In much the same way as a symmetric solution during factor analysis would support the results obtained, so a symmetric clustering pattern would tend to indicate a "good" solution, as opposed to one which results from the nature of the algorithm rather than the nature of the sample distribution.

2.7 A Worked Example

Two of the clustering techniques outlined in the previous section will be applied to a cluster analysis of the hypothetical data set used in the earlier worked example, in order to simply demonstrate how the algorithms work. The two methods to be used are, firstly, agglomerative or "bottom up" clustering using the single linkage (nearest neighbour) criterion and, secondly, divisive or "top down" clustering employing the complete linkage (furthest neighbour) criterion. The latter is of academic interest only at this stage, having never before been applied to a chemical analytical problem.

For both types of algorithms the point of departure is the similarity matrix. This has been established according to equation 2.3.3, and it contains simply the Euclidian distance between every pair of the twelve observations in the three dimensional data space.

	Case											
Case	1	2	3	4	5	6	7	8	9	10	11	12
1	0											
2	4.2	0										
3	1.7	4.6	0									
4	4.1	5.4	2.4	0								
5	3.3	5.4	3.5	5.1	0							
6	4.6	4.6	4.2	4.9	2.4	0						
7	8.6	10.5	8.8	9.8	5.4	5.4	0					
8	9.4	10.6	9.5	10.5	6.2	6.5	1.7	0				
9	10.3	11.5	10.2	10.7	7.2	6.2	3.0	3.9	0			
10	10.9	11.7	10.6	11.2	7.5	6.4	3.2	2.2	3.0	0		
11	13.6	13.6	13.7	14.5	10.4	9.9	5.1	5.0	4.1	4.2	0	
12	14.4	13.5	14.5	15.1	11.4	10.3	6.5	7.0	4.6	6.4	2.8	0

From a cursory examination of the similarity matrix the following becomes obvious. First, the minimum distance between compound 2 and any other member of the data set is 4.2, whereas most others have minimum distances considerably smaller than this, i.e. compound 2 may be regarded as an outlier. Second, by a similar argument compounds 11 and 12, in close proximity to each other, may be seen to be outliers, although this is more clearly the case for 12 than for 11. Third, the data fall into two diffuse clusters, i.e. compounds 1 to 6 and 7 to 12 in general being less than 5 units apart, with the elements of each cluster generally separated by more than 7 units.

Indeed, a three dimensional representation of the data distribution confirms these results, as is shown in Figure 2.7.1, where it can be seen that the compounds 1 to 6, with the exception of 2, fall into a quite different octant to those of compounds 7 to 12.

Unfortunately, though, in practise the situation is seldom as unambiguous nor as simple as this, and usually requires a skilful blend of cluster analysis, factor analysis and graphical interpretation in order to make any sense at all.

(a) **Single linkage agglomerative clustering**

This technique begins by considering each observation to initially be in a cluster on its own, subsequently seeking to cluster together those compounds and/or clusters nearest to one another, until only one cluster remains. Applying this algorithm to the similarity matrix, it may be seen that compounds 1 and 3, and 7 and 8, form the closest pairs (1.7 units) and will thus be first to be clustered together.

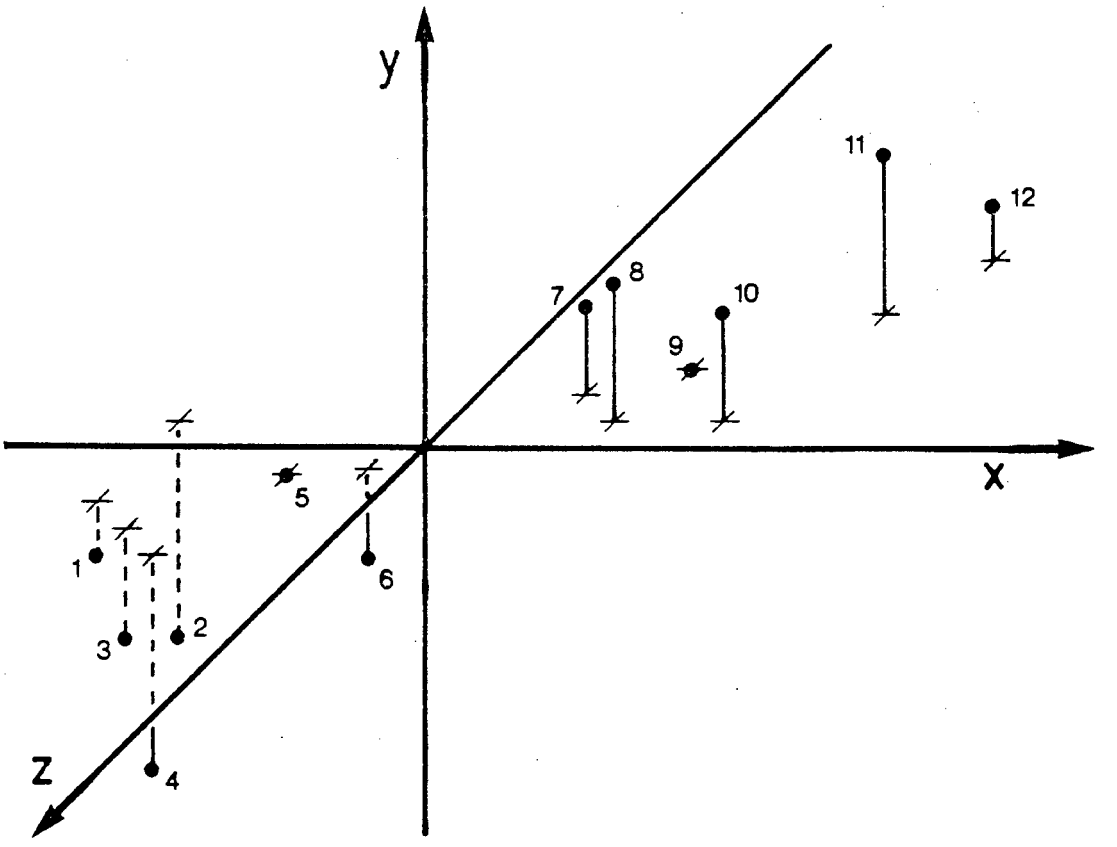
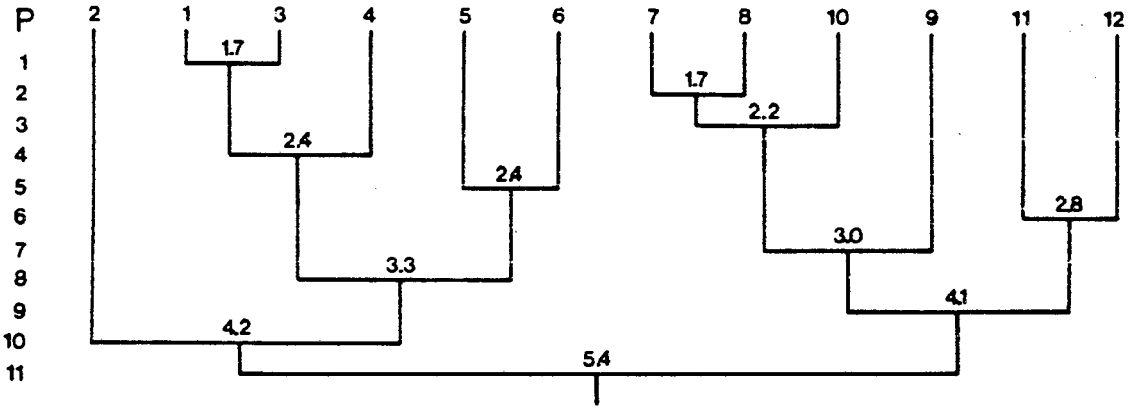


Figure 2.7.1 Diagram of the three dimensional data space of the hypothetical data set used.

Thereafter compound 10 is joined to the cluster (7,8), since its distance from the cluster (2.2 units) is less than the distance between any other pair of elements.

At the next stage, the fourth level of the clustering process, compound 4 will be joined to the cluster (1,3), since its distance to one of the elements (3) of that cluster is shorter than the distance between any other pair of compounds and/or clusters at that stage.

This procedure is then repeated for a total of 11 stages, until all the elements have been joined together to form one cluster. The history of this clustering process is best represented in the form of a *dendrogram*:



Three important points emerge from the dendrogram. Firstly, compound 2 is only merged to the cluster (1,3,4,5,6) at the penultimate level of clustering ($p = 10$) , thus easily identifying it as an outlier. Secondly, compounds 11 and 12, joined together at the sixth level, are finally only joined to cluster (7,8,9,10) at $p = 9$, thus also indicating that they may be treated as outliers of sorts. These observations echo the results obtained during the factor analysis of this data set, where these compounds were also clearly identifiable as outliers from the plot of factor 2 against 3 and, more clearly, from that of factor 1 against 2 (Figures 2.5.1(b) and (a) respectively). Finally, the dendrogram graphically illustrates the notion that the data consist of essentially two clusters (1,2,3,4,5,6) and (7,8,9,10,11,12), since the elements of both are kept separate from each other until the final cluster is formed.

In the general case of a data set with m observations there would be $(m-1)$ levels of clustering, and the algorithm would have difficulty in deciding at which stage to stop the process, i.e., at which stage are the clusters formed "meaningful" or "significant". In order to avoid confusion between this concept and that of "statistical significance" Massart [12] introduced the term "robust" cluster.

There have been some attempts at defining criteria for establishing the "correct" number of clusters. These are usually based on plots of some statistical or

semi-statistical measure, such as the average within-cluster distance, as a function of the number of clusters. Breaks in this curve are interpreted as indicating the emergence of robust clusters, or of the "correct" number of clusters. It must be pointed out, however, that this approach is still the subject of some debate [1,12] and does not yet seem to have been resolved.

Nevertheless, this method has been used successfully [1,13] and will therefore be demonstrated on this example. The measure chosen to indicate the break between "insignificant" and robust clusters is, for simplicity's sake, the nearest neighbour distance, since this would be expected to increase dramatically as the algorithm begins to cluster together robust, well separated clusters.

Figure 2.7.2 shows how the nearest neighbour distance, that is, the shortest distance between two nearest compounds and/or clusters, varies with the level of clustering p. Although the plot is not very striking, a discontinuity at the $p = 10$ level can be seen, implying that when the last two (= 12-10) clusters are joined there is an increase in "heterogeneity" since the clusters joined are relatively far apart. It could similarly be argued that there is a slightly less obvious break in the graph at the level $p = 8$, suggesting that there are four (= 12-8) clusters, these being the clusters (1,3,4,5,6), (7,8,9,10), (11,12) and (2). In this case, therefore, it would appear as if the "correct" number of clusters is either two or four.

Apart from this method of identifying the "correct" or robust clusters based on some statistical measure of similarity between clusters, there is another which has been devised by Massart and is premised on his definition of robustness. Thus, a cluster is defined as being "robust at a level p if at all higher p levels its elements do not intermingle with elements of other clusters formed at level p." [10] Furthermore, the lower the clustering level at which a robust cluster is formed, the more significant it is.

Whereas the former (statistical) method of identifying important clusterings is applicable to both hierarchical and non-hierarchical techniques, the latter method is applicable only to relocation clustering for two important reasons. First, since only one cluster is formed at each level during hierarchical clustering, it follows that the elements of any given cluster formed at level p cannot ever intermingle with those of another cluster formed at the same level, since none is formed. Second,

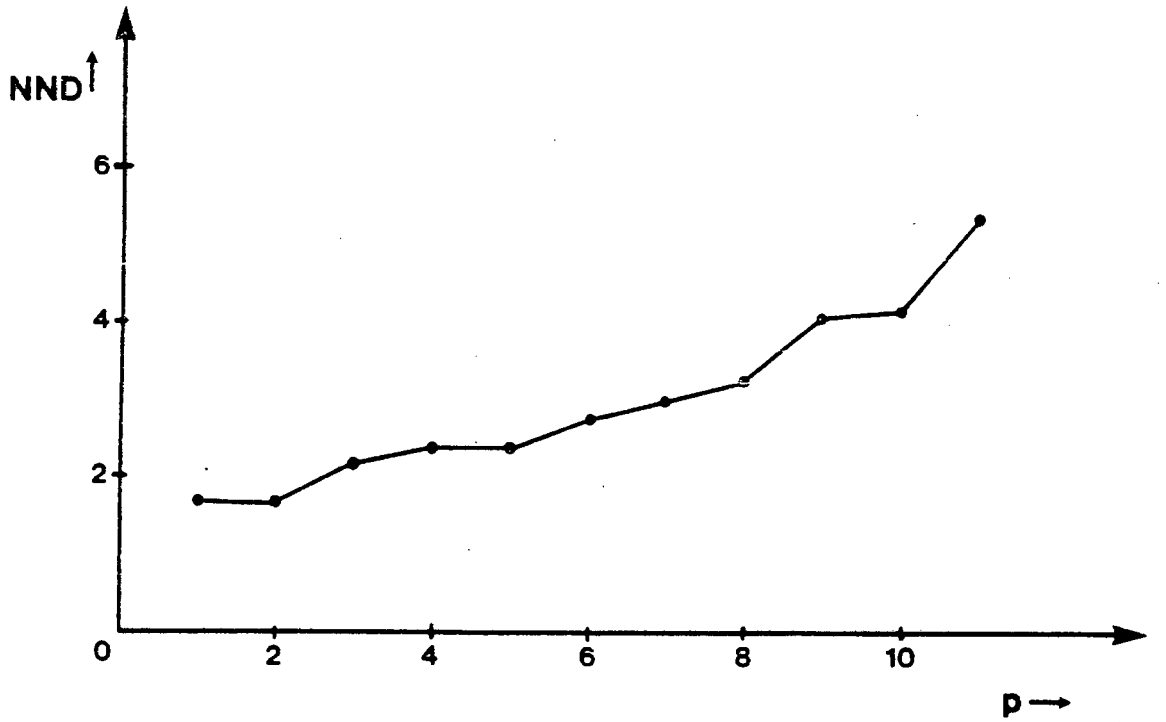


Figure 2.7.2 Plot of nearest neighbour distance versus level of clustering p .

ultimately all elements of a given data set will intermingle with all others in a single cluster, since agglomerative hierarchical clustering by its nature results in the fusion of all compounds into one cluster.

Consequently, the second method of identifying robust clusters is applicable only to relocation clustering and is based on an analysis of the actual membership of clusters, rather than on some statistical measure characterising the cluster. Essentially it involves fitting the data distribution to, say, a K cluster model (where K is some integer) and comparing the resultant cluster membership of the K clusters with the membership of the clusters formed for some other value of K . In the event of one or more clusters remaining essentially unchanged irrespective of how many clusters the data are fitted to, then those clusters can be said to be robust.

(b) Complete linkage divisive clustering

In this approach the data points are initially all assumed to be in one (all embracing) cluster, and in subsequent steps the algorithm seeks to split those elements furthest apart from each other in any given cluster, into two separate clusters, such

that these clusters are formed by the compounds closest to those elements originally split apart.

From the similarity matrix it may be seen that compounds 4 and 12, in fact, have the largest distance (15.1) between them. The algorithm will search the matrix for those elements closest to compounds 4 and 12, respectively, and will then divide them up into two clusters according to their proximity to compounds 4 and 12. At the first stage this will therefore result in two clusters (1,2,3,4,5,6) and (7,8,9,10,11,12).

At the second stage the algorithm searches for the largest distance amongst the various pairs of elements in the two clusters, and then splits that cluster which contains the most separated pair of elements. In this case these can be seen to be compounds 8 and 12 which are 7.0 units away from each other. The algorithm thus splits the cluster containing these two elements in such a way as to cluster around compound 8 those elements closest to it, and similarily for compound 12. Two clusters (7,8,9,10) and (11,12) emerge.

At the third stage the cluster (1,2,3,4,5,6) is split, since two of its members (2 and 4) are now furthest apart (5.4 units). This division gives rise to the clusters (2,6) and (1,3,4,5).

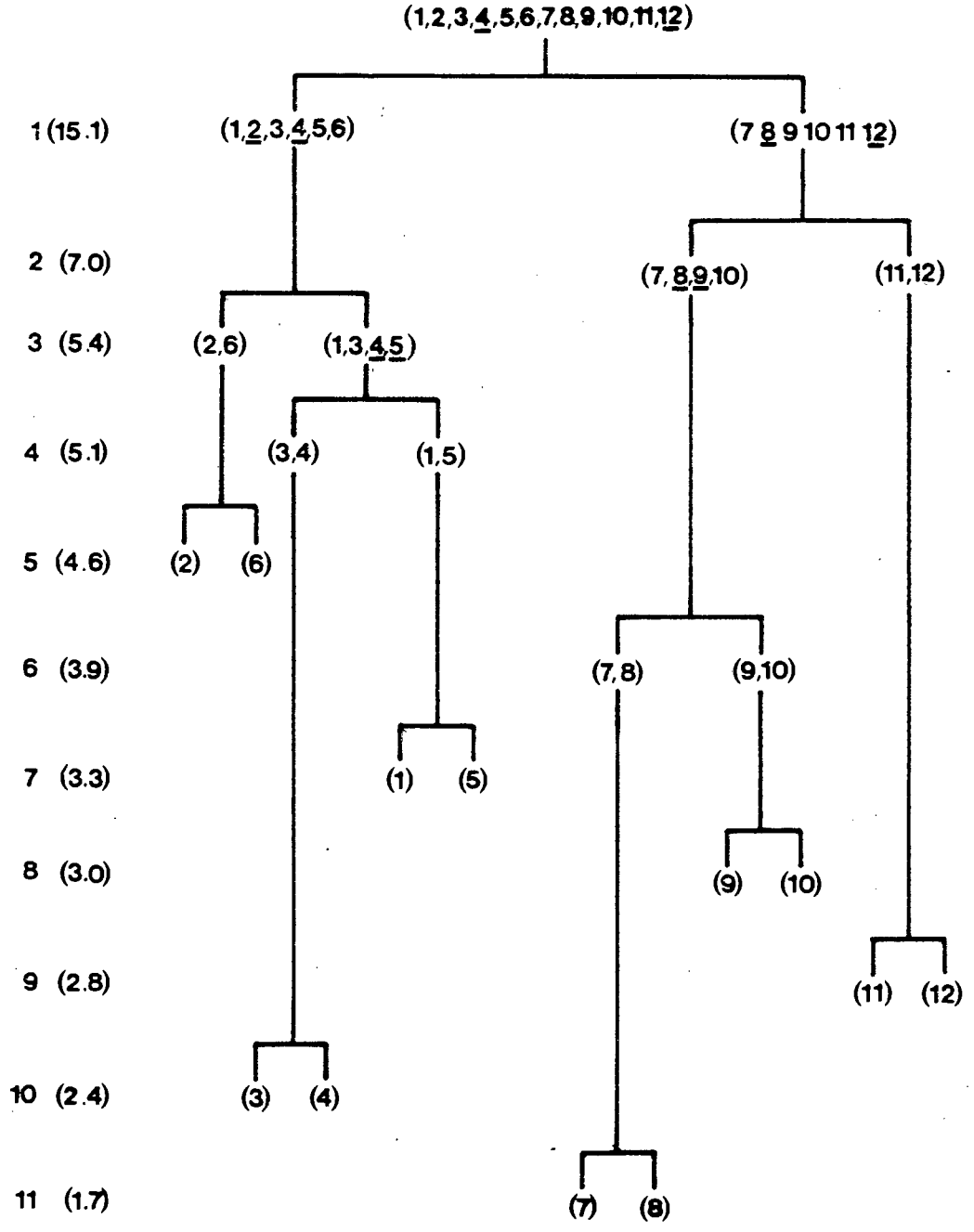
This step-wise subdivision of the data set can again be summarised in the dendrogram below.

The values in parentheses next to the clustering level p indicate the largest intercluster distance for the two clusters formed at that stage, while the elements furthest apart from each other in any given cluster are underlined.

The two dendrograms, the bottom up and the top down one, offer interesting comparisons. For example, comparing the clusters at the eighth level in the former with those at the third level in the latter, i.e., where there are four clusters in both cases, one can see quite clearly that the two algorithms give significantly different answers. Thus the agglomerative technique yields clusters (7,8,9,10), (11,12), (1,3,4,5,6) and (2) while the divisive method results in the clusters (1,3,4,5), (2,6), (7,8,9,10) and (11,12). Moreover, whereas the former reveals compound 2 as an outlier for nine successive clustering levels, the top down approach isolates both

compound 2 and 6 for only six successive clustering levels.

In conclusion, therefore, the various cluster analysis techniques offer a powerful tool for the multidimensional analysis of a data point distribution; they must be used with circumspection and an awareness of their particular shortcoming, though, and ought also to be backed up by other techniques or by a graphical analysis. In themselves the clustering algorithms do not yield an answer – this can only be arrived at through as judicious and objective an interpretation of their output as possible.



REFERENCES

1. D.L. Massart and L. Kaufman. *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, John Wiley and Sons, New York, 1983.
2. E.R. Malinowski and D.G. Howery. *Factor Analysis in Chemistry*, John Wiley and Sons, New York, 1980.
3. Murray-Rust, P., Motherwell, S. *Acta Cryst.* 1978, **B34**, 2518-2526.
4. Murray-Rust, P., Bland, R. *Acta Cryst.* 1978, **B34**, 2527-2533.
5. Murray-Rust, P., Motherwell, S. *Acta Cryst.* 1978, **B34**, 2534-2546.
6. Domenicano A., Murray -Rust, P., Vaciago, A. *Acta Cryst.* 1983, **B39**, 457-468.
7. Murray-Rust, P. *Acta Cryst.* 1982, **B38**, 2765-2771.
8. Murray-Rust, P., Raftery, J. *J. Mol. Graphics*, 1985, **3**, 50-59.
9. Murray-Rust, P., Raftery, J. *J. Mol. Graphics*, 1985, **3**, 60-68.
10. Smyers-Verberke, J., den Hartog, J.C., Dekker, W.H., Massart, D.L. *Analysis*, 1984, **12**, 486-489.
11. Willett, P. *J. Chem. Inf. Comput. Sc.*, 1984, **24**, 29-33.
12. Kaufman, L., Massart, D.L. *Anal. Chem.*, 1982, **54**, 911-917.
13. Norskov-Lauritsen, L., Bürgi, H.B. *J. Comput. Chem.*, 1985, **6**, 216-228.

APPENDIX

A vector $\overline{X}$ which is transformed into a multiple of itself by the matrix $\mathbf{M}$ is called an *eigenvector* of $\mathbf{M}$. $\overline{X}$ and $\mathbf{M}$ must fulfil the condition

$$\mathbf{M} \cdot \overline{X} = \lambda I \cdot \overline{X}$$

where λ is a scalar called the *eigenvalue* of $\mathbf{M}$ corresponding to the eigenvector $\overline{X}$, and I is the identity matrix.

The eigenvalues λ can be found from the *characteristic equation* of $\mathbf{M}$

$$|(\mathbf{M} - \lambda I)| = 0$$

which states that the determinant of the difference matrix in brackets must equal zero.

As an example consider the 3×3 matrix

$$\mathbf{M} = \begin{pmatrix} a1 & b1 & c1 \\ a2 & b2 & c2 \\ a3 & b3 & c3 \end{pmatrix}$$

Its characteristic equation would be

$$\left| \begin{pmatrix} a1 & b1 & c1 \\ a2 & b2 & c2 \\ a3 & b3 & c3 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right| = 0$$

$$\left| \begin{pmatrix} a1 & b1 & c1 \\ a2 & b2 & c2 \\ a3 & b3 & c3 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \right| = 0$$

$$\left| \begin{pmatrix} a1 - \lambda & b1 & c1 \\ a2 & b2 - \lambda & c2 \\ a3 & b3 & c3 - \lambda \end{pmatrix} \right| = 0$$

$$\begin{aligned} (a1 - \lambda)\big[(b2 - \lambda)(c3 - \lambda) - (b3 \times c2)\big] \\ - a2\big[b1(c3 - \lambda) - (b3 \times c1)\big] \\ + a3\big[b1 \times c2 - c1(b2 - \lambda)\big] = 0 \end{aligned}$$

This would yield a cubic equation in λ which could be solved for three real and non-negative roots if $\mathbf{M}$ is symmetrical!

Having obtained λ_1, λ_2 and λ_3 the corresponding eigenvectors $\overline{X}_1, \overline{X}_2$ and $\overline{X}_3$ can be obtained by solving

$$(\mathbf{M} - \lambda I) \cdot \overline{X} = 0$$

for the various values of λ, where $\overline{X}$ is the column vector

$$\overline{X} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Hence solving

$$\left(\begin{pmatrix} a1 & b1 & c1 \\ a2 & b2 & c2 \\ a3 & b3 & c3 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \right) \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

or

$$(a1 - \lambda)x + b1 \cdot y + c1 \cdot z = 0$$

$$a2 \cdot x + (b2 - \lambda) \cdot y + c2 \cdot z = 0$$

$$a3 \cdot x + b3 \cdot y + (c3 - \lambda) \cdot z = 0$$

for all three values of λ yields three eigenvectors

$$\overline{X}_1 = \begin{pmatrix} x1 \\ y1 \\ z1 \end{pmatrix} \qquad \overline{X}_2 = \begin{pmatrix} x2 \\ y2 \\ z2 \end{pmatrix} \qquad \overline{X}_3 = \begin{pmatrix} x3 \\ y3 \\ z3 \end{pmatrix}$$

Suppose now that n eigenvalues $\lambda_{1,\ldots n}$ and n corresponding eigenvectors $\overline{X}_{1,\ldots n}$ have been extracted from a $n \times n$ covariance matrix $\mathbf{C}$. Then, for all i we have, by definition

$$\mathbf{C} \cdot \overline{X}_i = \lambda_i \overline{X}_i.$$

Instead of using the individual $\overline{X}_i$ and λ_i, however, we can substitute the eigenvector matrix $\mathbf{E}$ composed of the column eigenvectors, and the eigenvalue matrix Λ, whose diagonal elements are the n eigenvalues and whose off-diagonal elements are zero. Thus

$$\mathbf{C} \cdot \mathbf{E} = \mathbf{E} \cdot \Lambda$$

Postmultiplying both sides of the equation by $\mathbf{E}^{-1}$, i.e., the inverse of $\mathbf{E}$, we obtain

$$\mathbf{C} \cdot \mathbf{E} \cdot \mathbf{E}^{-1} = \mathbf{E} \cdot \Lambda \cdot \mathbf{E}^{-1}$$

$$\mathbf{C} = \mathbf{E} \cdot \Lambda \cdot \mathbf{E}^{-1}$$

$$= \mathbf{E} \cdot \Lambda^{\frac{1}{2}} \cdot \Lambda^{\frac{1}{2}} \cdot \mathbf{E}^{-1}$$

Now, if the eigenvector matrix is composed of normalized eigenvectors, then the inverse of $\mathbf{E}$ is just the transpose of $\mathbf{E}$, i.e.

$$\mathbf{E}^{-1} = \mathbf{E}^{1}$$

Then, by equation 2.4.1 we have

$$\mathbf{C} = \mathbf{F} \cdot \mathbf{F}'$$

In other words, if the eigenvectors extracted from the covariance matrix $\mathbf{C}$ are normalized prior to computing the factor matrix $\mathbf{E}$, then $\mathbf{C}$ should be recoverable from $\mathbf{F}$ and its transpose $\mathbf{F}'$ by multiplication of these two.

CHAPTER 3

Data Search, Description of Conformation and Data Preparation

3.1 Data Search

The Crystallographic Date Centre of Cambridge University maintains a set of computer files which contain a database relating to the structures of organic and organometallic compounds and metal complexes as determined by X-ray or neutron diffraction. [1] This database will hereinafter be referred to as the Cambridge Structural Database (CSD). The CSD consists essentially of three major files: (i) the Bibliographic File (BIB), which houses bibliographic information, (ii) the Chemical Connectivity File (CONN), which contains a coded representation of the chemical structural diagram, and (iii) the structural Data File (DATA) with information on details such as the unit cell parameters, the space group, atomic coordinates. Individual entries in the CSD are identified by means of a reference code (refcode) which essentially consists of six alphabetic characters forming an acronym of the compound name.

There are two types of searches which can be effected. First, a search of the BIB file can be made for various kinds of bibliographic information. Second, the CONN file may be searched for a specific chemical fragment — a four-coordinate platinum complex with four tri-coordinated sulphur atoms, for example. Search results (hits) are produced as a file of refcodes for those entries which satisfy the input query. These refcodes are then used to withdraw crystallographic data on the hits from the DATA file. Once a subfile containing these data has been created, it may be acted on by either the geometrical analysis programme (GEOM78), the display program (PLUTO78) or by suitable user-designed software.

The CSD has become an essential resource for comparative structural studies on large numbers of molecular compounds. [2] In general, such studies initially require the definition of the basic molecular fragment to be examined, so that matching fragments in the CONN file can be identified and their respective refcodes used to withdraw crystallographic data from the DATA file. Thereafter the particular geometric parameters describing the molecular or fragment geometry which are to be compared with each other may be computed.

(a) Fragment definition

The choice of fragment for which the CONN file is to be searched needs to

be a judicious compromise between a definition which is too narrow in its scope, thus precluding entries whose particular distortions might be of extreme interest and importance, and one which is so broadly stated as to include fragments or data which add nothing to the analysis – apart, that is, from effort and time spent on it. In many cases this is an easy choice to make. For example, in their contribution to the series of papers "Chemical Reaction Paths" [3] , Bye, Schweizer and Dunitz were interested in examining the geometry of Ph_3PX fragments. In this case they simply had to search the file for the occurrence of Ph_3PX fragments. With us, however, the fragment definition was no such simple matter, since the nature of the problem which we wished to investigate was quite different.

In attempting to examine the molecular conformation of d^8 five-coordinate metal complexes the question must be asked: What is a five-coordinate compound? The answer to this question is not as straight forward nor as simple as it may at first seem. It has by now been well established that for certain metal complexes there is a relatively smooth progression from a tetra-coordinate to a penta-coordinate state and, in some cases, from the latter to a hexa-coordinate state. Bürgi has shown that this is the case for cadmium [4] , Britton and Dunitz for tin, germanium and lead [5] , and we have shown it for zinc [6] and nickel [7] – to name but a few. Consequently, the choice of whether a given metal complex is to be labelled as "five-coordinate" or "four-coordinate" is sometimes made on the basis of subjective criteria depending on whether an author of a paper is looking for (or expecting) one coordination number rather than another. Unfortunately, though, when such a case is included in the CSD, the author's possibly subjective evaluation or prejudice is simultaneously included in the data, especially in the chemical connectivity file CONN.

As a result of this, it is quite reasonable to assume that there may well be entries of metal complexes which are listed as four-connected, but which could alternatively be regarded as five-coordinate on the application of a different (subjective) set of criteria. Obviously this complicates the search for appropriate fragments, since it is precisely those fragments on the borderline between the two coordination numbers which offer unique insights — according to the structure correlation hypothesis they map out the pathway for the addition of a fifth ligand to a four-coordinate metal

centre. (We will hereinafter refer to such compounds as "four-plus-one" coordinate.) Clearly the inclusion of four-coordinate entries which happen to have a fifth atom in some proximity to the metal would be appropriate only for compounds for which it can somehow be verified that the extra atom is there by virtue of some interaction with the metal centre, rather than as a trivial consequence of its attachment to one of the four other ligands.

In the light of these considerations we decided on a two-pronged approach to the problem of data retrieval. Firstly, those entries listed as five-coordinate in the CSD for each of nickel, palladium, platinum, rhodium and iridium were to be withdrawn. Secondly, all entries for these metals listed as four-coordinate would be withdrawn and then subsequently searched for fragments which might alternatively be classified as five-coordinate, or as "approaching five-coordination", i.e., the four-plus-one coordinate entries.

In order to effect the latter search, some *a priori* concept of a fragment fitting the general description of "five-coordinate" had to be developed. For this we chose to rely on experience gathered during our previous studies of five-coordination. [6,7] It was felt that the study of penta-coordination in nickel was particularly useful for the following reasons: (i) nickel is in the same group as palladium and platinum, (ii) its four-coordinate complexes are predominantly square-planar, as are those of the other metals, (iii) Ni(II) has similar d^8 valence shell electron configuration to Pd(ii), Pt(ii), Rh(I) and Ir(I).

In this case we decided to define a five-coordinate fragment on the basis of bond angles contained within it. To formulate such a definition and to ensure an efficient search we needed to anticipate the results of the search to some extent. In other words we needed to anticipate what the bond angles in a "typical" five-coordinate molecular fragment might be.

By far the most commonly adopted conformations amongst penta-coordinate main group and transition element complexes are the trigonal bipyramid (TBP) and the square-based (or rectangular) pyramid (SQP), or slightly distorted forms of either. Indeed, it is now well established that the TBP and SQP are readily interconvertible via the Berry mechanism. [8] This involves simultaneous in-plane bends of the axial ligands (atoms 1 and 5 in Figure 3.1.1) and two of the equatorial

ligands (atoms 2 and 4), with the third equatorial ligand (atom 3) acting as a "pivot" for what has inappropriately been called a pseudorotation mechanism. When these in-plane bends are continued beyond the SQP level a new TBP is formed as shown in Figure 3.1.1.

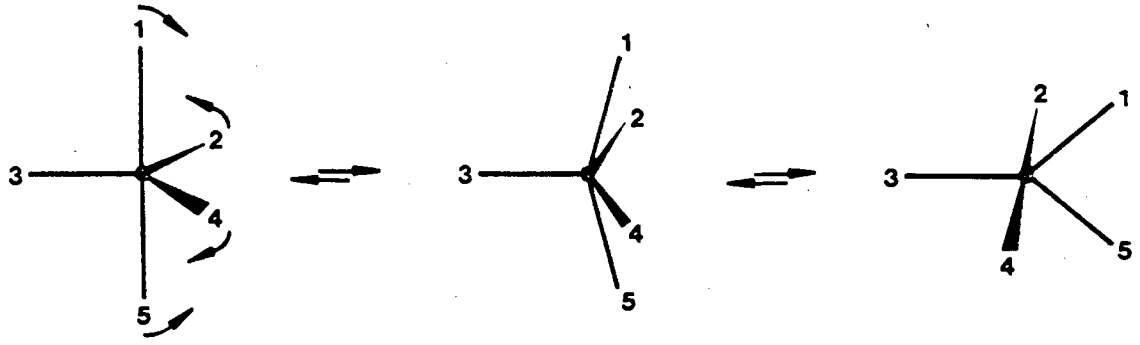


Figure 3.1.1. Diagram of the Berry mechanism

Thus, the primary types of distortions away from idealised trigonal bipyramidal or square-based pyramidal conformation which are observed amongst pentacoordinate complexes are distortions "along" the Berry coordinate. These include mainly a reduction in the angle between the axial ligands (θ_{15}) , with a concomitant opening of that between two equatorial ligands (θ_{24}) in the TBP, or, conversely, a widening of one trans-basal angle (θ_{15}) and a reduction of the other (θ_{24}) in the SQP. Over the last two decades, in particular, these general distortions have been well documented for both main group and transition element complexes, and Holmes has recently written a comprehensive review of the area. [9]

In addition to the broad angular deformations described above there have also been many observations of elongated or shortened interatomic distances, although these obviously contribute far less to distortions away from the TBP or SQP, than do angular displacements. We have shown, for example, that the apical bond length (d_3) in square pyramidal nickel complexes may vary over a large range without sig-

nificantly distorting the conformation of the molecular fragment away from SQP. [7] (In fact, an "ideal" SQP can, of course, have an apical bond of any length.)

The considerations outlined above, together with the empirical observations mentioned, led to the "angular" definition of a five-coordinate molecular fragment as indicated in Table 3.1.1. Included also in this table are the intra-fragment angles as defined for the TBP, and one example (of many possible ones) of these angles for an "ideal" SQP. The values for the SQP which we indicate have been used by Holmes [9] as well as by us in our earlier studies, [6,7] and they are obtained by placing M at the centre of mass of a $[L_5]$ square pyramid, with the distances from the central M to the five L atoms all equal. They are listed here merely for the purpose of comparison with the ranges defined by us for this study, since it has been suggested that the values of the intra-fragment angles of a SQP will depend on the d-orbital electron configuration. [10,11,12]

Table 3.1.1. Definition of angular range for a typical five-coordinate fragment (in degrees). Angles $\theta_{i,j}$ refer to angles between atoms i and j according to the labelling scheme shown in Figure 3.1.1, i.e., a TBP with axial ligands 1 and 5, and a SQP with apical ligand 3 and transbasal ligand pairs 1 and 5, and 2 and 4.

	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}	θ_{25}	θ_{35}	θ_{45}	θ_{15}
TBP	90	90	90	120	120	120	90	90	90	180
SQP	86	105	86	105	150	105	86	105	86	150
RANGE	75-108	80-110	75-108	70-132	105-175	70-132	66-107	66-108	66-107	150-180

In addition to the limits imposed on the intra-molecular angles for the purpose of identifying the four-plus-one coordinate entries, we also decided to limit the distance between the central metal and its five ligand donour atoms to no more than 3.4 Å. It has been found [5,6,7,13] that distances of this order represent the maximum at which an interatomic interaction of sorts can conceivably be postulated, this distance in most cases being larger than the sum of the van der Waal's radii. Clearly, one could have included fragments with interatomic distances greater than this, but experience has shown that the statistical scatter becomes totally random beyond this point. [13]

(b) Data Search and Retrieval

Identification of five-coordinate entries:

A search of the July 1984 version of the CONN file, at that stage containing 42381 entries, revealed 124 compounds containing nickel in an exactly five-coordinate environment, 11 palladium, 17 platinum, 50 rhodium and 59 iridium entries.

These hits were then manually sorted, rejecting those listed *wi*thout *a*tomic *c*oordinates (= wac), those with *dis*ordered structures (= dis), *all*yl compounds (= all), those with a *uni*molecular error (= uni) and compounds whose coordination number was incorrect or whose oxidation state, as judged by the entry in the CSD, precluded a d^8 electron configuration (= rej). Allyl complexes were rejected on the basis that a definite atom or point of ligation of the allyl ligand would be difficult to identify.

The number of entries rejected for each metal, followed by the reasons for their rejection in parenthesis, is given below:

Ni: 27 (16 wac, 3 dis, 2 uni, 6 rej)

Pd: 6 (3 wac, 3 rej)

Pt: 9 (4 wac, 3 rej, 2 all)

Rh: 28 (19 wac, 2 dis, 5 rej, 2 all)

Ir: 34 (16 wac, 5 dis, 6 rej, 7 all)

In order to test the angular definition of a five-coordinate fragment (as set out in Table 3.1.1) we used the data set of known five-coordinate compounds for a test run. The key words and parameters for the program GEOM are shown in Table 3.1.2.

Table 3.1.2. Keywords and parameters used during search for four-plus-one coordinate entries.

```
INPUT...FRAG PT 5 CONNECTED
INPUT...C
INPUT...C  THIS GEOM RUN ATTEMPTS TO FIT
INPUT...C  A GENERAL 5 COORD GEOMETRY TO
INPUT...C  THE 5 CONNECTED PT ENTRIES.
INPUT...C
INPUT...C  IT IS DIFFERENT FROM GEOM5GEN
INPUT...C  IN THAT IT LIMITS ALL ANGLES
INPUT...C  TO THE RANGES GIVEN.
INPUT...C
INPUT...AT1 AA
INPUT...AT2 AA
INPUT...AT3 AA
INPUT...AT4 AA
INPUT...AT5 AA
INPUT...AT6 PT
INPUT...BO 6 1
INPUT...BO 6 2
INPUT...BO 6 3
INPUT...BO 6 4
INPUT...BO 6 5
INPUT...TEST DIST 6 1 1.5 3.4
INPUT...TEST DIST 6 2 1.5 3.4
INPUT...TEST DIST 6 3 1.5 3.4
INPUT...TEST DIST 6 4 1.5 3.4
INPUT...TEST DIST 6 5 1.5 3.4
INPUT...TEST ANG 1 6 5 150 181
INPUT...TEST ANG 2 6 4 105 175
INPUT...TEST ANG 1 6 2 75 108
INPUT...TEST ANG 1 6 4 75 108
INPUT...TEST ANG 1 6 3 80 110
INPUT...TEST ANG 5 6 2 65 107
INPUT...TEST ANG 5 6 4 65 107
INPUT...TEST ANG 5 6 3 65 108
INPUT...TEST ANG 2 6 3 70 132
INPUT...TEST ANG 3 6 4 70 132
INPUT...END
INPUT...NOPRINT
INPUT...DEF D1 6 1
INPUT...DEF D2 6 2
INPUT...DEF D3 6 3
INPUT...DEF D4 6 4
INPUT...DEF D5 6 5
INPUT...DEF DEL51 5 6 1
INPUT...DEF DEL52 5 6 2
INPUT...DEF DEL53 5 6 3
INPUT...DEF DEL54 5 6 4
INPUT...DEF DEL12 1 6 2
INPUT...DEF DEL13 1 6 3
INPUT...DEF DEL14 1 6 4
INPUT...DEF DEL23 2 6 3
INPUT...DEF DEL34 3 6 4
INPUT...DEF DEL24 2 6 4
INPUT...DEF *RFACT
INPUT...DEF *YEAR
INPUT...END
* NO ATOM COORDS GIVEN
```

Essentially what the program did was number the ligands in such a way as to restrict the values of the interatomic angles found in the fragment to the ranges set out in Table 3.1.1. In other words, it attempted to superimpose onto the penta-coordinate molecular fragments the ligand numbering scheme employed in Figure 3.1.1 and Table 3.1.1, according to which a distortion away from a TBP towards a SQP manifests itself through a reduction in the angle θ_{15} and a concomitant widening of the angle θ_{24}, or conversely, a SQP distorts towards a TBP by a widening of θ_{15} and a reduction of θ_{24}.

The program managed to embrace all 157 usable five-coordinate entries, thereby indicating that our working angular definition for a penta-coordinate geometry certainly encompassed all those molecular geometries traditionally defined as five-coordinate, at least insofar as nickel, palladium, platinum, rhodium and iridium were concerned. Moreover, the program managed in almost all cases to label those atoms containing the largest interatomic angle as atoms 1 and 5, and those with the second largest as atoms 2 and 4, respectively, in line with the convention adopted in Table 3.1.1 and Figure 3.1.1. Not only does the program therefore offer a means of searching for a five-coordinate fragment, but it also has the added advantage of beginning to superimpose onto that fragment a uniform ligand numbering scheme.

Identification of four-plus-one coordinate entries:

A search of the July 1984 version of the CSD yielded 624 entries containing nickel in a four-coordinate environment, 428 containing palladium, 656 platinum, 165 rhodium and 69 iridium containing entries. This database was searched for the presence of six-atom fragments, i.e., four-plus-one coordinate fragments, whose geometry corresponded to that outlined in Table 3.1.1. GEOM gave 25 hits for the nickel complexes, 22 for palladium, 26 for platinum, 15 for rhodium and 11 for iridium. These were sorted along the same criteria as used for the five-coordinate entries, finally yielding 7 nickel, 10 palladium, 1 platinum, 3 rhodium and 1 iridium entry which could be considered as four-plus-one coordinate.

In four cases we found that ostensibly five-coordinate complexes had inadvertently been recorded as four-coordinate in the CONN file. We surmise that this may be due to peculiarities of these compounds which complicate the automated data-checking routines of the CSD: two rhodium compounds (refcodes PFPRHB

and RCOHPH) contain hydrogen atoms as fifth ligands, while another rhodium compound (BSAIRH) and an iridium compound (BEZPIF) consist of dimers, one nucleus of which is four-coordinated.

For the rest, the four-plus-one coordinated compounds fall into four classes, depending on what type of interaction the fifth, previously unrecognised bond comprises:

(i) There are six examples (AEBXNI, HMACNI, NIACTD, MAZTNI, NIHPOR, NIMAHP) of macrocyclic compounds where a counterion might be construed to be "bonded" to the essentially square planar complex – in one case (NIMAHP) [14] the oxygen of a perchlorate ion lies at a distance of "only" 2.77 Å from the nickel, and the authors were undecided as to whether this constituted a "bound" or an "ionic" perchlorate. Interestingly, all six of these compounds are nickel complexes, perhaps reflecting the relative difficulty of synthesising macrocyclic complexes containing ions from beyond the first transition series.

(ii) Four cases exhibit intermolecular contacts of less than 3.4 Å. TPTAMP [15] is an intercalation complex between chloroterpyridineplatinum(II) and adenosine-5'-monophosphate, where the nitrogen of the intercalated base approaches the platinum of the complex (3.39 Å). TPYPDC [16a] exhibits a "significant interaction" between the palladium of one molecule and a nitrogen atom on a neighbouring one (3.13 Å). OXMPDS and OXOXPD [16b,c] exhibit similar interactions between hydrogen-bonded dimers.

(iii) In seven cases there are intramolecular close approaches of significance. Six of these consist of interactions between the central metal atom and potential ligand atoms on (potentially) multi-dentate ligands (CHESNI (Ni), BARLIP, BOLTOL, BESFOU, MCPMPD, XTZPPD (all Pd)). BOLTOL [17a], for example, is a stereo-chemically non-rigid complex containing a F_6acac (= hexaflouroacetylacetonate) ligand whose one oxygen atom is 2.01 Å from the palladium nucleus, while the other is at a distance of 2.70 Å. The authors point to this bond and describe the structure as "distorted square pyramidal with a weak (apical) interaction". In fact, in all six of these examples the authors are at pains in deciding whether their complexes are to be considered four or five-coordinated, leading to such descriptions as "possibly pseudo-five-coordinate"

(MCPMPD [17b]). The remaining member of this group (BALNOR) contains a close approach between a palladium and an oxygen atom in a tetranuclear complex.

(iv) The complex IMPPPD (Pd) [18] exhibits a bridging iodine atom at a distance of 3.29 Å from the metal nucleus, which the authors refer to as a "short palladium-iodine contact" in an "approximately square planar" complex.

At this point an objection might be raised against the inclusion of macrocyclic compounds, based on the limited ability of these complexes to distort along the Berry coordinate due to the structural rigidity of the macrocyclic ligand. Our rationale for not excluding macrocycles at this stage was that if their geometries were indeed significantly different from those of the other five-coordinate complexes, then this would become apparent from the statistical analysis. This would therefore represent a far more objective criterion for their exclusion, than the simple fact that they are macrocycles would.

Finally, therefore, the total number of entries in the CSD which might be construed as being five-coordinate (or, at least, four-plus-one coordinate) is 104 for nickel, 15 for palladium, 9 for platinum, 25 for rhodium and 26 for iridium. Of these, six nickel, three rhodium and five iridium entries represent dimeric compounds with two geometrically distinct metal atoms, while one of the nickel hits, being tetrameric, contains four independant and distinct nuclei. This leads to a data base containing a total of 196 unique five-coordinate metal centres: 113 with nickel, 15 with palladium, 9 with platinum, 28 with rhodium and 31 with iridium.

Table 3.1.3 lists the interatomic angles and distances and ligand atom types found for the compounds comprising the data base to be used in the subsequent analysis.

TABLE 3.1.3. Refcodes, metal atom, ligand atom types L_1 to L_5, bond distances d_1 to d_5, and bond angles $\theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}, \theta_{23}, \theta_{24}, \theta_{25}, \theta_{34}, \theta_{35}, \theta_{45}$ for the 196 compounds of the data set. Ligand numbering corresponds to that shown in Figure 3.1.1 and developed in the text and in Table 3.1.1.

