Contribution from the Department of Chemistry, University of the Western Cape, Bellville 7530, South Africa, and Laboratorium für Chemische und Mineralogische Kristallographie, Universität Bern, Freiestrasse 3, CH-3012 Bern, Switzerland

Molecular Geometry of d^8 Five-Coordination. 2. Cluster Analysis, Archetypal Geometries, and Cluster Statistics

Thomas P. E. Auf der Heyde*,† and Hans-Beat Bürgi‡

Received April 5, 1989

The geometries of 196 metal complexes containing ML_5 molecular fragments have been represented by 2352 and 1568 points, respectively, in two 12-dimensional spaces spanned by the symmetry coordinates for a trigonal bipyramid (TBP; T-space) and a square pyramid (SQP; S-space). This paper, the second of a three-part series, employs the multivariate statistical technique of cluster analysis in order to probe the distribution of these points in the 12-dimensional data spaces. Two techniques have been employed: nonhierarchical *K*-means clustering and hierarchical clustering employing Ward's criterion. *K*-means clustering shows T-space to be characterized by four clusters: one whose average, or archetype, is a TBP of D_{3h} symmetry, and three isometric clusters characterized by a slightly distorted "SQP" of C_{2v} symmetry with trans-basal angles of 161 and 169°, respectively. S-Space is similarly characterized by four clusters: one with a "flattened" SQP (fSQP) archetype of C_{4v} symmetry with trans-basal angles of 171° and an apical bond 0.7 Å in excess of the standard M-L bond length, one characterized by an "elevated" SQP (eSQP) also of C_{4v} symmetry with trans-basal angles of 163° and whose apical bond is only 0.03 Å in excess of its standard length, and two isometric clusters characterized by slightly distorted TBPs of C_{2v} symmetry. It is shown that the SQP cluster in T-space is divided into the fSQP and eSQP clusters in S-space and that, consequently, the results from the two data spaces are identical. In all cases, the results are shown to be logically consistent with well-established chemical principles for these conformations. Hierarchical clustering, although not applied in as much detail as *K*-means clustering, yields essentially similar results. Structural characteristics for the archetypal conformations (TBP, fSQP, eSQP) are presented and analyzed in the light of previous theoretical and less comprehensive empirical studies and are shown to be largely consistent with these. Univariate statistics for the TBP cluster suggest there is a large variance in the axial bonds and in the axial-equatorial angles, i.e. the pyramidality of the TBP—the "umbrella" coordinate. The variance in the position of the apical ligand appears much greater for the fSQP than the eSQP. Bivariate statistics point to the principal importance of distortions along an S_N2 coordinate for the TBP, with the Berry coordinate being slightly less clearly mapped. For both the fSQP and the eSQP the "glue" coordinate—shortening of some bond distances at the expense of others—appears most significant, while only the eSQP maps the Berry coordinate; the fSQP fails to exhibit the distortions traditionally associated with this coordinate.

Introduction

In the previous paper of this series,[1] 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 has been described by two sets of 12 nonredundant symmetry coordinates. These sets refer to the two most common idealized 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 12-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 idealized TBP or to an idealized SQP. We have named these T-space and S-space, respectively. The origin of T-space represents an idealized ML_5 TBP with D_{3h} symmetry whose M-L bond lengths equal their "idealized single-bond" values. The conceptualization of S-space follows similar lines, with the exception that an idealized 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 12-dimensional space. This line reflects the one degree of freedom that the C_{4v} symmetry allows the values of the trans-basal angles in a SQP.

In order to represent the symmetries of the deformation spaces, each observed conformation was transformed into its isometric partners by the application of the 12 and 8 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. This artifact may be profitably exploited in the multivariate analysis that 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.

Univariate and bivariate statistics for the data distribution in T-space indicate the importance of distortions mimicking the S_N2 coordinate, the Berry coordinate, and a "constant-amount-of-glue" coordinate when the individual molecular geometries are regarded as distorted TBP. The "glue" coordinate is one which suggests that a constant amount of bonding is associated with the metal, as a result of which a lengthening of one bond brings about a shortening of others and vice versa. On the other hand, when the structures are considered as distorted SQP, the distribution in S-space appears to mimic distortions reminiscent of a reversible association coordinate at a square-planar center.

The aim of the ensuing multidimensional analysis is essentially 2-fold: (i) The *distribution* of the representative points in 12-D space will be probed by 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 by 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. In this paper, we will report the results of cluster analysis of the data distribution in T- and S-spaces, showing their implications for the "average" or "typical" molecular geometries of five-coordinate d^8 metal complexes.

In a previous paper,[2] we have attempted to condense various publications dealing with factor and cluster analysis, illustrating their essential characteristics by a simplified outline of their philosophical and mathematical bases and by applying them to a simple, hypothetical three-dimensional database. We will therefore dispense with a detailed description of cluster analysis, except to point out that it is a technique which reveals the presence of clusters or clouds of data points, by grouping together similar points in the data space.[3]

* To whom correspondence should be addressed.
† University of the Western Cape.
‡ Universität Bern.