1	AACANI10	NI	N	N	O	O	O	2.025	2.030	2.010	1.997	1.982
		87.72	100.14	83.92	157.34	102.42	159.76	83.65	97.19	102.16	97.17	
2	ASPHNI	NI	C	AS	AS	AS	P	1.875	2.318	2.314	2.338	2.207
		84.04	85.48	86.38	178.46	119.63	118.10	94.72	120.28	94.36	95.02	
3	ATSZNI10	NI	N	CL	S	S	N	2.131	2.305	2.289	2.315	2.130
		91.32	83.39	94.37	175.87	120.41	133.23	92.65	106.36	93.60	83.75	
4	AZOCNI10	NI	N	CL	CL	CL	N	2.218	2.280	2.302	2.314	2.234
		92.52	88.94	89.32	175.92	119.48	118.23	91.42	122.29	88.19	89.75	
5	BACBEM	NI	P	P	BR	P	P	2.225	2.236	2.587	2.215	2.232
		83.91	92.28	92.48	171.98	105.59	158.94	93.93	95.27	95.74	86.77	
6	BAPKEI	NI	P	P	P	P	C	2.210	2.257	2.252	2.259	2.035
		96.41	96.38	95.77	179.40	119.54	119.77	83.06	117.26	84.14	84.26	
7	BEBPNI	NI	P	BR	BR	P	P	2.227	2.465	2.505	2.183	2.225
		88.57	85.39	98.22	167.40	115.88	123.95	86.86	120.10	86.06	94.07	
8	BEVTAX	NI	P	P	BR	P	P	2.170	2.167	2.689	2.270	2.078
		91.69	100.12	87.48	163.12	103.12	165.85	88.78	90.92	96.19	87.97	
9	BGLNIA10	NI	N	N	N	N	N	1.870	1.864	2.723	1.863	1.880
		81.33	101.89	97.66	173.48	101.95	172.06	98.11	85.97	84.59	82.00	
10	BGLNIA10	NI	N	N	N	N	N	1.881	1.873	2.601	1.867	1.872
		82.02	102.19	97.03	172.91	102.81	170.16	97.39	86.98	84.85	82.34	
11	BIDSUC	NI	P	P	BR	P	P	2.215	2.264	2.542	2.264	2.215
		97.44	93.36	81.57	173.29	98.38	163.24	81.57	98.38	93.36	97.44	
12	BIDSUC	NI	P	P	BR	P	P	2.255	2.296	2.612	2.296	2.255
		96.09	93.60	83.39	172.81	94.12	171.76	83.39	94.12	93.60	96.09	
13	BIDSUC	NI	P	P	BR	P	P	2.242	2.272	2.573	2.272	2.242
		97.63	94.22	81.60	171.56	95.21	169.59	81.60	95.21	94.22	97.63	
14	BIDSUC	NI	P	P	BR	P	P	2.229	2.228	2.542	2.228	2.229
		95.34	91.97	84.18	176.06	96.99	166.02	84.18	96.99	91.97	95.34	
15	BIKBUS	NI	N	N	I	N	N	1.859	1.869	2.841	1.870	1.866
		81.04	95.31	97.86	167.17	97.03	168.84	97.20	94.12	97.52	81.38	
16	BIKLIQ	NI	P	P	CL	N	N	2.167	2.175	2.701	2.035	2.005
		89.12	88.35	90.82	175.98	93.31	165.85	94.18	100.83	93.74	85.41	
17	BIQNIC	NI	CL	CL	N	N	CL	2.306	2.372	2.041	2.035	2.400
		90.22	101.82	91.21	164.02	121.18	157.54	83.32	80.38	94.01	89.31	
18	BIZMEC	NI	N	BR	N	N	N	2.139	2.439	2.023	2.141	2.146
		88.15	98.18	88.81	166.57	120.87	140.00	87.40	99.05	94.98	86.47	
19	BIZMIG	NI	BR	N	N	N	N	2.490	1.969	2.073	2.068	1.944
		88.25	102.80	87.74	163.29	89.63	170.22	89.53	99.96	93.75	91.74	
20	BMPANI	NI	N	BR	N	BR	N	2.047	2.474	2.012	2.493	2.050
		90.67	82.99	93.75	163.66	112.00	149.13	90.47	98.87	81.47	93.61	
21	BMPNIB10	NI	P	BR	P	P	P	2.247	2.515	2.290	2.257	2.244
		83.60	93.29	92.20	167.11	111.08	126.56	83.73	122.35	93.47	93.43	
22	BMPONI	NI	P	BR	P	P	P	2.181	2.457	2.239	2.187	2.180
		86.76	90.40	92.10	172.26	112.24	123.44	85.51	124.31	92.73	91.98	
23	BPPENI	NI	BR	P	BR	P	N	2.332	2.172	2.699	2.172	2.006
		93.38	101.17	93.38	169.77	93.12	169.75	85.99	93.12	89.06	85.99	
24	BPYRNI	NI	N	N	N	N	N	1.853	1.844	2.348	1.844	1.853
		98.60	96.81	80.00	166.39	95.81	168.38	80.00	95.81	96.81	98.60	
25	BTCYTN	NI	S	S	S	S	S	2.170	2.240	2.414	2.168	2.234
		93.68	95.01	92.56	167.22	101.24	156.52	77.13	100.72	95.49	92.62	
26	BUDHEN	NI	I	N	I	N	C	2.627	2.050	2.613	2.038	1.898
		95.58	102.95	95.68	168.81	102.01	151.97	81.89	100.34	88.23	82.11	
27	BUKZAI	NI	S	S	BR	P	P	2.231	2.223	2.645	2.205	2.187
		85.66	102.05	86.28	163.32	98.98	167.53	84.84	91.99	92.96	100.57	
28	BUSNOS	NI	N	N	N	N	N	1.846	1.867	2.285	1.835	1.888
		81.68	93.30	95.34	165.98	93.87	166.09	96.61	99.88	100.70	82.98	
29	CAPPNC10	NI	N	N	CL	N	N	2.052	2.103	2.338	2.054	2.107
		94.97	97.68	91.68	158.18	99.76	159.46	72.06	98.58	101.71	95.20	
30	CEBPNI	NI	C	P	P	P	C	1.816	2.244	2.267	2.250	1.852
		90.20	89.57	92.79	177.03	121.74	126.38	87.75	111.81	93.33	86.71	
31	CENNIC	NI	N	N	N	CL	N	2.131	2.064	2.117	2.495	2.132
		96.36	81.97	92.10	172.73	90.95	170.95	81.13	93.42	91.22	90.86	

```
32 CMBPNI     NI P  C  P  C  P    2.178 1.847 2.321 1.872 2.180
    86.68 100.71  86.10 161.20 107.62 140.40  87.83 111.98  98.10  86.72
33 CMPMNI     NI N  CL N  CL CL   2.099 2.319 2.034 2.318 2.456
    94.90  87.55  90.25 173.79 103.51 144.21  82.63 112.09  98.56  88.48
34 CMPNIM     NI C  P  P  P  C    1.861 2.220 2.293 2.227 1.873
    90.91  89.46  91.57 176.17 116.37 132.78  87.24 110.81  94.37  87.23
35 CMTPPN     NI CL S  S  S  P    2.227 2.269 2.242 2.290 2.114
    92.36  91.76  90.66 178.49 120.56 112.06  88.71 127.15  88.62  87.95
36 CNTPNI     NI C  P  P  P  C    1.836 2.223 2.261 2.223 1.860
    88.61  92.37  87.78 176.49 116.75 127.01  90.03 116.21  91.14  90.46
37 CPEANI     NI CL P  P  P  N    2.169 2.299 2.194 2.223 1.965
    94.15  89.40  93.49 178.21 114.01 118.80  87.59 126.71  89.50  86.03
38 CPEPNI     NI C  P  P  P  C    1.869 2.205 2.290 2.189 1.893
    89.83  93.35  87.43 170.75 112.94 133.52  88.66 113.54  95.65  87.00
39 CPHENI     NI CL CL N  N  CL   2.379 2.394 2.059 2.069 2.318
    81.99  97.58  91.80 162.13 120.78 157.16  89.45  81.73 100.27  90.07
40 CRTNCN01   NI C  C  C  C  C    1.895 1.855 2.140 1.869 1.888
    91.43  97.20  85.95 161.76 101.33 161.26  86.69  97.42 100.96  90.03
41 CYPSNI     NI N  N  I  N  N    1.859 1.874 2.834 1.867 1.898
    81.54 101.88  95.35 164.22  94.83 166.56  97.96  98.61  93.89  81.47
42 DIPHNI10   NI I  P  P  P  N    2.709 2.218 2.218 2.218 2.127
    90.35  90.35  90.35 180.00 120.00 120.00  89.65 120.00  89.65  89.65
43 DMAENI     NI BR N  N  N  N    2.467 2.135 2.135 2.135 2.104
    95.78  95.78  95.78 180.00 119.00 119.00  84.22 119.00  84.22  84.22
44 DMPADP10   NI S  N  N  S  S    2.583 2.025 1.973 2.300 2.419
    89.49  93.31  81.67 165.76  82.45 145.73  97.15 130.84 100.01  85.81
45 DPASNI10   NI I  AS AS AS N    2.994 2.338 2.359 2.351 2.307
    93.77  93.64  98.02 177.94 119.08 119.96  85.84 118.56  84.80  83.91
46 DPENIA     NI I  P  P  P  N    3.019 2.256 2.292 2.255 2.258
    92.70  96.00  98.48 176.28 121.46 115.53  84.13 120.06  84.04  84.67
47 DPENIB     NI I  AS AS AS N    2.861 2.356 2.360 2.349 2.469
    94.93  99.58 100.65 175.76 120.64 114.42  80.90 118.64  82.00  81.90
48 DTBNIT10   NI S  S  S  S  S    2.222 2.223 2.776 2.219 2.212
    77.76  95.48 102.34 172.50 101.20 165.72 100.65  93.02  92.02  77.35
49 EDCRCN     NI C  C  C  C  C    1.867 1.872 2.167 1.868 1.839
    90.05  96.76  87.29 159.31 100.86 159.74  86.40  99.40 103.93  89.02
50 EDCRCN     NI C  C  C  C  C    1.827 1.913 1.992 1.902 1.850
    89.27  92.25  89.49 172.83 107.34 141.17  91.30 111.48  94.40  85.59
51 ESPNIQ     NI S  S  N  S  S    2.422 2.413 2.061 2.392 2.424
    83.03  99.64  92.51 161.30 100.70 149.01  91.66 110.28  98.96  82.82
52 ETBPNI     NI P  CL CL P  P    2.213 2.344 2.350 2.192 2.205
    85.93  85.85  98.15 167.01 113.43 127.41  87.20 119.15  86.79  94.78
53 ICFRNI     NI AS AS I  I  C    2.309 2.330 2.624 2.609 1.817
    93.54  89.37  92.96 174.26 122.08 122.36  91.94 115.20  89.15  82.66
54 IMPCNI     NI P  I  C  I  P    2.224 2.605 1.727 2.605 2.215
    90.16  89.97  90.16 179.67 123.88 112.24  89.66 123.88  90.36  89.66
55 IMPONI     NI P  I  I  P  P    2.189 2.658 2.664 2.166 2.185
    87.92  90.51  94.60 174.28 112.24 125.04  86.41 122.61  90.92  89.29
56 INOPNI     NI P  I  O  N  P    2.198 2.474 2.622 1.984 2.208
    91.98  86.38  88.63 175.02 113.69 170.18  91.94  76.12  94.82  86.97
57 INPNII     NI I  P  N  P  N    2.550 2.210 2.297 2.256 2.030
    88.67  97.95  94.64 176.42 115.69 133.36  87.89 109.88  84.46  87.00
58 IPBNIB     NI N  N  BR N  N    1.881 1.918 2.663 1.877 1.906
    81.92  94.92  95.76 165.12  97.07 165.59  96.86  97.31  99.94  81.71
59 IPESNI     NI P  I  I  S  P    2.192 2.543 2.794 2.190 2.185
    95.93  88.89  82.57 168.81 106.89 144.22  94.44 108.81  92.11  86.56
60 MASONI10   NI O  O  O  O  O    2.005 2.000 1.937 2.004 2.005
    87.07  96.40  87.36 163.15 101.99 151.95  89.20 105.94 100.45  88.25
61 MAZHNI     NI N  N  BR N  N    1.892 1.808 2.791 1.932 1.927
    83.00  94.90  96.90 165.20 103.20 163.70  82.50  93.00  90.90  96.60
62 MAZNIP     NI N  N  O  N  N    1.882 1.949 2.780 1.952 1.947
    84.50  83.00  83.50 177.50  86.00 163.10  97.20 105.00  99.00  94.40
```

```
63 MAZOCT     NI N  CL CL CL O    2.139 2.282 2.339 2.339 2.100
    95.41  91.53  91.53 171.35 115.98 115.98  93.24 127.36  84.67  84.67
64 MOPAON     NI N  BR N  BR O    2.013 2.416 2.018 2.398 2.319
    91.83  96.78  94.63 175.31 101.29 149.69  86.23 107.28  87.79  84.97
65 MPGANI     NI N  O  N  N  N    2.068 1.990 2.000 2.001 2.227
    85.43  95.64  95.34 164.86 126.28 131.12  79.51 102.35  94.47  93.47
66 MPNBNI     NI BR BR N  N  BR   2.458 2.468 2.021 2.034 2.649
    94.28  96.87  92.82 173.95 125.59 149.80  83.26  82.45  89.03  86.66
67 MPNCNI     NI CL CL N  N  CL   2.308 2.378 2.033 2.048 2.414
    91.51  98.15  91.51 168.50 121.65 155.40  82.59  82.00  93.35  89.87
68 MPPHNB10   NI P  BR BR BR P    2.263 2.349 2.374 2.339 2.273
    90.30  89.25  90.62 178.76 110.63 132.72  89.65 116.66  89.60  90.33
69 MTRENI10   NI N  N  N  N  N    2.043 2.135 2.077 2.092 1.971
    85.40  85.12  83.69 178.78 111.44 125.97  94.98 120.04  95.81  95.15
70 MTZNIT     NI N  N  N  N  N    2.053 2.058 2.020 1.993 2.096
    89.54  90.32  93.28 174.17 100.83 154.30  86.75 104.69  94.79  88.14
71 NBOSNI     NI S  S  S  S  S    2.139 2.159 3.280 2.168 2.149
    84.20  91.00  91.70 173.00  85.00 171.00  91.30 105.00  94.00  92.00
72 NICEAS10   NI N  N  N  O  O    1.997 2.199 1.977 1.954 1.929
    81.95 104.72  89.32 163.27 105.13 154.33  88.65 100.42  91.04  93.34
73 NICITA     NI O  O  O  O  O    2.001 1.986 2.062 2.053 1.952
    94.41  96.67  86.50 167.85  95.98 168.48  92.95  95.32  92.17  84.38
74 NIDCTP     NI S  S  S  S  S    2.235 2.234 2.428 2.156 2.151
    77.52  98.50  92.64 160.27 100.22 159.69  92.27  98.78  99.91  91.49
75 NIDSPI     NI I  I  S  P  S    2.514 2.567 2.789 2.120 2.189
    94.66  94.95  88.27 174.32 107.42 174.70  87.54  76.67  89.37  89.14
76 NIEACL10   NI O  N  N  O  N    1.975 2.167 2.158 1.974 2.087
   101.23 109.01  77.00 160.55 111.93 121.90  85.06 123.60  85.03  84.04
77 NIHNPB     NI H  P  P  P  N    1.570 2.197 2.213 2.207 2.063
    91.45  98.87  83.03 171.05 120.77 119.52  88.10 119.58  89.02  89.42
78 NIPPNS     NI S  P  P  P  N    2.336 2.247 2.248 2.249 2.316
    96.25  96.25  96.25 180.00 118.86 118.82  83.75 118.81  83.75  83.75
79 OXPHAD10   NI P  P  P  P  P    2.138 2.164 2.207 2.198 2.147
    91.49  89.27  89.21 177.87 119.37 117.46  90.12 123.17  88.73  91.29
80 PASNIB10   NI C  AS AS AS N    1.874 2.323 2.393 2.298 2.100
    92.60 101.13  88.26 173.28 113.76 122.30  87.65 122.58  84.88  85.93
81 PCNTNI     NI S  S  P  S  S    2.254 2.308 2.354 2.264 2.338
    88.19 101.24  83.84 150.78 103.90 149.29  82.90 106.73 107.90  89.74
82 PCNTNI     NI S  S  P  S  S    2.312 2.284 2.344 2.304 2.276
    89.09 107.84  82.58 151.26 105.74 149.30  84.80 104.95 100.85  88.47
83 PEACNI10   NI N  P  P  P  C    2.310 2.266 2.287 2.308 1.974
    84.36  84.09  83.65 173.89 126.16 113.82  94.13 116.81  92.07 102.36
84 PEAMNI     NI N  P  P  P  C    2.109 2.249 2.216 2.233 2.023
    87.05  87.87  86.58 178.01 119.36 117.12  94.93 122.80  91.18  92.47
85 PEANIC     NI C  P  P  P  N    1.740 2.217 2.223 2.205 3.250
   112.30 115.10 110.60 177.10 106.10 106.30  66.90 105.80  67.70  67.30
86 PEANNI     NI N  P  P  P  N    1.590 2.297 2.302 2.280 3.280
   118.40 113.60 108.50 174.30 103.10 106.50  66.80 105.70  65.80  66.70
87 PEASNI     NI S  P  P  P  N    2.130 2.336 2.278 2.328 2.071
    93.20  91.33  94.23 178.03 121.37 113.73  87.29 124.14  86.79  87.31
88 PEMENI     NI N  P  O  P  N    1.962 2.207 2.480 2.207 1.849
    86.92  80.77  87.56 175.33  96.68 167.85  93.37  93.11 103.81  91.26
89 PHASNI     NI P  AS AS AS N    2.346 2.400 2.411 2.412 2.584
   101.80 101.40  99.60 178.70 115.50 117.10  78.90 116.90  79.30  79.10
90 PMENSE     NI P  P  P  O  O    2.341 2.357 2.299 2.007 1.992
    88.51  92.07  99.19 160.99  90.84 153.23  93.37 114.26 106.80  71.31
91 PPMPNI     NI N  N  I  P  P    1.957 1.969 3.048 2.163 2.169
    94.17  86.91  86.00 175.90  92.44 154.74  82.52 112.77  90.80  98.03
92 PSNPEA10   NI SN P  P  P  N    2.539 2.274 2.288 2.302 2.168
    95.11  93.56  94.61 179.26 121.95 115.95  85.38 120.34  86.65  84.67
93 PSNPEA10   NI SN P  P  P  N    2.572 2.287 2.303 2.243 1.979
    90.74  93.30  95.38 174.21 121.73 116.80  84.66 120.58  86.17  89.85
```

```
 94 PTNNIB10  NI P  P  I  S  S   2.216 2.217 2.650 2.255 2.235
    99.87  91.96  86.27 166.79  92.98 163.05  85.05 102.66 100.07  85.79
 95 PTNNIB01  NI S  S  I  P  P   2.248 2.247 2.638 2.188 2.203
    85.13 103.56  86.57 161.21 100.46 166.53  84.72  91.78  93.81 100.16
 96 PYEENI10  NI N  N  N  N  N   2.101 2.078 2.011 2.061 2.123
    86.67 101.17  91.55 162.34  97.51 160.40  85.51 101.98  95.54  90.54
 97 QUMQNI    NI CL CL N  N  CL  2.427 2.404 2.059 2.048 2.302
    81.78  99.36  91.86 160.66 103.83 158.07  92.21  97.89  99.92  86.94
 98 SAIMNI10  NI N  O  N  O  N   2.010 1.953 2.084 1.954 1.995
    90.04  91.03  89.72 176.97 113.44 140.53  88.58 106.02  91.99  89.64
 99 SALDNI    NI N  O  N  O  N   2.033 1.982 2.061 2.009 2.025
    86.71  90.17  88.35 177.26 111.44 145.67  90.66 102.52  91.50  93.40
100 TCAPAN10  NI N  N  N  N  N   2.079 2.221 1.950 2.176 1.964
    82.52  98.55  82.61 160.99  98.87 157.80  93.82  99.57 100.44  94.89
101 TEDTEP    NI S  S  N  S  S   2.426 2.450 2.124 2.441 2.437
    82.93 103.75  89.54 153.17 102.35 155.23  92.62 102.35 103.05  83.52
102 TMCAZN    NI N  N  N  N  N   2.102 2.105 1.954 2.102 2.105
    84.37 102.03  94.03 161.77  96.04 161.77  91.49 102.03  96.04  84.37
103 TMPBNI    NI P  BR BR P  P   2.208 2.427 2.578 2.204 2.205
    86.62  87.70  96.31 167.29 117.57 132.32  86.21 110.11  86.34  96.27
104 TMPBNI    NI P  BR BR P  P   2.205 2.453 2.553 2.194 2.213
    87.45  87.19  95.52 169.29 113.18 134.56  86.49 112.25  87.08  95.03
105 TPENIB10  NI S  P  P  P  P   2.257 2.296 2.298 2.234 2.158
    97.11  95.34  92.53 177.34 114.45 124.81  84.83 118.51  85.47  84.86
106 TPIVNI    NI S  S  S  S  S   2.209 2.234 2.707 2.210 2.217
    76.99  99.05  99.97 168.65  92.56 170.10 103.39  97.24  92.28  77.70
107 AEBXNI10  NI N  N  O  N  N   1.874 1.877 3.019 1.839 1.844
    90.19 101.95  85.16 171.95  87.13 172.28  84.89  87.81  84.20 100.39
108 CHESNI    NI S  S  O  S  S   2.214 2.218 3.266 2.196 2.208
    82.61  89.41  97.08 179.27  71.39 174.05  97.93 114.56  91.23  82.33
109 HMACNI    NI N  N  S  N  N   1.924 1.880 3.284 1.879 1.923
    87.52  91.96  92.48 178.25  92.22 174.99  92.89  82.77  89.73  87.27
110 MAZTNI10  NI N  N  O  N  N   1.905 1.853 3.260 1.858 1.900
    92.14 102.61  87.25 178.60  78.45 173.19  86.92  95.05  76.19  93.56
111 NIACTD10  NI N  N  O  N  N   1.881 1.886 3.230 1.881 1.886
    95.80  89.21  84.24 168.63 114.67 168.63  86.40  76.69  79.76  95.80
112 NIHPOR    NI N  N  O  N  N   1.884 1.888 2.937 1.889 1.961
    87.77 107.81  88.66 167.66  77.48 172.22  92.91  97.07  84.33  92.06
113 NIMAHP    NI N  N  O  N  N   1.897 1.877 2.643 1.843 1.902
   101.81  99.36  83.94 162.15  78.95 172.27  89.43  95.11  96.36  86.29
114 BPINPA    PD BR P  BR P  P   2.529 2.269 3.017 2.318 2.273
    87.39  99.21  87.61 162.16  88.86 174.95  92.47  91.23  98.62  92.50
115 BRPINP    PD BR P  BR P  P   2.544 2.311 2.923 2.326 2.287
    86.03 104.77  86.85 155.92  86.33 171.39  95.55  90.82  99.31  92.93
116 EBZPPD10  PD P  BR BR P  P   2.356 2.555 2.936 2.294 2.303
    86.11  87.19  97.97 168.65 103.41 147.80  86.74 108.67  85.92  92.77
117 MPCPDS    PD P  CL CL P  P   2.325 2.434 2.956 2.265 2.345
    89.07  85.70  92.11 166.77  96.32 150.04  89.78 113.62  81.33  95.45
118 PDPASA    PD AS AS AS CL AS  2.408 2.331 2.860 2.331 2.375
    95.89  80.05  89.81 178.15  81.89 173.75  85.68 101.70  99.21  88.68
119 BALNOR    PD O  O  O  N  O   2.058 2.079 2.966 1.913 2.033
    88.61  86.07  92.51 175.03  81.40 172.93  86.42  91.70  92.97  92.40
120 BARLIP    PD P  P  O  O  P   2.321 2.208 2.654 2.110 2.343
    83.64  97.94  96.20 163.14 113.79 171.70  83.17  74.49  97.42  95.34
121 BESFOU    PD BR BR N  N  N   2.395 2.429 3.349 2.010 2.071
    93.13  84.75  93.94 170.91 107.85 172.80  93.80  74.15  98.68  79.03
122 BOLTOL    PD P  O  O  O  O   2.237 2.001 2.692 2.008 2.063
    91.94 108.23  88.27 176.27 102.58 174.19  91.29  82.84  72.85  88.34
123 IMPPPD    PD P  I  I  I  P   2.327 2.638 3.290 2.618 2.331
    90.16  93.25  89.84 175.65  94.39 166.84  88.89  98.74  91.06  90.12
124 MCPMPD    PD N  CL N  N  N   2.014 2.309 2.938 2.037 2.035
    91.65 102.15  87.91 176.95  76.76 167.04  88.86  90.67  75.04  90.91
```

```
125 OXMPDS     PD N  N  N  N  N   2.035 1.968 3.331 2.021 1.980
     77.76 100.94 107.71 173.64  92.18 174.40  96.03  85.65  77.72  78.47
126 OXOXPD     PD N  N  N  N  N   2.020 1.982 3.344 2.028 1.989
     78.97  83.02 107.53 174.64  84.34 173.49  95.71  96.64  97.17  77.79
127 TPYPDC     PD CL N  N  N  N   2.314 2.045 3.133 2.105 1.962
    100.92  94.15  96.26 170.91 101.39 159.01  79.35  89.23  94.71  81.79
128 XTZPPD     PD S  N  O  S  N   2.261 2.086 3.343 2.267 2.083
     87.66 100.91  84.22 164.67  76.61 167.05  97.73  95.03  66.76  87.76
129 BENLUB     PT SN SN SN SN SN  2.551 2.572 2.572 2.572 2.554
     90.23  90.23  90.23 180.00 120.00 120.00  89.77 120.00  89.77  89.77
130 BIYWEL     PT SN SN P  SN SN  2.626 2.585 2.337 2.595 2.645
     80.27 101.23  80.88 157.20  97.89 155.34  96.94 101.26 101.57  94.30
131 CNPLPT     PT N  N  N  C  N   2.007 2.062 2.761 1.931 2.053
     81.70 114.20  92.20 175.30 101.00 172.60  95.00  85.00  69.70  90.80
132 DCPTAC     PT CL CL S  S  S   2.326 2.331 2.860 2.245 2.245
     92.57  96.73  87.50 177.51 101.31 174.42  89.13  84.22  84.72  90.64
133 EBZPPT10   PT BR P  BR P  P   2.541 2.341 3.141 2.310 2.257
     85.35  98.87  86.88 154.53  86.24 166.91  98.58  84.57 106.47  92.96
134 MBFYPT     PT C  N  N  C  N   2.069 2.119 2.119 2.018 2.150
     92.48  92.48  88.48 178.47  83.11 157.08  86.37 119.75  86.37  92.97
135 MBZPBT10   PT BR P  BR P  P   2.535 2.313 3.027 2.308 2.248
     84.84  97.14  86.23 156.43  85.91 168.56  98.41  88.12 106.36  92.63
136 TMAGEP     PT GE GE GE GE GE  2.410 2.398 2.481 2.424 2.391
     87.78  92.58  90.83 175.70 111.12 141.47  89.67 107.41  91.57  89.01
137 TPTAMP     PT CL N  N  N  N   2.289 1.975 3.392 1.990 1.915
     99.10  90.48  97.39 178.28  77.03 163.51  80.93 103.03  87.84  82.59
138 BCHPIR     IR C  P  BR P  C   1.876 2.327 2.625 2.338 2.131
     93.45  88.34  93.39 176.79  93.67 170.87  86.06  92.57  94.85  86.77
139 BEGKUT     IR P  P  P  H  C   2.374 2.398 2.293 1.524 2.072
     98.76 100.84  81.11 156.19  83.39 166.53  93.74  83.42 100.68  91.34
140 BSCIIR10   IR P  S  I  C  S   2.318 2.373 2.741 1.777 2.384
     86.94  92.67  88.02 167.55 107.53 157.31  80.62  94.78  91.56 103.29
141 BSCIIR10   IR P  S  I  C  S   2.311 2.391 2.711 1.873 2.372
     92.42  92.46  87.96 170.15 102.02 165.95  80.47  91.99  95.68  97.27
142 BUKGAP     IR P  P  P  P  C   2.355 2.313 2.361 2.335 2.134
     94.35  96.25  94.46 178.51  98.92 155.85  86.99 102.41  84.18  84.05
143 CACDUF     IR P  P  P  C  C   2.301 2.278 2.288 2.162 2.090
     92.88  98.92  91.67 167.18  99.29 169.28  92.69  89.57  91.56  80.98
144 CACDUF     IR P  P  P  C  C   2.300 2.280 2.295 2.201 2.088
     95.77  99.61  91.02 167.99  98.92 168.71  91.65  88.80  88.53  80.25
145 CACFAN     IR P  P  P  C  C   2.299 2.282 2.286 2.146 2.096
     94.89  98.57  91.24 166.17  95.50 169.44  91.78  92.06  92.82  80.48
146 CAFKOJ     IR C  C  P  P  P   2.229 1.875 2.262 2.274 2.265
     88.50  84.50  90.60 176.64 125.92 118.78  89.90 114.86  94.05  92.77
147 CCSOPI     IR CL P  S  P  C   2.371 2.359 2.488 2.327 1.963
     87.29  97.38  89.28 172.81  92.57 169.42  92.82  97.80  89.80  89.33
148 CDPMIR     IR C  C  P  C  P   2.167 2.198 2.309 2.154 2.337
     93.40 108.94  83.51 154.11 114.74 157.35  88.70  87.33  93.42  84.71
149 CLTPIR     IR P  CL C  C  P   2.344 2.370 2.038 1.987 2.325
     89.68  92.21  88.40 176.87 118.15 120.99  87.58 120.86  90.40  91.72
150 CMPIRP     IR P  C  C  C  P   2.345 1.806 1.890 1.941 2.295
     91.89  80.32  93.02 177.43 125.58 115.87  87.24 118.26  98.23  89.53
151 CNOPIR     IR P  CL N  C  P   2.407 2.344 1.972 1.858 2.408
     88.90  90.07  91.14 175.71 101.30 161.29  86.85  97.41  91.28  92.72
152 ETCOIR     IR C  C  P  C  C   1.896 1.940 2.347 2.068 2.070
     90.87 100.59  92.12 160.95 101.87 157.67  93.14  99.32  96.80  77.34
153 HCPZIR     IR CL P  N  P  C   2.502 2.348 2.148 2.333 2.002
     87.81  97.27  90.20 174.36  98.99 167.42  93.45  93.59  77.11  89.74
154 HFBIRT     IR S  S  C  C  C   2.361 2.466 1.904 2.083 1.883
     84.52 101.04  90.43 158.78  99.40 164.74  91.76  95.70 100.18  87.75
155 HFBIRT     IR S  S  C  C  C   2.348 2.464 1.912 2.082 1.902
     83.28 103.02  89.43 157.58  97.54 162.74  92.59  99.34  99.36  88.21
```

```
156 HIRPXZ     IR I  P  N  P  C   2.787 2.315 2.074 2.326 2.012
     90.23  93.83  90.30 169.70  96.37 165.19  90.87  98.36  75.87  91.25
157 ICNPIR     IR C  P  N  P  I   1.699 2.366 1.900 2.352 2.667
     90.14 101.13  91.90 157.61  94.04 168.19  87.56  96.98 101.25  86.13
158 IRCNIR     IR P  S  P  C  S   2.252 2.384 2.301 1.887 2.419
     85.08  94.24  90.14 162.12 112.82 148.20  78.24  98.87  98.10 100.63
159 MCPEIR     IR P  C  P  P  P   2.306 1.945 2.345 2.324 2.338
     89.26  97.59  83.23 176.23 106.91 132.70  87.16 120.36  84.59  98.35
160 NPHZIR     IR P  N  P  C  C   2.401 1.912 2.322 1.749 2.091
     88.57  99.52  96.58 159.80 112.71 152.59  77.51  93.03  99.34  89.59
161 NRBIRB     IR CL P  C  C  C   2.368 2.391 1.971 2.141 1.831
     84.50  99.21  85.68 170.44  91.15 167.75  92.46  97.63  89.89  96.06
162 POIRID     IR C  C  P  P  P   2.067 2.133 2.301 2.280 2.307
     81.44  88.28  91.91 167.59  93.54 165.28  88.92  99.41 100.16  95.63
163 PTCIRB10   IR C  S  P  P  S   1.859 2.307 2.333 2.330 2.377
     93.18  90.70  93.86 165.30 119.87 124.79  72.17 114.71  95.52  95.60
164 TAZIRP10   IR P  C  P  N  N   2.384 1.807 2.342 1.940 1.971
     89.17 107.96  93.10 153.26  88.75 161.22  96.02 108.17  98.38  73.98
165 TCTPIR     IR P  C  C  C  P   2.367 1.945 1.906 1.863 2.367
     92.07  90.17  88.02 175.40 111.94 123.35  92.07 124.71  90.17  88.02
166 TCTPIR     IR P  C  C  C  P   2.373 1.969 1.930 1.872 2.373
     88.30  93.27  88.91 173.35 110.35 130.11  88.30 119.54  93.27  88.91
167 BEZPIF     IR S  S  C  C  P   2.370 2.526 1.616 2.118 2.311
     87.17  99.07  85.80 168.47 113.09 140.58  89.30 106.32  92.41  89.99
168 BEZPIF     IR S  S  C  C  P   2.387 2.518 1.964 2.072 2.305
     90.76  96.94  86.39 169.73 104.99 143.88  86.49 111.11  93.33  89.99
169 ACPLRH     RH P  P  C  CL P   2.388 2.285 1.971 2.391 2.384
     93.70  97.03  82.29 167.90  91.28 156.17  95.95 112.49  90.05  85.94
170 BARDED10   RH CL P  P  P  N   2.363 2.342 2.442 2.324 2.134
     94.70  94.00  93.71 179.95 100.68 151.99  85.32 105.32  86.04  86.25
171 BINNUH     RH N  N  C  N  N   2.031 2.039 1.962 2.031 2.026
     89.20  94.82  89.66 174.60  90.22 174.52  90.28  95.21  90.56  90.35
172 BNPORH     RH BR BR N  P  P   2.538 2.539 2.038 2.272 2.327
     87.80  89.30  90.66 176.54  90.88 165.70  93.06 103.32  94.03  87.69
173 BOXNEH     RH P  P  C  C  P   2.379 2.364 1.986 2.055 2.354
     99.46  91.85  82.03 163.01 100.68 168.39  96.54  90.75  90.75  81.16
174 BOXNEH     RH P  C  C  O  P   2.355 1.984 1.991 2.235 2.363
     86.60  88.56  93.63 173.13  91.69 162.45  87.49 105.86  88.11  93.06
175 BSAIRH     RH I  S  C  P  S   2.671 2.403 1.993 2.292 2.347
     98.79  89.57  90.97 165.23 106.14 160.61  80.27  90.56 104.90  86.04
176 BZPRRH10   RH CL CL C  P  P   2.363 2.388 1.995 2.290 2.272
     89.54 105.20  87.76 163.48 101.03 169.75  85.69  89.22  91.24  94.15
177 CCRHAF     RH P  P  S  C  P   2.364 2.346 2.432 1.884 2.363
     91.78  92.98  89.35 177.50  92.53 173.53  88.73  93.77  89.45  89.87
178 CLPSRB10   RH P  C  C  CL P   2.346 2.005 1.977 2.437 2.349
     87.03  87.51  97.65 170.63  91.43 161.02  83.69 107.09  91.33  91.58
179 CPSORH     RH CL P  S  P  P   2.412 2.352 2.325 2.345 2.242
     89.04 101.15  88.24 165.90  94.65 168.46  90.86  96.88  92.92  89.05
180 CPZRHF10   RH CL P  N  P  P   2.403 2.368 1.961 2.353 2.274
     88.53  98.95  88.55 165.18  93.57 172.61  88.95  93.58  95.78  92.15
181 CSMRHC     RH SB CL C  C  SB  2.586 2.381 2.000 1.964 2.584
     88.08  91.50  91.20 178.25 139.70 143.10  90.18  77.20  89.90  90.30
182 CSTPRH     RH CL P  S  P  C   2.356 2.371 2.450 2.367 1.847
     86.64  99.39  89.19 170.91  93.55 167.56  91.67  98.69  89.63  90.61
183 DPMCRH10   RH P  CL C  C  P   2.327 2.575 2.104 1.832 2.326
     86.28  95.82  89.83 167.55 102.89 144.35  89.28 112.76  96.50  87.02
184 DPMCRH10   RH P  CL C  C  P   2.340 2.607 2.034 1.836 2.339
     91.63  89.00  89.05 176.72 103.85 137.80  89.08 118.35  93.94  88.27
185 DPMRHD     RH P  CL C  C  P   2.327 2.590 2.051 1.840 2.324
     88.49  93.70  89.01 171.84 102.63 138.83  88.81 118.54  94.40  87.95
186 DPMRHD     RH P  CL C  C  P   2.335 2.523 2.096 1.813 2.340
     85.96  95.98  89.80 166.55 103.60 146.00  86.93 110.39  96.78  89.73
```

```
187 DPPBRH     RH P  P  C  C  P    2.361 2.433 1.922 1.886 2.366
     97.55  84.01  86.13 165.69 111.04 118.11  96.52 130.70  88.48  89.59
188 DPPCRH     RH P  P  P  C  P    2.322 2.359 2.383 1.894 2.306
     70.16 104.39  96.80 169.27  98.09 142.70 100.69 119.20  70.67  93.90
189 FPHPRH     RH C  O  O  C  N    2.164 2.217 2.236 2.062 2.106
     87.90 109.05  81.11 162.89  80.85 159.39  85.69 119.19  85.57  99.83
190 IPMARH     RH I  S  C  P  S    2.709 2.323 2.005 2.324 2.269
     88.08  99.52  90.90 161.95  96.37 169.62  88.51  93.98  98.47  89.28
191 NEPCRH10   RH CL P  N  P  P    2.408 2.374 1.909 2.374 2.281
     88.02 104.23  88.02 159.35  94.64 170.55  90.35  94.64  96.43  90.35
192 NSTPRH     RH P  O  N  P  O    2.285 2.078 1.913 2.324 2.090
     95.55  95.09 101.00 163.57  96.69 156.01  68.98  98.99  92.32  92.28
193 PRPEMR10   RH S  S  C  P  P    2.328 2.312 2.003 2.349 2.345
     87.10 102.11  87.27 165.10 100.72 166.64  85.82  92.31  92.08  96.76
194 BSAIRH     RH I  S  C  P  S    2.671 2.403 1.993 2.292 2.347
     98.79  89.57  90.97 165.23 106.14 160.61  80.27  90.56 104.90  86.04
195 PFPRHB     RH P  P  P  P  H    2.154 2.351 2.339 2.326 1.609
     99.84  98.02 100.03 169.60 113.79 118.18  80.31 120.33  91.36  71.14
196 RCOHPH     RH C  P  P  P  H    1.829 2.338 2.316 2.315 1.607
     94.75 104.02  97.80 168.40 115.84 120.54  74.61 116.72  85.12  84.07
```

3.2. Description of Molecular Geometry

(a) Introduction

A number of ways of describing the observed structure of a molecule or molecular fragment have been devised. They differ from each other in terms of the particular applications for which they have been developed. Thus, for example, molecules within a unit cell have classically been defined by the positional coordinates of their atomic nuclei, with the origin of the coordinate system being made coincident with the origin of the cell. A slightly modified version of this, introduced when analysing some aspect or other of the *isolated* molecule, has been the definition of the nuclear positions within an inertial system referred to the centre of mass of the molecule, the coordinate axes being coincident with the principal axes of inertia.

Yet another, very commonly used method describes molecular geometry with the aid of internal coordinates, i.e., intra-molecular inter-nuclear distances and angles, and this has enabled a ready comparison between two similar molecules or fragments from separate crystalline environments, which could not, of course, have been compared with each other simply on the basis of their Cartesian coordinates. In the case where a given molecular conformation can be related to some reference geometry, usually of high symmetry, a comparison of the internal coordinates of the observed structure with those of the reference molecule can give some insight into both the *type* and *degree* of the molecular distortion.

The concept of internal coordinates has proven to be particularly useful in the analysis of molecular vibrations, in that the normal modes of vibration of a molecule can be regarded as the resultant of a number of instantaneous atomic displacements along some basis vectors related to the internal coordinates. For example, a normal mode corresponding to a bond stretching vibration may be regarded as resulting from a displacement of two atoms in opposite directions along a vector coincident with the bond joining the two. Furthermore, of course, the individual normal modes each form a basis for an irreducible representation of the molecule.

A fourth method, which in fact has its historical roots in vibrational analysis, where it has also found most of its applications, is that of *symmetry adapted linear combinations of positional or internal coordinates*, also known as *symmetry coordi-*

nates. These are linear combinations of the internal coordinates of the molecule, chosen in such a way that they are orthonormal and form bases for irreducible representations of the symmetry group of the molecule.

Symmetry coordinates are essential to normal coordinate analysis; [19] they enable the factorisation and consequently the solution of the secular equation

$$|\mathbf{FG} - \lambda I| = 0$$

wherein $\mathbf{F}$ represents an inverse matrix of force constants, which consequently reflects the potential energies of the vibrations, $\mathbf{G}$ is a matrix composed of the masses and spatial relationships of the atoms, thereby reflecting the kinetic energy component, I is the identity matrix and λ is defined by

$$\lambda = 4\pi^2 c^2 \nu^2$$

Since the symmetry coordinates in general are chosen so as to transform according to the irreducible representations, each symmetry coordinate will therefore belong to a specific symmetry species of the molecular point group. Furthermore, the symmetry coordinates are incorporated into the $\mathbf{F}$ and $\mathbf{G}$ matrices in the determinantal equation above in order to block factor the product matrix $\mathbf{F} \cdot \mathbf{G}$ and hence simplify the secular equation. Consequently the various solutions of λ (and therefore the frequencies, ν, of the corresponding vibrations) can be related back to particular symmetry species, eg. A_1, B_1, E etc., since these have been included in the $\mathbf{F}$ and $\mathbf{G}$ matrices in the form of the symmetry coordinates. In this way the correspondence between an observed vibrational spectrum and the vibrational modes of the molecule giving rise to it can be established.

Whereas symmetry coordinates have historically been applied mainly to the analysis of vibrational problems in the manner outlined above [19,20] Bürgi [4] together with Murray-Rust and Dunitz [21–23] has recently begun applying this method of describing molecular geometry, and has used it for the analysis of static distortions of molecular fragments.

According to this approach the distortion of an observed structure away from a more symmetrical reference molecule, with the same atomic connectedness or constitution, is described in terms of a *total displacement vector* $\overline{D} = d_j p_j$, where the

d'_js are components along some set of displacement coordinates, p_j . Consider, for example, a hypothetical triatomic molecule whose internal coordinates are r_a, r_b and θ_{ab} , i.e., the two interatomic distances and the enclosed bond angle, respectively. Suppose further that this structure is to be compared to a reference molecule whose internal coordinates are $r_1 = r_2 = r_{ref}$ and $\theta = \theta_{ref}$. Then, for the observed structure, the displacement components (d_j) along the three coordinates (p_j) spanning the parameter space of the reference molecule, viz r_1, r_2 and θ , would be $(r_a - r_{ref}), (r_b - r_{ref})$ and $(\theta_{ab} - \theta_{ref})$, respectively. These, then, are the components of the displacement or distortion vector $\overline{D}$, whose origin in the three dimensional space is the representative point for the reference molecule, and whose tip corresponds to the point representing the observed structure. Of course, the *length* of this vector can give some indication of the relative distortion away from the reference point. Moreover, where two observed structures are to be compared with each other, the angle between their respective distortion vectors may indicate the relative orientations of the distortions in the two cases.

Now , instead of using internal coordinates to construct the displacement vector, symmetry coordinates may just as readily be used. However, since the number of internal coordinates sometimes exceeds the number of degrees of freedom, the choice of independent and redundant symmetry coordinates is not always unique. In spite of this possibility, though, the use of symmetry coordinates in describing molecular conformation offers distinct advantages over other methods, such as straight-forward internal coordinates.

In common with the method of internal coordinates it affords a means of comparing observed structures with each other, as well as in relation to an arbitrary, usually highly symmetrical standard. In contrast, though, the use of symmetry coordinates enables a more rigorous and far less subjective analysis of the "type" of distortion which a molecule has undergone, since, instead of focussing subjectively on one or other of the internal coordinates, a change in a given symmetry coordinate relates *directly* to the symmetry of the distortion. This is so because each symmetry coordinate belongs to a specific symmetry species of the point group (since it transforms according to that irreducible representation); a displacement along it therefore indicates directly the symmetry properties of that distortion.

For example, Murray-Rust, Bürgi and Dunitz have investigated two types of distortion from perfect tetrahedral symmetry – those preserving approximate C_{2v} symmetry and those preserving approximate C_{3v} symmetry. [21] They have shown how a distinction can be made between these groups by making use of symmetry coordinates in order to describe the nuclear configuration of MX_4 molecules. Moreover, they cite an example of a sila-alkane which was originally described as approximating a trigonal bipyramidal configuration (with the fifth, axial ligand at infinity) on the basis of the distortion made evident by the internal coordinates. On an examination of the picture painted by the symmetry coordinates, though, a different, probably more plausible interpretation was made, with significantly different chemical implications.

(b) **Construction of Symmetry Coordinates**

Symmetry coordinates, as pointed out earlier, are linear combinations of internal coordinates which transform according to the irreducible representations of the particular point group. They can generally be described by the relationship

$$S_i = \sum_k U_k \cdot r_k = N(a \cdot r_1 + b \cdot r_2 + \cdots + x \cdot r_k)$$

where S_i is the i-th symmetry coordinate composed of a combination of the k members of a symmetrically complete set of internal coordinates r_k, U_k is the coefficient of the k-th internal coordinate and N is the normalisation factor $(= (a^2 + b^2 + \cdots)^{-\frac{1}{2}})$.[24] By symmetrically complete set is meant a set of internal coordinates which is complete for all relevant symmetry operations of the point group, i.e. the members of the set transform amongst themselves.

The general relationship between the symmetry coordinates and the internal coordinates can also be expressed in terms of matrix notation as

$$\overline{S} = \mathbf{U} \cdot \bar{r}$$

where $\overline{S}$ is the column matrix (or vector) composed of the i symmetry coordinates, and $\mathbf{U}$ is the transformation matrix which transforms the column matrix $\bar{r}$ of k internal coordinates into $\overline{S}$.

We will examine the structures of the five-coordinate metal complexes in our data set in relation to the two idealized penta-coordinate molecules – the trigonal

bipyramid (TBP) and the square base pyramid (SQP) with D_{3h} and C_{4v} point group symmetries, respectively. Before deriving the symmetry coordinates for these two species, however, it is necessary to find how their internal coordinates transform, i.e., to find how often the various irreducible representations (symmetry species) appear in the reducible representation.

Using standard group theoretical techniques [19,20,24] the following genuine normal models of vibration may be derived

$$\text{TBP} \; : \; 2A_1' + 2A_2'' + 3E' + E''$$

$$\text{SQP} \; : \; 3A_1 + 2B_1 + B_2 + 3E$$

The contributions to these modes coming from the bond distance and bond angle internal coordinates, i.e., the bond stretches and the bends, respectively, are

$$\text{TBP} \; : \; \text{stretches} \; = 2A_1' + A_2'' + E' \quad \text{bends} \; = 3A_1' + A_2'' + 2E' + E''$$

$$\text{SQP} \; : \; \text{stretches} \; = 2A_1 + B_1 + E \quad \text{bends} \; = 3A_1 + 2B_1 + B_2 + 2E$$

Since in both cases there are fifteen internal coordinates (five distances and ten angles), but only twelve degrees of freedom $(3 \cdot 6 - 6)$, it follows that there will be three redundant coordinates. The symmetry of these can be inferred from a comparison of the twelve genuine modes of vibration with the fifteen contributions from the internal coordinates. Thus, for the TBP there will be three redundancies of A_1' symmetry, while in the case of the SQP two will be of A_1 and one of B_1 symmetry, respectively. Rather than deriving symmetry coordinates from the twelve independent internal coordinates, it is easier instead to derive these from the entire set of fifteen internal coordinates and then, afterwards, search for the redundancies amongst the symmetry species containing the redundant coordinates.

Symmetry coordinates are explicitly described [24] by

$$S_{j\alpha}^{i} = N \frac{\ell_i}{h} \sum_{R} D^{i}(R)_{\alpha\alpha}(Rr_k)$$

wherein $S_{j\alpha}^{i}$ is the j-th symmetry coordinate in the i-th irreducible representation corresponding to an element in the α–th row, N is the normalisation factor, ℓ_i is the dimension of the i-th irreducible representation (= 1 : for all A and B species; = 2 : for all E; = 3 : for all T), h is the order of the group, $D^{i}(R)_{\alpha\alpha}$ is the element in the α–th row and column of a matrix corresponding

to the operation R in the i-th irreducible representation, Rr_k is the k-th internal coordinate (or, alternatively, linear combination of internal coordinates) which is subjected to the operation R, and the sum is taken over all operations R of the point group.

In the case of all one dimensional representations, i.e., all representations of A or B symmetry, the values of $D^i(R)_{\alpha\alpha}$ will simply be equal to the characters (either -1 or +1) of the point group operations in the particular irreducible representation for which the symmetry coordinates are being derived. They are easily obtained from character tables. In the case of two or three dimensional representations, i.e. those of symmetry species E or T, the case is slightly more complicated, in that the characters derived from the character tables represent the *sum* of the individual diagonal elements $(D^i(R)_{\alpha\alpha})$ of the matrices corresponding to the operations R in the particular irreducible representations (i).

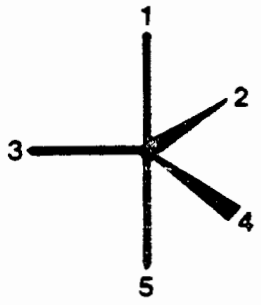
However, the transformation matrices under the E and T representations of the various point groups have been derived and tabulated [25,26], and can consequently be used in order to explicitly derive the $(\ell_j - 1)-$th partner of $S^i_{j\alpha}$ according to

$$S^i_{j\beta} = N\frac{\ell_i}{h}\Sigma_R D^i(R)_{\beta\alpha}(RS^i_{j\alpha})$$

Here $S^i_{j\beta}$ is the j-th symmetry coordinate in the i-th irreducible representation with respect to an element of the $\beta-$th row (i.e., for example, the second coordinate derived from the first $(S^i_{j\alpha})$ of a two dimensional representation), and $D^i(R)_{\beta\alpha}$ is an element in the $\beta-$th row and k-th column of the transformation matrix corresponding to the operation R under the i-th irreducible representation.

Figures 3.2.1 and 3.2.2 show the two reference molecules of D_{3h} and C_{4v} symmetry, respectively, and give the derived symmetry coordinates as well as their symmetry species. A full derivation for both the TBP and the SQP is appended (Appendix 1).

As pointed out previously the distortion of any observed structure from either of these two reference geometries can be expressed in terms of the total displacement vector $\overline{D} = d_j p_j = [d_j(\text{obs}) - d_j(\text{ref})]p_j$, where $d_j(\text{obs})$ and $d_j(\text{ref})$ represent the internal parameters of the observed and reference molecules, respectively. That is, d_j reflects the *displacement* of the observed structure along the displacement



Species A_1'

$$S_1 = 2^{-\frac{1}{2}}(r_1 + r_5)$$

$$S_2 = 3^{-\frac{1}{2}}(r_2 + r_3 + r_4)$$

$$S_{01} = \Delta\theta_{15} \qquad (R)$$

$$S_{02} = 3^{-\frac{1}{2}}(\Delta\theta_{24} + \Delta\theta_{23} + \Delta\theta_{34}) \qquad (R)$$

$$S_{03} = 6^{-\frac{1}{2}}(\Delta\theta_{13} + \Delta\theta_{12} + \Delta\theta_{14} + \Delta\theta_{35} + \Delta\theta_{25} + \Delta\theta_{45}) \qquad (R)$$

Species A''

$$S_3 = 2^{-\frac{1}{2}}(r_1 - r_5)$$

$$S_4 = 6^{-\frac{1}{2}}(\Delta\theta_{12} + \Delta\theta_{13} + \Delta\theta_{14} - \Delta\theta_{25} - \Delta\theta_{35} - \Delta\theta_{45})$$

Species E'

$$S_{5a} = 6^{-\frac{1}{2}}(2r_3 - r_2 - r_4)$$

$$S_{5b} = 2^{-\frac{1}{2}}(r_2 - r_4)$$

$$S_{6a} = 6^{-\frac{1}{2}}(2\Delta\theta_{24} - \Delta\theta_{34} - \Delta\theta_{23})$$

$$S_{6b} = 2^{-\frac{1}{2}}(\Delta\theta_{34} - \Delta\theta_{23})$$

$$S_{7a} = 12^{-\frac{1}{2}}(2\Delta\theta_{13} - \Delta\theta_{12} - \Delta\theta_{14} + 2\Delta\theta_{35} - \Delta\theta_{25} - \Delta\theta_{45})$$

$$S_{7b} = \frac{1}{2}(\Delta\theta_{12} - \Delta\theta_{14} + \Delta\theta_{25} - \Delta\theta_{45})$$

Species E''

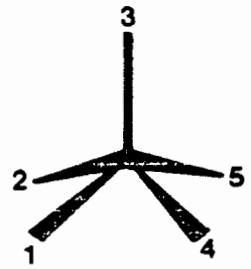
$$S_{8a} = 12^{-\frac{1}{2}}(2\Delta\theta_{13} - \Delta\theta_{12} - \Delta\theta_{14} - 2\Delta\theta_{35} + \Delta\theta_{25} + \delta\theta_{45})$$

$$S_{8b} = \frac{1}{2}(\Delta\theta_{12} - \Delta\theta_{14} - \Delta\theta_{25} + \Delta\theta_{45})$$

Figure 3.2.1. Diagram of idealized TBP showing numbering system, and corresponding symmetry coordinates. Coordinates marked (R) are considered redundant.

---

coordinate p_j . For this reason it is the *difference* (Δ) between the angles of the



Species A_1	$S_1 = r_3$	
	$S_2 = \frac{1}{2}(r_1 + r_2 + r_4 + r_5)$	
	$S_3 = 2^{-\frac{1}{2}}(\Delta\theta_{15} + \Delta\theta_{24})$	
	$S_{01} = \frac{1}{2}(\Delta\theta_{13} + \Delta\theta_{23} + \Delta\theta_{34} + \Delta\theta_{35})$	(R)
	$S_{02} = \frac{1}{2}(\Delta\theta_{12} + \Delta\theta_{14} + \Delta\theta_{25} + \Delta\theta_{45})$	(R)
Species B_1	$S_4 = \frac{1}{2}(r_1 + r_5 - r_2 - r_4)$	
	$S_5 = 2^{-\frac{1}{2}}(\Delta\theta_{15} - \Delta\theta_{24})$	
	$S_{04} = \frac{1}{2}(\Delta\theta_{13} + \Delta\theta_{35} - \Delta\theta_{23} - \Delta\theta_{34})$	(R)
Species B_2	$S_6 = \frac{1}{2}(\Delta\theta_{12} + \Delta\theta_{45} - \Delta\theta_{14} - \Delta\theta_{25})$	
Species E	$S_{7a} = 2^{-\frac{1}{2}}(r_1 - r_5)$	
	$S_{7b} = 2^{-\frac{1}{2}}(r_4 - r_2)$	
	$S_{8a} = 2^{-\frac{1}{2}}(\Delta\theta_{13} - \Delta\theta_{35})$	
	$S_{8b} = 2^{-\frac{1}{2}}(\Delta\theta_{34} - \Delta\theta_{23})$	
	$S_{9a} = 2^{-\frac{1}{2}}(\Delta\theta_{12} - \Delta\theta_{45})$	
	$S_{9b} = 2^{-\frac{1}{2}}(\Delta\theta_{14} - \Delta\theta_{25})$	

Figure 3.2.2. Diagram of idealized SQP showing numbering system, and corresponding symmetry coordinates. Coordinates marked (R) are considered redundant.

idealized molecules and the observed ones which appear in the expressions for the symmetry coordinates. Similarily, r_1, r_2 etc. refer to the *deviations* of the observed bond lengths from some standard ones. In practice, however, the actual values of d_j(ref) are important only for the totally symmetric coordinates, since they cancel out for all others.

The choice of redundant (or dependant) coordinates, as indicated in Figures 3.2.1 and 3.2.2, was based on the following criteria. First, we have already shown that in the case of the TBP the redundancies are all three of A_1' symmetry, while for the SQP two are of A_1 and one of B_1 symmetry. Therefore the search for redundant coordinates was limited to these representations. Second, focussing attention on S_{02} and S_{03} for the TBP, it can be seen that for infinitesimal angular displacements, the positive and negative deviations from $120°$ and $90°$, respectively, will cancel, so that these two symmetry coordinates can be considered redundant. Third, the choice between S_1, S_2 and S_{01} was made on the basis of wanting to retain all five of the coordinates involving bond length changes. Moreover, S_{01} may easily be seen to be dependant on the displacement along S_4 or S_7, i.e. S_{01} is a function of S_4 and S_7, whereas S_1 and S_2 are clearly independant of any other coordinates. Fourth, in the case of the SQP the coordinates S_{01} and S_{02} can clearly be seen to be dependent on S_3 as well as on each other. The same argument applies to S_5 and S_{04}.

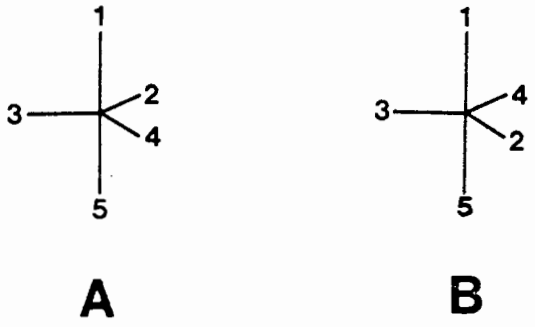
The above symmetry coordinates have been chosen in a self-consistent way, that is, the particular linear combinations of r's and θ's transforming as two dimensional representations match in pairs. For example, S_{5a} and S_{6a} for the TBP are both transformed into themselves by the same twofold axis, C_2, and mirror planes, σ_h and σ_v. However, the above choice is in no way unique and several other sets of coordinates could have been chosen. As Murray-Rust, Bürgi and Dunitz have pointed out [21], "it is immaterial which set of symmetry coordinates is chosen, so the choice should be made according to convenience." This suggestion was adhered to in our particular choices. Finally, the set of coordinates which we have selected for the TBP is the same as that used previously by Hoskins and Lord [27] in their vibrational analysis of the spectra of PF_5 and AsF_5. This is no coincidence, though, since we wished to investigate, amongst others, the same

chemical phenomenon as they did –exchange of axial and equatorial ligands in the TBP via the Berry mechanism involving a SQP intermediate.

3.3. Ligand Numbering

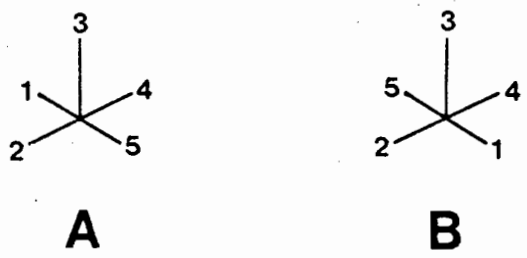
In order to relate the observed structures to the two reference geometries via the symmetry coordinates derived above, we need to superimpose the ligand numbering system (used as the basis for the derivation) onto the arrangement of the five ligand atoms about the central metal atom in the observed molecules. In other words, we want to number those five coordinated ligand atoms in such a manner as to most closely approximate the reference structures, i.e., so as to result in the shortest possible displacement vector. This is no trivial problem, since there are $5! = 120$ distinct ways of distributing the five labels (or sites) characterising the TBP and SQP frameworks amongst these ligands, i.e., there are 120 different *permutational isomers*. However, not all of these would be distinct from each other in the sense of their *absolute* distortion away from either of the reference geometries, although they would correspond to different *orientations* of that distortion.

Consider, for example, a slightly distorted trigonal bipyramidal molecule whose ligands could be labelled in either of, say, the following two ways



In this case the internal coordinates $r_1, r_3, r_5, \theta_{13}, \theta_{35}, \theta_{24}$ and θ_{15} are the same for both (A) and (B), whilst r_2 and r_4, θ_{12} and θ_{14}, θ_{25} and θ_{45}, θ_{23} and θ_{34} would be interchanged in going from (A) to (B). It may be seen that as a consequence of this the symmetry coordinates $S_1, S_2, S_3, S_4, S_{5a}, S_{6a}, S_{7a}$ and S_{8a} are identical for the two isomers, with S_{5b}, S_{6b}, S_{7b} and S_{8b} having identical magnitudes but opposite sign for (A) and (B), respectively.

Clearly, the lengths of the distortion vectors for the permutational isomers (A) and (B) will be identical since they are determined by the squares of their twelve components (the symmetry coordinates). However, their relative orientations in symmetry coordinate space will differ as a result of the different values of the particular coordinates S_{5b}, S_{6b}, S_{7b} and S_{8b} . A similar example may be quoted for a given, slightly distorted square pyramidal molecule whose ligands could be labelled in either of, say, the following two ways.



A **B**

Here different values for the coordinates S_6, S_{8a}, S_{9a} and S_{9b} would emerge, leading to correspondingly different orientations of the deformation vectors, whose lengths, though, would be identical to each other.

On closer examination of the two examples quoted above the following emerges. First, in the case of the TBP permutational isomers whose degrees of distortion are identical will all have *the same pair of axial ligands*, and will differ in the relative arrangements of their equatorial ligands. Second, for the SQP equally distorted permutational isomers will have *the same apical ligand and similar pairs of ligands trans to each other in the base of the pyramid.* In fact, for the TBP it turns out that there are ten possible ways of combining the five ligands into pairs of axial ligands; the number of permutations divided by the order of the point group $(120/12 = 10)$. We will call isomers whose absolute displacements from the reference molecule are equal *distortionally equivalent.* For the SQP there are fifteen groups of distortionally equivalent isomers $(120/8 = 15)$. Figures 3.3.1 and 3.3.2 give one example from each of the ten, respectively, fifteen groups of distortionally equivalent isomers of the TBP and SQP.

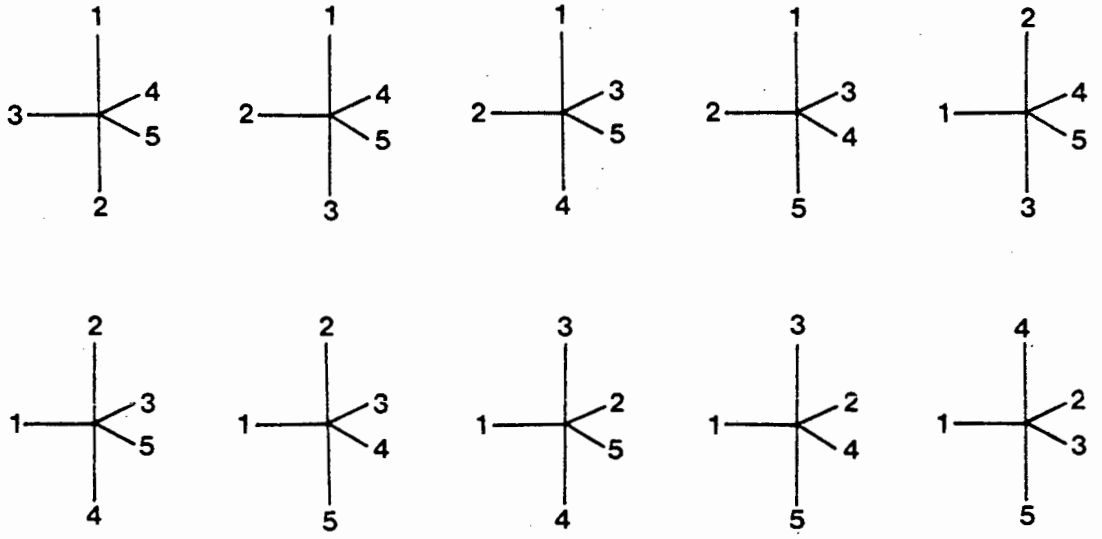


Figure 3.3.1. Diagram showing one member (of twelve) of each of the ten groups of distortionally equivalent TBP isomers. The members shown here all have their equatorial ligands numbered ascendingly in a clock-wise orientation.

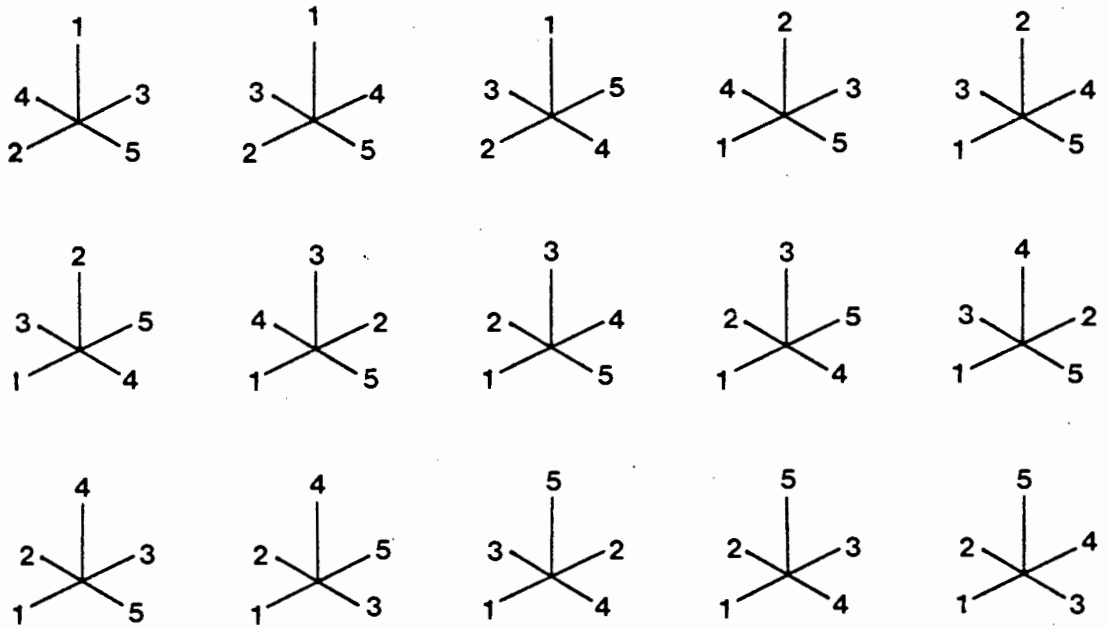


Figure 3.3.2. Diagram showing one member(of eight) of each of the fifteen groups of distortionally equivalent SQP isomers.

Consequently, the ligand numbering problem reduces down to investigating which distortionally equivalent group of isomers the observed structure falls into. We therefore need to fit the observed molecules to only ten or fifteen possible isomers, depending on whether we are attempting to view them as distorted TBP or distorted SQP, respectively, and then find that fit which yields the shortest displacement vector for a given observed structure.

In our case we decided to determine the observed distortion away from the idealized molecules not in the twelve dimensional space spanned by the complete set of symmetry coordinates, but rather in a space of reduced dimensionality defined by the symmetry coordinates composed of bond angle displacements only. The rationale behind this simplification is based on the following consideration. We are mainly interested in distinguishing whether a given structure is closer to the TBP or the SQP, i.e., distinguishing between a trigonal bipyramidal and a square pyramidal description of an observed molecule. For this decision interatomic distances, and hence symmetry coordinates involving these, are irrelevant. For the TBP, therefore, the observed distortion is estimated by the length of the seven dimensional displacement vector whose components are the displacement of the observed structures along the coordinates $S_4, S_{6a}, S_{6b}, S_{7a}, S_{7b}, S_{8a}$ and S_{8b} .

Where an observed structure is to be related to a SQP, however, a complication arises in that the trans-basal angles (θ_{15} and θ_{24}) are undefined for the SQP of C_{4v} symmetry, i.e., the symmetry of the point group allows them to adopt any value between $0°$ and $360°$ – so long as they are identical to each other. For this reason the displacement of the observed structures along the S_3 coordinate cannot be evaluated since there are no values $\theta_{15}(\text{ref})$ and $\theta_{24}(\text{ref})$ from which the deviations ($\Delta\theta_{15}$ and $\Delta\theta_{24}$) of the observed values can be estimated. (Earlier it was pointed out that for all symmetry coordinates apart from the perfectly symmetrical ones, the values of $d_j(\text{ref})$ cancel out – S_3 is of type A_1 , though.) Consequently we decided to estimate the observed distortion by means of a six dimensional vector whose components are the displacements of the observed structures along the coordinates $S_5, S_6, S_{8a}, S_{8b}, S_{9a}$ and S_{9b} . Of course, this is by no means a perfect solution, but short of subjectively hypothesising what the "ideal" trans-basal angle might be, it seems to be the only one.

The procedure for determining that system of numbering of the five ligands in each individual observed structure, which most closely approximates the one used as a basis (Figures 3.2.1 and 3.2.2) may consequently be outlined as follows:

(i) Determine the length of the displacement vector for the particular permutation of the ligands as it appears originally in the data matrix.

(ii) Permute this arrangement through all the distortionally *non*-equivalent groups and calculate the length of the displacement vector at each stage.

(iii) Store that permutation which corresponds to the least distortion.

These calculations and permutations were accomplished by means of a FORTRAN program written specifically for these tasks. A copy of this program is appended (Appendix 2).

3.4. Expansion of Data Set

The task of correctly labelling the ligands in the observed structures is not complete with the identification of the group of distortionally equivalent permutation isomers to which the structure belongs. What the above process has done is to merely identify the pair of axial ligands with respect to a distorted TBP, and to fix the apical ligand and the two pairs of trans-basal ligands for a distorted SQP. (Note: We speak of *axial* ligands in the case of a TBP, and of an *apical* ligand when referring to a SQP.) Within each group of distortionally equivalent isomers there are twelve or eight possible isometric arrangements of the ligands for the TBP and SQP, respectively. These are related to each other by the symmetry operations of the D_{3h} and C_{4v} point groups. Figures 3.4.1 and 3.4.2 illustrate this point highly schematically, since it must be borne in mind, that the twelve, respectively eight, asymmetric units containing the distortionally equivalent isomers exist in seven, respectively six, dimensional space, and not in two dimensional space as shown in the figures.

These Figures also show how the representative point for an observed structure can be found in either of the twelve or, respectively, eight asymmetric units of the TBP or SQP parameter space, while being equidistant from the origin in all cases. The origin in Figure 3.4.2 ought to be seen not as a point, but rather as a line perpendicular to the plane of the diagram and corresponding to the coordinate

S_3 , along which the SQP has a degree of freedom.

Which of the subspaces in the D_{3h} and C_{4v} parameter spaces the representative point for an observed structure is placed into depends essentially on the order in which the ligands appear originally in the data matrix. The reason for this is that the permutation which appears originally is simply one of either twelve or eight distortionally equivalent possibilities, and it therefore fixes the particular asymmetric unit within which the permutations outlined in Section 3.3 above take place. Since the order in which the CSD lists the ligands depends in part on the order in which the original authors listed them it follows that the starting permutations for the algorithm of Section 3.3, and thus the representative points for the least distorted TBPs or SQPs, will be randomly distributed amongst the available subspaces. This has implications for the subsequent statistical analysis of the data point distribution.

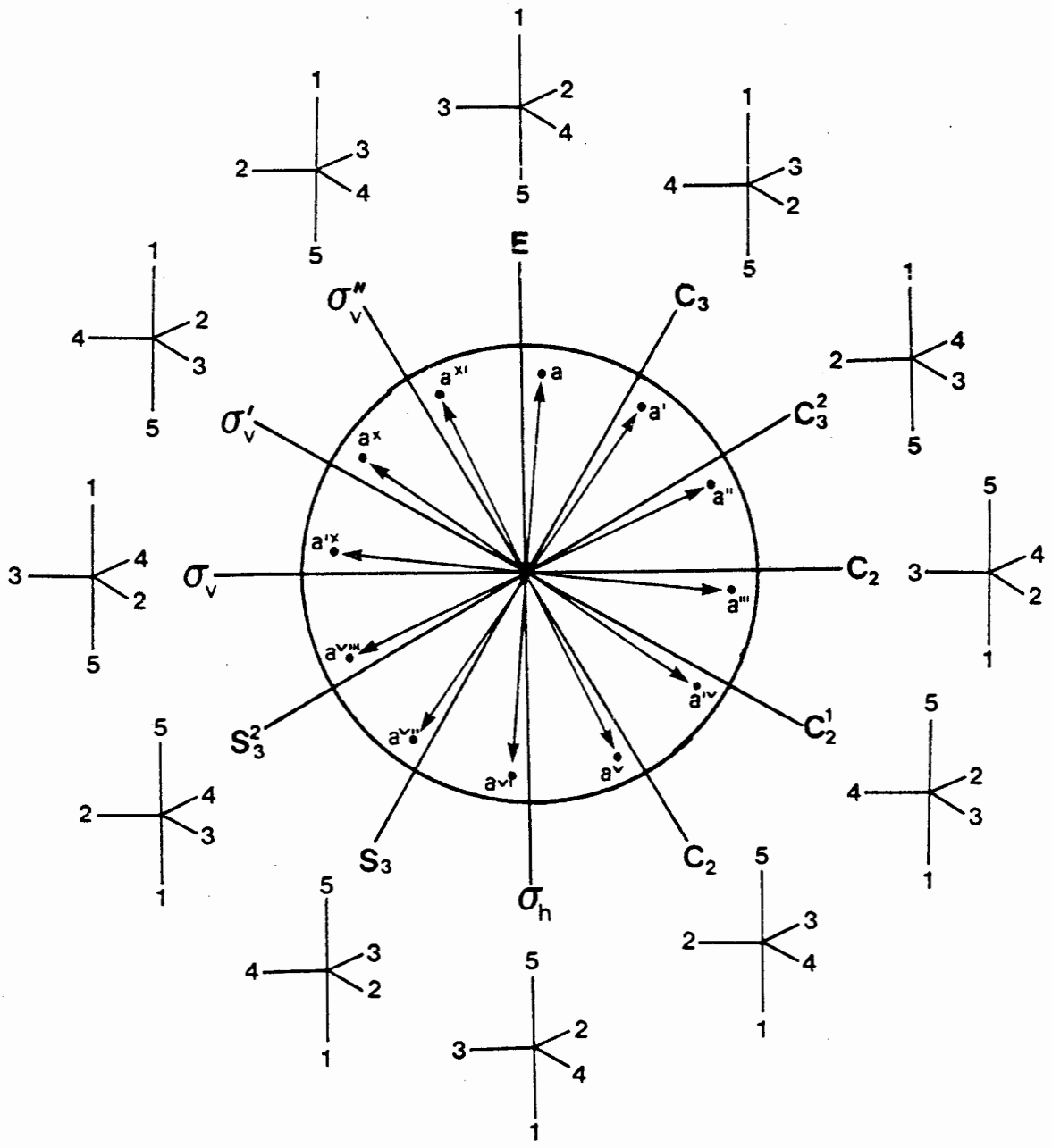


Figure 3.4.1. Diagram showing schematically how the twelve isometric members of a distortionally equivalent group are generated from the D_{3h} reference structure $(= E)$, and how the representative point for an observed structure may be placed $(a, a', a'' \ldots)$ in either of twelve asymmetric units composing the parameter space, while constantly being equidistant from the origin. Note that the pair of axial ligands is constant.

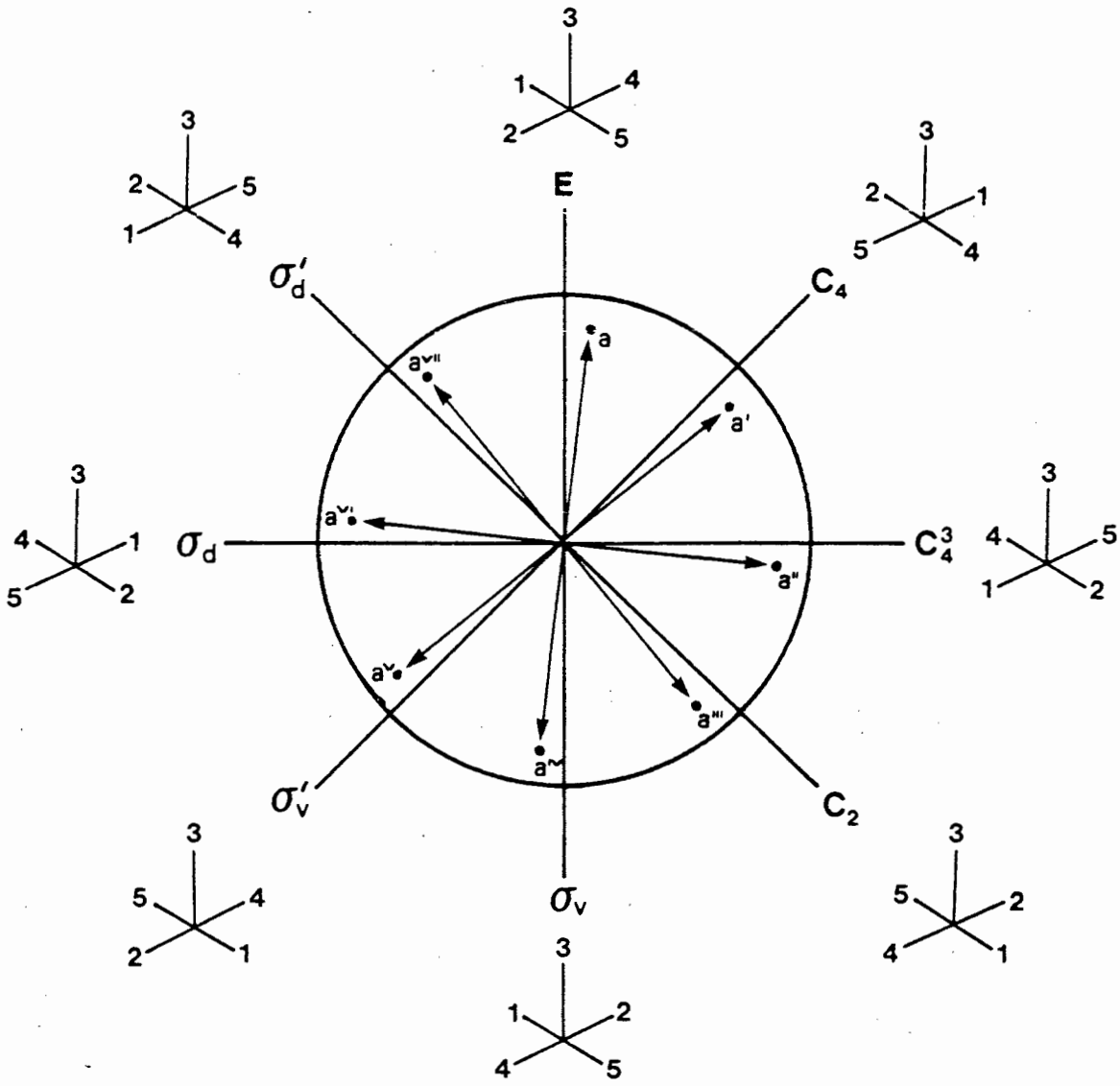


Figure 3.4.2. Diagram showing schematically how the eight isometric members of a distortionally equivalent group are generated from the C_{4v} reference structure $(= E)$, and how the representative point for an observed structure may be placed in either of the eight asymmetric units composing the parameter space $(a, a', a'' \ldots)$, while constantly being equidistant from the "origin" (see text). Note that the apical ligand is constant as are the pairs of trans-basal angles.

Murray-Rust has examined this problem of labelling in some detail, [28] and we will briefly outline his arguments here. Consider the case of a linear triatomic

molecular fragment XYZ, whose structure is completely described by the two interatomic distances between X and Y, and Y and Z, as shown in Figure ~~4.3~~ 3.4.3(a). If we were examining this fragment in a large number of different crystalline or molecular environments in the absence of some systematic method of labelling, then we would expect the *distributions* of r_1 and r_2 to be approximately equal, since there would be no way of distinguishing between them. This is shown in Figure 3.4.3(b). Thus, although there is no symmetry relating r_1 to r_2 in any *particular* fragment, there clearly is symmetry relating the *distributions* of r_1 and r_2. In the ideal case, if we had an infinite number of observations in which we could randomly label the two distances, we would get identical distributions of r_1 and r_2; distributions which would be related to each other by a symmetry operation – the diagonal mirror line $r_1 = r_2$.

Unfortunately, however, in the real case the number of data points is rather limited and the number of symmetry related asymmetric units much larger, so that the observations would be randomly spread through a far greater space than they were in the hypothetical case listed above. For example, in the D_{3h} parameter space there are twelve symmetry related asymmetric units, which would each contain no more than sixteen, on average, of the 196 observations in our data set. Clearly, this does not make a statistical analysis of the data distribution particularly feasible.

In such a case where there are limited data there are two ways out of this predicament. These will be exemplified with reference to the example quoted above. The first is to introduce a labelling scheme based on the geometry of the molecular fragment; one could, say, consistently label the shorter of the two distances r_2. The result of this is to limit all the representative points to one asymmetric unit, as shown in Figure 4.3(c). Although this may seem a good solution, it is that only in the case where none of the data points lie close to the edge of the sub-space. If they do, then the results of the statistical analysis will not reflect the essentially symmetrical distributions of r_1 and r_2, and will consequently paint a distorted picture of the most favoured geometry. Moreover, in more complicated examples, such as ours, it may prove impossible to devise such a unique labelling scheme.

The second approach starts from the realization that the data points are randomly distributed and that, in general, there will be symmetry relationships between

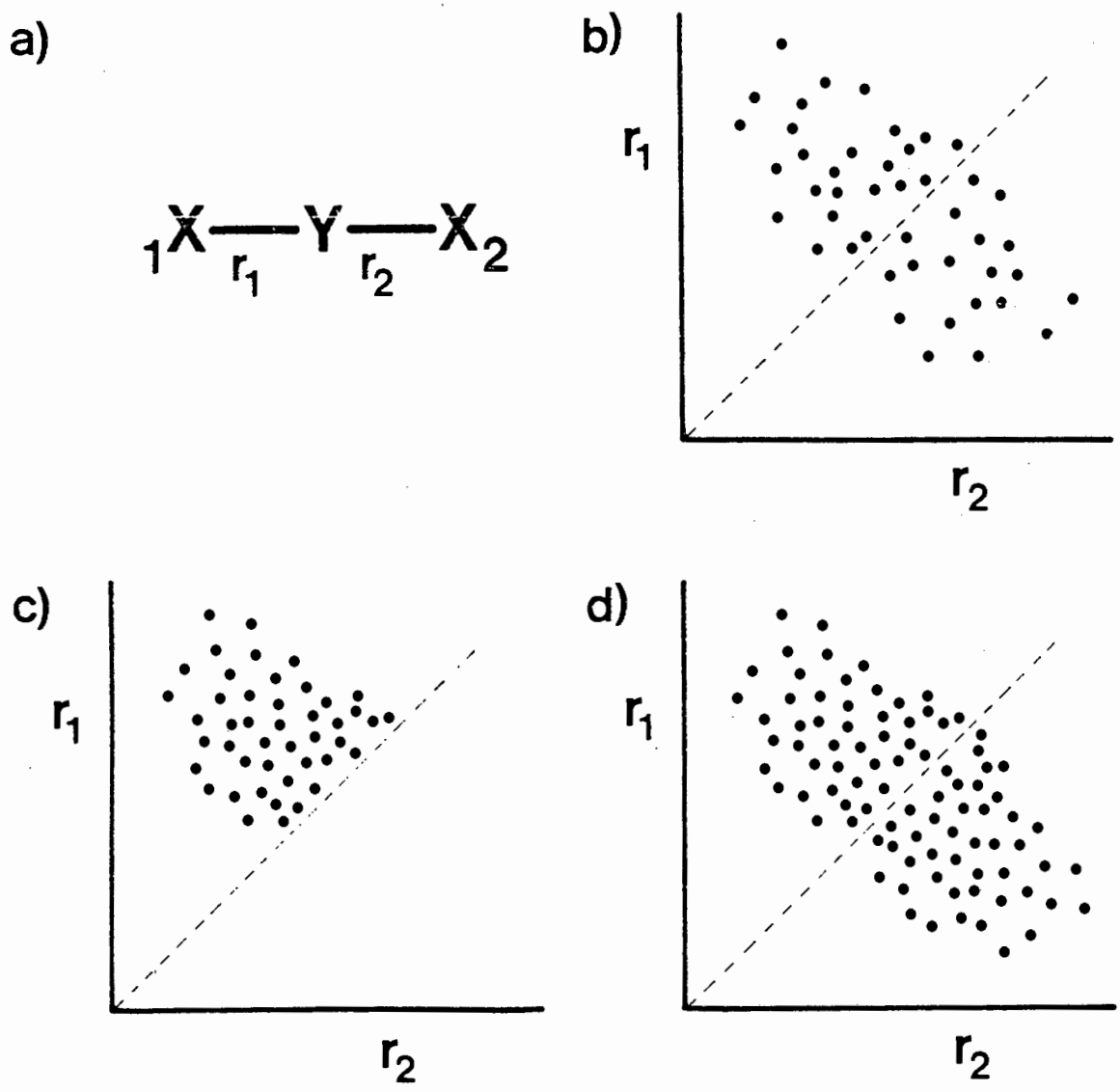


Figure 3.4.3 (a) Parameters describing a linear $X - Y - X$ fragment. (b) Hypothetical scattergram of r_1 and r_2 for a linear XYX fragment whose atoms have been randomly labelled 1 and 2 for each individual point. (c) Hypothetical scattergram of r_1 and r_2 for the same data as in (b), but with the atoms labelled so that $r_1 \geq r_2$. (d) Hypothetical scattergram of r_1 and r_2 for the same data as in (b) and (c), but with the atom labels permuted. This corresponds to the *expansion* of the data set.

---

the distributions of the various parameters within the parameter space. Therefore,

if there is *a priori* knowledge of the symmetry relationships in the n-dimensional case, then the symmetry elements giving rise to these relationships can be used to *generate* the symmetrical arrangements of the data point distribution. For example, Figure 3.4.3(c) can be reflected in the diagonal mirror line to give Figure 3.4.3(d), which obviously represents the true relationship between r_1 and r_2, whereas the former reflects our labelling scheme superimposed onto the observed structures.

In our case the eight or twelve distortionally equivalent isomers in the C_{4v} or D_{3h} parameter spaces, respectively, (hereinafter referred to as *S-space*, for square pyramidal space, and *T-space*, for trigonal bipyramidal space) are related to each other by the symmetry operations of the respective point groups as shown in Figures 3.4.1 and 3.4.2. In other words, by starting off with an arbitrary isomer its distortionally equivalent brothers may be generated by subjecting the original permutation to the symmetry operations of that data space. Similarly, the *complete* data distribution in the twelve dimensional spaces may be obtained by subjecting each of the representative points to the transformations of the symmetry elements.

Murray-Rust [28] and Bürgi and Norskov-Lauritsen [29] have demonstrated how this approach can be used. What it amounts to, in our case, is an expansion of the number of data points by a factor equal to the order of the point groups to which we are referring the observed structure. Thus, if the compounds are to be referred to the TBP, the number of data points is multiplied by 12 (one for each of the twelve operations of the group) while in the case of the SQP the factor is 8. Consequently we end up with 2352 $(= 12 \times 196)$ and 1568 $(= 8 \times 196)$ data points in T- and S-space, respectively!

The expansion of the permuted data set put out by the algorithm described in Section 3.3 was accomplished by a FORTRAN program written specifically for this purpose. A copy of it is appended (Appendix 3).

3.5 Scaling and Standardization

(a) Standard bond lengths

Our data set contains not only a mixture of interatomic distances and angles, but it also mixes atoms of different kinds, i.e., bonds not of the same *type*. Thus it contains bonds between nickel and oxygen, palladium and iodine, platinum and tin

etc. Both these factors can be expected to have considerable impact on any statistical analysis we might attempt; for this reason we need to intelligently standardize and/or scale the data prior to any such analysis.

The first problem is to put all the interatomic distances on a common scale. This is important since we want to be able to compare nickel complexes with those of, say, platinum and, furthermore, since we need to have some measure of "standard" bond lengths, from which the deviations in the observed structures can be calculated in order that we may obtain values for the totally symmetric coordinates S_1 and S_2 in both D_{3h} and C_{4v} parameter space. By "standard" bond length is meant some empirically obtained value for the average distance (in some statistical sense) between a given pair of atoms.

Klebe and Bürgi [30] have examined the results obtained from two different empirical bond strength (n)-bond length (d) relationships, and have found that the "standard" distances between a given pair of atoms derived from the two expressions, were the same (within three standard deviations) in all but three of eighteen cases. The two empirical models were those of Pauling (I) and of Brown and Shannon (II). These relate the standard (d_{io}) and the observed (d_i) values for the i-th type of bond to their respective empirical bond strengths n_o and n_i .

$$\text{(I)} \quad n_i/n_o = exp[-(d_i/d_{io})/c]$$

$$\text{(II)} \quad n_i/n_o = (d_i/d_{io})^{-N}$$

The data set used in that study consisted of four-, five- and six-coordinated Mg, Al, Si and P compounds and values of d_{io}, c and N (both empirical constants) were determined by minimizing the expression

$$\text{(III)} \quad (CN_0 - \sum_{i=1}^{CN} n_i)^2$$

separately for the different elements, but considering all coordination numbers (CN) simultaneously; CN_0 is a reference coordination number and n_0 is taken as 1.

We have proceeded similarily, employing the program written by Klebe and Bürgi, and have obtained values for standard bond lengths between the various metals and their ligand donour atoms for a reference coordination number of five.

The results are presented in Table 3.5.1. The input data consisted of the set of five-coordinate complexes extracted from the CSD as explained earlier, as well as the four-coordinate CSD entries for each type of metal as at July 1984. Just as in the previous study we see that almost invariably the results obtained from expression I (a) are within three standard deviations of those from II (b), although in some cases this means little, since the standard deviations are large.

In general, the results obtained for the nickel, palladium, platinum and rhodium data sets are good as judged by the standard deviations, these being less than 0.05 Angström in many cases. Moreover, in the majority of cases the standard deviations for the reference bond lengths obtained from II are smaller than those from I, most noticeably so in the case of the iridium complexes. Here the large σ^2 values result from the way in which the different types of ligand atoms are combined in the complexes composing the data set; the four-coordinate iridium complexes, for example, contain only carbon and phosphorus ligands. As a consequence of this, the algorithm cannot establish a reference value for, say, a bond to bromine, since it needs to have this type of bond appear in both four and five coordination in order to minimize III over various values of CN .

Also listed in Table 3.5.1 are values for standard bond lengths obtained from Pauling. [31] These empirical values may well be somewhat dated, but they still generally fall into or slightly below the range of values which we have found. The fact that many of Pauling's standard bond lengths are lower than ours may nonetheless be significant, insofar as his often pertain to four-coordinate complexes especially in the case of platinum and palladium, while ours have been derived for five coordination. For the sake of comparison we have also included in Table 3.5.1 standard bond lengths obtained from expression I for a reference coordination number of four.

It may be seen that for palladium and platinum our reference values for $CN_0 = 4$ are generally well within three standard deviations of Pauling's, while for nickel and rhodium they are on the whole slightly smaller. These observations support the suggestion made above. The almost nonsensical values obtained for the standard bond lengths for four-coordinate iridium are the result of the same problem which plagued the derivation of these values for the five-coordinate metal from equation I – an insufficient dataset.

In the light of the overall better results obtained from II, we decided to use these as our reference bond lengths, i.e., as the standard bond length between a given pair of atoms in a five-coordinate complex, to which the observed interatomic distances would be related.

(b) Scaling and angle measurements

The description of the geometry of the observed structures involves the use of both linear and angular displacement coordinates. Consequently there is a mixture of variables of considerably different magnitudes: Angström and degrees. Obviously, the larger the magnitude and/or variation of a given variable, the greater its relative influence on the statistical analysis and the more skewed the results potentially become. In order to overcome this problem it is often good practice to scale the variables.

In our case we decided to express the angular displacements in radians related to a circle of a radius r whose length equals the *average* metal-to-ligand distance in the dataset. In other words, an angle θ measured in degrees would then correspond to a radial displacement α measured in Angström according to

$$\alpha = 2\pi \cdot r \cdot (\theta/360^0).$$

We obtained a mean r of 2.231 Å from the average values for the dataset, of the bond lengths $d_1 = 2.226$ Å, $d_2 = 2.227$ Å $d_3 = 2.388$ Å, $d_4 = 2.152$ Å and $d_5 = 2.161$ Å referred, respectively, to a SQP. The exact value of r is not of extreme importance so long as it is reasonably representative of the distances which appear in the data. As a consequence of this scaling a change in a given angle of ten degrees, for example, would correspond to a radial displacement of 0.390 Angström – a value much closer to the distances or bond distance increments typical of the dataset than the original angular displacement.

We decided to use the scaled, raw data for the cluster and factor analyses, rather than the standardized $z-$scores . The reasons for this were, firstly, that the scaled data are all of similar magnitude and, secondly, because of the somewhat inconsistent manner in which transformation by the z transform treats the variance in the dataset. (This point has already been discussed under the section giving an outline of the statistical techniques used in this study.) This approach is not

unique, and in fact has been employed by Willett [32] in an evaluation of different relocation clustering algorithms to a dataset comprising variables as diverse as molar refractivity and heats of vaporization. Indeed, he points out that transformation of "fragment attributes leads to a noticeable drop in the utility of the subsequent classifications".

Table 3.5.1 Standard bond lengths d_{io} and standard deviations for five coordinate metal complexes. (a) Results for expression I, (b) results for expression II, (c) values taken from Pauling, (d) results for expression I with $CN_0 = 4$, m = number of distances used in refinement, N_{obs} = number of structures in dataset.

		Ni	Pd	Pt	Rh	Ir
H	a)	1.469 (325)			1.262 (390)	0.951 (1758)
	b)	1.370 (296)			1.478 (113)	1.633 (210)
	c)	1.51			1.62	1.62
	d)	I.233 (325)			1.098 (390)	0.314 (1758)
C	a)	1.958 (30)		2.193 (43)	2.100 (28)	1.925 (664)
	b)	1.895 (30)		1.892 (112)	2.077 (29)	1.988 (29)
	c)	1.98		2.08	2.09	2.09
	d)	1.722 (30)		1.989 (43)	1.937 (28)	1.287 (664)
N	a)	2.067 (14)	2.166 (10)	2.221 (11)	2.121 (43)	1.996 (626)
	b)	2.068 (13)	2.263 (18)	2.231 (50)	2.024 (30)	1.910 (79)
	c)	1.91	2.01	2.01	2.02	2.02
	d)	1.831 (14)	2.007 (10)	2.017 (11)	1.957 (43)	1.357 (626)
O	a)	2.038 (27)	2.132 (13)		2.189 (52)	
	b)	2.028 (32)	2.180 (26)		2.165 (48)	
	c)	1.87	1.97		1.98	
	d)	1.802 (27)	1.974 (13)		2.026 (52)	
P	a)	2.236 (26)	2.407 (23)	2.502 (38)	2.333 (50)	2.568 (392)
	b)	2.271 (17)	2.394 (38)	2.531 (144)	2.307 (20)	2.347 (27)
	c)	2.31	2.41	2.41	2.42	2.42
	d)	2.000 (26)	2.249 (23)	2.297 (38)	2.169 (50)	1.930 (392)
S	a)	2.365 (16)	2.478 (13)	2.480 (25)	2.445 (77)	2.310 (656)
	b)	2.333 (19)	2.667 (73)	2.243 (158)	2.339 (54)	2.339 (42)
	c)	2.25	2.35	2.35	2.36	2.36
	d)	2.129 (16)	2.320 (13)	2.276 (25)	2.281 (77)	1.672 (656)
Cl	a)	2.348 (43)	2.489 (14)	2.497 (16)	2.399 (84)	2.127 (1076)
	b)	2.351 (35)	2.513 (78)	2.639 (230)	2.383 (49)	2.483 (98)
	c)	2.20	2.30	2.30	2.31	2.31
	d)	2.112 (43)	2.331 (14)	2.293 (16)	2.235 (84)	1.489 (1076)
Ge	a)			2.420 (64)		
	b)			2.421 (22)		
	c)			2.53		
	d)			2.216 (64)		
As	a)	2.445 (50)	2.419 (41)			
	b)	2.472 (43)	2.448 (40)			
	c)	2.42	2.52			
	d)	2.181 (50)	2.261 (41)			

Br	a)	2.437 (52)	2.558 (34)	2.497 (97)	2.455 (187)	2.264 (1450)
	b)	2.410 (44)	2.645 (58)	2.458 (220)	2.538 (69)	2.628 (204)
	c)	2.35	2.45	2.45	2.46	2.46
	d)	2.201 (52)	2.400 (34)	2.292 (97)	2.292 (187)	1.626 (1450)
Sn	a)	2.688 (188)		2.565 (60)		
	b)	2.597 (212)		2.564 (22)		
	c)	2.61		2.71		
	d)	2.45 (188)		2.361 (60)		
Sb	a)				2.432 (217)	
	b)				2.490 (73)	
	c)				2.73	
	d)				2.267 (217)	
I	a)	2.507 (75)	2.731 (49)		2.208 (505)	2.302 (1096)
	b)	2.546 (54)	2.806 (53)		2.619 (129)	2.636 (116)
	c)	2.54	2.64		2.65	2.65
	d)	2.243 (75)	2.573 (49)		2.044 (505)	1.664 (1096)
m		690	211	245	235	171
N_{obs}		145	49	59	52	35

REFERENCES

1. Allen, F.H., Bellard, S., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.F. and Watson, D.G. *Acta Cryst.*, 1979, **B35**, 2331-2339.
2. A complete list of references to such studies may be obtained from the Cambridge Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England.
3. Bye, E., Schweizer, W.B., Dunitz, J.D. *J. Am. Chem. Soc.* 1982, **104**, 5893-5898.
4. Bürgi, H.B. *Inorg. Chem.* 1974, **12**, 2321-2325.
5. Britton, D., Dunitz, J.D. *J. Am. Chem Soc.* 1981, **103**, 2971-2979.
6. Auf der Heyde, T.P.E., Nassimbeni, L.R. *Acta Cryst.* 1984, **B40**, 582-590.
7. Auf der Heyde, T.P.E., Nassimbeni, L.R. *Inorg. Chem.* 1984, **23**, 4525-4532.
8. Berry, R.S. *J. Chem Phys.* 1960, **32**, 933-941.
9. Holmes, R.R. *Prog. Inorg. Chem.* 1984, **32**, 119-235.
10. Rossi, A.R., Hoffman, R. *Inorg. Chem.* 1975, **14**, 365-374.
11. Burdett, J. *Adv. Inorg. Chem. Radiochem.* 1978, **21**, 113-146.
12. Gillespie, R.J. *J. Chem. Soc.* 1963, 4679-4685.

13.(a) Rosenfield, R.E. Jr., Murray-Rust, P. *J. Am. Chem. Soc.* 1982, **104**, 5427-5430.

13.(b) Rosenfield, R.E. Jr., Parthasarathy, R., Dunitz, J.D. *J. Am. Chem. Soc.* 1984, **99**, 4860-4862.

13.(c) Guru Row, T.N., Parthasarathy, R. *J. Am. Chem. Soc.* 1981, **103**, 477-479.

14. Dewar, R., Fleischer, E. *Nature* 1969, **222**, 372-373.
15. Wong, Y., Lippard, S.J. *J.C.S. Chem. Comm.* 1977, 824-825.

16.(a) Intille, G.M., Pfluger, C.E., Baker, W.A. *J. Cryst. Mol. Struct.* 1973, **3**, 47-54.

16.(b) Endres, H. *Acta Cryst.* 1980, **B36**, 1347-1350.

16.(c) Endres, H., Weiss, J. *Acta Cryst.* 1981, **B37**, 1360-1363.

17.(a) Siedle, A.R., Newmark, R.A., Pignolet, L.H. *J. Am. Chem. Soc.*1982, **104**, 6584-6590.

17.(b) March, F.C., Fergusson, J.E., Robinson, W.T. *J.C.S. Dalton*, 1972, 2069-2076.

18. Bailey, N.A., Mason, R. *J. Chem. Soc. (A)*, 1968, 2594-2605.

19. E.B. Wilson, J.C. Decius and P C. Cross. *Molecular Vibrations – The Theory of Infrared and Raman Vibrational Spectra*, McGraw Hill, London, 1955.

20. F.A. Cotton. *Chemical Applications of Group Theory*, Second Edition, Wiley-Interscience, John Wiley and Sons, New York, 1971.

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

22. Murray-Rust, P., Bürgi, H.B., Dunitz, J.D. *Acta Cryst.* 1978, **B34**, 1793-1803.

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

24. D. Wald. *Gruppentheorie für Chemiker*, 1. Aufl., Weinheim, Deerfield Beach, Florida: VCH, 1985.

25. R.McWeeny. *Symmetry*, Pergamon Press, Oxford, 1963.

26. H.Poulet and J.P. Mathieu. *Vibration Spectra and Symmetry of Crystals*, Gordon and Breach, New York, 1976.

27. Hoskins, L.C., Lord, R.C. *J. Chem. Phys.* 1967, **46**, 2402-2412.

28. Murray-Rust, P. *Acta Cryst.* 1982, **B38**, 2765-2771.

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

30. Klebe, G., Bürgi, H.B. *19th Int. Congr. and Gen. Ass.,IUCR*, Hamburg 1984.

31. L. Pauling. *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, 1940.

32. Willett, P. *J. Chem. Inf. Comput. Sci.* 1984, **24**, 29-33.

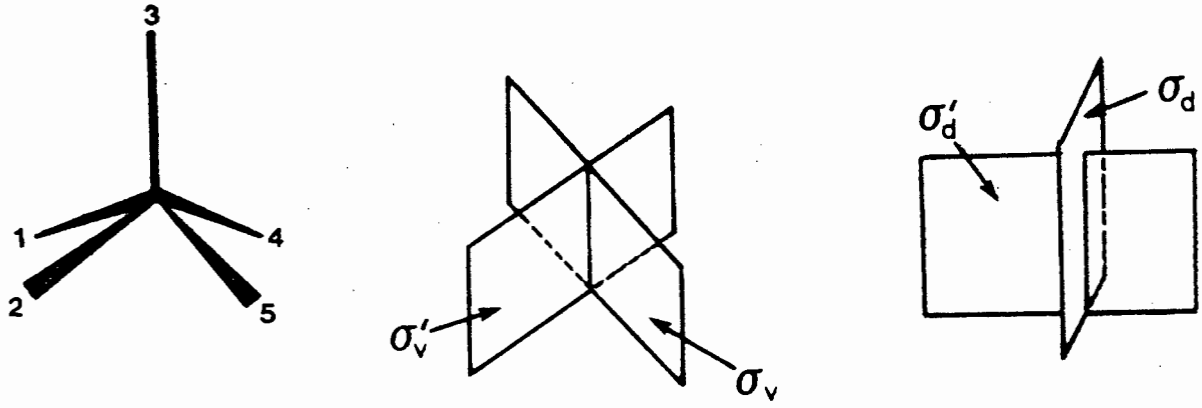
APPENDIX 1 – Derivation of Symmetry Coordinates

1. An explicit description for the derivation of symmetry coordinates is

$$S^i_{j\alpha} = N\frac{\ell_i}{h}\sum_R D^i(R)_{\alpha\alpha}(Rr_k)$$

where the operation R is performed on the members (r_k) of a symmetrically complete set of internal coordinates. Consequently, a useful first step in the derivation of symmetry coordinates is to examine how a typical member of such a set is transformed into others within the same group, under the various operations of the point group.

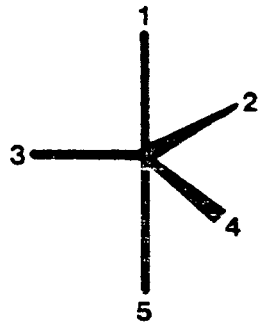
2. SQP –point group C_{4v}

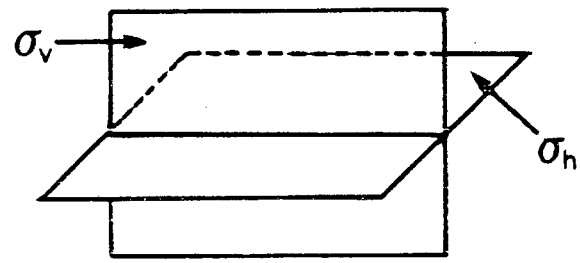


Symmetrically complete sets are: $(r_3);(r_1,r_2,r_4,r_5);(\theta_{13},\theta_{23},\theta_{34},\theta_{35});(\theta_{15},\theta_{24});$ $(\theta_{12},\theta_{25},\theta_{45},\theta_{14})$.

	Operation							
Typical member	E	C_4	C_4^3	C_2	σ_v	σ'_v	σ_d	σ'_d
r_3	r_3	r_3	r_3	r_3	r_3	r_3	r_3	r_3
r_1	r_1	r_4	r_2	r_5	r_1	r_5	r_4	r_2
θ_{15}	θ_{15}	θ_{24}	θ_{24}	θ_{15}	θ_{15}	θ_{15}	θ_{24}	θ_{24}
θ_{13}	θ_{13}	θ_{34}	θ_{23}	θ_{35}	θ_{13}	θ_{35}	θ_{34}	θ_{23}
θ_{12}	θ_{12}	θ_{14}	θ_{25}	θ_{45}	θ_{14}	θ_{25}	θ_{45}	θ_{12}

3. TBP – point group D_{3h}





Symmetrically complete sets are: $(r_1, r_5); (r_2, r_3, r_4); (\theta_{15}); (\theta_{23}, \theta_{34}, \theta_{24});$ $(\theta_{12}, \theta_{13}, \theta_{14}, \theta_{25}, \theta_{35}, \theta_{45})$.

				Operation								
Typical member	E	C_3	C_3^2	C_2	C_2'	C_2''	σ_h	S_3	S_3^2	σ_v	σ_v'	σ_v''
r_1	r_1	r_1	r_1	r_5	r_5	r_5	r_5	r_5	r_5	r_1	r_1	r_1
r_3	r_3	r_2	r_4	r_3	r_2	r_4	r_4	r_2	r_4	r_3	r_2	r_4
θ_{15}	θ_{15}	θ_{15}	θ_{15}	θ_{15}	θ_{15}	θ_{15}	θ_{15}	θ_{15}	θ_{15}	θ_{15}	θ_{15}	θ_{15}
θ_{24}	θ_{24}	θ_{34}	θ_{23}	θ_{24}	θ_{34}	θ_{23}	θ_{24}	θ_{34}	θ_{23}	θ_{24}	θ_{34}	θ_{23}
θ_{13}	θ_{13}	θ_{12}	θ_{14}	θ_{35}	θ_{25}	θ_{45}	θ_{35}	θ_{25}	θ_{45}	θ_{13}	θ_{12}	θ_{14}

4. The next step, typically, is to apply the so-called *projection operator* (P_α^i) to the members of each of the groups of internal coordinates, and to evaluate the sum over all operators (R) of the point group, viz.

$$P_\alpha^i = \frac{\ell_i}{h} \sum_R D^i(R)_{\alpha\alpha} \cdot R$$

In this expression P_α^i is the projection operator corresponding to the $\alpha-$th row of the i-th irreducible representation and, as before, ℓ_i , is the dimension of the i-th irreducible representation, h is the order of the group and $D^i(R)_{\alpha\alpha}$ is the element in the $\alpha-$th row and column of the transformation matrix corresponding to operator R in the i-th irreducible representation. As pointed out previously (see section 2), $D^i(R)_{\alpha\alpha}$ is equal to the character of operation R under the i-th

irreducible representation for all one-dimensional representations, and can hence be obtained from character tables. Furthermore, in practice the factor ℓ_i/h can be ignored until normalisation takes place at the end of the derivation.

5. SQP – Let us apply the projection operator P^{A_1} to the coordinate r_1, bearing in mind the simplifications outlined in point 4, above:

$$\begin{aligned} P^{A_1}(r_1) &\approx (1)E \cdot r_1 + (1)C_4 \cdot r_1 + (1)C_4^3 \cdot r_1 + (1)C_2 \cdot r_1 + (1)\sigma_v \cdot r_1 + (1)\sigma_v' \cdot r_1 \\ &\quad + (1)\sigma_d \cdot r_1 + (1)\sigma_d' \cdot r_1 \\ &= r_1 + r_4 + r_2 + r_5 + r_1 + r_5 + r_4 + r_2 \\ &= 2r_1 + 2r_2 + 2r_4 + 2r_5 \\ &\approx r_1 + r_2 + r_4 + r_5 \end{aligned}$$

The numbers in parentheses are the characters of A_1. The second approximation above can be made since the absolute values of the coefficients in the linear combination, as opposed to the relative ones, may be ascertained later by the normalization procedure. Similarly, applying P^{A_1} to all the other internal coordinates:

$$\begin{aligned} P^{A_1}(r_3) &\approx r_3 + r_3 + r_3 + r_3 + r_3 + r_3 + r_3 + r_3 = 8r_3 \approx r_3 \\ P^{A_1}(\theta_{15}) &\approx \theta_{15} + \theta_{24} + \theta_{24} + \theta_{15} + \theta_{15} + \theta_{15} + \theta_{24} + \theta_{24} \approx \theta_{15} + \theta_{24} \\ P^{A_1}(\theta_{13}) &\approx \theta_{13} + \theta_{34} + \theta_{23} + \theta_{35} + \theta_{13} + \theta_{35} + \theta_{34} + \theta_{23} \approx \theta_{13} + \theta_{23} + \theta_{34} + \theta_{35} \\ P^{A_1}(\theta_{12}) &\approx \theta_{12} + \theta_{14} + \theta_{25} + \theta_{45} + \theta_{14} + \theta_{25} + \theta_{45} + \theta_{12} \approx \theta_{12} + \theta_{14} + \theta_{25} + \theta_{45} \end{aligned}$$

$$\begin{aligned} P^{B_1}(r_1) &\approx (1)E \cdot r_1 + (-1)C_4 \cdot r_1 + (-1)C_4^3 \cdot r_1 + (1)C_2 \cdot r_1 + (1)\sigma_v \cdot r_1 \\ &\quad + (1)\sigma_v' \cdot r_1 + (-1)\sigma_d \cdot r_1 + (-1)\sigma_d' \cdot r_1 \\ &= r_1 - r_4 - r_2 + r_5 + r_1 + r_5 - r_4 - r_2 \approx r_1 + r_5 - r_2 - r_4 \\ P^{B_1}(r_3) &\approx r_3 - r_3 - r_3 + r_3 + r_3 + r_3 - r_3 - r_3 = 0 \\ P^{B_1}(\theta_{15}) &\approx \theta_{15} - \theta_{24} - \theta_{24} + \theta_{15} + \theta_{15} + \theta_{15} - \theta_{24} - \theta_{24} \approx \theta_{15} - \theta_{24} \\ P^{B_1}(\theta_{13}) &\approx \theta_{13} - \theta_{34} - \theta_{23} + \theta_{35} + \theta_{13} + \theta_{35} - \theta_{34} - \theta_{23} \approx \theta_{13} + \theta_{35} - \theta_{23} - \theta_{34} \\ P^{B_1}(\theta_{12}) &\approx \theta_{12} - \theta_{14} - \theta_{25} + \theta_{45} + \theta_{14} + \theta_{25} - \theta_{45} - \theta_{12} = 0 \end{aligned}$$

$$P^{B_2}(r_1) \approx r_1 - r_4 - r_2 + r_5 - r_1 - r_5 + r_4 + r_2 = 0$$

$$P^{B_2}(r_1) \approx r_3 - r_3 - r_3 + r_3 - r_3 - r_3 + r_3 + r_3 = 0$$

$$P^{B_2}(\theta_{15}) \approx 0$$

$$P^{B_2}(\theta_{13}) \approx 0$$

$$P^{B_2}(\theta_{12}) \approx \theta_{12} - \theta_{14} - \theta_{25} + \theta_{45} - \theta_{14} - \theta_{25} + \theta_{45} + \theta_{12} \approx \theta_{12} + \theta_{45} - \theta_{14} - \theta_{25}$$

In order to derive the functions which constitute partners under the two dimensional E representation, the individual elements of the transformation matrices for each operation R in the irreducible representation would need to be known. These have been tabulated, for example, by McWeeny, and Poulet and Mathieu. However, since these are not always readily at hand, we will employ a method outlined by Cotton, which restricts attention to the pure rotational symmetry about the principal axis. For the SQP we will use the group C_4 which has four operators. This approach views the two dimensional symmetry species as being, in fact, composed of two one dimensional ones, whose characters are complex numbers. Thus, application of the projection operator for the group C_4 to, say, r_1 yields:

$$P_1^E(r_1) \approx r_1 + ir_4 - ir_2 - r_5$$

$$P_2^E(r_1) \approx r_1 - ir_4 + ir_2 - r_5$$

These two complex functions are now simply seen as representing basis functions from which new linear combinations with real, as opposed to complex coefficients can be derived by addition and subtraction. Hence, adding the two functions above gives

$$\begin{array}{l} r_1+ir_4-ir_2-r_5 \\ +(r_1-ir_4+ir_2-r_5) \\ \hline 2r_1 - 2r_5 \end{array}$$

The first linear combination is therefore

$$r_1 - r_5$$

Next, subtracting the original functions

$$\begin{array}{l} r_1+ir_4-ir_2-r_5 \\ -(r_1-ir_4+ir_2-r_5) \\ \hline 2ir_4 - 2ir_2 \end{array}$$

$$2ir_4 - 2ir_2 = 2i(r_4 - r_2)$$

Consequently the second linear combination of internal coordinates, which should be orthogonal to the first and therefore its proper partner in forming a basis for the E representation, thus has the form

$$r_4 - r_2$$

Repeating these procedures for the other complete groups of coordinates yields

$$P_1^E(r_3) = 0$$

$$P_2^E(r_3) = 0$$

$$P_1^E(\theta_{15}) = 0$$

$$P_2^E(\theta_{15}) = 0$$

$$P_1^E(\theta_{13}) = \theta_{13} - \theta_{35}$$

$$P_2^E(\theta_{13}) = \theta_{34} - \theta_{23}$$

$$P_1^E(\theta_{12}) = \theta_{12} - \theta_{45}$$

$$P_2^E(\theta_{12}) = \theta_{14} - \theta_{25}$$

Finally, the fifteen functions derived above need to be normalized, resulting in

the fifteen symmetry coordinates:

$$S_1 = \frac{1}{2}(r_1 + r_2 + r_4 + r_5)$$
$$S_2 = r_3$$
$$S_3 = 2^{-\frac{1}{2}}(\theta_{15} + \theta_{24})$$
$$S_4 = \frac{1}{2}(\theta_{13} + \theta_{23} + \theta_{34} + \theta_{35})$$
$$S_5 = \frac{1}{2}(\theta_{12} + \theta_{14} + \theta_{25} + \theta_{45})$$
$$S_6 = \frac{1}{2}(r_1 + r_5 - r_2 - r_4)$$
$$S_7 = 2^{-\frac{1}{2}}(\theta_{15} - \theta_{24})$$
$$S_8 = \frac{1}{2}(\theta_{13} + \theta_{35} - \theta_{23} - \theta_{34})$$
$$S_9 = \frac{1}{2}(\theta_{12} + \theta_{45} - \theta_{14} - \theta_{25})$$
$$S_{10} = 2^{\frac{1}{2}}(r_1 - r_5)$$
$$S_{11} = 2^{\frac{1}{2}}(r_4 - r_2)$$
$$S_{12} = 2^{\frac{1}{2}}(\theta_{13} - \theta_{25})$$
$$S_{13} = 2^{-\frac{1}{2}}(\theta_{34} - \theta_{23})$$
$$S_{14} = 2^{-\frac{1}{2}}(\theta_{12} - \theta_{45})$$
$$S_{15} = 2^{\frac{1}{2}}(\theta_{14} - \theta_{25})$$

6. TBP – Let us repeat the procedure outlined above for the one dimensional representations of the D_{3h} point group:

$$P^{A_1'}(r_1) \approx r_1 + r_5$$
$$P^{A_1'}(r_3) \approx r_3 + r_2 + r_4$$
$$P^{A_1'}(\theta_{15}) \approx \theta_{15}$$
$$P^{A_1'}(\theta_{24}) \approx \theta_{24} + \theta_{23} + \theta_{34}$$
$$P^{A_1'}(\theta_{13}) \approx \theta_{13} + \theta_{12} + \theta_{14} + \theta_{35} + \theta_{25} + \theta_{45}$$

$$P^{A_2''}(r_1) \approx r_1 - r_5$$
$$P^{A_2''}(r_2) \approx 0$$
$$P^{A_2''}(\theta_{15}) \approx 0$$
$$P^{A_2''}(\theta_{24}) \approx 0$$
$$P^{A_2''}(\theta_{13}) \approx \theta_{12} + \theta_{13} + \theta_{14} - \theta_{25} - \theta_{35} - \theta_{45}$$

In order to derive the two dimensional basis functions we will employ yet another technique outlined by Cotton, which also circumvents the need to know each of the elements making up the transformation matrices, although it is based on these. It essentially entails deriving the first partner via the projection operator method using the characters for the E' and E'' representations of the D_{3h} point group, and subsequently transforming this into a linear combination of itself and its partner. The second partner is then found by adding the first, generating coordinate to the linear combination. Again, the question of normalization of the basis functions will be attended to at the end.

Applying the projection operator $P^{E'}$ to the groups of internal coordinates gives

$$P^{E'}(r_1) = 0$$

$$P^{E'}(r_3) = 2r_3 - r_2 - r_4$$

$$P^{E'}(\theta_{15}) = 0$$

$$P^{E'}(\theta_{24}) = 2\theta_{24} - \theta_{34} - \theta_{23}$$

$$P^{E'}(\theta_{13}) = 2\theta_{13} - \theta_{12} - \theta_{14} + 2\theta_{35} - \theta_{25} - \theta_{45}$$

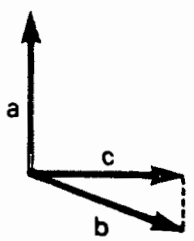
These functions are then transformed into others which are neither ± 1 times the first by, for example, subjecting them to rotation about the C_3 axis. Thus

$$C_3(2r_3 - r_2 - r_4) \rightarrow 2r_2 - r_4 - r_3$$

$$C_3(2\theta_{24} - \theta_{34} - \theta_{23}) \rightarrow 2\theta_{34} - \theta_{23} - \theta_{24}$$

$$C_3(2\theta_{13} - \theta_{12} - \theta_{14} + 2\theta_{35} - \theta_{25} - \theta_{45}) \rightarrow 2\theta_{12} - \theta_{14} - \theta_{13} + 2\theta_{25} - \theta_{45} - \theta_{35}$$

Now, the two functions $(2r_3 - r_2 - r_4)$ and $(2r_2 - r_4 - r_3)$ can be thought of as two vectors $\bar{a}$ and $\bar{b}$, respectively, which are related to each other by a three-fold axis in the space spanned by the symmetry coordinates, i.e., they lie at 120^0 degrees to each other rather like



The vector $\bar{c}$, orthogonal as it is to $\bar{a}$ (the generating coordinate), would represent a possible partner to it, and can be obtained from $\bar{a}$ and $\bar{b}$ by classical vector algebra as follows:

$$\begin{aligned} c &= b + \frac{1}{2}a \\ &= 2r_2 - r_4 - r_3 + \frac{1}{2}(2r_3 - r_2 - r_4) \\ &= \frac{3}{2}r_2 - \frac{3}{2}r_4 = \frac{3}{2}(r_2 - r_4) \end{aligned}$$

Thus the linear combination $(r_2 - r_4)$ is orthogonal to $(2r_3 - r_2 - r_4)$ and consequently represents the latter's partner in the E' representation. The partners to the other two functions derived above can be obtained by a similar procedure, as can that for the E'' coordinate. After appropriate normalization, then, the fifteen symmetry coordinates obtained by the methods above are:

$$\begin{aligned} S_1 &= 2^{-\frac{1}{2}}(r_1 + r_5) \\ S_2 &= 3^{-\frac{1}{2}}(r_3 + r_2 + r_4) \\ S_3 &= \theta_{15} \\ S_4 &= 3^{-\frac{1}{2}}(\theta_{24} + \theta_{23} + \theta_{34}) \\ S_5 &= 6^{-\frac{1}{2}}(\theta_{13} + \theta_{12} + \theta_{14} + \theta_{35} + \theta_{25} + \theta_{45}) \\ S_6 &= 2^{-\frac{1}{2}}(r_1 - r_5) \\ S_7 &= 6^{-\frac{1}{2}}(\theta_{13} + \theta_{12} + \theta_{14} - \theta_{35} - \theta_{25} - \theta_{45}) \\ S_8 &= 6^{-\frac{1}{2}}(2r_3 - r_2 - r_4) \\ S_9 &= 2^{-\frac{1}{2}}(r_2 - r_4) \\ S_{10} &= 6^{-\frac{1}{2}}(2\theta_{24} - \theta_{34} - \theta_{23}) \\ S_{11} &= 2^{-\frac{1}{2}}(\theta_{34} - \theta_{23}) \\ S_{12} &= 12^{-\frac{1}{2}}(2\theta_{13} - \theta_{12} - \theta_{14} + 2\theta_{35} - \theta_{25} - \theta_{45}) \\ S_{13} &= \frac{1}{2}(\theta_{12} - \theta_{14} + \theta_{25} - \theta_{45}) \\ S_{14} &= 12^{-\frac{1}{2}}(2\theta_{13} - \theta_{12} - \theta_{14} - 2\theta_{35} + \theta_{25} + \theta_{45}) \\ S_{15} &= \frac{1}{2}(\theta_{12} - \theta_{14} - \theta_{25} + \theta_{45}) \end{aligned}$$

APPENDIX 2 — Fortran programs for permuting ligand sites for TBP and SQP.

```
1000 c     ****************************************************************
1010 c     SQPDEF - This program reads in the internal coordinates for
1020 c     five coordinate complexes, calculates the symmetry coordinates
1030 c     based on a C2v symmetry and then establishes the deformation
1040 c     away from the idealised structure by calculating the length
1050 c     of a deformation vector representing that structure in the
1060 c     6 dimensional space spanned by the non-totally symmetric
1070 c     angle deformation symmetry coordinates only.
1080 c     It then permutes the ligands (and the internal coordinates) and
1090 c     repeats the calculation, until it finds the LEAST distorted
1100 c     permutation of the complex, which it stores.
1110 c     input unit=10, output unit=11.
1120 c     ****************************************************************
1130
1140       real d(5),tempd(5),del(5,5),tmpdel(5,5),s(12),temps(12),
1150 c           **distances** ****angle matrix**** *symmetry coords*
1160      &    defvec,tmpvec
1170 c         *defmtn vector*
1180       character*2 atyp(5),lig(5),metal
1190 c                  ***ligands***
1200       character*8 refcod
1210       integer i,j,k,l,m,x,y,z,tempk,templ,tempm
1220 c
1230 5     format(a8,2x,a2,1x,5(a2,1x),1x,5(f5.3,1x)/10(f6.2,1x))
1240 15    format(a8,2x,a2,1x,5(a2,1x),1x,5(f5.3,1x)/10(f6.2,1x)/
1250      &       8(f7.3,2x)/5(f7.3,2x))
1260
1270 1     read(10,5,end=9999) refcod,metal,(atyp(i),i=1,5),(tempd(k),k=1,5),
1280
1290 c     *reading in one half of the angle matrix*
1300
1310      & tmpdel(1,2),tmpdel(1,3),tmpdel(1,4),tmpdel(1,5),
1320      & tmpdel(2,3),tmpdel(2,4),tmpdel(2,5),tmpdel(3,4),
1330      & tmpdel(3,5),tmpdel(4,5)
1340
1350 c     *completing the other half of the angle data matrix*
1360
1370       do 22 i=1,4
1380         do 11 j=2,5
1390           if (j.gt.i) then
1400             tmpdel(j,i)=tmpdel(i,j)
1410           endif
1420 11      continue
1430 22    continue
1440
1450       defvec=1000.
1460       tmpvec=0.
1470
1480 c     *begin permuting the ligands*
1490
1500       do 100 I=1,5
1510         do 90 j=2,1,-1
1520           if(j.ne.i)then
1530             do 80 k=2,5
1540               if((k.ne.i).and.(k.ne.j))then
1550                 do 70 l=2,5
```

```
1560                    if((l.ne.i).and.(l.ne.j).and.(l.ne.k))then
1570                      do 60 m=2,5
1580                        if((m.ne.i).and.(m.ne.j).and.(m.ne.k)
1590      &                     .and.(m.ne.l))then
1600
1610 c                        *storing lig3,4 and 5*
1620
1630                          tempk=k
1640                          templ=l
1650                          tempm=m
1660
1670 c                  *calculate symcos and defmtn vector and return*
1680
1690                          call symco(tempd,tmpdel,temps,tmpvec,
1700      &                           i,j,tempk,templ,tempm)
1710
1720                          if (tmpvec.lt.defvec) then
1730
1740 c                        *store this permutn*
1750
1760                            call store(lig,atyp,d,tempd,del,tmpdel,
1770      &                           defvec,tmpvec,i,j,tempk,templ,
1780      &                           tempm,s,temps)
1790                          endif
1800
1810 c                       *swop ligs 3 and 4*
1820
1830                          tempk=l
1840                          templ=k
1850
1860                          call symco(tempd,tmpdel,temps,tmpvec,
1870      &                           i,j,tempk,templ,tempm)
1880
1890                          if (tmpvec.lt.defvec) then
1900
1910                            call store(lig,atyp,d,tempd,del,tmpdel,
1920      &                           defvec,tmpvec,i,j,tempk,templ,
1930      &                           tempm,s,temps)
1940                          endif
1950
1960 c                        *swop ligs 4 and 5*
1970
1980                          tempk=k
1990                          templ=m
2000                          tempm=l
2010
2020                          call symco(tempd,tmpdel,temps,tmpvec,
2030      &                           i,j,tempk,templ,tempm)
2040
2050                          if (tmpvec.lt.defvec) then
2060
2070                            call store(lig,atyp,d,tempd,del,tmpdel,
2080      &                           defvec,tmpvec,i,j,tempk,templ,
2090      &                           tempm,s,temps)
2100                          endif
2110
2120                          go to 90
2130                        endif
2140 60                   continue
2150                    endif
2160 70                 continue
2170                  endif
```

```
2180 80          continue
2190           endif
2200 90      continue
2210 100   continue
2220
2230 c     *end of lig permutn and store*
2240
2250       write(11,15) refcod,metal,(lig(x),x=1,5),(d(y),y=1,5),
2260      &             del(1,2),del(1,3),del(1,4),del(1,5),
2270      &             del(2,3),del(2,4),del(2,5),del(3,4),
2280      &             del(3,5),del(4,5),
2290      &                   (s(z),z=1,12),defvec
2300
2310 c     *start reading next data point*
2320
2330       go to 1
2340
2350 9999  end
2360
2370
2380
2390
2400
2410 c     ******************************************************
2420       subroutine symco(tempd,tmpdel,temps,tmpvec,i,j,tempk,
2430      &                                    templ,tempm)
2440 c     ******************************************************
2450
2460 c     *This subroutine calculates the 12 symcos and the corresponding
2470 c     *defmtn vector and passes these back to the main pgm
2480
2490       real tempd(5),tmpdel(5,5),temps(12),tmpvec
2500       integer i,j,tempk,templ,tempm
2510
2520       temps(1)=(tempd(tempk))
2530       temps(2)=(tempd(i)+tempd(j)+tempd(templ)+tempd(tempm))/2
2540       temps(3)=(tmpdel(i,tempm)+tmpdel(j,templ))/sqrt(2)
2550       temps(4)=(tempd(i)+tempd(tempm)-tempd(j)-tempd(templ))/2
2560       temps(5)=(tmpdel(i,tempm)-tmpdel(j,templ))/sqrt(2)
2570       temps(6)=(tmpdel(i,j)+tmpdel(templ,tempm)
2580      &          -tmpdel(i,templ)-tmpdel(j,tempm))/2
2590       temps(7)=(tempd(i)-tempd(tempm))/sqrt(2)
2600       temps(8)=(tempd(templ)-tempd(j))/sqrt(2)
2610       temps(9)=(tmpdel(i,tempk)-tmpdel(tempk,tempm))/sqrt(2)
2620       temps(10)=(tmpdel(tempk,templ)-tmpdel(j,tempk))/sqrt(2)
2630       temps(11)=(tmpdel(i,j)-tmpdel(templ,tempm))/sqrt(2)
2640       temps(12)=(tmpdel(i,templ)-tmpdel(j,tempm))/sqrt(2)
2650
2660       tmpvec=sqrt(temps(5)**2+temps(6)**2+temps(9)**2
2670      &          +temps(10)**2+temps(11)**2+temps(12)**2)
2680 c
2690       return
2700       end
2710
2720
2730
2740 c     ************************************************************
2750       subroutine store(lig,atyp,d,tempd,del,tmpdel,
2760      &                 defvec,tmpvec,i,j,tempk,templ,
2770      &                              tempm,s,temps)
2780 c     ************************************************************
2790
```

```
2800 c     *This subroutine stores all the data for the permutns
2810 c     *which are calculated prior to it being called
2820
2830       real d(5),tempd(5),del(5,5),tmpdel(5,5),s(12),temps(12),
2840      &   defvec,tmpvec
2850       character*2 atyp(5),lig(5)
2860       integer i,j,x,tempk,templ,tempm
2870
2880 c     *store this permutn*
2890
2900       lig(1)=atyp(i)
2910       lig(2)=atyp(j)
2920       lig(3)=atyp(tempk)
2930       lig(4)=atyp(templ)
2940       lig(5)=atyp(tempm)
2950       d(1)=tempd(i)
2960       d(2)=tempd(j)
2970       d(3)=tempd(tempk)
2980       d(4)=tempd(templ)
2990       d(5)=tempd(tempm)
3000       del(1,2)=tmpdel(i,j)
3010       del(1,3)=tmpdel(i,tempk)
3020       del(1,4)=tmpdel(i,templ)
3030       del(1,5)=tmpdel(i,tempm)
3040       del(2,3)=tmpdel(j,tempk)
3050       del(2,4)=tmpdel(j,templ)
3060       del(2,5)=tmpdel(j,tempm)
3070       del(3,4)=tmpdel(tempk,templ)
3080       del(3,5)=tmpdel(tempk,tempm)
3090       del(4,5)=tmpdel(templ,tempm)
3100
3110 c     *store all 12 symcos*
3120
3130       do 44 x=1,12
3140         s(x)=temps(x)
3150 44    continue
3160
3170 c     *store the defmtn vector*
3180
3190       defvec=tmpvec
3200
3210       return
3220       end
3230
```

```
1000 c     ***********************************************************************
1010 c     TBPDEF - This program reads in the internal coordinates for
1020 c     five coordinate complexes, calculates the symmetry coordinates
1030 c     based on a D3h symmetry and then establishes the deformation
1040 c     away from the idealised structure by calculating the length
1050 c     of a deformation vector representing that structure in the
1060 c     7 dimensional space spanned by the symmetry coordinates
1070 c     involving bond angle deformations only.
1080 c     It then permutes the ligands (and the internal coordinates) and
1090 c     repeats the calculation, until it finds the LEAST distorted
1100 c     permutation of the complex, which it stores.
1110 c     input unit=10, output unit=11
1120 c     ***********************************************************************
1130
1140       real d(5),tempd(5),del(5,5),tmpdel(5,5),s(12),temps(12),
1150 c          *distances*   ****angle matrix***  *symmetry coords*
1160      &    defvec,tmpvec
1170 c         *deformtn vector*
1180       character*2 atyp(5),lig(5),metal
1190 c                  ***ligands***
1200       character*8 refcod
1210       integer i,j,k,l,m,x,y,z
1220
1230 5     format(a8,2x,a2,1x,5(a2,1x),1x,5(f5.3,1x)/10(f6.2,1x))
1240 15    format(a8,2x,a2,1x,5(a2,1x),1x,5(f5.3,1x)/10(f6.2,1x)/
1250      &         8(f7.3,2x)/5(f7.3,2x))
1260
1270 1     read(10,5,end=9999) refcod,metal,(atyp(i),i=1,5),(tempd(k),k=1,5),
1280
1290 c     *reading in one half of the matrix containing the angles*
1300
1310      &            tmpdel(1,2),tmpdel(1,3),tmpdel(1,4),tmpdel(1,5),
1320      &            tmpdel(2,3),tmpdel(2,4),tmpdel(2,5),tmpdel(3,4),
1330      &            tmpdel(3,5),tmpdel(4,5)
1340
1350 c     *completing the other half of the angle data matrix*
1360
1370       do 22 i=1,4
1380         do 11 j=2,5
1390           if (j.gt.i) then
1400             tmpdel(j,i)=tmpdel(i,j)
1410           endif
1420 11      continue
1430 22    continue
1440
1450       defvec=1000.
1460       tmpvec=0.
1470
1480 c     *begin to permute the ligands*
1490
1500       do 50 I=1,4
1510 100     do 60 j=2,5
1520           if(j.gt.i)then
1530             do 70 k=1,5
1540               if((k.ne.i).and.(k.ne.j))then
1550                 do 80 l=1,5
```

```
                    if((l.ne.i).and.(l.ne.j).and.(l.ne.k))then
                      do 90 m=1,5
                          if((m.ne.i).and.(m.ne.j).and.(m.ne.k)
     &                        .and.(m.ne.l))then

c     *call to sub which calculates the symcos and defmtnvec*

                       call symco(tempd,tmpdel,temps,tmpvec,i,j,k,l,m)

                          if (tmpvec.lt.defvec) then

c                           *store this permutn*

                            lig(1)=atyp(i)
                            lig(2)=atyp(j)
                            lig(3)=atyp(k)
                            lig(4)=atyp(l)
                            lig(5)=atyp(m)
                            d(1)=tempd(i)
                            d(2)=tempd(j)
                            d(3)=tempd(k)
                            d(4)=tempd(l)
                            d(5)=tempd(m)
                            del(1,2)=tmpdel(i,j)
                            del(1,3)=tmpdel(i,k)
                            del(1,4)=tmpdel(i,l)
                            del(1,5)=tmpdel(i,m)
                            del(2,3)=tmpdel(j,k)
                            del(2,4)=tmpdel(j,l)
                            del(2,5)=tmpdel(j,m)
                            del(3,4)=tmpdel(k,l)
                            del(3,5)=tmpdel(k,m)
                            del(4,5)=tmpdel(l,m)

c                           *storing all 12 symcos*

                            do 44 x=1,12
                             s(x)=temps(x)
44                          continue

c                           *storing defmtn vector*

                            defvec=tmpvec

c                           *end of store*

                          endif
                          go to 60
                        endif
90                    continue
                    endif
80                continue
                  endif
70              continue
                endif
60            continue
50          continue

c     *end of ligand permutn*

            write(11,15) refcod,metal,(lig(x),x=1,5),(d(y),y=1,5),
     &                   del(1,2),del(1,3),del(1,4),del(1,5),
```

```
2180      &              del(2,3),del(2,4),del(2,5),del(3,4),
2190      &              del(3,5),del(4,5),
2200      &              (s(z),z=1,12),defvec
2210
2220 c     *begin reading in next data point*
2230
2240       go to 1
2250
2260 9999  end
2270
2280
2290
2300
2310
2320 c     ******************************************************
2330       subroutine symco(tempd,tmpdel,temps,tmpvec,i,j,k,l,m)
2340 c     ******************************************************
2350
2360 c     *This subroutine calculates the 12 symcos and the corresponding
2370 c     *defmtn vector and passes these back to the main pgm
2380
2390       real tempd(5),tmpdel(5,5),temps(12),tmpvec
2400       integer i,j,k,l,m
2410
2420       temps(1)=(tempd(i)+tempd(m))/sqrt(2)
2430       temps(2)=(tempd(j)+tempd(k)+tempd(l))/sqrt(3)
2440       temps(3)=(tempd(i)-tempd(m))/sqrt(2)
2450       temps(4)=(tmpdel(i,j)+tmpdel(i,k)+tmpdel(i,l)-tmpdel(j,m)
2460      &          -tmpdel(k,m)-tmpdel(l,m))/sqrt(6)
2470       temps(5)=(2*tempd(k)-tempd(j)-tempd(l))/sqrt(6)
2480       temps(6)=(tempd(j)-tempd(l))/sqrt(2)
2490       temps(7)=(2*tmpdel(j,l)-tmpdel(k,l)-tmpdel(j,k))/sqrt(6)
2500       temps(8)=(tmpdel(k,l)-tmpdel(j,k))/sqrt(2)
2510       temps(9)=(2*tmpdel(i,k)-tmpdel(i,j)-tmpdel(i,l)
2520      &         +2*tmpdel(k,m)-tmpdel(j,m)-tmpdel(l,m))/sqrt(12)
2530       temps(10)=(tmpdel(i,j)-tmpdel(i,l)+tmpdel(j,m)-tmpdel(l,m))/2
2540       temps(11)=(2*tmpdel(i,k)-tmpdel(i,j)-tmpdel(i,l)
2550      &         -2*tmpdel(k,m)+tmpdel(j,m)+tmpdel(l,m))/sqrt(12)
2560       temps(12)=(tmpdel(i,j)-tmpdel(i,l)-tmpdel(j,m)+tmpdel(l,m))/2
2570
2580       tmpvec=sqrt(temps(4)**2+temps(7)**2+temps(8)**2+temps(9)**2
2590      &           +temps(10)**2+temps(11)**2+temps(12)**2)
2600
2610       return
2620       end
2630
```

APPENDIX 3 — Fortran programs for expansion of data set according to D_{3h} and C_{4v} symmetry.

```
1000 c     ****************************************************************
1010 c     SQPEXP - This program reads the internal coordinates put out by
1020 c     SQPDEF (the least distorted C4v permutation of the 5-coordinate
1030 c     complex) and then permutes these into 8 symmetry equivalent points
1040 c     in the 12-D parameter space, according to the symmetry operations
1050 c     of the C4v point group.  It then also calculates the symmetry
1060 c     coordinates.  Output of the expanded  data is onto unit=11,
1070 c     and input is from unit=10.
1080 c     ****************************************************************
1090
1100       real d(5),del(5,5)
1110 c          *distances*
1120       character*2 lig(5),metal
1130 c                   ***ligands***
1140       character*8 refcod
1150       integer i,j,k,l,m
1160
1170 5     format(a8,2x,a2,1x,5(a2,1x),1x,5(f5.3,1x)/10(f6.2,1x)//)
1180
1190 1     read(10,5,end=9999) refcod,metal,(lig(i),i=1,5),(d(k),k=1,5),
1200
1210 c     *reading in one half of the matrix containing the angles*
1220
1230      &           del(1,2),del(1,3),del(1,4),del(1,5),
1240      &           del(2,3),del(2,4),del(2,5),del(3,4),
1250      &           del(3,5),del(4,5)
1260
1270 c     *completing the other half of the angle data matrix*
1280
1290       do 22 i=1,4
1300         do 11 j=2,5
1310           if (j.gt.i) then
1320             del(j,i)=3.1416*2.231*del(i,j)/180.0
1330             del(i,j)=del(j,i)
1340           endif
1350 11      continue
1360 22    continue
1370
1380       i=1
1390       j=2
1400       k=3
1410       l=4
1420       m=5
1430
1440       call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1450
1460       j=4
1470       l=2
1480
1490       call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1500
1510       i=2
1520       j=1
1530       l=5
1540       m=4
1550
```

```
1560        call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1570
1580        j=5
1590        l=1
1600
1610        call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1620
1630        i=4
1640        j=1
1650        l=5
1660        m=2
1670
1680        call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1690
1700        j=5
1710        l=i
1720
1730        call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1740
1750        i=5
1760        j=2
1770        l=4
1780        m=1
1790
1800        call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1810
1820
1830        j=4
1840        l=2
1850
1860        call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1870
1880
1890 c      *begin reading in next data point*
1900
1910        go to 1
1920
1930 9999   end
1940
1950
1960
1970
1980
1990 c      ********************************************************
2000        subroutine prntit(refcod,metal,lig,d,del,i,j,k,l,m)
2010 c      ********************************************************
2020
2030 c      This subroutine calculates the symmetry coordinates for the
2040 c      permutation specified when it is called, and then prints the
2050 c      ligands,internal- and symmetry coordinates for that permutn.
2060
2070        real d(5),tempd(5),del(5,5),tmpdel(5,5),s(12),defvec
2080        integer i,j,k,l,m,x
2090        character*2 atyp(5),lig(5),metal
2100        character*8 refcod
2110
2120 5      format(a8,2x,a2,1x,5(a2,1x),1x,5(f5.3,1x)/10(f6.3,1x))
2130 15     format(8(f7.3,2x)/5(f7.3,2x))
2140
2150        tempd(1)=d(i)
2160        tempd(2)=d(j)
2170        tempd(3)=d(k)
```

```
2180        tempd(4)=d(l)
2190        tempd(5)=d(m)
2200        atyp(1)=lig(i)
2210        atyp(2)=lig(j)
2220        atyp(3)=lig(k)
2230        atyp(4)=lig(l)
2240        atyp(5)=lig(m)
2250
2260        tmpdel(1,5)=del(i,m)
2270        tmpdel(1,2)=del(i,j)
2280        tmpdel(1,3)=del(i,k)
2290        tmpdel(1,4)=del(i,l)
2300        tmpdel(2,3)=del(j,k)
2310        tmpdel(2,4)=del(j,l)
2320        tmpdel(2,5)=del(j,m)
2330        tmpdel(3,4)=del(k,l)
2340        tmpdel(3,5)=del(k,m)
2350        tmpdel(4,5)=del(l,m)
2360
2370        s(1)=(tempd(3))
2380        s(2)=(tempd(1)+tempd(2)+tempd(4)+tempd(5))/2
2390        s(3)=(tmpdel(1,5)+tmpdel(2,4))/sqrt(2)
2400        s(4)=(tempd(1)+tempd(5)-tempd(2)-tempd(4))/2
2410        s(5)=(tmpdel(1,5)-tmpdel(2,4))/sqrt(2)
2420        s(6)=(tmpdel(1,2)+tmpdel(4,5)
2430       &           -tmpdel(1,4)-tmpdel(2,5))/2
2440        s(7)=(tempd(1)-tempd(5))/sqrt(2)
2450        s(8)=(tempd(4)-tempd(2))/sqrt(2)
2460        s(9)=(tmpdel(1,3)-tmpdel(3,5))/sqrt(2)
2470        s(10)=(tmpdel(3,4)-tmpdel(2,3))/sqrt(2)
2480        s(11)=(tmpdel(1,2)-tmpdel(4,5))/sqrt(2)
2490        s(12)=(tmpdel(1,4)-tmpdel(2,5))/sqrt(2)
2500
2510        defvec=sqrt(s(5)**2+s(6)**2+s(9)**2
2520       &          +s(10)**2+s(11)**2+s(12)**2)
2530
2540        write(11,5) refcod,metal,(atyp(x),x=1,5),(tempd(x),x=1,5),
2550       &            tmpdel(1,2),tmpdel(1,3),tmpdel(1,4),tmpdel(1,5),
2560       &            tmpdel(2,3),tmpdel(2,4),tmpdel(2,5),tmpdel(3,4),
2570       &            tmpdel(3,5),tmpdel(4,5)
2580
2590        write(11,15) (s(x),x=1,12),defvec
2600
2610        return
2620        end
2630
```

```
1000 c     ***********************************************************************
1010 c     TBPEXP - This program reads the internal coordinates put out by
1020 c     TBPDEF (the least distorted D3h permutation of the 5-coordinate
1030 c     complex) and then permutes these into 12 symmetry equivalent points
1040 c     in the 12-D parameter space, according to the symmetry operations
1050 c     of the D3h point group.  It then also calculates the symmetry
1060 c     coordinates.  Output of the expanded data is onto unit=11,
1070 c     input is from unit=10.
1080 c     ***********************************************************************
1090
1100
1110       real d(5),del(5,5)
1120 c          *distances*
1130       character*2 lig(5),metal
1140 c                  ***ligands***
1150       character*8 refcod
1160       integer i,j,k,l,m,x,z
1170
1180 5     format(a8,2x,a2,1x,5(a2,1x),1x,5(f5.3,1x)/10(f6.2,1x)//)
1190
1200 1     read(10,5,end=9999) refcod,metal,(lig(i),i=1,5),(d(k),k=1,5),
1210
1220 c     *reading in one half of the matrix containing the angles*
1230
1240      &           del(1,2),del(1,3),del(1,4),del(1,5),
1250      &           del(2,3),del(2,4),del(2,5),del(3,4),
1260      &           del(3,5),del(4,5)
1270
1280 c     *completing the other half of the angle data matrix*
1290
1300       do 22 i=1,4
1310         do 11 j=2,5
1320           if (j.gt.i) then
1330            del(j,i)=3.1416*2.231*del(i,j)/180.0
1340            del(i,j)=del(j,i)
1350           endif
1360 11      continue
1370 22    continue
1380
1390       i=1
1400       j=2
1410       k=3
1420       l=4
1430       m=5
1440
1450       call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1460
1470       do 33 z=1,2
1480         x=j
1490         j=k
1500         k=l
1510         l=x
1520       call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1530 33    continue
1540
1550       x=j
```

```
1560        j=k
1570        k=x
1580
1590          call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1600
1610        do 44 z=1,2
1620          x=l
1630          l=k
1640          k=j
1650          j=x
1660          call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1670 44     continue
1680
1690        i=5
1700        j=4
1710        k=3
1720        l=2
1730        m=1
1740
1750        call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1760
1770        do 55 z=1,2
1780          x=l
1790          l=k
1800          k=j
1810          j=x
1820          call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1830 55     continue
1840
1850        x=l
1860        l=k
1870        k=x
1880        call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1890
1900        do 66 z=1,2
1910          x=j
1920          j=k
1930          k=l
1940          l=x
1950          call prntit(refcod,metal,lig,d,del,i,j,k,l,m)
1960 66     continue
1970
1980 c      *begin reading in next data point*
1990
2000        go to 1
2010
2020 9999   end
2030
2040
2050
2060
2070
2080 c      ******************************************************
2090        subroutine prntit(refcod,metal,lig,d,del,i,j,k,l,m)
2100 c      ******************************************************
2110
2120 c      This subroutine calculates the symmetry coordinates for the
2130 c      permutations specified when it is called, and then prints the
2140 c      ligands,internal- and symmetry coordinates for that permutation.
2150
2160        real d(5),tempd(5),del(5,5),tmpdel(5,5),s(12),defvec
2170        integer i,j,k,l,m,x
```

```
      character*2 atyp(5),lig(5),metal
      character*8 refcod

5     format(a8,2x,a2,1x,5(a2,1x),1x,5(f5.3,1x)/10(f6.3,1x))
15    format(8(f7.3,2x)/5(f7.3,2x))

      tempd(1)=d(i)
      tempd(2)=d(j)
      tempd(3)=d(k)
      tempd(4)=d(l)
      tempd(5)=d(m)
      atyp(1)=lig(i)
      atyp(2)=lig(j)
      atyp(3)=lig(k)
      atyp(4)=lig(l)
      atyp(5)=lig(m)

      tmpdel(1,5)=del(i,m)
      tmpdel(1,2)=del(i,j)
      tmpdel(1,3)=del(i,k)
      tmpdel(1,4)=del(i,l)
      tmpdel(2,3)=del(j,k)
      tmpdel(2,4)=del(j,l)
      tmpdel(2,5)=del(j,m)
      tmpdel(3,4)=del(k,l)
      tmpdel(3,5)=del(k,m)
      tmpdel(4,5)=del(l,m)

      s(1)=(tempd(1)+tempd(5))/sqrt(2)
      s(2)=(tempd(2)+tempd(3)+tempd(4))/sqrt(3)
      s(3)=(tempd(1)-tempd(5))/sqrt(2)
      s(4)=(tmpdel(1,2)+tmpdel(1,3)+tmpdel(1,4)-tmpdel(2,5)
     &         -tmpdel(3,5)-tmpdel(4,5))/sqrt(6)
      s(5)=(2*tempd(3)-tempd(2)-tempd(4))/sqrt(6)
      s(6)=(tempd(2)-tempd(4))/sqrt(2)
      s(7)=(2*tmpdel(2,4)-tmpdel(3,4)-tmpdel(2,3))/sqrt(6)
      s(8)=(tmpdel(3,4)-tmpdel(2,3))/sqrt(2)
      s(9)=(2*tmpdel(1,3)-tmpdel(1,2)-tmpdel(1,4)
     &        +2*tmpdel(3,5)-tmpdel(2,5)-tmpdel(4,5))/sqrt(12)
      s(10)=(tmpdel(1,2)-tmpdel(1,4)+tmpdel(2,5)-tmpdel(4,5))/2
      s(11)=(2*tmpdel(1,3)-tmpdel(1,2)-tmpdel(1,4)
     &        -2*tmpdel(3,5)+tmpdel(2,5)+tmpdel(4,5))/sqrt(12)
      s(12)=(tmpdel(1,2)-tmpdel(1,4)-tmpdel(2,5)+tmpdel(4,5))/2

      defvec=sqrt(s(4)**2+s(7)**2+s(8)**2+s(9)**2
     &          +s(10)**2+s(11)**2+s(12)**2)

      write(11,5) refcod,metal,(atyp(x),x=1,5),(tempd(x),x=1,5),
     &            tmpdel(1,2),tmpdel(1,3),tmpdel(1,4),tmpdel(1,5),
     &            tmpdel(2,3),tmpdel(2,4),tmpdel(2,5),tmpdel(3,4),
     &            tmpdel(3,5),tmpdel(4,5)

      write(11,15) (s(x),x=1,12),defvec

      return
      end
```

CHAPTER 4
DATA ANALYSIS AND DISCUSSION

CHAPTER 4
DATA ANALYSIS

We have extracted structural data on 196 five-coordinate complexes of nickel, palladium, platinum, rhodium and iridium, all with a d^8 electron configuration. For each the geometry of the central ML_5 fragment, the metal atom with its five ligand donor atoms, has been precisely described by two sets of twelve non-redundant symmetry coordinates. These sets correspond to the two most common idealised five-coordinate conformations, viz. the trigonal bipyramid (TBP) and the square based (or rectangular) pyramid (SQP), respectively. We now regard each observed ML_5 molecular fragment as being represented by a point in a twelve-dimensional (12-D) space spanned by the symmetry coordinates. There are two such spaces within which the representative points are distributed, depending on whether the observed structures are being related to an idealised TBP or to an idealised SQP. We have named these T- and S-space, respectively.

T-space can be understood simply as being composed of twelve dimensions, with its coordinates corresponding to the set of twelve non-redundant symmetry coordinates derived for a TBP. Its origin represents an idealised ML_5 TBP with D_{3h} symmetry whose M-L bond lengths do not exceed their "idealised single bond" values. The conceptualization of S-space follows similar lines, with the exception that an idealised ML_5 SQP with C_{4v} symmetry and "ideal" M-L bond lengths is not represented simply by a point, but rather by a continuous line in twelve-dimensional space. This line reflects the one degree of freedom which the C_{4v} symmetry allows the values of the trans-basal angles in a SQP.

In order to circumvent the problem of developing a unique system of nomenclature which enables one to distinguish two isometric structures from each other, each observed conformation was transformed into its isometric partners by the application of the twelve and eight symmetry operations of the D_{3h} and C_{4v} point groups, respectively. As a consequence of this the number of data points in T-space expands to 2352, while that in S-space becomes 1568. Moreover, the expansion of the data set has introduced elements of symmetry into the twelve-dimensional data distribution. This artefact may be profitably exploited in the multi-variate analysis

which follows, since whatever symmetry is present in the data set must emerge also during the factor and cluster analysis, thus yielding an indicator as to the reliability of the results obtained.

The aim of the ensuing multi-dimensional analysis is essentially two-fold: (i) the *distribution* of the representative points in 12-D space will be probed using cluster analysis. This serves to establish the clustering pattern (if any), i.e. to establish the essential or kernel conformations around which the observed structures cluster. (ii) The *shape* of these clusters will be examined using factor analysis so as to establish the coordinates along which the data point clouds expand, i.e. in order to map out those coordinates along which the conformations distort most significantly, and along which the various clusters are joined to each other.

4.1 Univariate and Bivariate Statistics

(a) Univariate statistics

For two reasons the amount of information which may be gleaned from univariate statistics is limited. Firstly, had they been obtained from the unexpanded data base, then they would have little meaning due to the lack of a systematic method of labelling the five ligand atoms. Secondly, where they are derived for the expanded data set, they will reflect artificial manipulations which result in symmetry related parameters all having identical means and variances. Nevertheless, the variances of the individual symmetry coordinates do offer interesting insights into the underlying distortions of the ML_5 fragment. Table 4.1.1 gives the standard deviation and variance of each symmetry coordinate in T-space, as well as the percentage of the total variance in the bond distance- and angle-containing symmetry coordinates, respectively, which each can account for. Table 4.1.2 gives the same information for S-space.

In the interpretation of these results we shall make use of the graphical representation of symmetry coordinates, a technique in common use in standard texts on vibrational analysis. This involves representing the degree of freedom of the internal coordinates in a symmetry coordinate by appropriate arrows on a diagram of the molecular fragment. Usually these arrows are drawn in a length proportional to the coefficient of the internal coordinate in the symmetry coordinate,

Table 4.1.1 Standard deviations (σ) , variances (σ^2) , percentage of variance in bond distance symmetry coordinates (% BV) and percentage of variance in angular coordinates (% AV) for T-space data distribution.

	σ [Å]	σ^2 [Å ^{2}]	% BV	% AV
S1	.171	.029	15.0	—
S2	.134	.018	9.2	—
S3	.133	.018	9.2	—
S4	.358	.128	—	3.6
S5a	.254	.065	33.3	—
S5b	.254	.065	33.3	—
S6a	1.236	1.529	—	44.4
S6b	1.236	1.529	—	44.4
S7a	.233	.054	—	1.6
S7b	.233	.054	—	1.6
S8a	.272	.074	—	2.2
S8b	.272	.074	—	2.2

but in our case we chose to represent the degree of freedom by singly- or doubly headed arrows, instead, since the coefficients in our case are always either one or two. Figure 4.1.1 graphically illustrates the symmetry coordinates representing a TBP referred both to a TBP, as well as a SQP* Only *independent* distortions are shown, i.e. those which do not follow naturally from others. For example in $S_4(= 6^{-\frac{1}{2}}(\theta_{12} + \theta_{13} + \theta_{14} - \theta_{25} - \theta_{35} - \theta_{45}))$ the decrease in the angles *below* the equatorial plane perforce results from the simultaneous increase of all three *above* the plane. Consequently only the latter are detailed for S_4 in Figure 4.1.1. The symmetry coordinates representing a C_{4v} SQP are graphically illustrated in Figure 4.1.2, referred both to a SQP and a TBP. †

* This is done to facilitate the interpretation of data relating to both TBPs and SQPs which might exist in T-space; a point which will become clear in due course.

† This is done for similar reasons as before.

Table 4.1.2 Standard deviations (σ) , variances (σ^2) , percentage of variance in bond distance symmetry coordinates (% BV) and percentage of variance in bond angle symmetry coordinates (% AV) for S-space data distribution.

	σ [Å]	σ^2 [Å ^{2}]	% BV	% AV
S1	.317	.101	59.8	—
S2	.177	.031	18.6	—
S3	.492	.242	—	18.0
S4	.109	.012	7.2	—
S5	.801	.642	—	47.6
S6	.268	.072	—	5.4
S7a	.110	.012	7.2	—
S7b	.110	.012	7.2	—
S8a	.361	.131	—	9.7
S8b	.361	.131	—	9.7
S9a	.253	.064	—	4.8
S9b	.253	.064	—	4.8

Considering first the variance in the bond increment symmetry coordinates for the TBP, a comparison of Table 4.1.1 with Figure 4.1.1 reveals that there is a fair amount of variance in the lengths of the axial bonds (S_1) but an even greater one in those of the equatorial bonds (S_{5a}) , with two of the bonds becoming shorter as the third lengthens, or vice versa. Possibly related to this is the variance in the bond angle symmetry coordinates of S_{6a} and S_{6b} , the largest by far, with $6a$ indicating that there is a large variance in the equatorial angles, with one opening up while the other two become smaller, or vice versa. For S-space the largest contribution to the bond increment symmetry coordinates comes from S_1 , which suggests that there is a large variance in the apical distance of the SQP. For the bond angle symmetry coordinates the largest contribution comes from S_5 indicating a large variance in the trans-basal angles, with one angle opening up as the other closes.

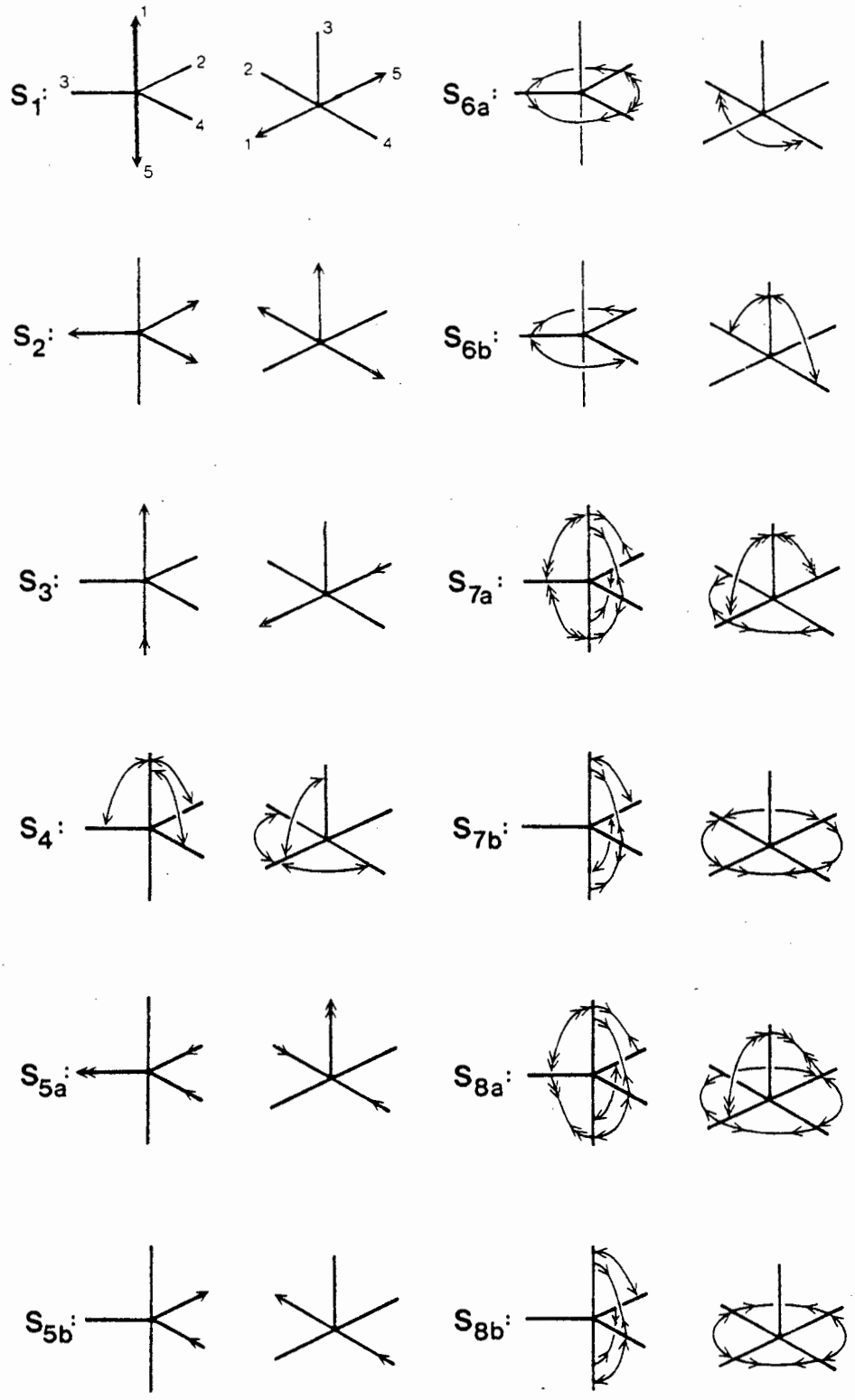


Figure 4.1.1 Graphic representation of D_{3h} symmetry coordinates referred to a TBP and a SQP. Single headed arrows indicate changes in internal coordinates whose coefficient inthe symmetry coordinate is one, double headed arrows represent those with coefficient equal to two. Only *independent* distortions are shown.

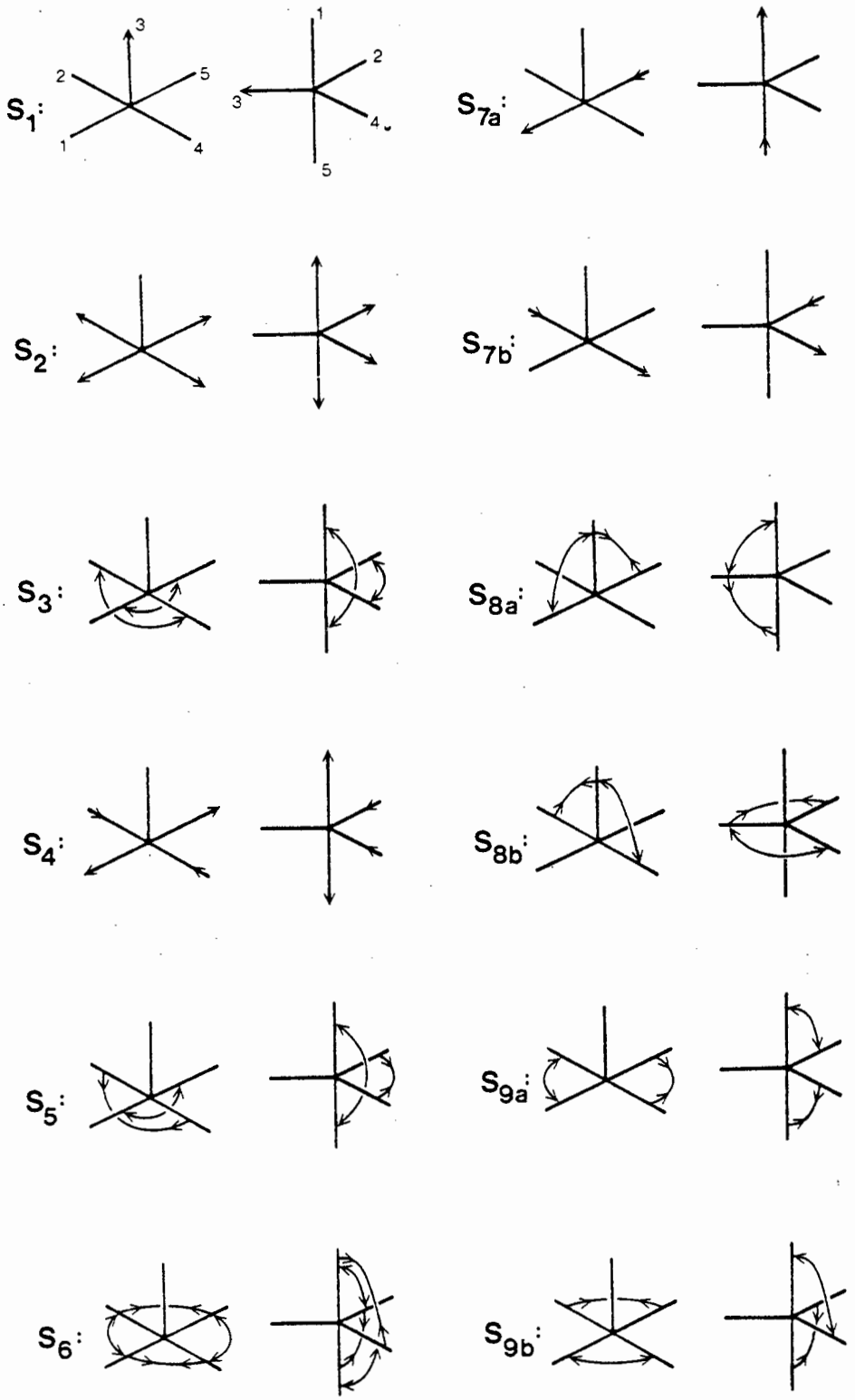


Figure 4.1.2 Graphic representation of C_{4v} symmetry coordinates referred to a SQP and a TBP. Single headed arrows indicate changes in internal coordinates whose coefficient in the symmetry coordinate is one, double headed arrows represent those with coefficient equal to two. Only *independent* distortions are shown.

(b) **Bivariate statistics**

The correlations between the symmetry coordinates can shed light on a number of things. Recall that they are orthonormal and transform according to the irreducible representations of the point group (... and consequently belong to specific symmetry species). It follows, therefore, that symmetry coordinates belonging to different symmetry species should ideally be uncorrelated. This implication may be used to test, firstly, whether the coordinates which we have derived do in fact form bases for the irreducible representations and, secondly, whether our manipulations of the data set exchanged parameters which ought not to have been interchanged with each other.

Tables 4.1.3 and 4.1.4 give the correlation matrix for the symmetry coordinates in T- and S-space, respectively. From these it may be seen quite clearly that symmetry coordinates from different symmetry species (see Figures 3.2.1 and 3.2.2) are at least linearly uncorrelated, while those of one symmetry are correlated, often quite highly. In T-space, for example, S_1 and S_2 , both belonging to the A_1' representation, have a correlation of -0.55, while they are uncorrelated with any other symmetry coordinates (all of which belong to a different species). Similarly, in S-space the coordinates S_1, S_2 and S_3 (all A_1) are highly correlated, whilst they are linearly uncorrelated with any others, all of which are of different symmetry.

We may therefore conclude that for the data set *as a whole* the symmetry coordinates which we have chosen are not linearly correlated across different symmetry species, neither in T- nor in S-space. Furthermore, we may see that the routines written to handle the data expansion according to the symmetry operations of the two point groups have operated correctly.

Of more interest, though, is an investigation of the particular distortions which pairs of correlated symmetry coordinates represent. For example, in T-Space $S_3 (= 2^{-\frac{1}{2}}(r_1 - r_5))$ represents the lengthening of one axial bond (r_1) and a shortening of the other (r_5) , while $S_4 (= 6^{-\frac{1}{2}}(\theta_{12} + \theta_{13} + \theta_{14} - \theta_{25} - \theta_{35} - \theta_{45}))$ is indicative of an "umbrella" type distortion of the TBP whereby the axial-equatorial angles *above* the equatorial plane are increasing, while those *below* are decreasing. In this case, however, because of the negative correlation between S_3 and S_4 $(r = -.67)$ the umbrella distortion is reversed. The *combination* of the two distortions,

Table **4.1.3** Correlation matrix for symmetry coordinates in T-space. Symmetry species to which the coordinates belong are also shown.

	A_1'		A_2''		E'						E''	
	S_1	S_2	S_3	S_4	S_{5a}	S_{5b}	S_{6a}	S_{6b}	S_{7a}	S_{7b}	S_{8a}	S_{8b}
S_1	1.00											
S_2	-0.55	1.00										
S_3	0.00	0.00	1.00									
S_4	0.00	0.00	-0.67	1.00								
S_{5a}	0.00	0.00	0.00	0.00	1.00							
S_{5b}	0.00	0.00	0.00	0.00	0.00	1.00						
S_{6a}	0.00	0.00	0.00	0.00	0.59	0.00	1.00					
S_{6b}	0.00	0.00	0.00	0.00	0.00	0.59	0.00	1.00				
S_{7a}	0.00	0.00	0.00	0.00	-0.01	0.00	0.45	0.00	1.00			
S_{7b}	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.45	0.00	1.00		
S_{8a}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	
S_{8b}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

which are shown in Figure 4.1.3, is reminding of a classical S_N2 coordinate, i.e. a bi-molecular nucleophilic substitution reaction at a tetrahedral centre. Another correlation of interest is that between S_{5a} and S_{6a} $(r = .59)$. These distortions, also represented in Figure 4.1.3, mirror part of the Berry intramolecular exchange coordinate along which one of the equatorial ligands (atom three in our case) acts as a pivot for the distortion, its distance (r_3) to the metal increasing, while the other two equatorial ligands move closer towards the metal. Concomitantly they move further apart from each other, increasing the angle between them (θ_{24}) while decreasing the other two equatorial-equatorial angles (θ_{23} and θ_{34}). The last correlation in T-space on which we shall focus is that between S_1 and S_2 $(r = -.55)$; its corresponding distortions are also shown in Figure 4.1.3. Here it can be seen that increases in the axial distances (r_1 and r_5) bring about decreases in the

Table 4.1.4 Correlation matrix for symmetry coordinates in S-space. Symmetry species to which the coordinates belong are also shown.

	A_1			B_1		B_2	E					
	S_1	S_2	S_3	S_4	S_5	S_6	S_{7a}	S_{7b}	S_{8a}	S_{8b}	S_{9a}	S_{9b}
S_1	1.00											
S_2	-0.78	1.00										
S_3	0.55	-0.48	1.00									
S_4	0.00	0.00	0.00	1.00								
S_5	0.00	0.00	0.00	0.39	1.00							
S_6	0.00	0.00	0.00	0.00	0.00	1.00						
S_{7a}	0.00	0.00	0.00	0.00	0.00	0.00	1.00					
S_{7b}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00				
S_{8a}	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.27	0.00	1.00			
S_{8b}	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	-0.27	0.00	1.00		
S_{9a}	0.00	0.00	0.00	0.00	0.00	0.00	-0.37	0.37	0.08	-0.08	1.00	
S_{8b}	0.00	0.00	0.00	0.00	0.00	0.00	-0.37	-0.37	0.08	0.08	0.00	1.00

equatorial distances (r_2, r_3, r_4) , as if there were only a fixed amount of "bonding power" to the metal, this being apportioned equally between the five ligands — if one (or two) are removed this leaves more "glue" for the others, which consequently bond more firmly.*

In S-space there are only three major correlations. The first, between S_1(=

* This phenomenon whereby there appears to be a constant amount of bonding electron density associated with the metal atom is reminiscent of Paulings' constant bond order concept, according to which the sum of the bond orders around a given atom stays conserved even as it forms or breaks bonds to other atoms (Pauling, L. The Nature of the Chemical Bond, Cornell University Press: Ithaca, New York, 1960.) We have chosen to term this coordinate the "constant amount of glue" coordinate as a play on the German word "Klebe"; G. Klebe wrote the program for the determination of standard bond lengths (Chapter 3) based on this concept of constant bond order, and "Klebe" also means "glue".

r_3) and $S_2(= \frac{1}{2}(r_1 + r_2 + r_4 + r_5))$ is by far the most important ($r = -.78$) and represents essentially the S-space equivalent of that between S_1 and S_2 in T-space — the "constant amount of glue" distortion. The second, between S_1 and S_3 ($r = .55$) indicates that an increase in the apical distance is correlated with an increase in the trans-basal angles of the SQP, i.e. a flattening of the SQP. The third, between S_2 and S_3 ($r = -.48$) shows that the flattening of the SQP (S_3) is correlated with a shortening of the metal-basal ligand bond distances. Taken together, the three distortions mirror those of a classical reversible association reaction of a square planar centre. They are illustrated in Figure 4.1.4. Finally, the data also exhibit a small distortion along the equivalent of the Berry coordinate in S-space. The correlation ($r = 0.39$) between S_4 and S_5 represents a part of the Berry distortion coordinate, although in this case it accounts for only 15 percent of the data variance.

One final point of interest is the amount of variance which each pair of variables can maximally describe. This may be judged from the square of the correlation coefficient, as pointed out in Section 2.5. From an inspection of the correlation matrix 4.1.3 it can be seen that in the case of T-space the linear relatedness of any two symmetry coordinates can maximally account for 45 percent of the sample variance (S_3 and S_4), while for S-space (Table 4.1.4) the figure is 61 percent (S_1 and S_2).

4.2 Choice of Clustering Techniques

As pointed out in Sections 2.6 and 2.7 different clustering techniques and, above all, different linkage criteria can quite conceivably yield different analyses, and for this reason one technique should always be supported by a second. In this case we chose to probe the 12-D data distribution by means of, firstly, non-hierarchical or relocation clustering employing the K-means method and, secondly, hierarchical agglomerative clustering using Ward's criterion.

The choice of relocation clustering in the first instance was dictated by practical considerations. At the University of Cape Town three statistical analysis packages offering clustering algorithms are available: CLUSTAN [1] , SPSS [2] and BMDP [3] . Of these the latter seemed the most versatile and user friendly, offering four different cluster analysis programmes, two of which could be used for the purpose of clustering

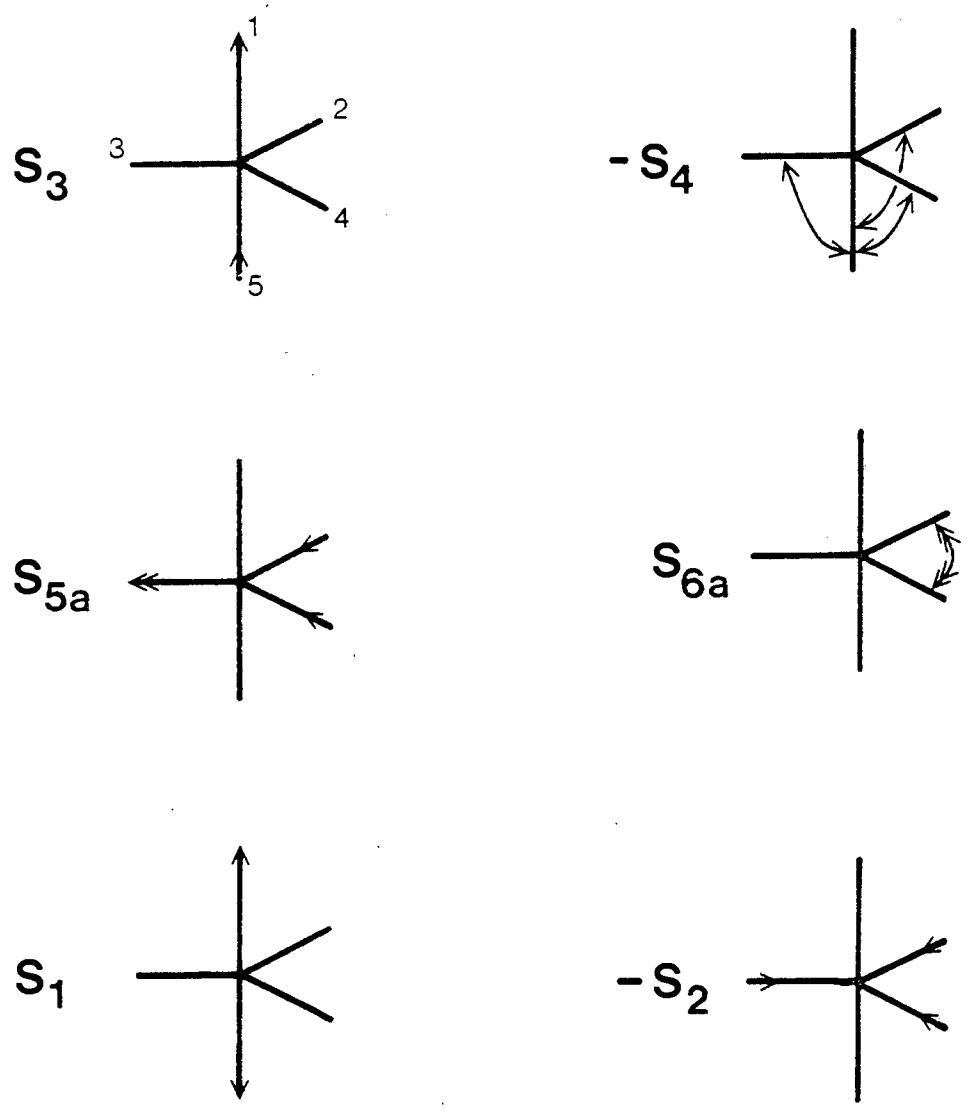


Figure 4.1.3 Distortions corresponding to some pairs of correlated symmetry coordinates in T-space. Where the correlation is negative, the inverse of one of the symmetry coordinates is represented, eg. $-S_4$. It is argued that the *correlated distortions* correspond to (1) the S_N2 coordinate, (2) the Berry coordinate and (3) the "constant amount of glue" coordinate.

---

together observations (cases) characterised by continuous variables (programmes P2M and PKM), one which clusters together simultaneously variables and cases characterised by discrete variables (P3M) and one which merely clusters together variables (P1M).

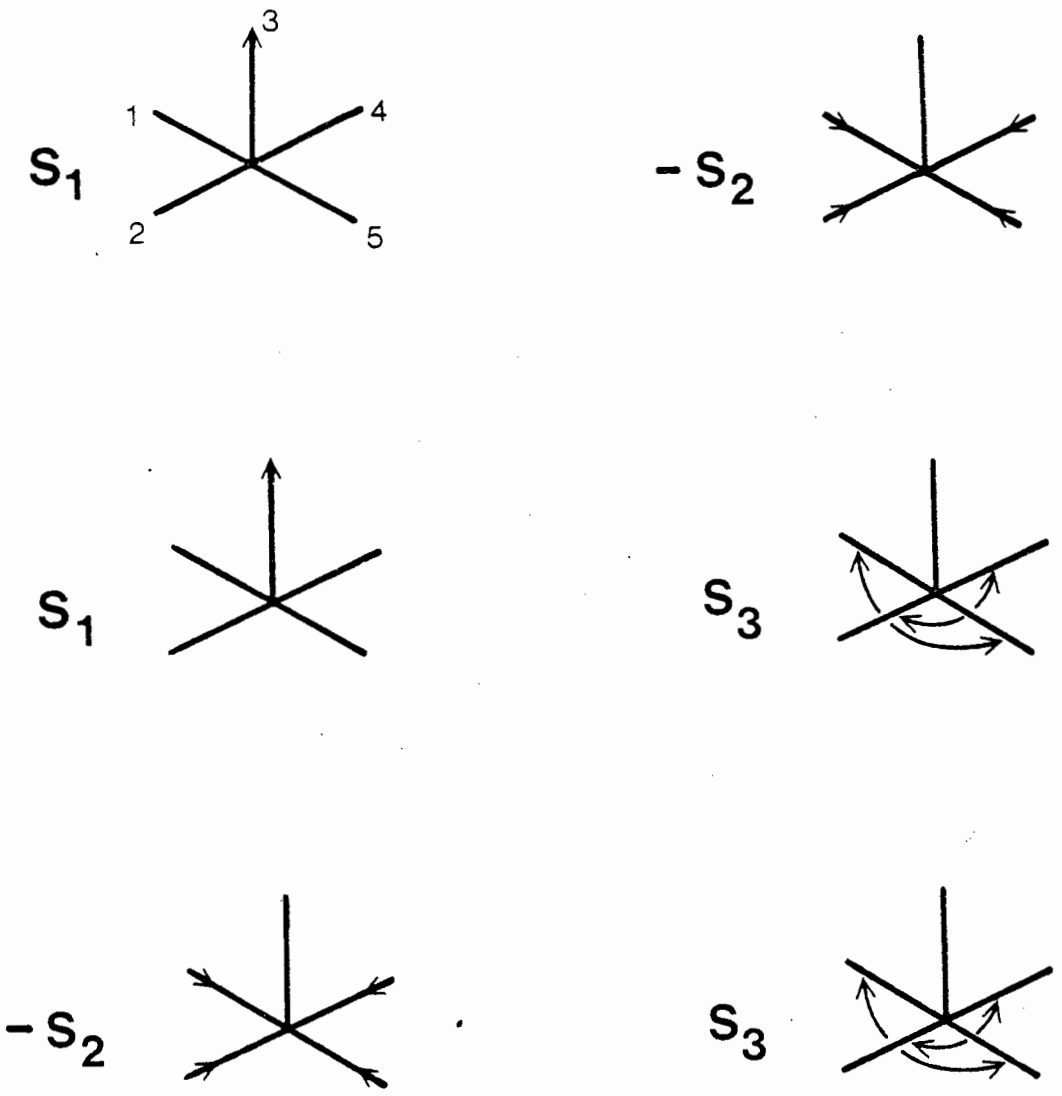


Figure 4.1.4 Distortions corresponding to some pairs of correlated symmetry coordinates in S-space. Where the correlation is negative, the inverse of one of the symmetry coordinates is represented, eg. $-S_2$. It is argued that the *sum* of the *correlated distortions* corresponds to a reversible association coordinate for a square planar centre.

P2M offers essentially a bottom-up (agglomerative) hierarchical clustering with a choice of four distance measurements and three linkage criteria. Output is in the form of a dendrogram and/or a distance (similarity) matrix, as well as several inter- and intracluster statistics. Unfortunately there are no plotting routines associated

with this programme, so that there is no option of visually displaying the results of the cluster analysis in any form other than a dendrogram or a distance matrix. These forms are quite inappropriate in our case, however, since the output for our data set comprising 2352 data points in T-space, for example, would run over circa 1600 pages in the case of the dendrogram, or circa 350 pages in that of the matrix.

PKM, on the other hand, provides some reasonably comprehensive line printer plotting techniques which enable one to display the analysis results visually in a number of different ways. More importantly, though, programme PKM employs non-hierarchical clustering techniques, specifically K-means clustering, and therefore produces no dendrogram or distance matrix. Instead, it partitions the data set into K clusters, and then allocates each data point to the cluster whose center is closest to it at the completion of each run. (K-means clustering has already been described in detail in Section 2.6.)

PKM affords the user the opportunity of specifying either the number of clusters into which the data are to be sorted (K), or the cluster centrotypes around which the data points are to be clustered. The calculations can be performed either on the raw data or on the standardised data, and the programme offers four standardisation procedures: standardisation to unit variance, unit covariance, unit within-cluster variance and unit within-cluster covariance. Since the effect of standardisation is to change the definition of the distance measure, it follows that these five options yield five different distance measures. If the raw data are used, for example, the corresponding distance measure would be the Euclidean distance, whereas if the data were standardised to unit within-cluster covariance, the Mahalanobis distance would be calculated.

4.3 Non-hierarchical Cluster Analysis

(a) Similarity measure and robustness criteria

In our case we decided to use the unstandardised data with angle measurements scaled to radial displacements of the order of one Angström, as outlined in Section 3.5. *This in effect meant that we were choosing the Euclidean distance as a measure of similarity.* Furthermore, two criteria were chosen for determining the "goodness" or "robustness" of a particular result. The first criterion derives from

the symmetry introduced into the expanded data set (as has already been discussed in Sections 2.4 and 2.6) and its basic assumption is that whatever symmetry is present in the data set must be reflected in the results of both the cluster and the factor analysis. The second criterion borrows from Massart's concept of robustness and it is related to the actual cluster affiliation of a given data point (see Section 2.7). Firstly, a cluster in T- or S-space is considered robust, if its members at a level K (i.e. K clusters) do not intermingle at higher levels K with members of other clusters formed at the first level K. Secondly, and in this study perhaps more importantly, *a cluster in T-space is considered robust if its members are similarly clustered together in S-space*, and vice versa. This latter criterion derives from the idea that conformations clustered together as, say, TBP's in T-space ought to be similarly defined (and therefore clustered together) in S-space, otherwise the classification must have been an artefact of the clustering algorithm, rather than an inherently rigorous classification.

(b) **Initial attempts**

Since we had no *a priori* concept of the likely clustering pattern which our data set would exhibit in 12-D space, we had no prior knowledge of what the possible cluster centrotypes might be, and by the same token could not estimate the number of clusters which might be formed. It was therefore necessary to begin with a hit-and-miss method, and first attempts centred around fitting a twelve cluster model to the data in T-space and a eight cluster model around that in S-space. The rationale behind these attempts was that we expected *at least* that number of clusters as a consequence of our artificial data manipulation, whereby the original data sets were expanded twelve- and eight-fold, respectively.

However, judging by both the symmetry and the affiliation criteria our early attempts at arriving at some sensible clustering were leading nowhere. Values of K from 8 to 30 were used in PKM, but we were not able to find any coherent or symmetrical clustering pattern. It gradually became clear that the larger the value of K, the more complicated the picture became, and we soon realised that the assumption on which we were basing the higher K values was not necessarily correct.

Consider, for example, the situation where the data points are quite densely

clustered together approximately in the centre of one asymmetric unit of a given multi-dimensional space, as shown in Figure 4.3.1(a). Then, when these points are subjected to the type of data expansion which we have applied in our case, they would be transformed into their isometric partners in the adjacent asymmetric units, as shown in Figure 4.3.1(b). In this case the final number of clusters formed in the entirety of the parameter space would be equal (or close) to the number of symmetry operations of the point group (four in this instance).

However, where the original cluster is close to a symmetry element, as in Figure 4.3.1(c), subsequent expansion does not necessarily lead to a situation where the number of clusters formed is equal to the order of the point groups. In fact, the number of clusters may be considerably smaller, depending on the data distribution prior to expansion, as shown in Figure 4.3.1(d).

In light of the above consideration it was decided to attempt to fit models involving smaller K's to the data set, starting with K = 2 through to 10, and in the process employing our two robustness criteria outlined earlier. Initially the results were plotted using the default mode of PKM which plots bivariate scattergrams of numbers indicating cluster membership onto a plane through the centers of the three most populous clusters. In this way a plot very similar to that depicted in Figure 4.3.2 emerged for the T-space data set with K = 4. Its rather obvious three-fold symmetry immediately drew our attention.

(c) **T-space**

On examining the result obtained for K=4 more closely it became clear that a highly symmetric clustering pattern had emerged, which in fact mirrored the three-fold symmetry of the D_{3h} point group. The algorithm had yielded three identical but well separated clusters (clusters number one, two and four) situated at the corners of an equilateral triangle in the 12-D space, with the last, also well separated cluster (number three) placed in the centre of the triangle.* These clusters will from now on be referred to as T1, T2, T4 and T3, respectively. Table 4.3.1 gives some inter- and intra-cluster statistics, while Figure 4.3.2 shows a projection

* This three-dimensional description of the arrangement of the clusters relative to each other is meant merely to enable a visualisation of the result, since such a description can obviously be of only very limited use in describing a twelve-dimensional picture.

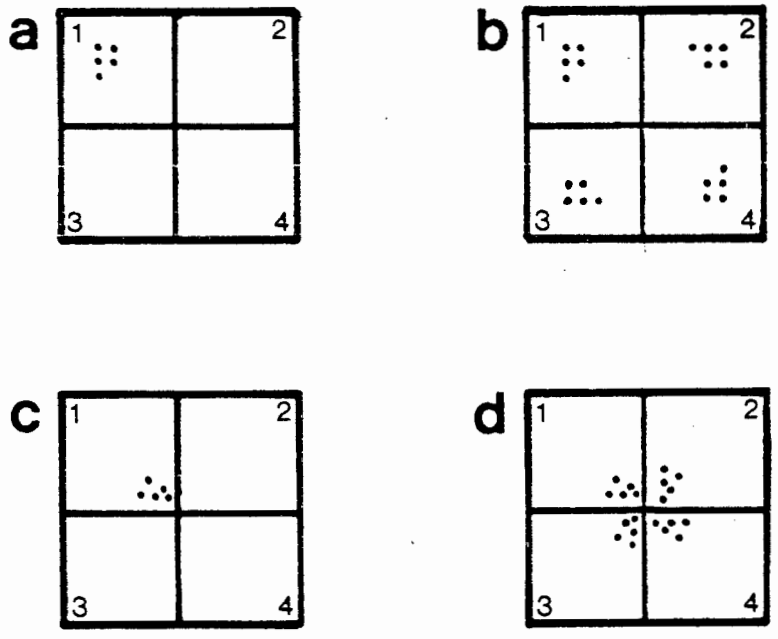


Figure 4.3.1. Diagram schematically showing relationship between data distribution *prior* to expansion by the symmetry elements of the point group ((a) and (c)), and number of clusters finally existing in entirety of the parameter space ((b) and (d)). Point group used is C_4 and the numbers 1, 2, 3 and 4 refer to the four asymmetric units comprising the hyperspace. They are here projected onto two dimensions.

---

of the four clusters onto the plane defined by the first two factors extracted from the T-space data set. This plane was chosen above the default setting of PKM in order to effect the maximum cluster separation possible, so as to enhance the visual interpretation.

The interest in obtaining an optimal clustering lies, of course, not merely in the number of clusters nor their symmetrical relation, but rather in finding the *cluster centrotype*, or archetypal conformation forming the centre of the cluster, around which the observed molecular geometries aggregate. The programme PKM allows the user to attach a label indicating cluster affiliation to the end of each observations'

Table 4.3.1. Inter- and intra-cluster statistics for T-space. Average size = average distance of cluster members from centre of cluster [Å]. Distance matrix gives inter-cluster distances [Å].

Cluster	T1	T2	T3	T4	
Number of members:	548	548	708	548	(= 2352)
Average size:	0.885	0.885	0.727	0.885	

Distance matrix:	T1	T2	T3
T2	3.457		
T3	2.002	2.002	
T4	3.457	3.457	2.002

record. This enables the user to subsequently analyse individual clusters for such statistics as the mean and variance of the variables *for that cluster*. Using this option we determined the average internal angles of the ML_5 fragments composing the cluster centrotypes. These are given in Table 4.3.2, together with diagrams depicting the centrotype conformation and the corresponding permutation of the ligand atoms.

Quite clearly the T1, T2 and T4 centrotypes correspond to a slightly distorted SQP of C_{2v} symmetry with trans-basal angles of $169°$ and $161°$, while that of T3 corresponds to a slightly distorted TBP of D_{3h} symmetry with its axial angle slightly closed to $174°$.* Moreover, the three SQP conformers coincide with the three possible SQPs which may be formed via the Berry mechanism from the TBP corresponding to the T3 centrotype, as shown in Figure 4.3.3. According to this scheme, starting from the TBP in the centre, either of three SQPs may potentially be formed, depending on which of the three equatorial ligand atoms (2, 3 or 4) acts as the pivot for the distortion to subsequently become the apical atom in the SQP.

It would appear, therefore, as if the observed conformations, when referenced to a perfect TBP, fall into two groups — those which are more trigonal bipyramidal

* The average axial angle of the centrotype TBP cannot be $180°$, since the individual observed angles are all less than $180°$ with only three exceptions.

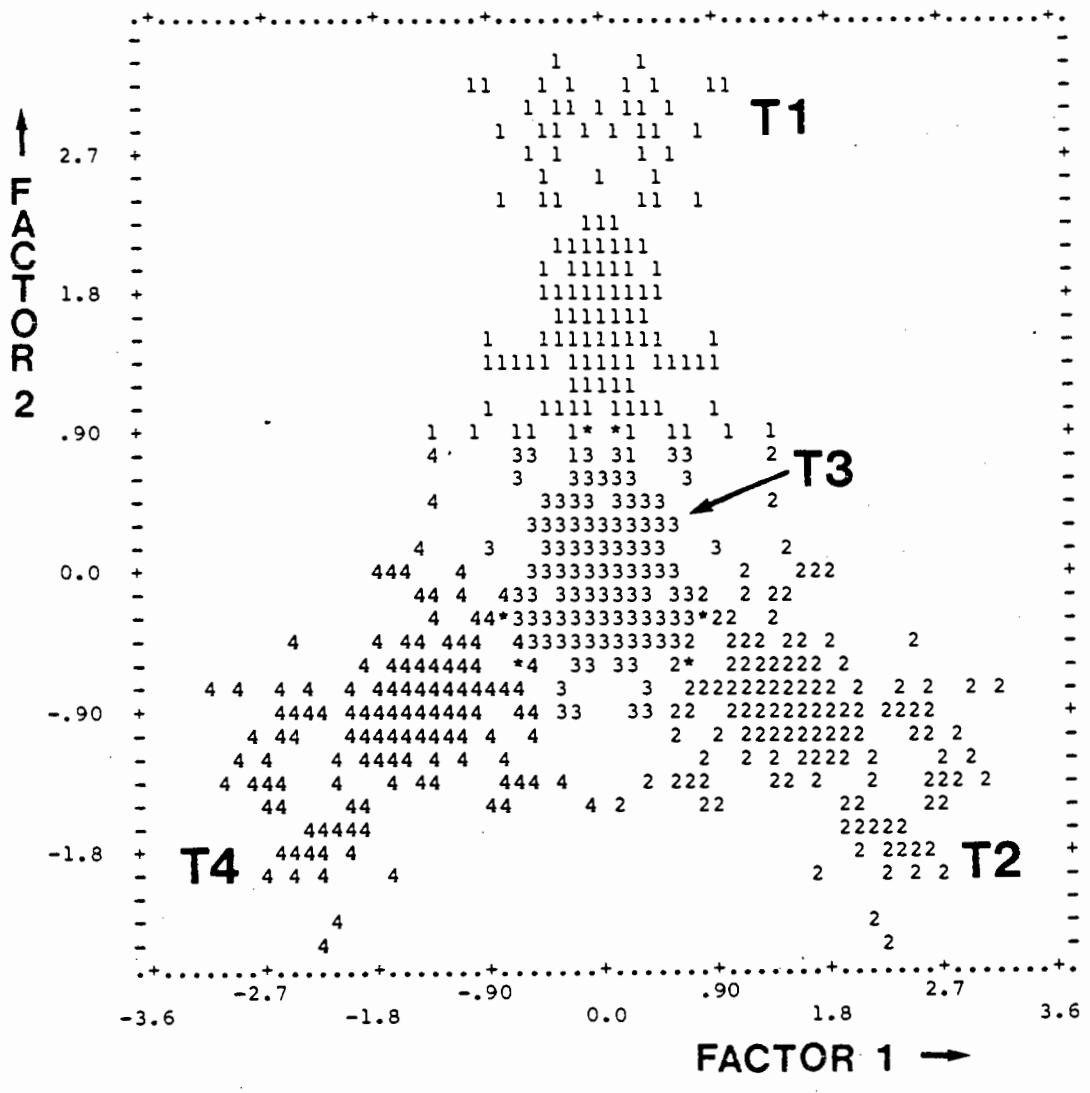


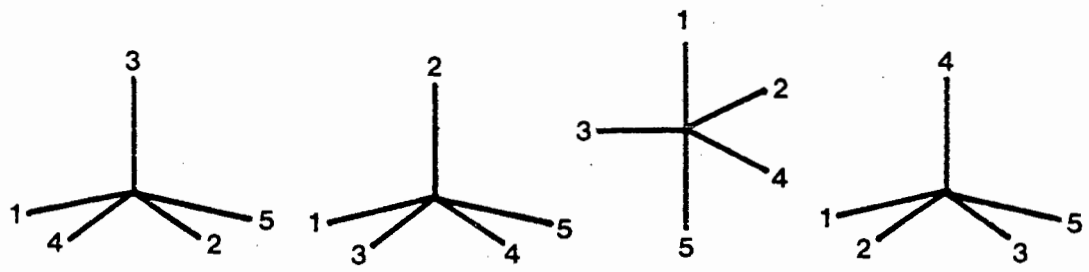
Figure 4.3.2 Projection of T-space clusters onto plane defined by the first two factors. Numbers indicate cluster affiliation. T1, T2, T3 and T4 is the notation used for the various clusters throughout this study. Three-fold symmetry is slightly distorted by the line plotter.

---

in nature and those which tend more towards a square pyramidal geometry. This is hardly surprising in view of what is known about five-coordination. What is more striking, though, is the fact that the observed molecular fragments have been sorted into these two groups (completely) *automatically* by the algorithm.

The emergence of three archetypal SQPs corresponding to those expected when

Table 4.3.2. Internal angles [°] of cluster centrotypes in T-space. Absolute conformations of T1, T2 and T4 are identical, but the ligand sites are permuted. Estimated standard deviations are given in parenthesis. M is the number of cluster members, N is the number of refcodes in the cluster.



	Cluster			
Angle	T1	T2	T3	T4
θ_{12}	89(6)	94(7)	90(7)	89(6)
θ_{13}	94(7)	89(6)	90(7)	89(6)
θ_{14}	89(6)	89(6)	90(7)	94(7)
θ_{15}	169(6)	169(6)	174(5)	169(6)
θ_{23}	99(10)	99(10)	119(7)	161(9)
θ_{24}	161(9)	99(10)	119(7)	99(10)
θ_{25}	89(6)	94(7)	90(7)	89(6)
θ_{34}	99(10)	161(9)	119(7)	99(10)
θ_{35}	94(7)	89(6)	790(7)	89(6)
θ_{45}	89(6)	89(6)	90(7)	94(7)
M	548	548	708	548
N	137	137	59	137

considering the possible distortions of the archetypal TBP indeed tempts the conclusion that the data map out the expected (and often empirically observed) reaction pathways whereby an archetypal TBP distorts into a SQP along a coordinate maintaining C_{2v} symmetry.* In our case the square pyramidal conformer formed

* Of course, the presence of precisely these three isometric SQPs results from the data expan-

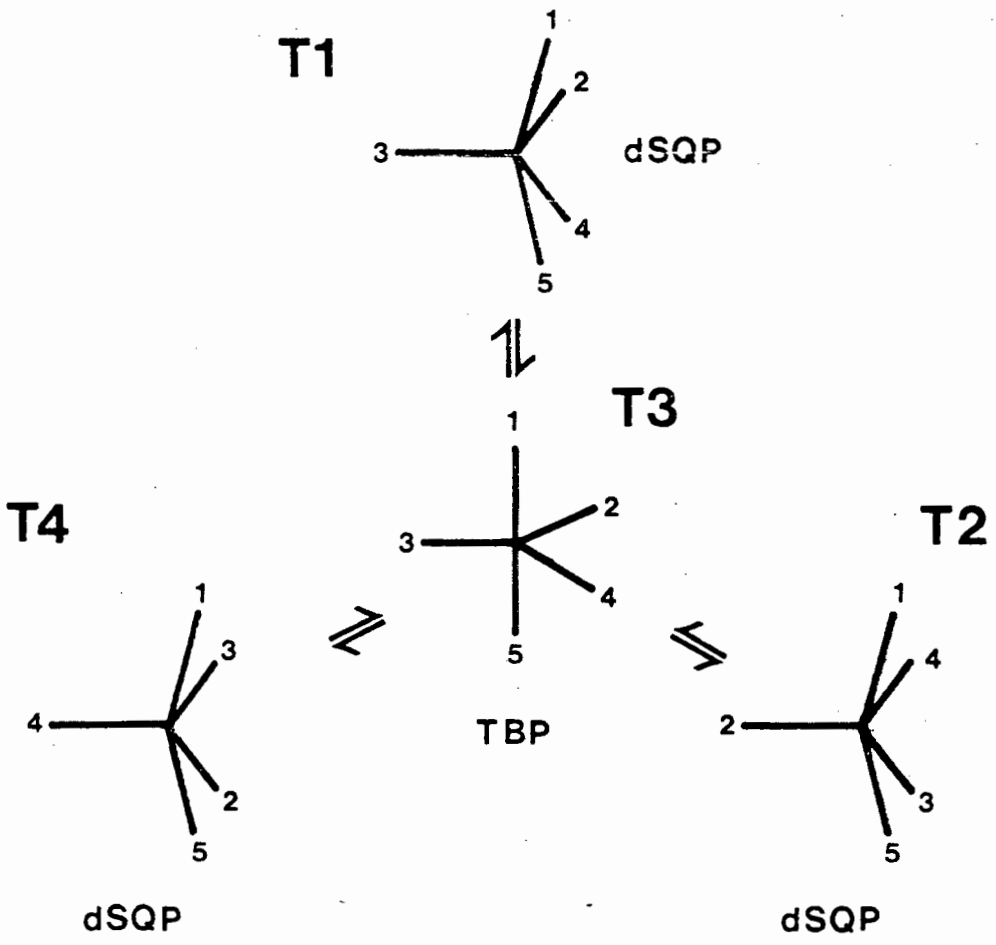


Figure 4.3.3. Diagram showing three possible distortions of the central TBP (corresponding to the T3 centrotype) into three SQPs (corresponding to the T1, T2 and T4 centrotypes).

---

has preserved only the C_{2v} symmetry of the distortion coordinate and the D_{3h} TBP, but not yet assumed a C_{4v} geometry. This, however, may be the result of attempting to fit molecular geometries approaching C_{4v} symmetry into a D_{3h} framework (since we are referring them to a TBP) — the highest symmetry that these compounds could exhibit under such circumstances is C_{2v} , since there is no

---

sion, but this fact does not detract from the argument developed subsequently.

C_4 symmetry element present in the D_{3h} point group.

(d) **S-space**

The critical test of whether the clustering pattern observed in T-space for K = 4 is meaningful, lies of course, in how well this would compare with the results of a similar analysis in S-space. It turns out, in fact, that the two analyses complement each other extremely well.

For K = 4 the algorithm yields a symmetrical Y-shaped clustering pattern, with two identical clusters (clusters number one and three) situated at the tips of the bi-pronged fork, and two unequal clusters (numbers four and two) placed at the centre and the bottom tip of the Y, respectively. These clusters will henceforth be referred to as S1, S3, S4 and S2, respectively. The four clusters are well separated in comparison to their average size, except for clusters S2 and S4, where the distance between their centres is of the same order as the average size of the cluster. The reason for this will become clear in due course. Table 4.3.3 shows some relevant inter- and intra-cluster statistics.

The Y-shaped arrangement of the four clusters is best* represented by a plot in which the data are projected onto the plane composed by the first and fourth factors extracted from the S-space data set, as shown in Figure 4.3.4. This particular projection shows up only the two-fold symmetry of the data space, and not the expected four-fold one. The reason for this is that the four-fold symmetry is associated with the sub-space made up of the *first three* factors, and consequently will not show up in a plot involving a lower factor. We have found, nevertheless, that this particular plot is the most informative and makes the most sense of the statistics in Table 4.3.3.

Table 4.3.4 gives the internal angles characterising the various cluster centrotypes of S1, S2, S3 and S4, and shows schematic diagrams of their conformation. Although we do not intend at this stage to comment on differences between the bond lengths of the centrotypes, it is important to point out that there is a sub-

* In this case we have chosen the subspace containing the factor 1/factor 4 plane to represent the relative arrangement of the four clusters, since this plane is most suited to making three-dimensional sense of the 12-D picture.

Table 4.3.3. Inter- and intra-cluster statistics for S-space. Average size = average distance of cluster members from centre of cluster [Å]. Distance matrix gives inter-cluster distances [Å].

Cluster		S1	S2	S3	S4	
Number of members:		276	312	276	704	(= 1568)
Average size:		0.697	0.932	0.697	0.719	
Distance matrix:		S1	S2	S3		
	S2	1.836				
	S3	2.531	1.836			
	S4	1.441	0.833	1.441		

stantial difference in the apical bond distance between the SQPs characteristic of S2 and S4. We have attempted to reflect this in the diagrams of Table 4.3.4. From this table it can readily be seen that clusters S1 and S3 represent slightly distorted TBPs with C_{2v} symmetry which are marginally displaced along the Berry coordinate towards a SQP, as judged by the slight reduction in the axial angle and two of the equatorial angles, and a concomitant opening of the remaining one. S2 is characterised by a "flattened SQP" (henceforth fSQP) of C_{4v} symmetry with the central metal almost coplanar with the four basal ligand atoms and the apical ligand at a considerable distance. S4, on the other hand, is characterised by a more "conventional elevated SQP" (henceforth eSQP) also of C_{4v} symmetry, whose metal, though, is clearly out of the basal plane (as judged by the trans-basal angles) and whose apical ligand is closer to the metal than was the case for the S2 centrotype.

Here again a clear picture which is logically consistent with well established chemical principles emerges, when these cluster centrotypes are viewed in the correct sequence, as shown in Figure 4.3.5. Let us begin with the S2 centrotype fSQP, which could be viewed also as an "early intermediate"in the reversible addition of a fifth ligand to a square-planar, four-coordinate metal centre. From here there is a gradual progression to the eSQP (S4 centrotype) as the apical ligand moves closer towards the metal. Finally, the "true" eSQP intermediate has the possibility of distorting

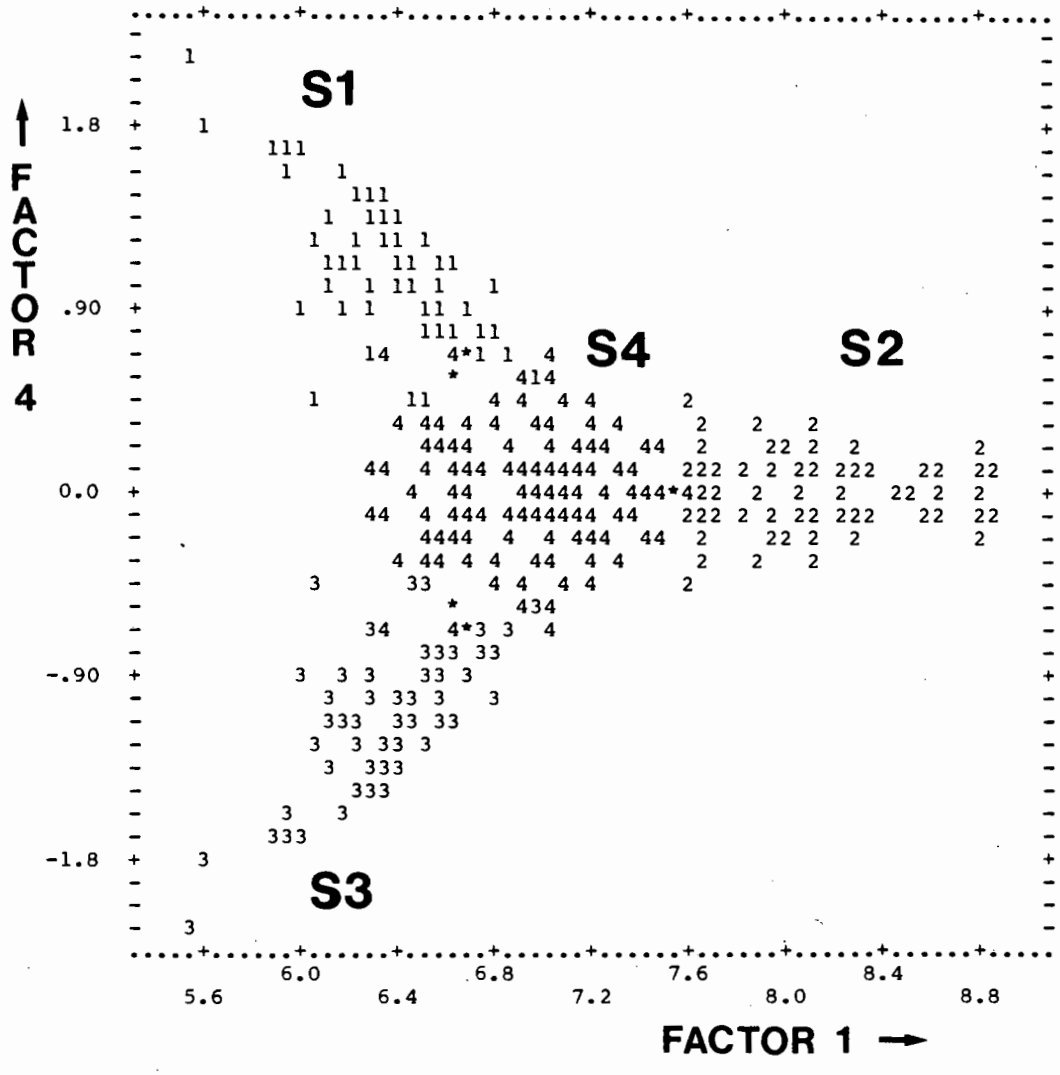


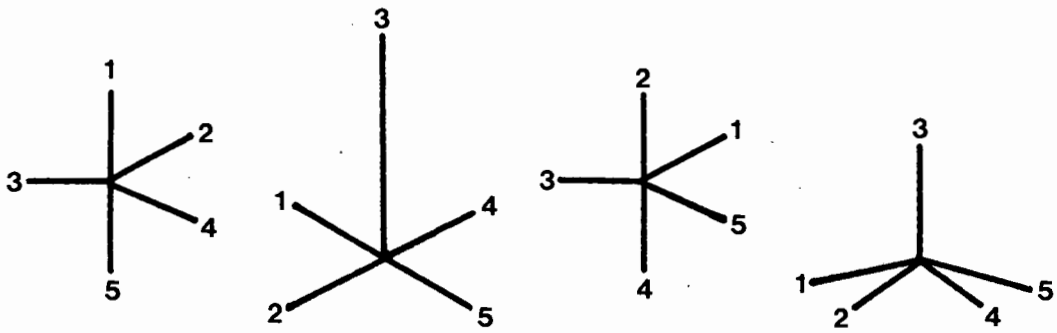
Figure 4.3.4. Projection of S-space clusters onto plane defined by first and fourth factors. Numbers indicate cluster affiliation. S1, S2, S3 and S4 is the notation used for the respective clusters throughout this study.

---

into either of two TBPs, depending on which of the trans-basal angles (θ_{15} or θ_{24}) opens up toward 180^{o} .

Similar to the case with the C_{2v} SQP centrotypes in T-space the TBP centrotypes of S1 and S3, corresponding to the two TBPs which could be formed from the S4 eSQP, do not attain the expected D_{3h} symmetry, instead. Again this may well be the result of attempting to fit observed conformations approaching D_{3h}

Table 4.3.4. Internal angles (°) of cluster centrotypes in S-space. Absolute conformations of S1 and S3 are identical, but the ligands are interchanged. Estimated standard deviations are given in parenthesis. M is the number of cluster members, N is the number of refcodes in the cluster.



	Cluster			
Angle	S1	S2	S3	S4
θ_{12}	89(6)	90(7)	89(6)	89(5)
θ_{13}	92(7)	92(10)	115(7)	97(7)
θ_{14}	89(6)	90(7)	89(6)	89(5)
θ_{15}	174(5)	171(5)	128(9)	163(7)
θ_{23}	115(7)	92(10)	92(7)	97(7)
θ_{24}	128(9)	171(5)	174(5)	163(7)
θ_{25}	89(6)	90(7)	89(6)	89(5)
θ_{34}	115(7)	92(10)	92(7)	97(7)
θ_{35}	92(7)	92(10)	115(7)	97(7)
θ_{45}	89(6)	90(7)	89(6)	89(5)
M	276	312	276	704
N	69	39	69	88

symmetry into a C_{4v} framework. The highest symmetry which the two could have in common is C_{2v} .

(e) T-space versus S-space

At first glance the compatibility of the results from T-space with those from S-

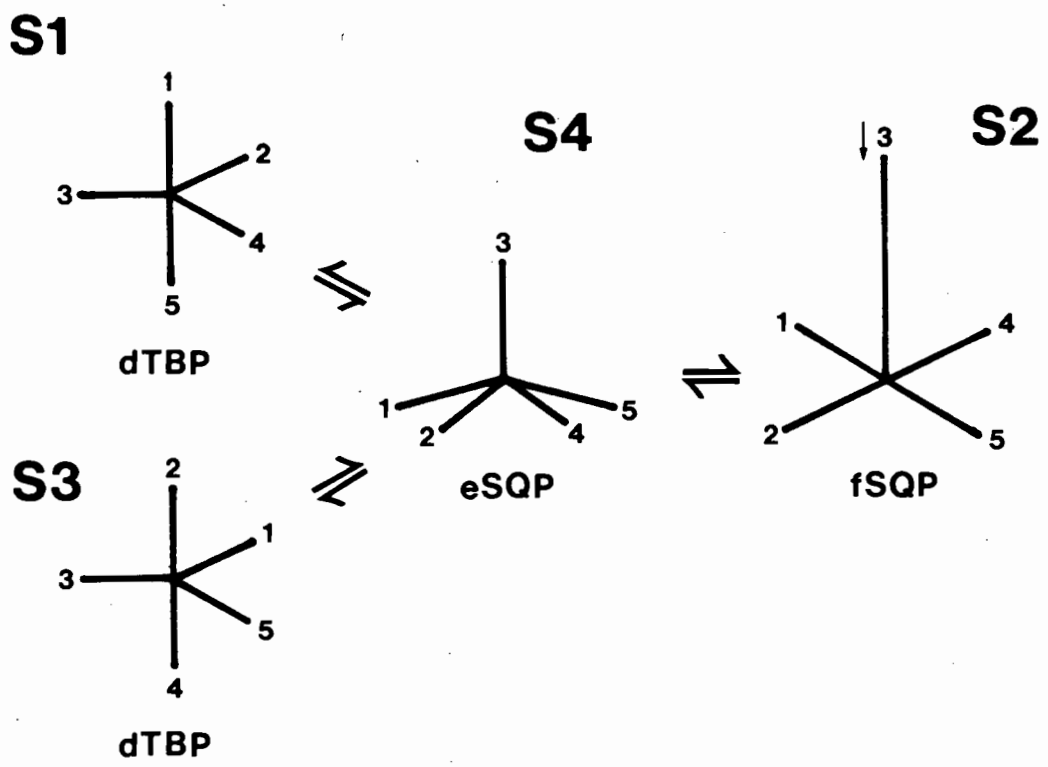


Figure 4.3.5. Diagram showing possible distortions of the fSQP (corresponding to the S2 centrotype) through an eSQP (S4 centrotype) to either of two dTBPs (S1 and S3 centrotypes).

---

space may not be obvious — after all, the one analysis gives us one archetypal TBP with D_{3h} symmetry and three identical C_{2v} SQPs, while the second yields two C_{2v} TBPs and two non-identical C_{4v} SQPs. Notwithstanding this, the results are in fact similar.

In essence what the algorithm has produced in both cases is a breakdown of the dataset into two groups — those tending towards a TBP conformation and those tending towards the SQP. The apparent differences stem from the different symmetries of the data spaces. In T-space, whose symmetry is that of the D_{3h} point group, those observed geometries tending towards the TBP will naturally aggregate around the archetypal D_{3h} TBP at the origin. On the other hand, *those compounds whose geometries approach the SQP, will need to cluster around*

an archetype whose symmetry is as close as possible to that of a perfect SQP (C_{4v}), *while still being an element of the data space of* D_{3h} *symmetry.* Such an archetype would correspond, as it were, to an intersection between the sets of subgroups of the D_{3h} and C_{4v} point groups. Moreover, it would correspond to the highest symmetry subgroup which is an element of the set composing this intersection. This subgroup is the C_{2v} point group. Consequently the square pyramidally distorted geometries aggregate about an archetypal C_{2v} "square" pyramid instead of one with C_{4v} symmetry.

Similar reasoning may be applied in explaining the absence of an archetypal D_{3h} TBP in S-space — the symmetry of the data space simply does not allow for its existence. Instead, those geometries approaching a TBP conformation are forced to cluster around a C_{2v} "TBP".

Proof for the above explanation may be obtained from a comparison of the actual cluster membership in each of the two spaces. Before going on to do this, however, we need to briefly discuss the existence in S-space of two C_{4v} SQP archetypes. As mentioned previously an "ideal" SQP of C_{4v} symmetry may have both an apical bond length and identical trans-basal angles of any value. Theoretically, therefore, both "flattened" SQPs with large trans-basal angles and apical bond lengths, and "elevated" SQPs with smaller angles and bond distances may exist. Clearly, objective descriptions of these two conformations would be extremely difficult to formulate — when is a SQP flattened and when is it elevated? Indeed, in our earlier study of nickel complexes [4] we found a smooth transition from the one to the other. We believed that this was the result not only of an absence of criteria differentiating between the fSQP and the eSQP, but that it also mirrored the (most likely) rather small energetic differences between the two, leading to an even spread of observed geometries between these two conformations.

Nevertheless, in this study the algorithm employed managed to differentiate between the two, splitting up the square-pyramidally disposed structures into two clusters corresponding to the fSQP (S2) and the eSQP (S4). This speaks volumes for the power and accuracy of this form of analysis. To enlarge on this point let us review the observed structures as indeed representing various points on the Born-Oppenheimer energy hyper-surface of the square pyramidal ML_5 fragment. The

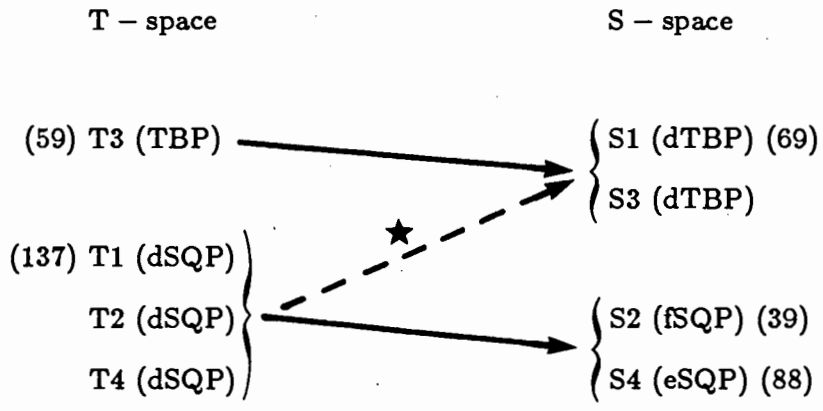
two points representing the archetypal fSQP and eSQP are then very likely to be separated by only a very low energy "pass" — relative to each other they would represent no more than two rather flat "dimples" on the energy surface. Consequently, the spread of observed molecular structures around these points will be large and diffuse, with no clear boundaries established between the two distributions. Purely *graphical* techniques would consequently be unlikely to suffice as a means of differentiation. This point becomes clear when one compares the average sizes of S2 and S4 (0.697 Å and 0.932 Å) with the distance between them (0.833 Å) — the clouds of data points quite possibly diffuse into each other. Despite this, the *statistical* technique employed in this case evidently succeeded in fixing the positions of the archetypal conformations somewhere in the centre of the cloud of data points surrounding them.

Let us continue now, to comparing the actual cluster membership in the two data spaces. This is outlined schematically in Table 4.3.5. As can be seen the entire membership of T3, representing a TBP in T-space, is transferred to the identical clusters S1 and S3, representing the TBP in S-space. Similarly almost the entire membership of the identical clusters T1, T2 and T4 representing the SQP in T space, is transferred to either of S2 or S4, both representing the SQP in S-space. With only ten exceptions, therefore, the observed molecular geometries are similarly classified in the two data spaces. In terms of both the criterion relating to the expected symmetry of the clustering pattern, and that of the cluster membership in the two spaces, we may therefore safely assume that the clusters formed at the K = 4 level are robust. The only difference between the result obtained for T-space and that for S-space is that the SQP cluster in the former (T1, T2, T4) has been split in two in the latter, according to whether the coordination polyhedron more closely approximates to the fSQP or the eSQP. This, as we have argued, is facilitated by the change in symmetry of the data space from D_{3h} to C_{4v} in going from T- to S-space. In other words, ***the closer the symmetry of the data space to that of the archetypal geometry which a given observed molecule approaches, the more enhanced the classification.***

(f) T-space versus S-space continued

Although, as we have seen, the algorithm sorted the square pyramidally dis-

Table 4.3.5. Analysis of cluster membership of trigonal bipyramidal and square pyramidal clusters in T- and S-space. The only structures which are defined differently in the two data spaces are those indicated by the asterisk. Numbers in parenthesis indicate number of refcodes for each cluster.



★ = CMBPNI, CMPMNI, CSMRHC, DPPCRH, DPMCRH10, EDCRCN, IPESNI, SALDNI, TMAGEP, BEZPIF.

posed structures into two groups, viz. the eSQP and fSQP, this was not a simple task, and indeed it needed 21 iterations to arrive at the final clustering for K = 4 in S-space. This ought to be compared to only nine iterations which were needed to arrive at the final clustering for K = 4 in T-space, in spite of the fact that T-space contains 784 more points for classification than does S-space. The difficulty of separating out the eSQPs from the fSQPs becomes even more obvious when one considers that only eleven iterations are needed to achieve an optimum clustering in S-space when K = 3. At this value of K the algorithm collapses S2 (fSQP) and S4 (eSQP) into one cluster containing 996 members, while retaining S1 and S3 (TBP) as two clusters with 288 and 284 members, respectively. Figure 4.3.6 shows a scatterplot of this clustering pattern projected onto the same plane as that in Figure 4.3.4, with which it should be compared. Table 4.3.6 gives the internal angles corresponding to the centrotypes of the three clusters which we will refer to as $S^3 1, S^3 2$ and $S^3 3$.

The very slight dissymmetry of this clustering is evidenced by the unequal number of cluster members for $S^3 1$ and $S^3 3$. These clusters are represented by

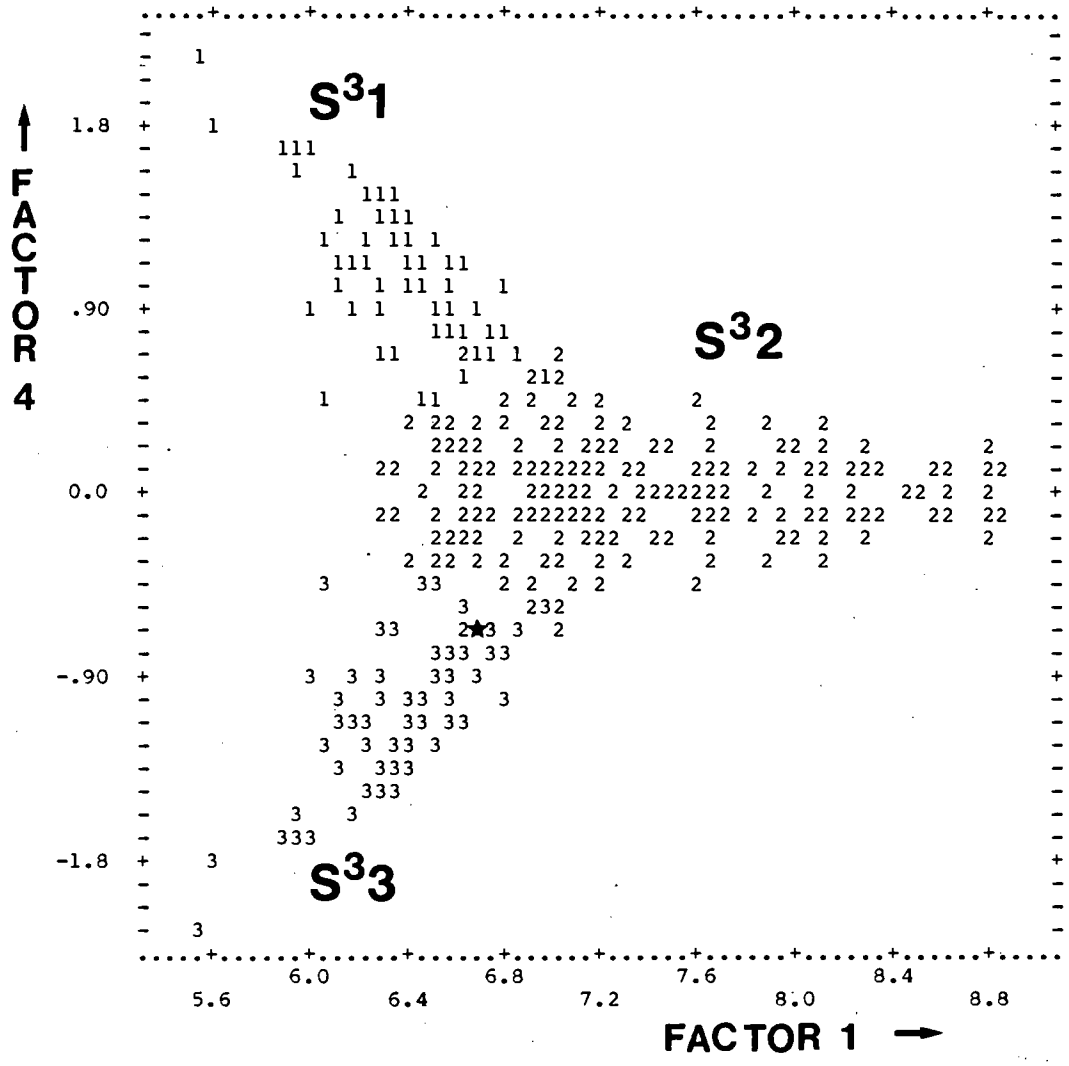
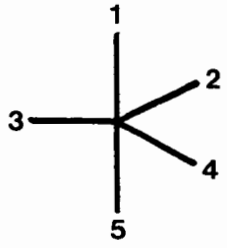


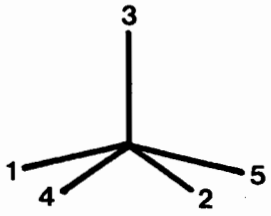
Figure 4.3.6. Projection of S-space clusters for K = 3 onto the plane defined by factor 1 and factor 4 extracted from S-space for K = 4. Numbers indicate cluster affiliation, and the asterisk indicates the superposition of points belonging to different clusters.

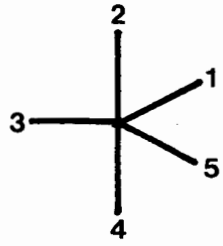
---

distorted TBP centrotypes of C_{2v} symmetry which are almost isometric partners to each other as those of S1 and S3 are when K = 4. Indeed the "typical" TBPs identified in S-space for K = 3 and K = 4 are very similar, as a comparison of Tables 4.3.4 and 4.3.6 show. Also the membership of S1 (dTBP) for K = 4 and S^31 (dTBP) for K = 3 are almost identical with the membership of the former

Table 4.3.6. Internal angles [°] of cluster centrotypes for K = 3 in S-space. Estimated standard deviations are given in parenthesis. M is the number of cluster members, N is the number of refcodes in the cluster.







	$S^3 1$	$S^3 2$	$S^3 3$
θ_{12}	89(6)	89(6)	89(6)
θ_{13}	92(7)	96(8)	115(8)
θ_{14}	89(6)	89(6)	89(6)
θ_{15}	174(5)	166(7)	129(9)
θ_{23}	114(8)	96(8)	92(7)
θ_{24}	129(10)	166(7)	174(5)
θ_{25}	89(6)	89(6)	89(6)
θ_{34}	114(8)	96(8)	92(7)
θ_{35}	92(7)	96(8)	114(8)
θ_{45}	89(6)	89(6)	89(6)
M	288	996	284
N	66	131	65

being taken over into that of the latter except for only three compounds which are lost to $S^3 2$ (SQP). The $S^3 2$ centrotype corresponds to an eSQP with almost perfect C_{4v} symmetry.* Its geometry can be seen to be intermediate between that of S2 (fSQP) and S4 (eSQP) for K = 4, as the comparison of Tables 4.3.4 and 4.3.6 will also show. Clearly, the reduction in the number of iterations from 21 (for K =

* Although Table 4.3.6 seems to indicate perfect C_{4v} symmetry for $S^3 2$, it does so only because the θ_{ij}'s and e.d.s.'s are rounded off. In fact the symmetry is only very nearly C_{4v}.

4) to eleven (for K = 3) has been achieved by the clustering together of the eSQPs and the fSQPs into one cluster representing the average of the two. The cluster affiliations also reveal that the entire membership of clusters S2 and S4 have been merged with three additional compounds being added from S1 (= S3).

The dissymmetry in the clustering arises from the inclusion, on the one hand, of the compound MOPAON [5] in S^31 (dTBP) and its exclusion, on the other, from S^33 (ideally the isometric partner of S^31) — it is instead classified with S^32 (SQP). This unequal treatment most likely arises from a combination of a structural peculiarity of the molecule and a breakdown in symmetry brought about by the algorithm. MOPAON, which was already identified as an outlier in our previous examination of five-coordinate nickel [4], contains two six-membered metal-ligand rings which are well known to be sterically hindering in five-coordination. Its two largest internal angles at 175° and 150° tend to suggest that its geometry is intermediate between that of a TBP and a SQP, thereby perhaps explaining its inclusion as a TBP in S^31 and as a SQP in S^32. In fact, at 0.959 Å and 0.960 Å the representative point for MOPAON is almost equidistant from the centres of S^31 (size = 0.713 Å) and S^32 (size = 0.851 Å), respectively.

Nørskov-Lauritsen and Bürgi [6] have examined now different hierarchical clustering algorithms deal with a highly symmetric data distribution, and have shown how some algorithms break down symmetry at certain points in the clustering process. In our case we are not dealing with a hierarchical algorithm, but there seems no reason, *in a least-squares sense*, why two symmetry equivalent points should be classified unequally, unless some peculiarity of the algorithm necessitates this. Although PKM employs non-hierarchical procedures, there clearly is an order, firstly, to the subdivision of the data space into K sections, and secondly, to the allocation of individual members to the K clusters. Consequently the final classification will depend both on the initial assignment of compounds to clusters and on the order in which the structures are processed. Clearly the cluster named one by the programme is first (in some non-hierarchical sense of the word) to have its approximate centre fixed and be allocated its members, followed by that named two, and so on. In this case, MOPAON is classified together with cluster one, then with cluster two, but not with cluster three. Short of analyzing the PKM flow-chart we can only

speculate that MOPAONs' close proximity to both TBP and SQP coupled to the order in which the PKM algorithm allocates cluster centres and members, give rise to the dissymmetry which emerges for K = 3.* MOPAON is represented by the asterisk in Figure 4.3.6, which indicates a superposition of this compound with a member of cluster three.

In order to examine the possibility of sorting eSQPs from fSQPs in T-space we attempted to fit a model with K = 7 to the data in T-space. The rationale behind this was that a partitioning of each isometric SQP cluster in T-space (T1,T2, T4) into two clusters representative of the eSQP and fSQP, respectively, ought to result in a total of seven clusters — one TBP, three isometric eSQPs and three isometric fSQPs. After 30 iterations (the maximum for PKM) the data had been sorted into seven clusters with, respectively, 393, 412, 134, 530, 194, 206 and 483 members — clearly not symmetrical. However, a scatterplot of the cluster pattern onto the factor 1/factor 2 plane, as in Figure 4.3.2, exhibited the general features one would expect if T1, T2 and T4 in the latter figure were split in two. Encouraged by this we devised an alternative strategy and decided to specify the seven initial cluster centres and then let PKM optimise them. As "seeds" we chose (i) for cluster one the D_{3h} TBP centrotype of T1, (ii) for clusters two, four and six the C_{4v} SQP centrotype of S4 (eSQP) with the angles permuted so as to correspond to the three isometric SQPs shown in Figure 4.3.3, (iii) for clusters three, five and seven the C_{4v} centrotype of S2 (fSQP) also permuted according to Figure 4.3.3. Input simply consisted of the values of the twelve symmetry coordinates for the D_{3h} TBP evaluated for the various conformations listed in (i), (ii) and (iii) above.

After just ten iterations (only one more than for K = 4) the algorithm had managed to produce a symmetrical clustering pattern, of which a scatterplot onto the factor1/factor 2 plane of Figure 4.3.2 is shown in Figure 4.3.7. The refined centrotypes for the seven clusters $T^71, T^72, T^73, T^74, T^75, T^76$ and T^77 are given in Table 4.3.7.

An analysis of the cluster membership indicates that T^72 (eSQP) and T^73 (fSQP) together contain the membership of T1 (dSQP) with an additional four

* Willett[7] has shown that different *initial* clusters and different *orders* of processing the data can indeed affect the final classification, although this is not necessarily the case.

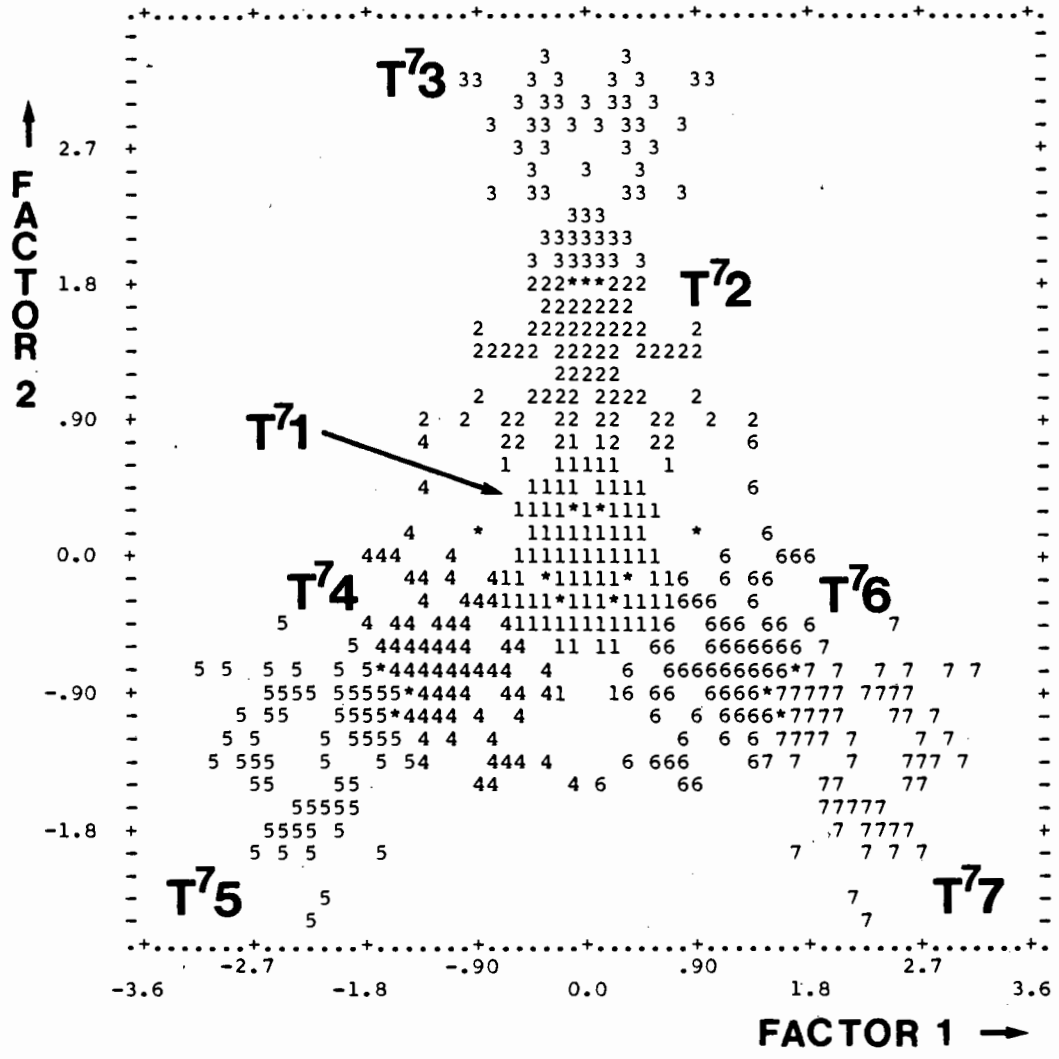
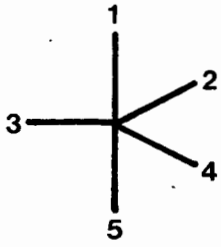


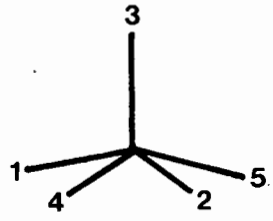
Figure 4.3.7 Projection of T-space clusters for K = 7 onto the plane defined by factor 1 and factor 2 extracted from T-space for k = 4. Numbers indicate cluster affiliation, and asterisks indicate the superposition of points belonging to different clusters.

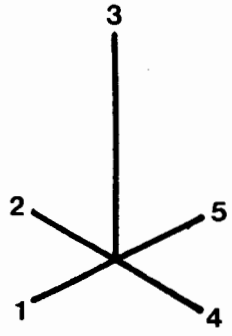
---

refcodes absorbed from T3. Obviously then, T1 has been split in two, and a comparison of Tables 4.3.2 and 4.3.7 shows that T1 (dSQP for K = 4) is intermediate between $T^7 2$ (eSQP) and $T^7 3$ (fSQP). Not only has the separation of eSQPs from fSQPs been achieved in T-space, but it has been greatly facilitated by seeding the data set with approximate cluster centrotypes.

Table 4.3.7 Internal angles [°] of refined cluster centrotypes for K = 7 in T-space. Estimated standard deviations are given in parenthesis. Angles for $T^7 4, T^7 6$ and $T^7 5, T^7 7$ may be obtained by permuting $T^7 2$ and $T^7 3$, respectively, according to T2 and T4 in Figure 4.3.3. M is the number of cluster members, N is the number of refcodes in the cluster.







	$T^7 1$	$T^7 2 (\approx T^7 4, T^7 6)$	$T^3 3 (\approx T^7 5, T^7 5, T^7 7)$
θ_{12}	90(7)	89(5)	90(6)
θ_{13}	90(7)	95(6)	91(8)
θ_{14}	90(7)	89(5)	90(6)
θ_{15}	175(5)	167(6)	173(4)
θ_{23}	119(7)	102(9)	95(8)
θ_{24}	119(7)	155(8)	168(5)
θ_{25}	90(7)	89(5)	90(6)
θ_{34}	119(7)	102(9)	95(8)
θ_{35}	90(7)	95(6)	91(8)
θ_{45}	90(7)	89(5)	90(6)
M	660	340	224
N	55	85	56

Interestingly, the refined archetypal eSQP and fSQP again display only C_{2v} symmetry, in spite of the fact that the cluster seeds were of C_{4v} symmetry. This serves to substantiate the argument made earlier that the symmetry of the data

space places restrictions on the symmetry to which a cluster centrotype may be refined.

The more specific classification of compounds as either eSQP or fSQP for K = 7 in T-space does not correspond as well with that in S-space as does the broader breakdown for K = 4 in T-space into simply TBP and SQP. The membership of $T^7 2$ (eSQP) is different from that of S4 (eSQP), as is that of $T^7 3$ (fSQP) from S2 (fSQP), with $T^7 3$ (fSQP) losing 14 compounds to S4 (eSQP), for example. As a result both of this and of the slight dissymmetry of the result for K = 3 for S-space, we would therefore argue that *the results obtained for K = 4 are the best* in terms of our two robustness criteria, even though K = 7 delivers slightly more detail in T-space than does K = 4.

(g) Outliers

We will here briefly examine both the group of ten compounds identified as divergent in Table 4.3.5, as well as the cluster outliers, that is compounds which lie at a much greater than average distance from their cluster centrotypes.

The ten refcodes listed above represent observed molecular geometries which approximate, on the one hand, to the SQP of C_{2v} symmetry in T-space but on the other to a TBP of C_{2v} symmetry in S-space. Accordingly one might expect these compounds to have a geometry which is intermediate between the TBP and the SQP, such that they may as readily be classified with the former as with the latter. Indeed, in examining the papers [8] in which these structures were originally reported, it becomes clear that the authors were at great pains in describing accurately what the structures were. Consequently these reports abound with descriptions such as "intermediate" and "distorted trigonal bipyramid". Most of them were published before attempts, such as those by Muetterties [9] and Holmes [10], to develop more precise conformational descriptions of five-coordinate complexes became more widely known. These methods incorporate the entire inner coordination polyhedron into establishing the degree of distortion, rather than subjectively focussing on the two largest angles (θ_{15} and θ_{24} in our numbering scheme) and comparing these to some "ideal" values which they have in the "ideal" conformations. Applying Holmes' method to the observed structures in order to determine the degree of distortion away from a TBP towards a SQP, as outlined in our earlier

study [4] of nickel complexes, we find that this ranges from 36 percent in the case of IPESNI [8g] to 59 percent in that of CMBPNI [8a] . Clearly, the speculation that these compounds are truly intermediate between the two extreme conformations is correct, and this may afford a reason as to why they are classified once as SQP, and another time as TBP.

Interestingly, a close examination of the inner coordination sphere geometry reveals that six of these structures approximate very closely to C_2 or C_{2v} symmetry. Of course, this is hardly surprising, since we have been arguing all along that compounds intermediate between SQP and TBP are likely to manifest C_{2v} distortion, since the energetically most favoured distortion coordinate is that maintaining C_{2v} symmetry. Nevertheless, it is gratifying to see this assertion supported by the evidence. Intriguingly, Ibers in his study [8f] of the square pyramidal form of the pentacyanonickelate anion $[Ni(CN)_5]^{3-}$ points out that "the observed C_2 geometry is apparently not due to crystal-packing interactions *but is rather an energy minimum of the free ion.*" This statement is all the more impressive when one sees it in the context of its date of formulation : 1968. At this stage the Berry mechanism was still largely unrecognised in inorganic chemistry, and was certainly totally ignored in transition metal chemistry. Whether or not Ibers had it in mind when the paper was written is unclear, since nowhere in the publication is there mention of the Berry mechanism.

We turn now to examining the cluster outliers — those structures which lie at greater than average distance from the centre of the cluster into which they have been classified. Part of the output from PKM is a histogram displaying the distance from the cluster centre to each observation. An example of this is shown in Figure 4.3.8. From these histograms one can estimate the density or homogeneity of a given cluster, and it is also fairly easy to identify cluster outliers from them.

On the whole, the clusters formed in both T-space and S-space are fairly homogeneous, with very few outliers. Below we examine separately each cluster, identifying outliers by their refcodes and indicating their distance (in Å) from the cluster centre in parenthesis behind the refcode. In eleven cases the outliers lie more than twice the average size* of the cluster away from its centre.

* The average size is defined as the average distance of all cluster members to its centre.

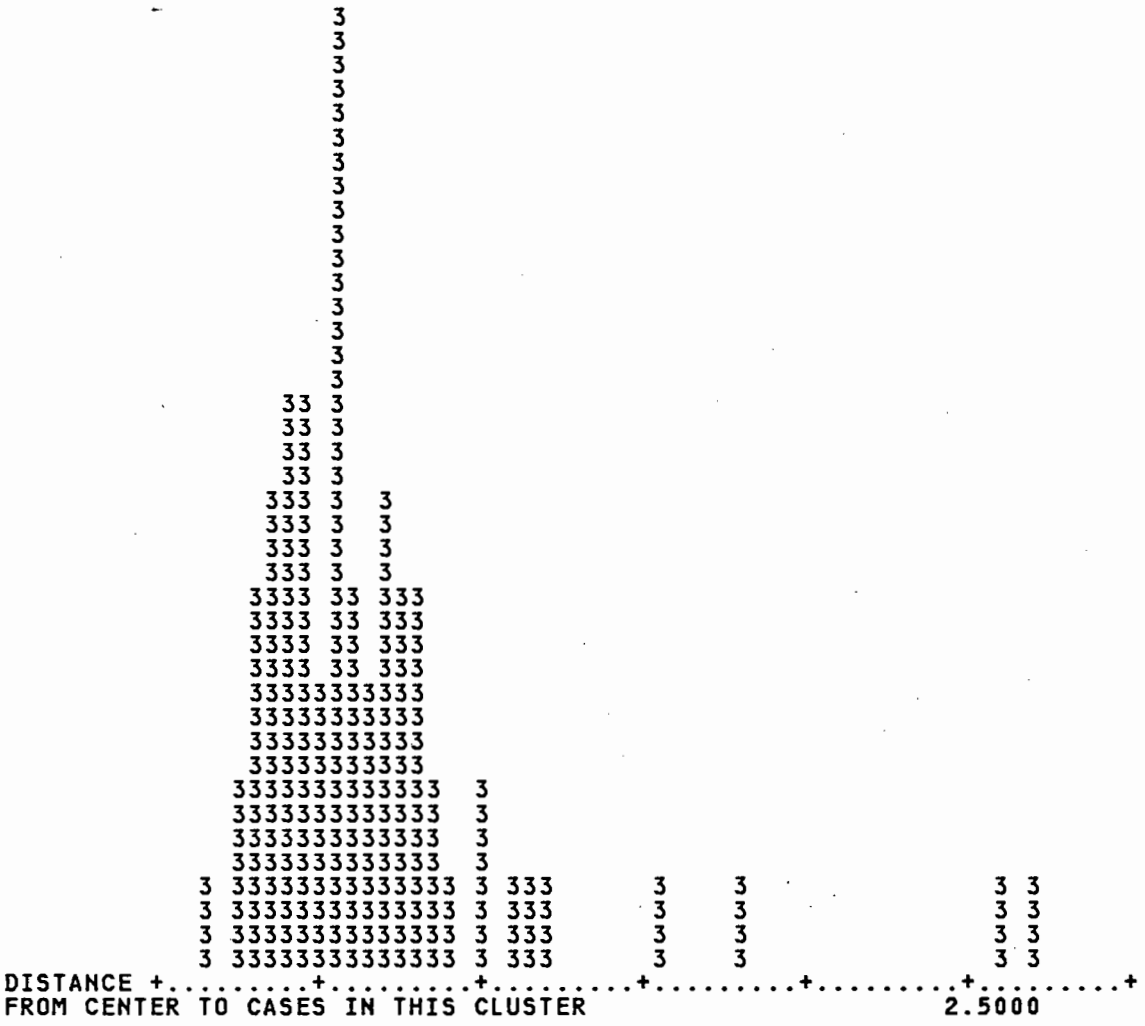


Figure 4.3.8. Histogram showing distance of cluster member to centre, in this case cluster S3. Each number represents one member.

(i) S1 (dTBP) = S3; size = 0.697 Å

DPPCRH(1.568) [8d] and CSMRHC(1.836) [8c] are both compounds which have previously been identified as having a geometry intermediate between a TBP and a SQP, thereby explaining why they are found on the fringes of this cluster. PEANIC(2.606) [11a] and PEANNI(2.741) [11b] are nickel carbonyl and nitrosyl complexes, respectively, with the tripodal ligand tris(2-diphenylphosphinoethyl)amine. They both exhibit extremely long axial interactions (when viewed

as dTBP) between nickel and the nitrogen of the base, with the metal displaced by almost one Angström from the equatorial plane. These distortions lead the authors to comment on a "tetrahedral distortion from TBP geometry ... maintaining C_{3v} symmetry". Quite likely it is this C_{3v} distortion which forces these compounds onto the outer edge of this cluster whose centrotype, afterall, exhibits C_{2v} symmetry.

(ii) S2(fSQP); size = 0.932 Å

CHESNI(1.481) [11c] is a four-plus-one coordinate compound with an extremely long (3.25 Å) apical interaction between nickel and the "dangling" oxygen atom of a bridging thiolate ligand in a hexameric complex. Its geometry consequently approximates a square planar one far more closely that it does a square pyramidal geometry, thereby possibly accounting for its distance from the cluster centre. CNPLPT(1.398) [11d] is a cyanobis(1,10-phenanthroline)platinum(II) complex whose "cation possesses no elements of symmetry", thereby perhaps explaining why it is an outlier.

(iii) S4(eSQP); size = 0.719 Å

DMPAPD10(1.606) [11e] contains a four-membered nickel-dithiophosphinato ring and the large and bulky 2,9-dimethyl-1,10-phenanthroline ligand, both of which contribute towards a considerable distortion of the geometry away from the SQP. This structure was also identified as an outlier in our previous study of nickel compounds [4], although in that case it was done merely on the basis of a graphical analysis.

MPNBNI(1.451) [11f] is a dimer which similarly contains a four-membered ring leading to a geometry which is quite "asymmetric".

FPHPRH(1.367) [11g] is also distorted towards "roughly C_s symmetry" by the presence of seven atoms "within bonding distance".

(iv) T1(dSQP) = T2, T4; size = 0.885 Å

CSMRHC(1.973) and DPPCRH(1.735) were previously identified as compounds intermediate between a TBP and a SQP, and as outliers of cluster S1. Here they again occur as outliers, quite likely for the same reason as before — their intermediacy. Other previous outliers which lie at large distances from the centre

of T1, presumably as a result of similar reasons as before are FPHPRH(1.394), MPNBNI(1.424), DMPADP(1.668) and CHESNI(1.724).

(v) T3(TBP); size4 = 0.727 Å

PEANNI(2.605) and PEANIC(2.462) are again the furthest outliers by far, most likely as a result of their strong distortion away from a TBP to a tetrahedron.

4.4 Hierarchical Cluster Analysis*

The nature of the output from the hierarchical cluster analysis programme P2M makes the processing of large data sets by this programme impractical. As a result we chose not to use the BMDP package but to follow, instead, the approach outlined by Nørskov-Lauritsen and Bürgi [6] in employing the procedure CLUSTER incorporated in the Statistical Analysis System (SAS). [12] In their pioneering paper Nørskov-Lauritsen and Bürgi describe the application of *Ward's method* as implemented in CLUSTER to the analysis of eight torsion angles in the molecular fragment $M(PPh_3)_2$. They describe how the history of the clustering process is documented by a number of descriptors which enable the user to decide when an optimal number of clusters has been formed in the agglomerative process, i.e. when the best compromise between the number of clusters — which should be small — and their information content — which should be large — has been reached.

The descriptor which they found "particularly useful for estimating the optimal number of clusters" is the cubic clustering criterion (CCC) [12b,13] . This criterion rests on an approximation to the expected value of the within-cluster sum of squares. The approximation, in turn, assumes that "a uniform distribution on a hyperrectangle will be divided into clusters shaped roughly like hypercubes." It is pointed out in the SAS User's Guide [12] that for certain large samples this assumption gives very accurate results. The exact definition of the CCC is "very complicated" [6] , but essentially it is a statistical measure which, when plotted against the number of clusters formed, can indicate an optimum number of clusters by a peak in the

* This part of the analysis was performed during a brief stay at the University of Berne. The time available did not permit a full analysis. Nevertheless, the results obtained adequately substantiate those obtained from relocation clustering.

graph. In this sense its use is rather similar to that of the nearest neighbour distance illustrated in Figure 2.7.2 of the worked example of section 2.7(a).

We will substantiate the results of the relocation cluster analysis by illustrating the essential outcome of the application of Ward's method to our data set.

(a) Similarity measure and robustness criteria

We chose the Euclidean distance in 12-D space as our similarity measure, in keeping with the approach adopted in the relocation cluster analysis. As an indicator of robustness we decided to rely on the CCC in this case, since time constraints prevented a full analysis of the cluster membership or the symmetry of the result.

(b) Results

Figures 4.4.1 and 4.4.2 represent plots of the CCC against the number of clusters formed for T-space and S-space, respectively. Pointers to the correct use of the CCC cited by the SAS User's Guide [12b] are: (i) peaks on the plot with CCC greater than two or three indicate good clusterings, (ii) very distinct non-hierarchical clusters show a sharp rise before the peak followed by a gradual decline, (iii) peaks with the CCC between zero and two indicate possible clusters, (iv) if the CCC increases continuously as the number of clusters increases, the distribution may be grainy. Also, the guide points out that "the power of the CCC seems to be at least as good as that of the human eye in two dimensions with 100 observations."

An inspection of the result obtained for T-space reveals that the most dramatic peak by far occurs at four clusters. This, of course, coincides exactly with the result obtained by relocation clustering. The value of the CCC (= 32) indicates a very robust result, and the sharp rise up to the peak suggests that the clusters are very distinct and of a non-hierarchical nature, at least from the level of twelve clusters upwards. Far less significant "peaks" occur at 12 and 29 clusters — these presumably result from the clustering together of clouds of isometric conformations which lie close to each other in parameter space (possibly along symmetry elements). Of interest is the lack of any peak at seven clusters; a result which is not in agreement with that obtained previously, where a solution with seven clusters was shown to be feasible.

The result for S-space, shown in Figure 4.4.2, is similarly encouraging. The

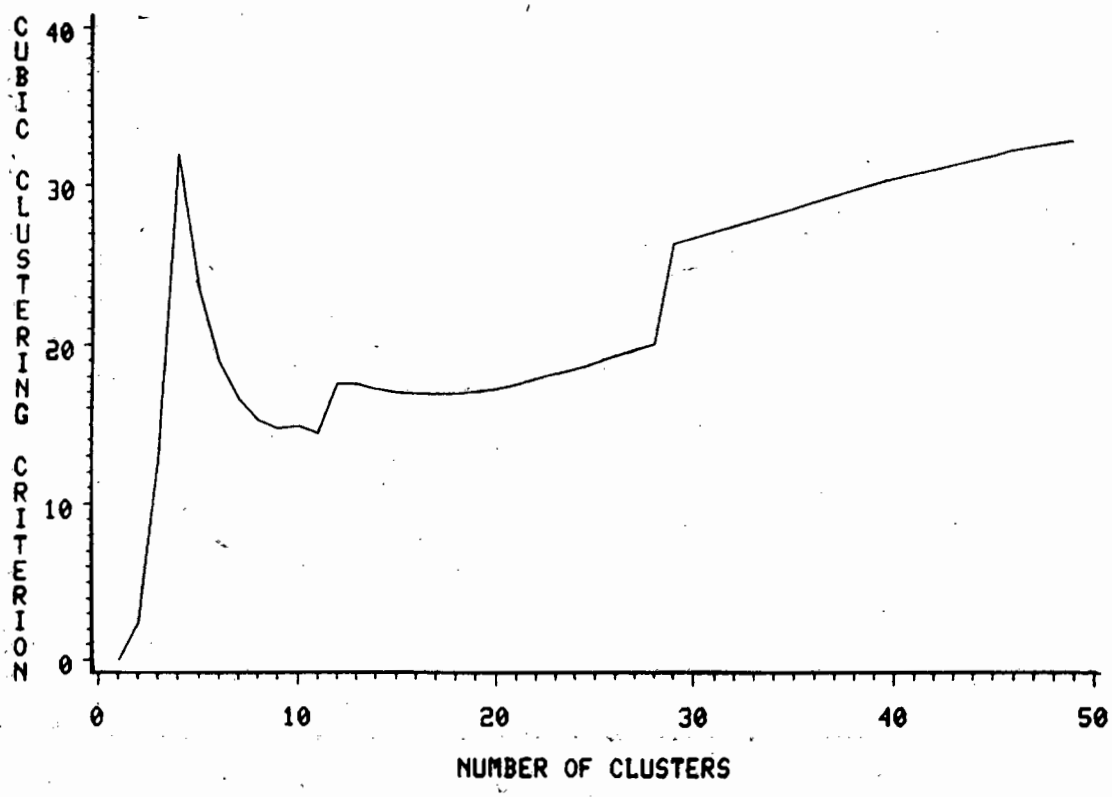


Figure 4.4.1 Cubic Clustering Criterion plotted against number of clusters in T-space.

---

main peak at four clusters clearly coincides with the result obtained by relocation clustering, while the (only) alternative at three clusters is synonymous with the results obtained for K = 3, where the clusters corresponding to the eSQP (S4) and the fSQP (S2) of the K = 4 result are merged into one. Interestingly, both the much gentler rise to the peak (as opposed to that in Figure 4.4.1) as well as the overall greater smoothness of the graph tend to suggest that the clusters in S-space lie much closer together than do those in T-space, resulting in less dramatic increases in homogeneity as they are successively clustered together. The lower value of the CCC (= 10.5) also seems to indicate a reduced robustness in the clustering, while its continuous increase with increasing number of clusters suggests a "grainy" data distribution. This apparent closer proximity of the clusters in S-space, which might alternatively be described as a greater *diffuseness* of the clusters, (on the

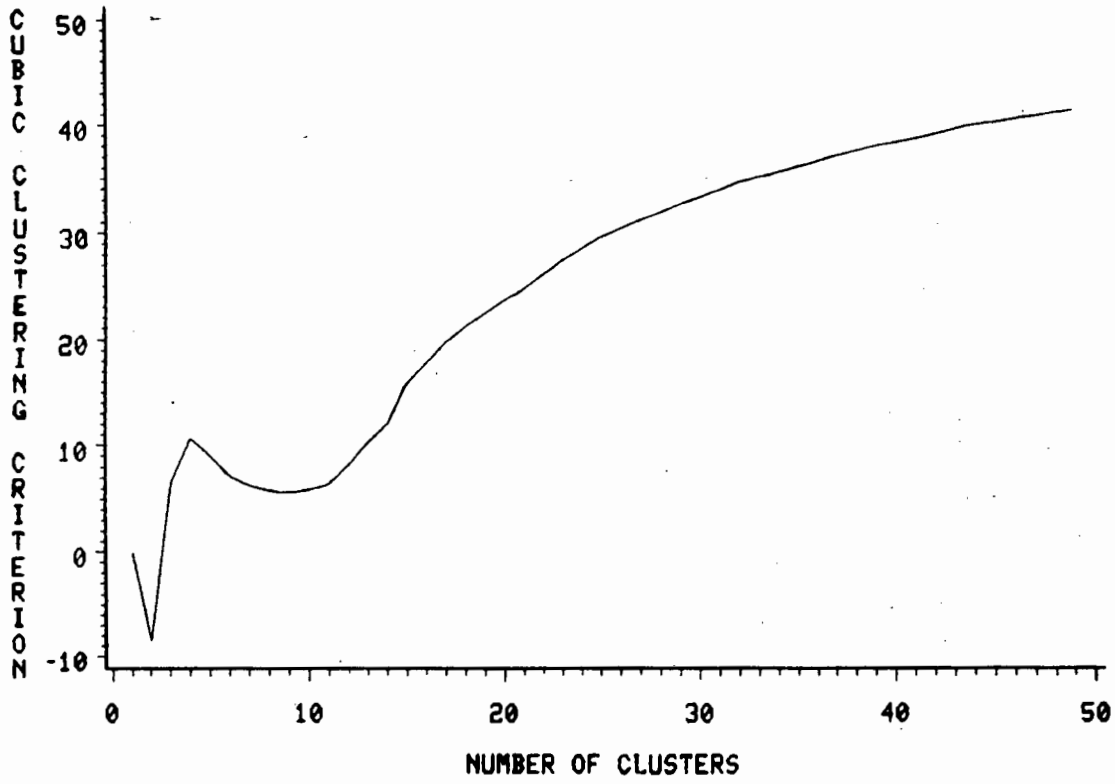


Figure 4.4.2. Cubic Clustering Criterion plotted against number of clusters in S-space.

---

strength of the suggestion of "graininess"), could result from the nature of the C_{4v} point group, i.e. the symmetry of the parameter space. C_{4v} symmetry allows a real, observed SQP any value of its two trans-basal angles, so long as they are equal to each other. This allows a *smear*, in other words, of permissible values or, more formally, the C_{4v} symmetry has one degree of freedom. So, whereas the symmetry of T-space forces a separation of the TBP and SQP clusters, but treats the SQP cluster simply as one (for K = 4), the symmetry of S-space allows the square pyramidally disposed structures to manifest the *range* in their trans-basal angles. It is this range, in turn, which leads to the gradual progression from fSQP to eSQP being more clearly observable in S-space than in T-space; in other words, the data in S-space are more "smeared out" than they are in T-space by virtue of the C_{4v} symmetry. Consequently the clusters in S-space are less distinct overall.

The closer proximity of the clusters in S-space relative to that in T-space was previously indicated during the discussion concerning the K-mean algorithms' ability to separate the eSQPs from the fSQPs, where the suggestion was made that these two clusters might diffuse into one another (section 4.3(e)). Indeed, a comparison of the average intercluster distance as obtained from K-means clustering (Tables 4.3.3 and 4.3.5) reveals a separation of 2.730 Å for T-space, as compared to 1.653 Å for S-space. Considering that the average cluster sizes are 0.846 Å and 0.761 Å, respectively, the enhanced classification in T-space over that of S-space becomes obvious.

4.5 Analysis of Structural Results

We will here extract further structural information from the results of *K-Means* clustering, having shown above that the results obtained from this method and those yielded by hierarchical cluster analysis using Ward's method are very nearly identical. The structure of five-coordinate transition metal complexes, in particular those approaching SQP conformation, has been the subject of much discussion and work. One of the more recent comprehensive reviews is that written by Holmes. [14] In it is an exhaustive overview of his work specifically with phosphorus and other main group element compounds, as well as a compilation of studies in five-coordinate transition metal complexes and a structural analysis of these compounds using his dihedral angle technique. His data base for nickel and platinum is rather small, though, comprising only 22 nickel and 4 platinum compounds with no other complexes containing d^8 metals. In his review, as well as in a subsequent paper [15], he attempts to compare the observed structures to ones predicted by Rossi and Hoffmann on the basis of extended Hückel studies on the PtL_5^{3-} system [16], and also offers interpretations of the observed conformations in terms of non-bonded repulsions between d-orbital electron density and bond electron density.

On the basis of an angular overlap model he suggests that high-spin d^8 compounds would favour a more SQP conformation, and his data appear to agree with this assertion. Table 4.5.1 outlines the cluster membership in T-space and S-space, dissected according to the metal. It can be seen that in both cases approximately twice as many compounds are classified as SQP than are TBP. When regarding the individual metals it becomes obvious that this tendency is more pronounced for the

second- and third-row elements, most of which are likely to be low-spin, than it is for the nickel compounds. Without knowledge about the spin states of the complexes, therefore, little can be definitely said about the conformational preferences of different spin states.

Table 4.5.1 Cluster membership broken up according to metals. Recall that T1 = T2 = T4 and S1 = S3.

	Ni	Pd	Pt	Rh	Ir	Total
T1(dSQP)	68	15	8	23	23	= 137
T3(TBP)	45	0	1	5	8	= 59
S1(dTBP)	50	0	2	8	9	= 69
S2(fSQP)	22	12	5	0	0	= 39
S4(eSQP)	41	3	2	20	22	= 88

Turning to the average coordination sphere geometry for the individual metals we find, not surprisingly, that it is very similar within each of the clusters, corresponding to the cluster averages which are presented as centrotypes in Tables 4.3.2, 4.3.4 and 4.3.6. We will consequently treat the various cluster archetypes as representative of the metals whose compounds are classified with that cluster, and refrain from commenting in each case on the individual metals. Table 4.5.2 gives the average bond distance increments d_1 to d_5 for the various clusters; it should be read in conjunction with Tables 4.3.2, 4.3.4 and 4.3.6. From these tables we can obtain average values of the internal coordinates for an "average" TBP (T1), an "average" fSQP (S2), an "average" eSQP (S4) and an "average" SQP ($S^3 2$ from the analysis for K = 3).

There has been much discussion concerning the value of the trans-basal angle (θ_{15} and θ_{24}) in an "ideal" SQP and the relative proportions of the axial to the equatorial bonds or the apical to the basal bonds in "ideal" TBPs and SQPs, respectively. It is generally agreed, however, that these will depend on the d -electron configuration. [14–17] Clearly, no such thing as an "ideal" SQP exists; instead there exists a range of square pyramidally disposed conformations ranging from a fSQP to

Table 4.5.2 Average bond distance increments [Å] for the clusters indicated. T1 = T2 = T4 and S1 = S3. Estimated standard deviations are given in parenthesis. $\overline{SQP}$ = average SQP, centrotype of S^32 .

	T1(dSQP)	T3(TBP)	S1(dTBP)	S2(fSQP)	S4(eSQP)	$S^32(\overline{SQP})$
d_1	-.049(116)	.030(205)	.024(192)	-.165(82)	-.008(88)	-.058(111)
d_2	-.057(106)	.-.025(83)	-.032(83)	-.165(82)	-.008(88)	-.058(112)
d_3	.219(353)	-.025(83)	.001(88)	.692(298)	.030(125)	.239(364)
d_4	-.057(106)	-.025(83)	-.032(83)	-.165(82)	-.008(88)	-.058(112)
d_5	-.049(116)	.030(205)	.024(192)	-.165(82)	-.008(88)	-.058(111)

an eSQP. We shall consequently compare the averages of both types of SQP, as well as that of an "average" SQP (i.e. one composed of a merger of the fSQP and eSQP) with the empirical results listed by Holmes [14,15] and the predictions of Rossi and Hoffmann. [16] Table 4.5.3 outlines this comparison. It lists the following observations made by Holmes (H) for TBP and SQP nickel complexes: (i) axial bonds are circa 0.05 Å shorter than equatorial bonds, (ii) the trans-basal angle for high-spin complexes is 161^o , for low-spin complexes 173^o , (iii) the apical bond is longer by circa 0.2 Å than basal bonds. The predictions by Rossi and Hoffmann (RH) are: (i) axial bonds are shorter than equatorial bonds, i.e. $d_{axial} - d_{equatorial} < 0$, (ii) the trans-basal angle is 164^o , (iii) the apical bond is longer than the basal bonds, i.e. $d_{apical} - d_{basal} > 0$.

With one notable exception the results of this study compare well with those of H and RH. While RH predicted that the axial bonds of a TBP are likely to be stronger and therefore shorter than the equatorial bonds, and H's findings seem to support this, we in fact find that the axial bonds are longer by 0.055 Å than the equatorial ones. A breakdown of the average bond distance increments for each metal of cluster T1 is given in Table 4.5.4. It shows that this is not the result, simply, of nickel "swamping" the other metals, but rather that rhodium and iridium also follow this pattern, while the one platinum compound in this group does not. At this point no explanation for this dichotomy between our findings, on the one hand, and those of H and RH on the other, can be offered, other than to point

Table 4.5.3 Comparison of results of this study with results taken from Holmes (H) [14,15] and predictions by Rossi and Hoffmann (RH). [16] d_{ax} = average axial bond increment, d_e = avg. equatorial increment, d_{ap} = avg. apical bond increment, d_{ap} = avg. apical bond increment, d_b = avg. basal bond increment. neg = negative, pos = positive. HS = high-spin, LS = low spin. $\overline{SQP}$ = "average" SQP as characterised by S^32 .

	TBP			SQP				
	T1(TBP)	H	RH	S2(fSQP)	S4(eSQP)	$S^32(\overline{SQP})$	H	RH
							161 HS	
θ_{15} [°]	174	—	—	171	163	166	173 LS	164
$d_{ax} - d_e$ [Å]	.055	-.05	neg	—	—	—	—	—
$d_{ap} - d_b$ [Å]	—	—	—	.857	.038	.297	.2	pos

out two differences between Holmes' earlier empirical study and ours. Firstly, the subset of compounds from which he draws the conclusion that axial bond distances are shorter than equatorial ones by 0.05 Å is a very select one — it includes only compounds with like ligands in two or more sites on TBPs which are less than 50 percent displaced towards SQP *according to his dihedral angle method.* In this case this comprises only three compounds, whereas our study averaged 59 compounds.* Secondly, in calculating the dihedral angles contained by the edges of the inner coordination sphere so as to establish the percentage distortion TBP → SQP the metal-ligand distances are brought to unity. This has the effect of *distorting* the inner coordination sphere, albeit minutely, which in turn influences the results.

Finally, two further intriguing observations can be made. Table 4.5.1 illustrates that the majority of palladium and platinum compounds adopt a more flattened square pyramidal conformation, whereas in the case of rhodium and iridium the tendency is very dramatically towards an eSQP. Whether this notable difference

* The estimated standard deviations in Table 4.5.4 show that there is a large variance particularly in the axial distance increments of nickel. Consequently it is quite feasible that a given subset of our data set might not reflect the average tendency, but rather that found by Holmes.

Table 4.5.4 Average bond distance increments within cluster T3 (TBP)[Å]. N is the number of compounds in the cluster. Estimated standard deviations are given in parenthesis.

	Ni	Pd	Pt	Rh	Ir
d_1	.038(229)	—	-.012(2)	.007(114)	.002(76)
d_2	-.022(69)	—	.008(0)	-.013(140)	-.054(103)
d_3	-.022(69)	—	.008(0)	-.013(140)	-.054(103)
d_4	-.022(69)	—	.008(0)	-.013(140)	-.054(103)
d_5	.038(229)	—	-.012(2)	.007(114)	.002(76)
N	45	0	1	5	8

in behaviour between the two groups of compounds is related to some inherent difference between the two groups of metals, or results, instead, from geometric constraints imposed by different sets of ligands needs to be further investigated. Possibly related to this is the final observation that the relative proportions of apical- to basal bond lengths for the eSQP containing rhodium and iridium is quite different from that of the eSQP containing nickel, even though the standard deviations are large in all cases. The tendency for platinum is similar to that for nickel, while that for palladium is closer to the other two metals, although for both platinum and palladium the values are based on only two and three structures, respectively. Table 4.5.5 illustrates that the apical bond is *longer* by 0.084 Å for nickel, while it is *shorter* by 0.049 Å for rhodium, and only longer by 0.01 Å for iridium.

4.6 Intracluster Statistics

Now that cluster analysis has shown the essential distribution of the data in 12-D space, it might be instructive to investigate *intra-cluster* univariate and bivariate statistics in a manner similar to that used for T- and S-spaces overall in section 4.1.

(a) Univariate statistics

T-Space

Table 4.6.1 gives the standard deviations, variances and percentage variance

Table 4.5.5 Average bond distance increments within cluster S4(eSQP)[Å]. N is the number of compounds in the cluster. Estimated standard deviations are given in parenthesis.

	Ni	Pd	Pt	Rh	Ir
d_1	-.024(87)	-.084(25)	.038(94)	.011(66)	.001(101)
d_2	-.024(87)	-.084(25)	.038(94)	.011(60)	.001(101)
d_3	-.060(118)	.337(77)	-.153(42)	-.038(80)	.001(84)
d_4	-.024(87)	-.084(25)	.038(94)	.011(60)	.001(101)
d_5	-.024(87)	-.084(25)	.038(94)	.011(60)	.001(101)
N	41	3	2	20	22

for each symmetry coordinate in clusters T1 (dSQP) and T3 (TBP), as well as the percentage of the total variance in, respectively, bond increment and bond angle symmetry coordinates explicable by each. T3 is shown to have a large variance in the bond increment coordinates S_1 and S_3 , which can be seen to imply large variances in the length of the axial bonds from the graphic representation of symmetry coordinates show in Figure 4.1.1. A similarly large variance in S_4 (the umbrella coordinate) clearly implies a large variation in the degree of pyramidality of the fragment composed of one axial and all three equatorial ligands, i.e. a large variation in the out-of-plane displacement of the metal atom from the equatorial plane. This variance therefore again hints at the importance of S_N2 -type distortions in the TBP.

For the typical C_{2v} SQP represented by T1 there is a large variation in the length of the apical bond (S_{5a}) and in the position of the apical ligand with respect to the other four ligands (S_{6a}, S_{6b}) . This is the interpretation when the D_{3h} symmetry coordinates are referred to, or "fitted" to a SQP as shown in Figure 4.1.1. Interestingly, whereas the symmetry of T-space is reflected in the variances of the degenerate coordinates $(S_5 - S_8)$ for cluster T3, the variances in T1 indicate that this cluster does not conform to the symmetry of the data space.

S-Space

Table 4.6.1 Standard deviations (σ) , variances (σ^2) , percentage of variance in bond increment coordinates (% BV) and percentage of variance in angular symmetry coordinates (% AV) for clusters T1 and T3 in T-space.

	T1(dSQP)				T3(TBP)			
	σ [Å]	σ^2 [Å ^{2}]	% BV	% AV	σ [Å]	σ^2 [Å ^{2}]	% (BV)	% (AV)
S_1	.150	.022	12.8	—	.191	.036	34.6	—
S_2	.140	.020	11.2	—	.084	.007	6.8	—
S_3	.067	.004	2.6	—	.219	.048	45.8	—
S_4	.245	.060	—	8.2	.534	.286	—	47.2
S_{5a}	.348	.121	69.5	—	.082	.007	6.4	—
S_{5b}	.082	.007	3.9	—	.082	.007	6.4	—
S_{6a}	.432	.187	—	25.6	.331	.110	—	18.1
S_{6b}	.440	.193	—	26.5	.331	.110	—	18.1
S_{7a}	.236	.056	—	7.6	.175	.031	—	5.0
S_{7b}	.196	.039	—	5.3	.175	.031	—	5.0
S_{8a}	.316	.100	—	13.7	.142	.020	—	3.3
S_{8b}	.309	.096	—	13.1	.142	.020	—	3.3

S Space

Table 4.6.2 illustrates univariate statistics for clusters S1, S2 and S4 in S-space. Here, in converse to the situation in T-space, the variances in S1 (dTBP) indicate that this cluster does not conform strictly to the symmetry (C_{4v}) of data space. In this case the large variance in S_{7a} implies a large variation in the axial bonds of the distorted TBP, while that in S_5 reflects the range of trans-basal angles available to the C_{2v} dTBP. The variance in S_{8a} suggests a large degree of freedom in the position of the pivot atom. These interpretations can be recognised from the graphic representation of the C_{4v} symmetry coordinates referred to a TBP, shown in Figure 4.1.2. For the flattened SQP a large variation in the apical bond distance becomes apparent from S_1 , while the variance in S_{8a}, S_{8b} and S_6 suggest a large flexibility in the position of the apical ligand, i.e. this ligand is not fixed to lying on an imaginary perpendicular line drawn through the centre of the basal plane of the fSQP. For the eSQP (S4) the bond distance variance is not limited mainly to the apical distance, as was the case for the fSQP, but is spread more evenly over all

five bonds; the majority is still associated with the apical bond, though, since S_2 comprises all four basal bonds, S_1 only the apical bond. All symmetry coordinates contribute approximately equally to the overall variance.

(b) **Bivariate statistics**

T-Space

The correlation matrices for clusters T1(dSQP) and T3 (TBP) are shown in Table 4.6.3. The method, introduced in Section 4.1, of *summing* the distortions representing the individual symmetry coordinates in a correlated pair in order to obtain the *correlated distortion*, will be applied here again in an interpretation of the intra-cluster correlation data. For T1 (dSQP) two major correlations appear: between S_1 and S_{5a} (r = -.812), and between S_2 and S_{5a} (r = .819). The first corresponds to the reversible departure of the apical ligand and the concomitant shrinking of the basal bonds in the dSQP. This is represented in Figure 4.6.1 where the D_{3h} symmetry coordinates are referred to a SQP (as illustrated in Figure 4.1.1) to facilitate the interpretation of the data. In other words, this represents the "constant amount of glue" coordinate already identified for T-space overall. The second major correlation, also graphically illustrated in Figure 4.6.1, essentially indicates that there is a large degree of freedom in the position of the apical ligand, a characteristic of square pyramidally disposed conformations which we have repeatedly encountered. These two correlations are each able to account for 66 percent of the variance in T1. It is interesting to note, also, that symmetry coordinates are correlated *across different symmetry species*, i.e. irreducible representations are mixed. Most likely, this arises out of the inability of T1 to conform fully to the symmetry of the D_{3h} data space.

For T3 (TBP), however, the correlations occur in blocks according to their symmetry, and there is no mixing of different irreducible representations. Only one major correlation ($r = .84$), accounting for 70 percent of sample variance in T3, characterises the distribution of data in this cluster. This correlation (between S_3 and S_4) is graphically represented in Figure 4.6.1 and it mirrors the S_N2 coordinate.

Three major correlated distortions, described as an S_N2 coordinate, a Berry

Table 4.6.2 Standard deviations (σ) , variances (σ^2) , percentage of variance in bond increment coordinates (% BV) and percentage of variance in angular coordinates (% AV) for clusters S_1, S_2 and S_4 .

	S1 (dTBP)				S2 (fSQP)			
	σ [Å]	σ^2 [Å ^{2}]	% BV	% AV	σ [Å]	σ^2 [Å ^{2}]	% BV	% AV
S_1	.087	.008	7.98	—	.298	.089	76.8	—
S_2	.133	.018	18.6	—	.116	0.014	11.77	—
S_3	.248	.061	—	10.4	.216	.046	—	5.7
S_4	.148	.022	22.8	—	.054	.003	2.5	—
S_5	.317	.101	—	17.0	.209	.044	—	5.3
S_6	.151	.023	—	3.9	.411	.169	—	20.6
S_{7a}	.203	.041	43.2	—	.072	.005	4.5	—
S_{7b}	.084	.007	7.5	—	.072	.005	4.5	—
S_{8a}	.345	.119	—	20.1	.489	.239	—	29.1
S_{8b}	.306	.094	—	15.8	.489	.239	—	29.1
S_{9a}	.311	.097	—	16.4	.205	.042	—	5.1
$S9b$	.311	.097	—	16.4	.205	.042	—	5.1

	S4 (eSQP)			
	σ [Å]	σ^2 [Å ^{2}]	% BV	% AV
S_1	.125	.016	33.3	—
S_2	.118	.014	29.8	—
S_3	.296	.087	—	16.4
S_4	.078	.006	12.9	—
S_5	.278	.077	—	14.5
S_6	.260	.068	—	12.7
S_{7a}	.075	.006	12.0	—
S_{7b}	.075	.006	12.0	—
S_{8a}	.319	.102	—	19.1
S_{8b}	.319	.102	—	19.1
S_{9a}	.219	.048	—	9.1
S_{9b}	.219	.048	—	9.1

Table 4.6.3 Correlation matrix for symmetry coordinates in clusters T1 (dSQP) and T3 (TBP). Symmetry species to which the coordinates belong are also shown.

T1	A_1'		A_2''		E'						E''	
	S_1	S_2	S_3	S_4	S_{5a}	S_{5b}	S_{6a}	S_{6b}	S_{7a}	S_{7b}	S_{8a}	S_{8b}
S_1	1.00											
S_2	-.61	1.00										
S_3	0.00	0.00	1.00									
S_4	0.00	0.00	-0.21	1.00								
S_{5a}	-0.81	0.82	0.00	0.00	1.00							
S_{5b}	0.00	0.00	0.00	0.00	0.00	1.00						
S_{6a}	-0.47	0.47	0.00	0.00	0.53	0.00	1.00					
S_{6b}	0.00	0.00	0.00	0.00	0.00	-0.02	0.00	1.00				
S_{7a}	0.33	-0.45	0.00	0.00	-0.47	0.00	-0.29	0.00	1.00			
S_{7b}	0.00	0.00	0.00	0.00	0.00	-0.26	0.00	-0.29	0.00	1.00		
S_{8a}	0.00	0.00	-0.05	0.28	0.00	0.00	0.00	0.00	0.00	0.00	1.00	
S_{8b}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

T3	A_1'		A_2''		E'						E''	
	S_1	S_2	S_3	S_4	S_{5a}	S_{5b}	S_{6a}	S_{6b}	S_{7a}	S_{7b}	S_{8a}	S_{8b}
S_1	1.00											
S_2	-0.26	1.00										
S_3	0.00	0.00	1.00									
S_4	0.00	0.00	-0.84	1.00								
S_{5a}	0.00	0.00	0.00	0.00	1.00							
S_{5b}	0.00	0.00	0.00	0.00	0.00	1.00						
S_{6a}	0.00	0.00	0.00	0.00	0.25	0.00	1.00					
S_{6b}	0.00	0.00	0.00	0.00	0.00	0.25	0.00	1.00				
S_{7a}	0.00	0.00	0.00	0.00	-0.11	0.00	0.36	0.00	1.00			
S_{7b}	0.00	0.00	0.00	0.00	0.00	-0.11	0.00	0.36	0.00	1.00		
S_{8a}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	
S_{8b}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

coordinate and a glue coordinate (for short) were identified for the data distribution in T-space overall. It would appear as if one has been "lost" during the subdivision of T-space into the TBP and SQP clusters; the glue coordinate has been seen to

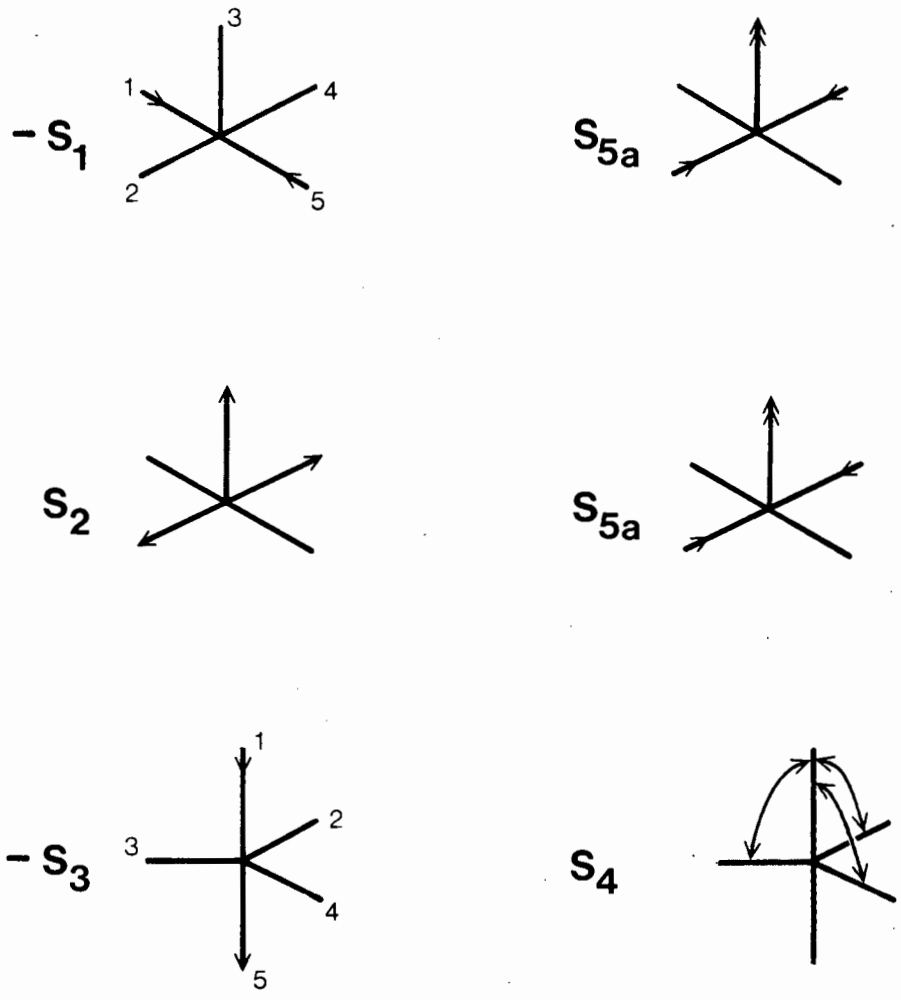


Figure 4.6.1 Graphical representation of distortions corresponding to correlated pairs of symmetry coordinates. The first two are from cluster T1 (dSQP) and the third from cluster T3 (TBP). Where the correlation is negative the inverse of one coordinate has been drawn, eg. $-S_1$.

---

be associated with T1 (dSQP), the S_N2 with T3 (TBP). However, in the case of the Berry coordinate both clusters have retained a component of it, thereby reducing its importance from that for T-space overall. T1 and T3 show correlations with coefficients of 0.53 and 0.25, respectively, between coordinates S_{5a} and S_{6a} . Chemically speaking, the association of the Berry coordinate with both the SQP

and TBP clusters in T-space is unremarkable, of course, since both conformations can exhibit the C_{2v} distortion.

S-space

Table 4.6.4 shows the correlation matrices for clusters S1, S2 and S4. As was the case previously with T1, cluster S1 (dTBP) shows correlations between coordinates of different symmetry species, these presumably arising out of an inability of this cluster to conform to the symmetry of the data space. The picture is complex, with a number of correlations above 0.50. First we will consider the correlations of S_{7a} with S_{8a}, S_{9a} and S_{9b} with coefficients of -0.701, -0.764 and -0.764, respectively. A graphic interpretation of these correlations shows that they accord to the S_N2 coordinate at a TBP in C_{4v} parameter space, with all three correlations together presenting a more coherent picture than just the correlation between S_{7a} and either one of the others. Related to these are the positive correlations between S_{8a} and S_{9a}, S_{9b} with another characteristic, the variance in the axial bonds, mirrored in the correlation between S_2 and S_4 (r = 0.647). These correlated distortions are graphically illustrated in Figure 4.6.2. Finally, lest it be lost again, a component of the Berry coordinate is present in the correlation (r = 0.393) between S_4 and S_5 .

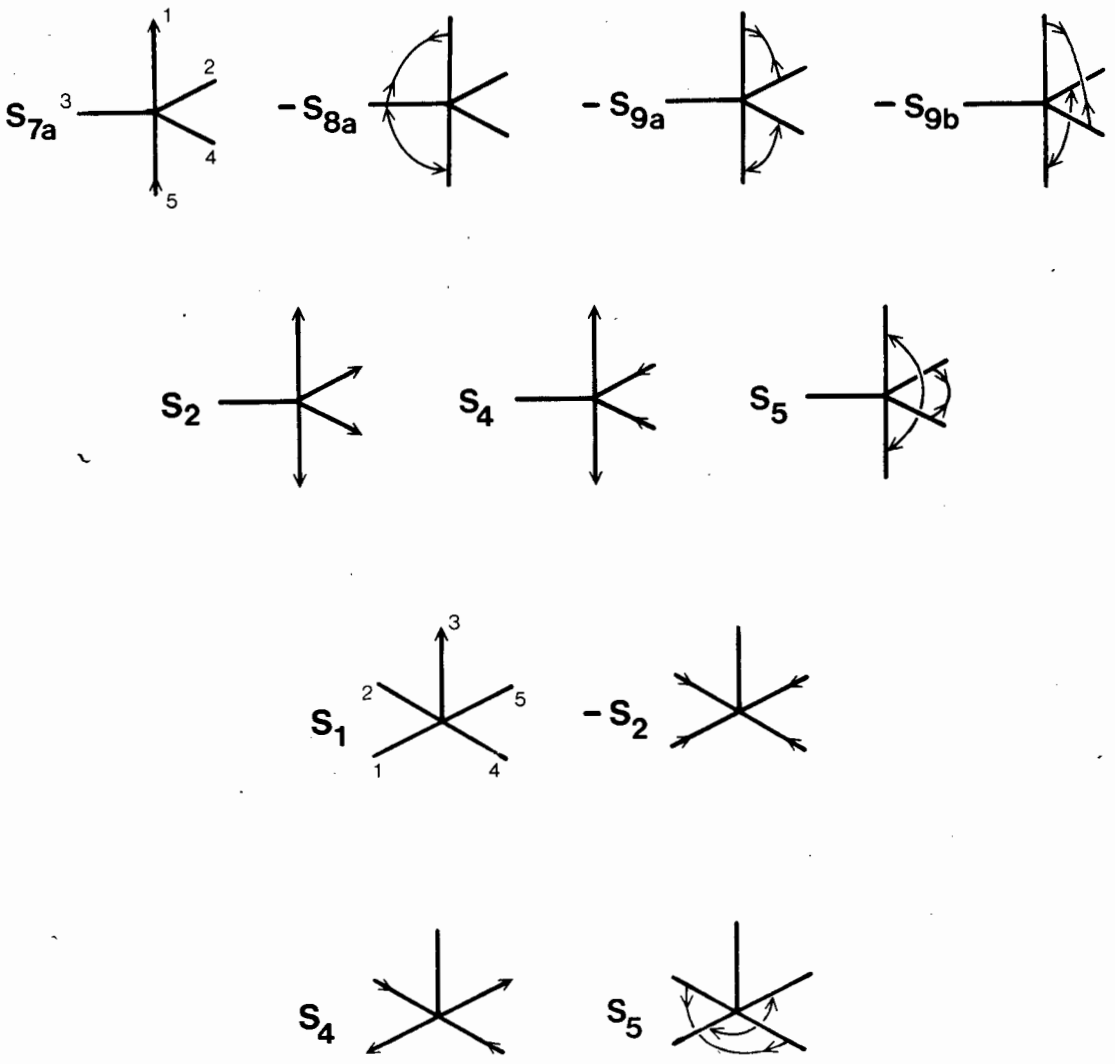


Figure 4.6.2 Graphical representation of distortions corresponding to correlated pairs of symmetry coordinates. Where the correlation is negative the inverse of one coordinate has been drawn, eg. $-S_{8a}$. Correlations are from clusters S1, S2 and S4 and are discussed in the text.

Table 4.6.4 Correlation matrix for symmetry coordinates in clusters S1 (d TBP), S2 (fSQP) and S4 (eSQP). Symmetry species to which the coordinates belong are also shown.

S1	A_1			B_1		B_2	E					
	S_1	S_2	S_3	S_4	S_5	S_6	S_{7a}	S_{7b}	S_{8a}	S_{8b}	S_{9a}	S_{9b}
S_1	1.00											
S_2	-0.29	1.00										
S_3	0.07	-0.30	1.00									
S_4	-0.24	0.65	-0.25	1.00								
S_5	-0.19	0.29	-0.57	0.39	1.00							
S_6	0.00	0.00	0.00	0.00	0.00	1.00						
S_{7a}	0.00	0.00	0.00	0.00	0.00	0.00	1.00					
S_{7b}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00				
S_{8a}	0.00	0.00	0.00	0.00	0.00	0.00	-0.70	0.00	1.00			
S_{8b}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.34	0.00	1.00		
S_{9a}	0.00	0.00	0.00	0.00	0.00	0.00	-0.76	0.05	0.68	0.03	1.00	
S_{9b}	0.00	0.00	0.00	0.00	0.00	0.00	-0.76	0.05	0.68	-0.03	0.71	1.00

S2	A_1			B_1		B_2	E					
	S_1	S_2	S_3	S_4	S_5	S_6	S_{7a}	S_{7b}	S_{8a}	S_{8b}	S_{9a}	S_{9b}
S_1	1.00											
S_2	-0.60	1.00										
S_3	0.26	0.06	1.00									
S_4	0.00	0.00	0.00	1.00								
S_5	0.00	0.00	0.00	-0.35	1.00							
S_6	0.00	0.00	0.00	0.00	0.00	1.00						
S_{7a}	0.00	0.00	0.00	0.00	0.00	0.00	1.00					
S_{7b}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00				
S_{8a}	0.00	0.00	0.00	0.00	0.00	0.00	-0.28	0.00	1.00			
S_{8b}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.28	0.00	1.00		
S_{9a}	0.00	0.00	0.00	0.00	0.00	0.00	0.03	-0.03	-0.05	0.05	1.00	
S_{9b}	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	-0.05	-0.05	0.00	1.00

	S_1	S_2	S_3	S_4	S_5	S_6	S_{7a}	S_{7b}	S_{8a}	S_{8b}	S_{9a}	S_{9b}
S_1	1.00											
S_2	-0.61	1.00										

S_3	0.13	-0.09	1.00									
S_4	0.00	0.00	0.00	1.00								
S_5	0.00	0.00	0.00	0.40	1.00							
S_6	0.00	0.00	0.00	0.00	0.00	1.00						
S_{7a}	0.00	0.00	0.00	0.00	0.00	0.00	1.00					
S_{7b}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00				
S_{8a}	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	1.00			
S_{8b}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	1.00		
S_{9a}	0.00	0.00	0.00	0.00	0.00	0.00	-0.22	0.22	-0.11	0.11	1.00	
S_{9b}	0.00	0.00	0.00	0.00	0.00	0.00	-0.22	-0.22	-0.11	-0.11	0.00	1.00

For both S2 (fSQP) and S4 (eSQP) the major correlation is that between S_1 and S_2 with coefficients of -0.602 and -0.611, respectively. This corresponds to the glue coordinate or, alternatively, can also be interpreted in terms of a reversible elimination from the apical position of a SQP, with a concomitant shortening of the basal angles, or vice versa. The correlation between S_4 and S_5 , if positive as in the case of S1, can be taken as a component of the Berry coordinate, with those trans-basal bonds destined to becoming axial bonds in the TBP lengthening as the angle between them increases, while the other pair of bonds shortens with a decrease in the angle contained by it. This type of distortion flies in the face of a structural generalisation which states that as two bonds shorten, so the angle between them increases for steric reasons. However, the "driving force" behind Berry intramolecular exchange is certain to be energetic in nature, with the result that this truism is violated. S4 (eSQP) manifests the positive correlation of S_4 with S_5 ($r = 0.403$) that is representative of the Berry coordinate, while S2 (fSQP) shows a negative correlation ($r = -0.352$) which, in fact, reflects the truism mentioned above. Consequently it would seem as if the flattened square pyramid needs to elevate prior to any distortion along the Berry coordinate. Lastly, the correlations for the two latter clusters are both blocked according to symmetry, with no correlations between coordinates of different symmetry.

Overall, therefore, the distortions identified for S-space as a whole have been retained on the subdivision into clusters. The major one, the glue factor, is associated with clusters S2 (fSQP) and S4 (eSQP), the Berry distortion with S1 (dTBP) and S4 (eSQP), but not with S2 (fSQP).

7. Factor Analysis

Cluster analysis has established the broad outlines of the data distribution; factor analysis will here be used in order to explore the shape of these distributions, i.e. to map the coordinates along which the clouds of data points expand. This will yield a picture of the generalised distortions which the ML_5 molecular fragment manifests in the solid state. Murray-Rust, in particular, has pioneered the application of this technique to the analysis of molecular structure. [18-21] He has shown [21] how the mathematical treatment of normal coordinate analysis closely parallels that of factor analysis. This arises from the possibility of factorising, in the same way, both the matrix of force constants which appears in the potential energy expression of the general valence force field, and the matrix of the covariances or correlations between the symmetry coordinates which forms the basis of the factor analysis. As a consequence of this parallel the most important factors will be closely related to soft normal coordinates. He has pointed out, however, that this parallel can easily be destroyed by rotation of the factor axes, if rotation does not occur *within* blocks of the matrix, since the rotated factors no longer lie along the eigenvectors (the directions of maximum variance).

(a) Setting the parameters

P4M, the BMDP factor analysis programme, can extract factors from either the covariance or the correlation matrix by a number of different methods including principal components. Several methods of rotation are available, including orthogonal and oblique rotation, and the number of factors, the number of iterations and the cut-off-points can be chosen by the user. The output contains, amongst others, univariate and bivariate variable statistics, eigenvalues and corresponding factor loadings, rotated factor loadings and scatter plots of these.

In all cases we chose the correlation matrix as the matrix to be factored, with the method for initial factor extraction as principal component analysis. Only factors whose eigenvalues exceeded unity (Kaiser's criterion) were retained and rotation was orthogonal.

(b) T-Space

The analysis of the correlation matrices for T-space, T1 and T3 revealed that

three major distortions characterised the data distribution. These were described in terms of adherence to an S_N2 coordinate, a Berry coordinate and a glue coordinate, with the S_N2 distortion associated with cluster T3 (TBP), the glue distortion with cluster T1 (dSQP) and the Berry coordinate being manifested in both clusters. Factor analysis will investigate the data for *simultaneous* correlations between two *or more* symmetry coordinates, as opposed to simply bivariate correlations. The task then is the reification of the factors, i.e. their translation into chemical terms. Rotation of the factors can aid this analysis, but does not necessarily do so. Consequently we shall only include rotation where it has indeed helped the interpretation of the factors.*

Results of the factor analysis are listed in Table 4.7.1, while Figure 4.7.1 graphically illustrated the corresponding distortion coordinates. The symmetry of the data distribution manifests itself in both the relation *between* factors representing the two partners of a degenerate representation (eg. F1 and F2 for T-space overall — they are both of E' symmetry†), as well as in the relation between symmetry equivalent coordinates *within* a factor, which have identical absolute loadings (eg. S_3 and S_4 in F3 of T-space overall — both belong to the A_2'' representation). In other words, the requirement that the symmetry inherent in the data distribution manifest itself in the results of factor analysis (as it did earlier in those of cluster analysis) has been met.

T-space overall:

For T-space overall the most important coordinate mapped is the Berry intramolecular exchange coordinate. This is characterised by the degenerate factor F1 and F2 whose partners together account for 28 percent of the sample variance. The second most important coordinate is the S_N2 coordinate mapped by F3 which is of A_2'' symmetry and accounts for 14 percent of the variance, while the constant-amount-of-glue coordinate appears in F4 (A_1' symmetry, 13 percent of variance). The degenerate factor F5 plus F6, of E' symmetry and accounting for 18 percent of variance, is interesting in that it maps a distortion which can be interpreted in

* As it turns out, there are only two cases where rotation gives chemically more meaningful results than do the unrotated factors.

† Such factors will henceforth be termed degenerate factors.

Table 4.7.1 Results of factor analysis of T-space. Only unrotated factors with eigenvalues greater than unity are shown. Sym = symmetry of factor, % Var = percentage of sample variance explained by each factor, Coordinate = general distortion mapped by the factor.

Cluster	Factors	Sym	% Var	Coordinate
T-Space	F1=0.743 S_{5a} +0.935 S_{6a} +0.554 S_{7a}	E'	14	Berry
	F2=0.743 S_{5b} +0.935 S_{6b} +0.554 S_{7b}	E'	14	Berry
	F3=-0.914 S_3 +0.914 S_4	A_2''	14	S_N2
	F4=-0.880 S_1 +0.880 S_2	A_1'	13	Glue
	F5=-0.605 S_{5b} +0.804 S_{7b}	E'	9	
	F6=-0.605 S_{5a} +0.804 S_{7a}	E'	9	
T1(dSQP)	F1=-0.838 S_1 +0.866 S_2 +0.940 S_{5a} +0.680 S_{6a} -0.605 S_{7a}	$A_1' + E'$	26	Addition/ Elimination
	F2=-0.530 S_3 +0.801 S_4 +0.670 S_{8a}	$A_2'' + E''$	12	Angular flexibility
	F3=-0.730 S_{5b} +0.359 S_{6b} +0.796 S_{7b}	E'	11	Angular flexibility
T3(TBP)	F1=-0.959 S_3 +0.959 S_4	A_2''	15	S_N2
	F2=0.358 S_{5b} +0.877 S_{6b} +0.707 S_{7b}	E'	12	Berry
	F3=0.358 S_{5a} +0.877 S_{6a} +0.707 S_{7a}	E'	12	Berry
	F4=0.793 S_1 -0.793 S_2	A_1'	10	Glue
	F5=0.874 S_{5a} +0.105 S_{6a} -0.573 S_{7a}	E'	9	
	F6=0.874 S_{5b} +0.105 S_{6b} -0.573 S_{7b}	E'	8	

two ways. These are related to whether this distortion is primarily associated with the TBP cluster (T3) or with the SQP cluster (T1). In the event of the former we could argue that this distortion resembles the incipient stages of the Berry intramolecular exchange coordinate, whereas in the latter it might be more appropriate to describe it as mirroring the incipient stages of a reversible addition/elimination reaction at a square planar centre.

T1:

Cluster T1 (dSQP), in contrast to T3 (TBP) and T-space overall, exhibits mixing of coordinates from different symmetry species in the same factor. This clearly is related to the inability of this cluster to conform to the symmetry of parameter

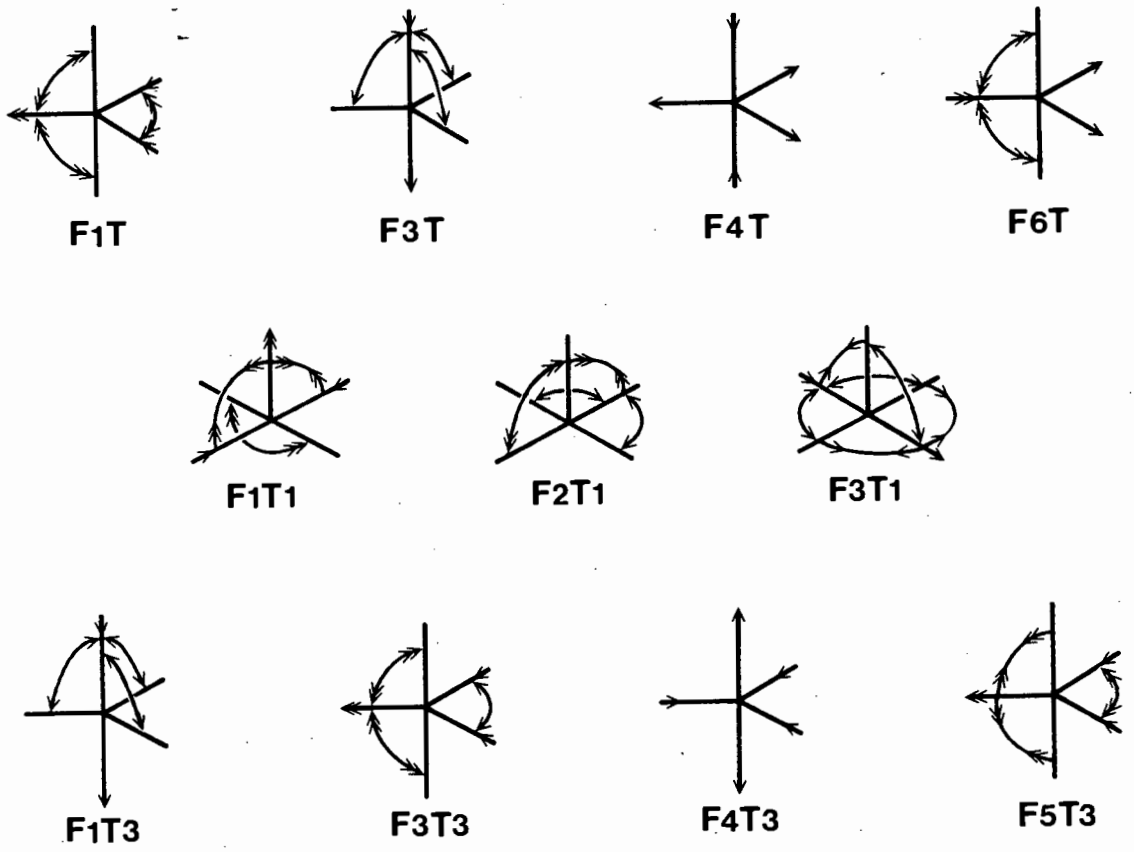


Figure 4.7.1 Graphic representation of factors describing T-space overall (FnT; n = 1,3,4,6), cluster T1 (FnT1; n = 1,2,3) and cluster T3 (FnT3; n = 1,2,4,5). Diagrams are constructed by superimposing the distortion due to each of the symmetry coordinates which are components of a factor onto *one* framework. Only *independent distortions* are indicated for the sake of visual simplicity. The number of heads to the arrows is a consequence of the superposition of symmetry coordinates, and is not related to their loading in a factor. The SQP framework was chosen for cluster T1 since its archetype is a dSQP. F2T, F5T, F2T3 and F6T3 are not illustrated since they simply form partners to F1T, F6T, F2T3 and F6T3 in the irreducible representation, and their inclusion would not simplify the interpretation.

---

space. Figure 4.7.1 graphically illustrates the distortions mapped by the factors, but superimposed this time onto a SQP framework in order to facilitate the reification

of the factors. F1 rather obviously resembles the expected distortions occurring during a reversible addition/elimination reaction at a square planar centre, and it conforms to C_{2v} symmetry. F2 and F3 can be interpreted as a manifestation of an easily deformable SQP molecular fragment, a fragment exhibiting a large degree of angular flexibility in no chemically obvious sense. It is important, though, that both factors preserve elements of C_{2v} symmetry as well; they preserve the $\sigma_v(xz)$ and $\sigma_v'(yz)$ elements, respectively.

T3:

The variance in T3 (TBP) is associated with the S_N2 distortion coordinate (F1, A_2'' symmetry, 15 percent of variance), the Berry coordinate (F2 plus F3, E' symmetry, 24 percent of variance) and the glue coordinate (F4, A_1' symmetry, 10 percent of variance). These distortions are all diagrammatically represented in Figure 4.7.1, as is F5. This factor features the same components (S_{5a}, S_{6a}, S_{7a}) as the Berry factor (F3), except that the loading for S_{7a} in F5 is negative, while it is positive in F3 and all other Berry coordinates. The opening up of an equatorial angle concomitant with a shortening of the two bonds containing it can be explained on the basis of simple steric requirements. The negative correlation of S_{7a} (the widening of the already large axial angle) with S_{5a} and S_{6a} might result from steric crowding brought about by the opening of the equatorial bond.

Relative importance of symmetry coordinates:

The graphic representation of distortions mapped by the factors is useful in their reification and translation into chemical terms, but it attaches equal absolute importance to each symmetry coordinate in a factor. The loadings of the coordinates can give some idea of the significance of a particular distortion, but the relative importance of the symmetry coordinates must be established by multiplying their standard deviations by their loading in a given factor. Table 4.7.2 reports the relative importance of each symmetry coordinate in the factors extracted for T-space overall and for clusters T1 (dSQP) and T3 (TBP). It affords slightly more insight into the details of the distortions mapped by some of the factors. For example, F1 for T-space which maps the Berry coordinate seems to be dominated by S_{6a}, the opening of the equatorial angle (θ_{24}) which accompanies the intramolecular exchange. A chemical interpretation of this might be that the relative change which

this angle has to undergo along the Berry coordinate is greater than that experienced by the axial angle (θ_{15}) . This would conform to the results of cluster analysis which indicate that an "average" SQP — which would presumably approximate to that on the Berry coordinate — has trans-basal angles $\theta_{15} = \theta_{24} = 166°$, implying that θ_{24} needs to change by $46°, \theta_{15}$ only by $14°$. Apart from F1 no other factor seems to contain symmetry coordinates whose importance is either vastly greater or smaller than that of the others in that factor.

Table 4.7.2 Relative importance of bond symmetry coordinates $(S_1, S_2, S_3, S_{5a}, S_{5b})$ [Å] and angular coordinates $(S_4, S_6 - S_8)$ [Å] in the factors for T-space (T) and clusters T1 and T3.

	S_1	S_2	S_3	S_4	S_{5a}	S_{5b}	S_{6a}	S_{6b}	S_{7a}	S_{7b}	S_{8a}	S_{8b}
F1T					0.189		1.156		0.129			
F2T						0.189		1.156		0.129		
F3T			-0.121	0.327								
F4T	-0.150	0.118										
F5T						-0.154				0.187		
F6T					-0.154				0.187			
F1T1	-0.125	0.121			0.327		0.294		-0.142			
F2T1			-0.036	0.197							0.212	
F3T1						-0.060		0.158		0.156		
F1T3			-0.210	0.512								
F2T3						0.029		0.291		0.124		
F3T3					0.029		0.291		0.124			
F4T3	0.151	-0.067										
F5T3					0.072		0.035		-0.100			
F6T3						0.072		0.035		-0.100		

For F2 in cluster T1 (dSQP) symmetry coordinate S_3 , mapping the simultaneous lengthening and shortening of a pair of trans-basal bonds, appears to be

much less significant than S_4 or S_{8a}, both of which reflect the angular flexibility referred to earlier. How much of this difference is due to the comparison of different kinds of symmetry coordinates (incremental versus angular) and how much of it might reflect real differences in the degree of distortion is difficult to judge. For T3 (TBP) the most striking difference in relative importance is that between S_{5a} and both of S_{6a} and S_{7a} in F3, the Berry factor. The data suggest that changes in the equatorial bond lengths — the pivotal bond becomes longer, the other two shorter — are far less important along the Berry coordinate than are the changes in angles (specifically the equatorial angle, θ_{24}, and the axial angle, θ_{15},). In other words, a given molecular conformation could easily be distorted along the C_{2v} exchange coordinate in terms of adjustments of its angles, without its bond lengths necessarily having to conform rigidly to the requirement that the pivotal (soon to be apical-) bond be much longer than the basal ones. This, of course, also conforms to chemical criteria, since we know that a range of SQPs with differing apical bond lengths exist, and that, in any case, this bond can adapt to any length without affecting the C_{2v} symmetry anywhere along the Berry coordinate. How much of the difference in the relative importance of S_{5a} and S_{6a}, S_{7a} is the result, again, of comparing bond symmetry coordinates with angular ones, and how much of it reflects real structural factors is difficult to assess, though.*

T-space minus nickel compounds;

Since nickel compounds represent the largest homogenous grouping in the data (113 out of 196 entries) it is of interest to establish whether the results of the analysis are in any way dominated by this grouping, in which case this would point to an inherent difference between five-coordination in nickel and that in the other metals. Table 4.7.3 illustrates the results of factor analysis of T-space *excluding* nickel entries.

The order of importance of the coordinates has altered somewhat from that of T-space including nickel (Table 4.7.1). In the latter the order was: (i) Berry

* One way of doing this, perhaps, could be the calculation of the relative importance of each internal coordinate by multiplication of the inverse symmetry coordinate matrix by the relative importance of the symmetry coordinates in each factor. This would still require, though, a comparison of bond distance increments and angles, albeit scaled angular displacements.

coordinate (28 percent of variance), (ii) S_N2 coordinate (14 percent), (iii) glue coordinate (13 percent), whereas the exclusion of nickel brings about the order: (i) Berry (26 percent), (ii) glue (18 percent), (iii) S_N2 coordinate (11 percent). Essentially, therefore, the glue coordinate becomes slightly more important, the S_N2 slightly less on exclusion of nickel. This correlates well with the facts established earlier. First, the glue coordinate seems to be more closely associated with the square pyramidally disposed structure, being an important component in the addition/elimination coordinate for a square-planar centre — the change in bond lengths, in particular that of the apical bond, is more important in this coordinate and the fSQP → eSQP distortion, than it is for the S_N2 coordinate at a TBP. Second, by far the majority of TBP compounds are nickel-containing (45 out of 59). Exclusion of the nickel entries would therefore increase the relative amount of the SQP structures in T-space, and consequently result in magnifying all those factors primarily associated with this conformation. It is interesting to note that the remaining 14 TBP conformations still make mapping of the S_N2 coordinate possible.

The relatively unaltered importance of the Berry coordinate as the primary distortion in T-space does not come as a surprise in view of the ability of both TBP and SQP conformations to manifest such C_{2v} distortions. In this light it is surprising, though, that the Berry coordinate (as mapped by S_{5a}, S_{6a} and S_{7a} in T-space) is not manifest in the T1 (dSQP) cluster. Perhaps this is related to the clusters' inability to conform fully to the symmetry of T-space.

F5 and F6 show the same inverse relation of S_{5a} to S_{6a} and S_{7b} (or, respectively, S_{5b}, S_{6b} and S_{7b}) as was observed earlier for F5T3 and F6T3; their loadings are of opposite sign, whereas in the Berry factors (F1T, F2T, F2T3 and F3T3) they are all of similar sign. What is more, is that this curious distortion is still as important *after* the exclusion of nickel entries as it is *prior* to. This would suggest that the distortion is related mainly to the remaining compounds, i.e. that it results from some distortion of the SQP conformation which happens to be mapped by this factor.

(c) **S-space**

Univariate and bivariate statistics for the sample distribution in S-space sug-

Table 4.7.3 Results of factor analysis of T-space excluding all nickel comounds. Sym = symmetry of factor, % Var = percentage of variance explained by the factor. Only factors with eigenvalues greater than unity are listed.

Factor	Sym	% Var	Coordinate
F1=-0.956 S_1 +0.956 S_2	A'	18	Glue
F2=0.739 S_{5a} +0.921 S_{6a} +0.397 S_{7a}	E'	13	Berry
F3=0.739 S_{5b} +0.921 S_{6b} +0.397 S_{7b}	E'	13	Berry
F4=-0.830 S_3 +0.830 S_4	A_2''	11	S_N2
F5=-0.586 S_{5a} +0.091 S_{6a} +0.880 S_{7a}	E'	9	
F6=-0.586 S_{5b} +0.091 S_{6b} +0.880 S_{7b}	E'	9	
F7=1.00 S_{8a}	E''	8	
F8=1.00 S_{8b}	E''	8	

gest that the variance lies primarily along a coordinate mirroring a reversible addition/elimination reaction at a square planar centre. Apparently two types of distortion together give rise to this coordinate; when the glue coordinate (S_1, S_2) is correlated to the flattening or elevation of the SQP (S_3) the addition/elimination coordinate arises. Some variance also appears to be associated with the Berry coordinate. This distortion is mapped by clusters S1 (dTBP) and S4 (eSQP), while the addition/elimination coordinate is associated with S2 (fSQP) and S4. Furthermore, an S_N2 coordinate also seems to be associated with S1.

Table 4.7.4 illustrates the results of factor analysis of S-space, while Figure 4.7.2 represents graphic illustrations of the factors. The symmetry of the data distribution re-emerges in the symmetry of the factors. For S2, S4 and S-space overall the factors do not mix symmetry coordinates from different irreducible representations, and the two partners of a degenerate factor have identical eigenvalues (and, hence, percentage variance), with the absolute loadings of identical symmetry coordinates in the two partners being equal. S1 mixes the species as did T1 in the case of T-space. Most likely this is also a consequence of this clusters' inability to conform fully to C_{4v} symmetry, exhibiting only C_{2v} symmetry instead.

Table 4.7.4 Results of factor analysis of S-space. Primed factors are rotated, unprimed unrotated. Only factors whose eigenvalues are greater than unity are listed. In each factor only those symmetry coordinates whose relative importance (Table 4.7.5) is greater than 0.01 [Å] are tabulated. Sym = symmetry of factor, % Var = percentage variance explained by each factor. Add/El = Addition/Elimination.

Cluster	Factors	Sym	% Var	Coordinate
S-Space overall	F1′ =0.917 S_1 -0.888 S_2 +0.765 S_3	A_1	14	Add/El
	F2′ =0.863 S_{7a} -0.516 S_{8a} -0.562 S_{9a} -0.562 S_{9b}	E	14	S_N2
	F3′ =0.863 S_{7b} -0.516 S_{8b} +0.562 S_{9a} -0.562 S_{9b}	E	14	S_N2
	F4′ =0.832 S_4 +0.832 S_5	B_1	12	Berry
	F5′ =1.000 S_6	B_2	9	
S1(dTBP)	F1=-0.914 S_{7a} +0.853 S_{8a} +0.888 S_{9a} +0.888 S_{9b}	E	26	S_N2
	F2=-0.439 S_1 +0.770 S_2 -0.651 S_3 +0.781 S_4 +0.731 S_5	$A_1 + B_1$	20	
	F3=0.821 S_{7b} -0.817 S_{8b}	E	11	
	F4=-0.542 S_1 +0.349 S_2 +0.610 S_3 +0.284 S_4 -0.454 S_5	$A_1 + B_1$	9	
S2(fSQP)	F1=0.917 S_1 -0.842 S_2 +0.292 S_3	A_1	14	Add/El
	F2=0.822 S_4 -0.822 S_5	B_1	11	
	F3=0.546 S_{7a} +0.546 S_{7b} -0.556 S_{8a} -0.566 S_{8b} +0.297 S_{9b}	E	11	Angular flexibility
	F4=0.546 S_{7a} -0.546 S_{7b} -0.556 S_{8a} +0.566 S_{8b} +0.297 S_{9a}	E	11	Angular flexibility
	F5=0.064 S_1 +0.396 S_2 +0.939 S_3	A_1	9	Elevation of SQP
S4(eSQP)	F1′ =0.888 S_1 -0.879 S_2 +0.291 S_3	A_1	14	Add/El
	F2′ =0.837 S_4 +0.837 S_5	B_1	12	Berry
	F3′ =0.740 S_{7a} +0.516 S_{8a} -0.541 S_{9a} -0.541 S_{9b}	E	12	Angular flexibility
	F4′ =0.740 S_{7b} +0.516 S_{8b} +0.541 S_{9a} -0.541 S_{9b}	E	12	Angular flexibility
	F5′ =1.000 S_6	B_2	9	

S-space overall:

The most dominant distortion mapped seems to be the S_N2 coordinate for a

trigonal bipyramidally disposed conformation. This coordinate manifests itself in the degenerate factor whose two partners are F2′ and F3′, and it accounts for 28 percent of variance. The next most important one is the addition/elimination coordinate mapped by F1′, accounting for 14 percent of sample variance. F4′ maps the Berry coordinate describing 12 percent of sample variance. The last factor F5′ (whose eigenvalue is greater than unity) does not correspond to any recognisable chemical coordinate, but gives the distinct impression that it could be related to the presence of a group of compounds containing either two rather rigid bidentate ligands with small bites, or a macrocycle of some kind.

In this case the unrotated factors F2 and F3 contained two symmetry coordinates whose loadings were small, but not quite zero, resulting in distortions which could not be sensibly interpreted. Rotation of the factor axes left F1, F4 and F5 unchanged, but considerably simplified the reification of F2′ and F3′.

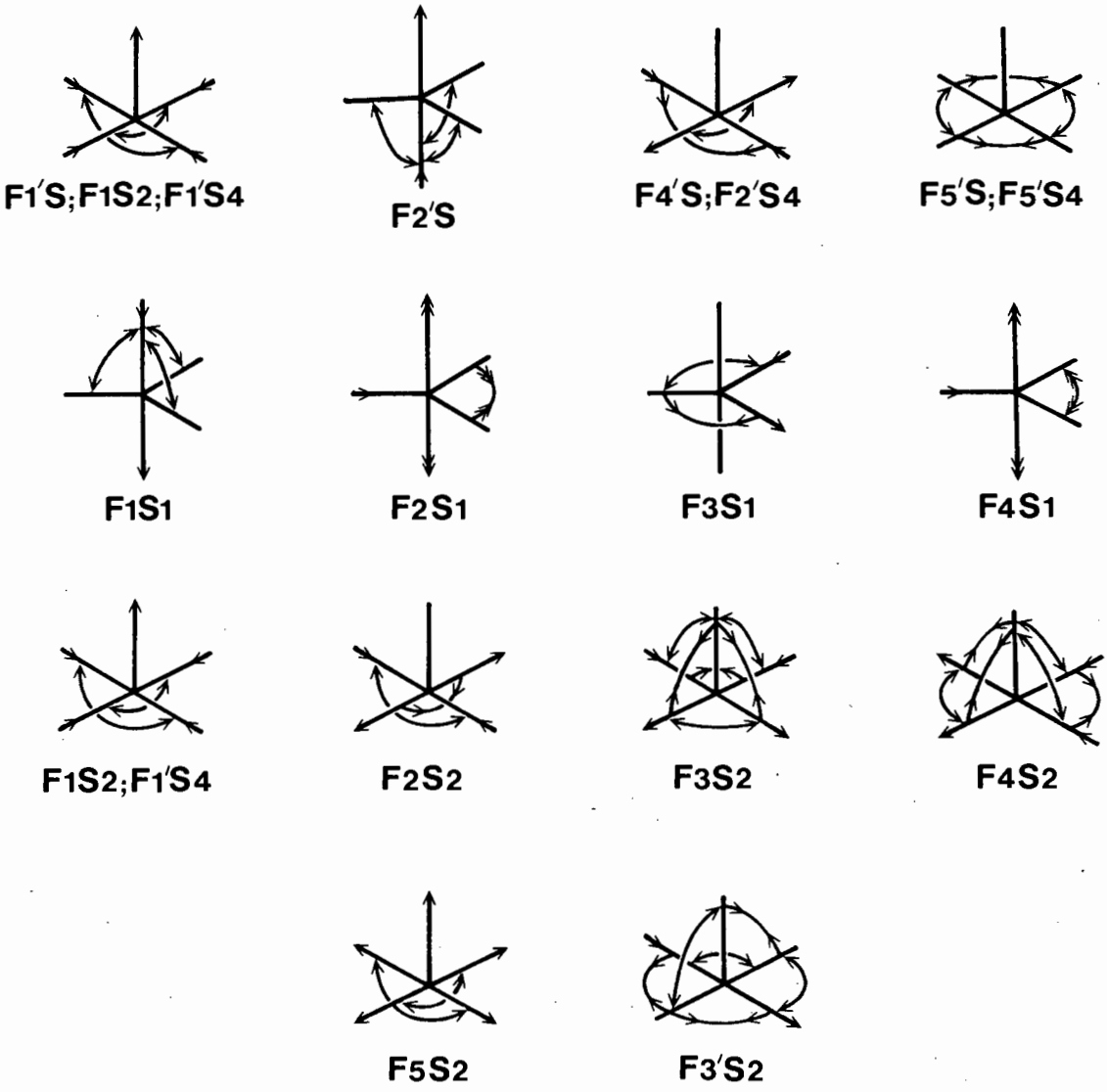


Figure 4.7.2 Graphic representation of factors describing S-space overall (FnS; n = $1', 2', 4', 5'$), cluster S1 (FnS1; n = 1,2,3,4), cluster S2 (FnS2; n = $1', 2', 3', 5'$) and cluster S4(FnS4; n = $1', 2', 3', 5'$) . Only those independent distortions are shown, whose relative importance in the factor exceeds 0.01 [Å]. F2′S is shown with reference to a TBP in order to facilitate its reification.

S1:

Not unexpectedly this cluster is characterised by the S_N2 coordinate, mapped by F1 of E symmetry and accounting for 26 percent of sample variance. The expected partner to F1 under the E representation is not immediately obvious, and the only possibility seems to be F3, which contains two of the four expected symmetry coordinates and accounts for 11 percent of variance. The last factor, a mixture of A_1 and B_1 representations, essentially maps the opening of an equatorial angle in the TBP with a concomitant shortening of the bond opposite this angle, and a lengthening of the axial bonds. The related coordinate, F2, differs in that while the axial bonds lengthen, the equatorial angle decreases. Alternatively, the positive correlation between S_4 and S_5 in this factor can be interpreted as mapping the Berry coordinate, while the factors' other components (S_1, S_2, S_3) trace the addition/elimination or glue coordinate. In this sense, then, the results obtained for T3 (TBP) and those of S1 (dTBP) would correspond : in the former the factors are (i) S_N2 , (ii) Berry and (iii) glue, while the S_N2 coordinate and a combination of Berry and glue distortions constitute the factors for S1. There seems no clear chemical description of the distortions mapped by F3 and F4. Most likely their complexity is not unrelated to the incompatibility of the symmetry of this cluster to that of S-space.

S2:

F1 here combines the glue coordinate (S_1 and S_2) with the flattening/elevation of the SQP (S_3) to map the addition/elimination coordinate. It is of A_1 symmetry and explains 14 percent of sample variance. The degenerate factor with components F3 and F4, accounting for 22 percent of variance, seems to map the angular flexibility which was referred to previously, whereby the position of the apical ligand relative to the ML_4 fragment (the metal with four basal ligands) is rather variable in terms of both its distance to the metal, and the angles between it and the ML_4 fragment. F5 (A_1 symmetry, 9 percent of variance) contains the same components as F1 (addition/elimination), but their loadings are of different sign. Consequently it appears that this factor simply maps the flattening/elevation of the fSQP (S_2) , since the distortion in the bond lengths (S_1, S_2) is not correlated as it should be for the addition/elimination coordinate. Finally, F2 is curious

in that it contains the Berry coordinate components (S_4 and S_5), but they are negatively correlated, instead of positively which they would be if mapping the intramolecular exchange coordinate. This was previously observed and commented on when bivariate statistics for S2 were analysed in Section 4.6. The most likely explanation is that this factor simply maps the response of the fSQP to a shortening or lengthening of its basal bonds, and that it is unrelated to the Berry coordinate, which is manifested only by the eSQP. It could also be taken as implying that the fSQP does not undergo Berry rotation, or is not an intermediate on the Berry coordinate. In other words, *a SQP which is flattened cannot undergo intramolecular exchange via the Berry mechanism, unless it first becomes elevated.*

S4:

This cluster also has 14 percent of its variance associated with the addition/elimination coordinate which is mapped by F1′ , as was the case for S2. The Berry coordinate which was identified by F4′ for S-space overall, but not picked up in S1 (dTBP), is mapped by F2′ and it accounts for 12 percent of sample variance. The degenerate factor represented by F3′ and F4′ , describing 24 percent of variance, contains similar components to those of F3S2 and F4S2, suggesting that it also mirrors the flexibility in the apical bond with respect to the residual ML_4 fragment. Finally, F5′ clearly is identical to F5S, the factor which possibly reflects the presence of a group of compounds with severely constraining ligands, either two bidentate ligands with small bites or a macrocycle.

In this case rotation was essentially employed in order to simplify F2, which contained four components with loadings below 0.09 while two components had loadings of the order of 0.8. On rotation the small loadings disappeared and the interpretation of F2′ was vastly simpler. F1 and F5 remained unchanged, while the previous slight asymmetry in the factor loadings of F3 and F4 was removed upon rotation.

In summary, the results obtained for T1 (dSQP) and those for S2 (fSQP) and S4 (eSQP) correlate well. In the former the addition/elimination and the angular flexibility coordinate only were identified, while for the SQP clusters in S-space these two distortions were mapped, with the addition of the Berry coordinate in the case of S4 (eSQP). Quite possibly the absence of the Berry coordinate in T1,

already commented upon earlier, is also related to the mixing of S2 (fSQP) and S4 (eSQP) in T1 (dSQP), with distortions manifested only by S4 being eclipsed on the merging of these two clusters. This phenomenon, most likely, is enhanced by the fact that in T-space only *one half* of the possible variance in the SQP conformations is manifested, effectively, since the *full* C_{4v} symmetry of this variation cannot emerge due to the constraints placed upon it by the symmetry of the D_{3h} data space.

Relative importance of symmetry coordinates:

There appear to be no significant differences in the relative importances of the various symmetry coordinates as collected in Table 4.7.5 in the factors extracted for S-space overall. What small differences there are, could well result from the comparison of bond and angular symmetry coordinates (eg. S_{7a} and S_{8b} in F2′S), with the possible exception of F4′ . This factor reflects the Berry coordinate, S_4 mapping the lengthening of (soon to become) axial bonds and the shortening of equatorial ones, while S_5 maps the opening and closing of the respective trans-basal angles which is concomitant with the changes in bond lengths. From the relative importances listed in Table 4.7.5 it might be argued that the change in bond lengths accompanying a distortion along the Berry coordinate is considerably less important than the angular changes occurring during this distortion. This conclusion tallies well with that made from an analysis of F1 for T-space in which the angular coordinate mapping the opening of the equatorial angle (S_{6a}) was also considerably more important than that describing the simultaneous bond length changes. For S1 all of F2, F3 and F4 manifest small differences in importance amongst the various symmetry coordinates. Quite possibly, though, these differences spring from the same problem which, in the first place, is responsible for the emergence of these complex factors which map no chemically meaningful distortion — that of the incompatibility of S1 with the symmetry of S-space.

The flexibility in the position of the apical ligand appears to be of primary importance in accounting for the variability amongst the fSQPs (S2). In F1 the distance of this ligand from the metal (S_1) is rather more important than that of the others (S_2) and than the variability in the trans-basal angles (S_3) . This suggests that the approach of the apical ligand to the residual ML_4 fragment along the addition/elimination coordinate is, at least initially, not accompanied by

Table 4.7.5 Relative importance of bond symmetry coordinates $(S_1, S_2, S_4, S_{7a}, S_{7b})$ [Å] and angular coordinates $(S_3, S_5, S_6, S_8, S_9)$ [Å] in the factors for S-space (S) and clusters S1, S2 and S4. Rotated factors are denoted by primes.

	S_1	S_2	S_3	S_4	S_5	S_6	S_{7a}	S_{7b}	S_{8a}	S_{8b}	S_{9a}	S_{9a}
F1′S	0.291	-0.157	0.376									
F2′S							0.095		-0.186		-0.142	-0.142
F3′S								0.095		0.186	0.142	-0.142
F4′S				0.091	0.667							
F5′S						0.268						
F1S1							-0.186		0.294		0.276	0.276
F2S1	-0.038	0.103	-0.161	0.115	0.232							
F3S1								0.070		-0.250		
F4S	-0.047	0.047	0.151	0.042	-0.144							
F1S2	0.274	-0.098	0.063									
F2S2				0.045	-0.171							
F3S2							0.039	0.039	-0.272	-0.272		0.061
F4S2							0.039	-0.039	-0.272	0.272		0.061
F5S2	0.019	0.046	0.202									
F1′S4	0.111	-0.103	0.086									
F2′S4				0.065	0.233							
F3′S4							0.055		0.165		-0.119	-0.119
F4′S4								0.055		0.165	0.119	-0.119
F5′S4						0.260						

much elevation in the fSQP. In F3 and F4 the angular flexibility of the apical ligand (S_{8a}, S_{8b}) dominates slightly, suggesting that the angle of approach (or departure) of the apical ligand is not limited strictly to the perpendicular line through the centre of the basal L_4 plane of the SQP. In F5 (the "elevation" coordinate) the

variability in the trans-basal angles (S_3) seems to be most important, while bond distances (S_1, S_2) play a much less significant role. This suggests, as does F1, that there is a degree to which the pyramidalisation at the metal is independent of the proximity of the apical ligand.

For the addition/elimination coordinate (F1′) in S4 (eSQP), however, the proximity of all five ligands to the metal and the degree of pyramidalisation appear to be more closely linked than in the fSQP, as indicated by the very similar relative importances of the bond distances (S_1, S_2) and angles (S_3) in this factor. This might suggest, that as the addition of the apical ligand proceeds, a point is reached at which the elevation of the metal from the basal L_4 plane is at a maximum, although the apical bond has not yet shortened to its minimum. At this juncture the relative importances switch over from what they are in S2 (fSQP), and changes in bond distances predominate over those in the degree of pyramidalisation. Alternatively, if one assumes the eSQP to be representative of the end of the addition/elimination coordinate, then the important role which this factor has in accounting for the variance in S4 (14 percent) would suggest that the Berry and the addition/elimination coordinate might well be adhered to simultaneously. Purely from symmetry considerations this would seem unlikely, though, since the symmetries of the two distortions are different — the addition/elimination coordinate is of A_1 symmetry, the Berry of B_1. For the Berry coordinate F2′ in S4 the bond distances (S_4) again seem less significant in describing the variance along this distortion than do the trans-basal (or, soon to be, axial and equatorial) bonds (S_5).

Overall, the reification of the factors extracted for S-space and its clusters has been rather more difficult than it was in the case of T-space. This is manifested, firstly, in the need for factor rotation and, secondly, in the considerably more complex nature of the distortions mapped by the factors. Quite likely this stems from the closer proximity and greater diffusion of the clusters in S-space; a complication arises in the extraction of "pure" factors from clouds of representative points when the distributions merge or overlap.

S-space minus nickel compounds:

The results of factor analysis on the S-space data excluding nickel compounds

are tabulated in Table 4.7.6. Where they differ, both rotated and unrotated factors are shown. Prior to rotation the order of the factors was (i) S_N2 (22 percent of variance), (ii) addition/elimination (18 percent) and (iii) Berry (11 percent). After rotation it altered to (i) addition/elimination (18 percent) and (ii) Berry (11 percent), with the S_N2 coordinate all but disappearing. Both these results ought to be compared to the results for S-space including nickel entries, where the order in factors is (i) S_N2 (28 percent), (ii) addition/elimination (14 percent) and (iii) Berry (12 percent).

Table 4.7.6 Results of factor analysis of S-space excluding all nickel compounds. Rotated factors are denoted by primes. Sym = symmetry of factor, % Var = percentage of variance accounted for by factor. Only the first four factors have been listed, and only compounds with loadings greater than 0.1 are included in the factors. Add/El = Addition/Elimination.

Factor	Sym	% Var	Coordinate
F1 = 0.962 S_1 - 0.935 S_2 + 0.614 S_3	A_1	18	Add/El
F2 = 0.828 S_{7b} - 0.247 S_{8b} + 0.516 S_{9a} -0.516 S_{9b}	E	11	S_N2
F2′ = 0.800 S_4 + 0.800 S_5	B_1		Berry
F3 = 0.829 S_{7a} -0.247 S_{8a} - 0.516 S_{9a} -0.516 S_{9b}	E	11	S_N2
F3′ = -0.500 S_{7a} -0.500 S_{7b} + 0.855 S_{9b}	E		
F4 = 0.799 S_4 + 0.799 S_5	B_1	11	Berry
F4′ = -0.500 S_{7a} + 0.500 S_{7b} + 0.855 S_{9a}	E		

With the exclusion of nickel entries, constituting the bulk of the TBPs in the data, factors associated primarily with with the SQP conformation, such as the addition/elimination coordinate ought to come into prominence, with those associated with the TBP falling by the wayside. As it turns out the exclusion has indeed made the addition/elimination coordinate more important, with the S_N2 coordinate still prominent in the unrotated factors, though. Curiously, rotation interchanges the factors and eliminates the S_N2 coordinate, resulting in the pattern expected on

the basis of the results from T-space. In this case, then, rotation yielded more meaningful results that did the unrotated factors.

REFERENCES

1. D. Wishart. *CLUSTAN Users' Guide, The Clustan Project*, University College, London, 1975.

2. N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner and D.H. Bent. *Statistical Package for the Social Sciences*, 2nd Edition, McGraw-Hill, New York, London, 1975.

3. W.J. Dixon, Editor. *BMDP Statistical Software*, 1983. Printing with Additions, University of California Press, Berkeley, 1983.

4. Auf der Heyde, T.P.E., Nassimbeni, L.R. *Inorg. Chem.* 1984, **23**, 4525-4532.

5. Orioli, P.L., Di Vaira, M. *J. Chem Soc. (A)* 1985, **6**, 216-228.

6. Nørskov-Lauritsen, L., Bürgi, H.B. *J. Comput. Chem.* 1985, **6**, 216-228.

7. Willett, P. *J. Chem. Inf. Comput. Sci.* 1984, **24**, 29-33.

8.(a) CMBPNI: Powell, H.M., Watkins, D.J., Witford, J.G. *J. Chem. Soc. (A)* 1971, 1803-1810.

8.(b) CMPMNI; Jansen, J.C., van Koningsveld, H., van Ooijen, J.A.C., van Reedijk, J. *Inorg. Chem.* 1980, **19**, 170-174.

8.(c) CSMRHC: Mague, J.T. *Inorg. Chem.* 1970, **9**, 1610-1618.

8.(d) DPPCRH: Pignolet, L.H., Doughty, D.H., Nowicki, S.C, Casalnuovo, A.L. *Inorg. Chem.* 1980, **19**, 2172-2177.

8.(e) DPMCRH10: Cowie, M. *Inorg. Chem.* 1979, **18**, 286-292.

8.(f) EDCRCN: Raymond, K.N., Corfield, P.W.R., Ibers, J.A. *Inorg. Chem.* 1968, **7**, 1362-1372.

8.(g) IPESNI: Fälth, L. *Chemica Scripta.* 1976, **9**, 167-170.

8.(h) SALDNI: Seleborg, M., Holt, S.L., Post, B. *Inorg. Chem.* 1971, **10**, 1501-1504.

8.(i) TMAGEP: Estes, E.D., Hodgson, D.J. *Inorg. Chem.* 1973, **12**, 2932-2935.

8.(j) BESPIF: Kalck, P., Bonnet, J.-J., Poilblanc, R. *J. Am. Chem. Soc.* 1982, **104**, 3069-3077.

9. Muetterties, E.L., Guggenberger, L.J. *J. Am. Chem. Soc.* 1974, **96**, 1748-1756.

10. Holmes, R.R. *Acc. Chem. Res.* 1979, **12**, 257-268.

11.(a) PEANIC: Ghilardi, C.A., Sabatini, A., Sacconi, L. *Inorg. Chem.* 1976, **15**, 2763-2767.

11.(b) PEANNI: Di Vaira, M., Ghilardi, C.A., Sacconi, L. *Inorg. Chem.* 1976, **15**, 1555-1561.

11.(c) CHESNI: Gould, R.O., Harding, M.M. *J. Chem. Soc. (A)* 1970, 875-881.

11.(d) CNPLPT: Wernberg, O., Hazell, A. *J.C.S. Dalton* 1980, 973-978.

11.(e) DMPADP10: Shetty, P.S., Fernando, Q. *J. Am. Chem. Soc.* 1970, **92**, 3964-3969.

11.(f) MPNBNI: Butcher, R.J., Sinn, E. *Inorg. Chem.* 1977, **16**, 2334-2343.

11.(g) FPHPRH: Hughes, R.P., Krishnamachari, N., Lock, C.J.L., Powell, J., Turner, G. *Inorg. Chem.* 1977, **16**, 314-319.

12.(a) SAS Institute Inc., *SAS User's Guide : Statistics*, 1982 Edition, SAS Institute Inc., Cary, NC 27511, 1982.

12.(b) SAS Institute Inc., SAS User's Guide : Basics, 1982 Edition, SAS Institute Inc., Cary, NC 27511, 1982.

13. Searle, W.S. *SAS Technical Report : A-108*, SAS Institute Inc., Cary, NC 27511, 1983.

14. Holmes, R.R. *Prog. Inorg. Chem.* 1984, **32**, 119-235.

15. Holmes, R.R. *J. Am. Chem. Soc.* 1984, **106**, 3745-3750.

16. Rossi, A.R., Hoffmann, R. *Inorg. Chem.* 1975, **14**, 365-372.

17.(a) Burdett, J.K. *Adv. Inorg. Chem. Radiochem.* 1978, **21**, 113-134.

17.(b) Gillespie, R.J. *J. Chem Soc.* 1963, 4679-4692.

18. Murray-Rust, P., Motherwell, S. *Acta Cryst.* 1978, **B34**, 2534-2546.

20. Domenicano, A., Murray-Rust, P., Vaciago, A. *Acta Cryst.* 1983, **B39**, 457-468.

21. Murray-Rust, P. *Acta Cryst.* 1982, **B38**, 2765-2771.

CHAPTER 5

Summary and Conclusion

5.1 Summary of Chemical Results

Cluster analysis has shown that the molecular geometry of five-coordinate d^8 complexes is characterised by three archetypal conformations — the trigonal bipyramid, a flattened square pyramid and an elevated square pyramid. Collectively the data exhibit distortions of these conformations which can be described in terms of adherence to an S_N2 -, a Berry-, an addition/elimination- and a glue coordinate, as the results of factor analysis have demonstrated.

TBP

Essential structural features of the TBP include large variances in the axial bond lengths (r_1 and r_5) and the degree of elevation of the metal atom out of the equatorial plane, with a correspondingly large range in the axial-equatorial angles. On average, the axial bonds are longer than the equatorial ones, while the average trans-axial angle (θ_{15}) of 174^o is indicative of very few structures in the data set adopting perfect TBP conformation. The geometries of the ML_5 fragments classified as TBP essentially map a coordinate which is reminiscent of the distortions of the TBP along a S_N2 reaction pathway: a lengthening of one axial bond, a shortening of the other and a concomitant umbrella type distortion of the angles between the equatorial ligands and the two axial ones. The second factor delineates a coordinate involving simultaneously the closure of the axial angle (θ_{15}) and an opening of one of the equatorial ones (θ_{24}) , an intramolecular distortion which parallels that of the Berry coordinate. Moreover, it appears that the increase in θ_{24} is considerably more important in distorting the TBP towards the SQP, than is θ_{15} or than are any changes in bond lengths. The third and least important factor of the three involves the simultaneous lengthening of the axial bonds and a shortening of the equatorial ones, or vice versa. We have called this the constant-amount-of-glue coordinate, or simply the glue coordinate. It reveals that the bonding electron density of, say, the axial bonds may only be increased at the expense of that of the equatorial bonds, or vice versa.

Of the metals constituting the data set nickel exhibits the greatest preponderance of TBP conformations (approximately 40 percent of the nickel entries), with rhodium and iridium following (18 percent and 26 percent), while palladium and platinum appear to adopt this conformation only under exceptional circumstances

(0 and 1 case, respectively).

fSQP

A large variance in the position of the apical ligand with respect to the residual ML_4 fragment characterises the fSQP. In other words, both the length of the apical bond and the angles between it and the other bonds are highly variable. The average apical bond for compounds classified with this group is considerably longer (by 0.8 Å) than the basal bonds, and the average trans-basal angle lies at 171^o . The data primarily cluster along a pathway which maps a combination of the glue coordinate — describing the simultaneous shortening of the apical and lengthening of the basal bonds, or vice versa — with a coordinate describing the elevation, or degree of pyramidalisation, of the fSQP. This combined coordinate is reminiscent of the distortions manifested by an ML_5 moeity constituted by the square planar ML_4 fragment and the fifth ligand, L , during a reversible addition/elimination reaction at a four-coordinate centre. The glue component of the addition/elimination coordinate is far more important for the fSQP than it was for the TBP; the relation between the apical bonding electron density and the basal one appears to be of a much greater sensitivity than that between the axial and the equatorial bonds. The data suggest, however, that at *large* distances the apical ligand does not influence the geometry around the metal in a very consistent way.

The Berry coordinate is not manifested by the fSQP data. Instead, it seems that the fSQP cannot undergo intramolecular exchange via the Berry mechanism unless a certain degree of pyramidalisation has been achieved.

The tendency towards this conformation is greatest for palladium and platinum, the majority of whose complexes adopt it, with approximately half the square-pyramidally disposed nickel complexes also conforming to it. Rhodium and iridium, on the other hand, do not appear in this cluster at all — their SQP complexes are all part of the eSQP cluster.

eSQP

This conformation is characterised by an apical bond length which is much nearer that of the basal bonds than was the case for the fSQP, and whose variance relative to that of the basal bonds is also much reduced. The structure of this

archetype is altogether much less variable than that of the fSQP, and its average trans-basal angle is 163^{o} . The distortion coordinates along which the members of this cluster lie map, firstly, the addition/elimination- and, secondly, the Berry coordinate, the latter in contrast to the fSQP. As was the case for the TBP, this coordinate is dominated by changes in the angles, changes in bond lengths being negligible. The least important factor describes the flexibility in the apical ligand which is reduced, though, over what it was in the fSQP.

The addition/elimination coordinate appears to differ slightly from that for the fSQP, in that changes in bond distances and the pyramidalisation of the ML_5 fragment are almost equally important. This suggests that the addition of a fifth ligand to the ML_4 moeity may be accompanied by increasing elevation of the SQP, in contrast to the results obtained for the fSQP, which indicated a decreased dependency between these two components. Consequently, the approach of the fifth ligand to the ML_4 fragment seems initially to bring about only a small elevation of the SQP, which becomes increasingly enhanced, however, as the proximity of the apical ligand increases. As mentioned earlier the tendency to adopt this conformation is greatest for rhodium and iridium, and much lower for palladium and platinum.

5.2 Reaction Coordinates of the ML_5 Fragment

Five-coordinate intermediates and/or transition states have been postulated and demonstrated for many ligand exchange reactions of square-planar molecules. They have historically been divided into three main groups — nucleophilic substitutions, electrophilic substitutions, and oxidative additions followed by reductive eliminations. It is becoming increasingly clear, however, that a relationship exists between many types of reaction mechanisms previously classified quite separately. [1–3] Often this relationship results from *geometrically similar* reaction pathways involving the formation of a five-coordinate species, be it a true intermediate in an associative nucleophilic substitution, an early transition state in an oxidative addition or a solvento species in a dissociative electrophilic substitution, for example. In this sense, then, many, if not most, reactions of four-coordinate complexes can be deemed at some stage to involve the formation of a five-coordinate species via an

essentially associative mechanism:

$$XML_3 + Y \rightleftharpoons [XML_3Y] \rightleftharpoons X + ML_3Y$$

where X and L represent any coordinated ligand atoms, and Y is either X, L or solvent. There is general agreement that this reaction involves attack (nucleophilic or electrophilic) at the metal by the incoming "ligand" Y, with the five coordinate adduct adopting TBP and SQP conformations at various stages along the reaction pathway. In most cases, though, geometrical details concerning conformational changes along the reaction have not been forthcoming. Cross [1-3] has reviewed the published work in this area with particular attention to the geometrical implications of the studied reactions, [2] but most of his conclusions have had to remain speculative in this regard, due to the classic difficulty of direct observation of the reacting species.

In his review Cross sketches the results of mechanistic and kinetic studies on nucleophilic substitutions of, mainly, square planar platinum and palladium complexes, on the one hand, and oxidative additions to the compounds of iridium and rhodium, on the other. He points out that this apparent bias towards the two types of reactions for the different metals reflects the nature of the available data, though to some extent it possibly also reflects inherent differences between the two groups of metals. We shall outline the broad conclusions drawn by Cross, adding to these the data we have earlier compiled on the dynamic stereochemistry of five-coordinate nickel, [4] as well as relevant studies published since his review appeared.

Figure 5.2.1 outlines the essential reaction pathways followed by the pentacoordinate intermediate or transition state $[XML_3Y]$, as suggested by the mechanistic and kinetic data documented both by Cross for rhodium, iridium, platinum and palladium, and by us for nickel. [4] The addition of an incoming ligand is proposed to take place along the perpendicular to the ML_4 fragment, giving rise to a square pyramidal species, though a distinction is never made between a fSQP and an eSQP. It is further suggested that this species often distorts along the C_{2v} Berry coordinate into a TBP, which may, in turn, distort towards yet another, isometric SQP. In the case of nickel compounds the dissociation of the TBP intermediate *via* an S_N2 coordinate has sometimes been indicated. This pathway has

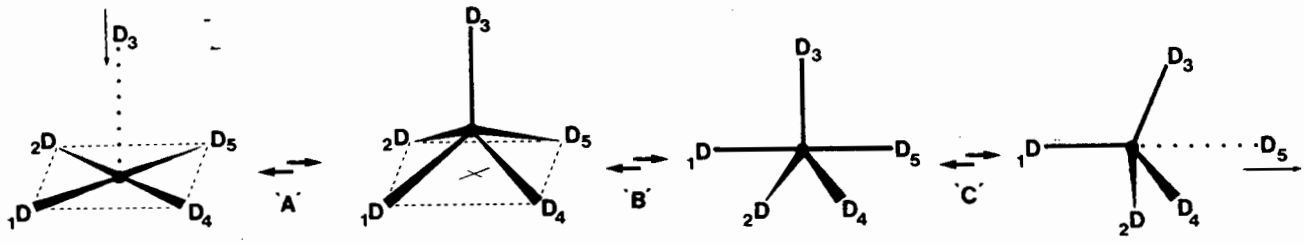


Figure 5.2.1 Diagram of essential reaction pathways of a five-coordinate fragment ML_5 . 'A' maps the reversible addition of a fifth ligand to a square planar centre, and the progression from a fSQP to a eSQP, 'B' maps the SQP → TBP transformation according to the Berry coordinate, 'C' maps a reversible S_N2 reaction at a tetrahedral centre.

---

also been invoked in a few examples of nucleophilic substitutions and isomerisations of tetrahedral nickel complexes. [4]

Until recently it remained uncertain whether the five-coordinate species possesses sufficient stability to be regarded as an actual intermediate, rather than merely as a phase of the activated complex. However, on the basis of a mass spectrometric study of the addition of chloride ion to $[PtCl_2(PEt_3)_2]$, Turco *et al* [5] have suggested the formation of a true five-coordinate adduct with a lifetime greater than 10^{-5} s. They leave open the question, though, of whether the intermediate they have found is in fact distinct from the transition state for the addition reaction, or not.

Often more concrete evidence for the formation of five-coordinate intermediates appears to be inferrable from studies of association reactions or ligand exchange at four-coordinate complexes. The addition of cyanide ion to $[PtL_2]^{2+}$ (L = 1,10-phenanthroline or 2,2'-bipyridine), for example, gives rise to an isolable five-coordinate species $[PtL_2(CN)]^+$, which is also indicated as the intermediate in the reaction leading to $[PtL(CN)_2]$. Using these two examples, Wernberg [6,7] has also shown that the rate of ligand substitution leading to the product $[PtL(Nu)_2]$ for various nucleophiles (Nu) is faster in the case of the bipyridyl complexes than in that of the phenanthroline compounds. He has rationalised this in terms of the bipyridyl ligand being more flexible and consequently a better leaving group than

phenanthroline. Quite likely, though, it is the lower rigidity of the bipyridine *in combination* with the nature of the 5-coordinate complex which results in the enhanced kinetics; he has suggested that the phenanthroline-containing intermediate exhibits intramolecular exchange in accordance with the Berry mechanism. [8] These examples highlight the crucial role which the five-coordinate adduct often plays as a result of its particular geometric features.

The character of the intermediate has also been put forward as a factor in explaining kinetic data obtained for olefin exchange in a series of *trans*-dichloro (η -olefin)(pyridine)platinum(II) complexes. Chottard [9] *et al* describe how this exchange is catalysed by the transient chelation of a hydroxy group on the pyridine resulting in the formation of a five-coordinate species.

Stereochemical non-rigidity and, in particular, intramolecular exchange via principally the Berry mechanism, has been documented in a number of cases for nickel, [4] although evidence for this reaction in complexes of the other metals is scanty. [1–3,10] Where it has been suggested it is often done simply on the basis of temperature dependant nmr studies and not on the basis of kinetic studies. Yamazaki, for example, suggests a fluxional palladium complex of a substituted phenanthroline on the basis of variable-temperature nmr experiments and describes its dynamic motion as an intramolecular rearrangement akin to the Berry mechanism. [11] A similar facile rearrangement between TBP and SQP conformers of a dicarbonylbis(triphenylphosphine) complex of rhodium has been suggested, again on the basis of nmr studies. [12] Although in many cases the Berry mechanism has been invoked without sufficient experimental support, there nevertheless are examples in which it seems highly plausible. [2,4,10]

Finally, some tetrahedral nickel complexes have been demonstrated to undergo nucleophilic substitution or ligand exchange reactions involving the simultaneous attack and departure by two ligands in the same way as the classic S_N2 reaction at tetrahedral carbon has been conceptualised.

5.3 Conclusion

Comparing the results of the statistical analysis with the mechanistic data which have been collected for the ML_5 fragment, it is difficult to avoid the con-

clusion that the static deformations manifested by the five-coordinate molecular fragments constituting our data set do indeed mirror those proposed to occur in solution for this fragment. In other words, the completely objective mathematical techniques applied to the data have yielded correlations between independent geometrical parameters (the symmetry coordinates) which map molecular distortions identical to those proposed to occur along certain reaction pathways for the ML_5 molecular fragment. These results, then, lend support to the structure correlation hypothesis: 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. [13]

In some instances this analysis has revealed trends which are perhaps implicit in available mechanistic data on the solution behaviour of ML_5 , although they have yet to be explicitly stated. For example, the metals studied here appear to fall into three categories. The first, containing platinum and palladium, hardly adopts TBP conformation and prefers the fSQP over the eSQP. In contrast, the second grouping consisting of rhodium and iridium adopts TBP conformation in roughly one fifth of cases, and prefers the eSQP, while the third, constituting nickel, is intermediate between the two, having a slight preference for the eSQP, over the TBP and fSQP. The difference between the chemistry of nickel and that of the other metals is already evident in the fact that its four-coordinate complexes can adopt either tetrahedral or square-planar conformation, while the latter only is exhibited by the others. This obviously has implications for the range of reaction pathways available to five-coordinate nickel, as the mapping of the S_N2 pathway by, predominantly, the nickel complexes shows. However, the presence — albeit it a reduced one — of this distortion coordinate in the data even *after* the removal of the nickel compounds indicates that this pathway might also be available to rhodium and iridium (the other constituents of the TBP cluster), even though this may not be evident from the chemistry of their four-coordination.

Platinum and palladium complexes have been shown to preferentially adopt the fSQP conformation, which does not manifest the Berry distortion. As a result one would expect only a few instances of intramolecular exchange via this mechanism for

five-coordinate complexes of these metals. The scarcity of experimental data which unambiguously indicate this type of fluxionality for platinum and palladium may, consequently, indicate that other mechanisms are operable in most intramolecular exchange reactions of these metals. Rhodium and iridium, on the other hand, have a very clear preference for the eSQP which does exhibit Berry distortions. This would suggest that the Berry coordinate represents a viable option for intramolecular exchange in these metals, in contrast to platinum and palladium.

The apparent predominance of nucleophilic substitution at palladium and platinum, as compared to that of oxidative addition at iridium and rhodium (indicated by Cross [2]), might be a reflection of inherent differences between these two groups of metals in much the same way as their *structural* differences might be. A thorough comparative study could perhaps shed some light on this issue.

The geometric options available to the ML_5 fragment, as revealed by this analysis, may be useful in the interpretation of hitherto unclear results from mechanistic or kinetic studies, or may offer deeper insights into existing explanations. For example, Dunitz and Meyer have explained the apparent reluctance of nucleophiles such as chloride ion, or O atoms of H_2O or CH_3OH , to approach square coplanar nickel in highly constrained ligand systems, such as macrocycles, on the basis of the likely demands which this proximity would make on the ligand to metal bonds as a consequence of shifts in the d electron configuration. [14] They argue that these ligands would be unable to accomodate a lengthening of the bonds which the change in electron configuration would induce. Our previous study on nickel, [4] in which each compound in the data set was explicitly investigated, revealed that the apical bond in compounds with such ligands was indeed much longer than average, but that they nevertheless formed formally penta-coordinate complexes. We have here shown that the *initial* approach of the fifth (apical) ligand towards the square-planar ML_4 is accompanied by only a small amount of pyramidalisation, but that as its proximity to the metal increases, a threshold seems to be reached, beyond which the elevation rises quickly. Macrocycles, and similar constrained ligands, will allow only a small elevation of the central metal atom, due to their reduced flexibility, thereby limiting the approach of the apical ligand. It would appear, therefore, as if the *angular* rigidity of the ligand is the constraining factor. This may seem a

trivial difference to the original explanation offered by Dunitz and Meyer, but it could conceivably be an important one in the design of metal-specific ligands, or homogenous catalysts, for example.

The details of the geometric reaction pathways available to ML_5 might also be put to use in the refinement of existing force fields. The cluster archetypes and the deformation coordinates represent *averaged* data in which the perturbing influences of particular crystalline environments have been averaged out. They may consequently be regarded, at least, as local minima and their geometric parameters can be used to refine force fields accordingly.

In our experience the techniques we have applied in this study have been objective and very powerful in extracting information from the data set. There are two observations, however, which need to be made. The first concerns the automated nature of the analysis — this represents an obvious advantage in that results can be objectively and speedily obtained. It also carries with it, though, the danger of alienating the analyst from the data since the automatic routines and algorithms remove the necessity of "hands-on" work; the often invaluable intimate contact with the data can thereby be lost. The second point relates to the nature of the information derived from the analysis. A data set such as ours contains a veritable treasure trove of information, while the powerful techniques offer us tools with which to excavate the hoard. Obviously, amongst much useful information there will be masses of trivia, and inevitably the question arises: where does one stop? The answer to this question is not at all simple. The possibility of refining the techniques so as to obtain more and more information, coupled to that of uncovering new secrets (however trivial they may be), represents a strong incentive to continuing the analysis almost *ad infinitum* (and *ad absurdum*).

In our case we have identified one further area which might be profitably investigated. The question is: what is the relation, if any, between the *type* of ligand atom and the *site* it occupies in the archetypal conformations. The aim would be to examine how the classification of Lewis acids and bases as hard or soft relates to concepts such as apicophilicity,* and in which way these concepts influence the

* To be specific, the term apicophilicity would refer to the tendency of a ligand to seek the apical site in the SQP, while axiophilicity would relate to a preference for the axial site of the

distortion coordinates mapped. This analysis would necessitate the development and deployment of algorithms capable of handling categorical variables as opposed to continuous variables.

McDowell and Streitwieser [15] have performed a theoretical study of model substituted phosphoranes $[PH_4X]$, deriving from this an order of ligand axiophilicity.† It would be interesting to see how their results for phosphorus correlate with the trends manifested in this data set. The influence of $\sigma-$donor capability of the ligand atoms on the dynamic stereochemistry of five-coordinate nickel has been examined by Tatsumi *et al* [16] using molecular orbital calculations on $[Ni(CH_3)_2A_3]$, where A is a weaker $\sigma-$donor than CH_3, eg. PR_3. The study reveals high potential energies in TBPs with strong donors in the equatorial plane, and minimum energies when they occupy both axial sites. For the SQP a *cis*-basal configuration of the strong $\sigma-$donor ligands leads to the most stable conformation, the *trans*-basal one being slightly higher in energy, while a $\sigma-$donor in the apical position leads to the largest potential energies. It would be very interesting indeed to examine the ligand configurations for complexes approximating the cluster archetypes, which presumably lie close to, or at local potential minima, and for complexes lying along the distortion coordinates mapped, which are likely to lie along "energy valleys", i.e. reaction pathways, in the potential energy hypersurface for ML_5.

TBP.

† They refer to it as apicophilicity, however, in spite of the fact that they are dealing with trigonal bipyramidal phosphorus.

REFERENCES

1. Anderson, G.K., Cross R.J. *Acc. Chem. Res.* 1984, **17**, 67-74.
2. Cross, R.J. *Chem. Soc. Rev.* 1985, **14**, 197-223.
3. R.J. Cross. *Mechanisms of Inorganic and Organometallic Reactions, Volume 2*, Chapter 5, Plenum Press, New York, 1984.
4. Auf der Heyde, T.P.E., Nassimbeni, L.R. *Inorg. Chem.* 1984, **23**, 4525-4532.
5. Turco, A., Morvillo, A., Vettori, U., Traldi, P. *Inorg. Chem.* 1985, **24**, 1123-1125.
6. Wernberg, O. *Acta Chem. Scand.* 1985, **A39**, 223-225.
7. Wernberg, O. *J. Chem. Soc. Dalton Trans.* 1986, 725-728.
8. Wernberg, O., Hazell, A. *J. Chem. Soc. Dalton Trans.* 1980, 973-978.
9. Guillot-Edelheit, G., Chottard, J.-C. *J. Chem. Soc. Dalton Trans.* 1984, 169-173.
10. A.J. Deeming. *Mechanisms of Inorganic and Organometallic Reactions, Volume 2,* Chapter 13, Plenum Press, New York, 1984.
11. Yamazaki, S. *Bull. Chem. Soc. Jpn.* 1987, **60**, 1155-1156.
12. Sanger, A.R. *Can. J. Chem.* 1985, **63**, 571-575.

13.(a) Bürgi, H.B. *Inorg. Chem.* 1973, **12**, 2321-2325.

13.(b) Murray-Rust, P., Bürgi, H.B., Dunitz, J.D. *J. Am. Chem. Soc.* 1975, **97**, 921-922.

14. Dunitz, J.D., Meyer, E.F. *Helv. Chim. Acta* 1971, **54**, 77-89.
15. McDowell, R.S., Streitwieser, Jr., A. *J. Am. Chem. Soc.* 1985, **107**, 5849-5855.
16. Tatsumi, K., Nakamura, A., Komiya, S., Yamamoto, A., Yamamoto, T. *J. Am. Chem. Soc.* 1984, **106**, 8181-8188.