(1) See: Auf der Heyde, T. P. E.; Bürgi, H.-B. *Inorg. Chem.* preceding paper in this issue.
(2) Auf der Heyde, T. P. E. Submitted for publication in *J. Chem. Educ.*
(3) Massart, D. L.; Kaufman, L. *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*; John Wiley and Sons: New York, 1983.

0020-1669/89/1328-3970$01.50/0

Experimental Section and Results

1. Choice of Clustering Techniques. As pointed out elsewhere,[2,3] 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, first, nonhierarchical or relocation clustering employing the K-means method and, second, 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 (where much of this work was done) three statistical analysis packages offering clustering algorithms are available: CLUSTAN,[4] SPSS,[5] BMDP.[6] Of these, the latter seemed the most versatile and user friendly, offering four different cluster analysis programs, three of which are hierarchical. 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 these programs, 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 1600 pages in the case of the dendrogram or circa 350 pages in that of the matrix.

BMDP program PKM, on the other hand, provides line printer plotting techniques that enable one to display the analysis results visually in a number of different ways. More importantly, though, program PKM employs nonhierarchical 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. 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.

2. Nonhierarchical (Relocation) Cluster Analysis. (a) Similarity Measure and Robustness Criteria. In our case we decided to use unstandardized data with angle measurements scaled to radial displacements on the order of 1 Å, as outlined in part 1 of this study.[1] *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 clustering. The first criterion derives from the symmetry introduced into the expanded data set, and its basic assumption is that whatever symmetry is present in the data set must be reflected in the results of the cluster analysis, whose algorithm does not take explicit account of symmetry. Second, *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, e.g., TBPs in T-space ought to be similarly defined (and therefore clustered together) in S-space; otherwise, the classification must have been an artifact due to the clustering algorithm or the reference structure, rather than an inherently rigorous classification.

(b) Initial Attempts. Since we had no a priori concept of the likely clustering pattern that 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 that might be formed. It was therefore necessary to begin with a hit-or-miss method, and first attempts centered around fitting a 12-cluster model to the data in T-space and an 8-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 12- and 8-fold, respectively.

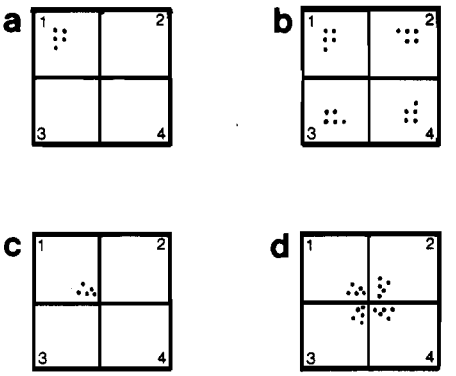

Figure 1. Diagram showing schematically the relationship between data distribution *prior* to expansion by the symmetry elements of the point group (a and c) and the number of clusters finally existing in the entirety of the parameter space (b and d). The point group used is C_4 and the numbers 1, 2,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

Table I. Inter- and Intracluster Statistics for T-Space[a]

cluster		T1	T2	T3	T4
no. of members[b]		548	548	708	548
av size		0.885	0.885	0.727	0.885
dist matrix		T1	T2	T3	
	T2	3.457			
	T3	2.002	2.002		
	T4	3.457	3.457	2.002	

[a] Average size = average distance of cluster members from center of cluster (Å). Distance matrix gives intercluster distances (Å). [b] Total number of members = 2352.

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 realized 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 center of one asymmetric unit of a given multidimensional space, as shown in Figure 1a. Then, when these points are subjected to the type of data expansion that we have applied in our case, they would be transformed into their isometric partners in the adjacent asymmetric units, as shown in Figure 1b. 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 1c, 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 1d.

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 $K = 10$, and in the process to employ our two robustness criteria outlined earlier. Initially, the results were plotted by 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 2 emerged for the T-space data set with $K = 4$. Its rather obvious 3-fold symmetry immediately drew our attention.

(c) T-Space. When we examined the result obtained for $K = 4$ more closely, it became clear that a highly symmetric clustering pattern had emerged, which in fact mirrored the 3-fold symmetry

(4) Wishart, D. *CLUSTAN User's Guide, The Clustan Project*; University College: London, 1975.
(5) Nie, N. H.; Hull, C. H.; Jenkins, J. G.; Steinbrenner, K.; Bent, D. H. *Statistical Package for the Social Sciences*, 2nd ed.; McGraw-Hill: New York, London, 1975.
(6) Dixon, W. J., Ed. *BMDP Statistical Software*, 1983 Printing with Additions; University of California Press: Berkeley, CA, 1983.

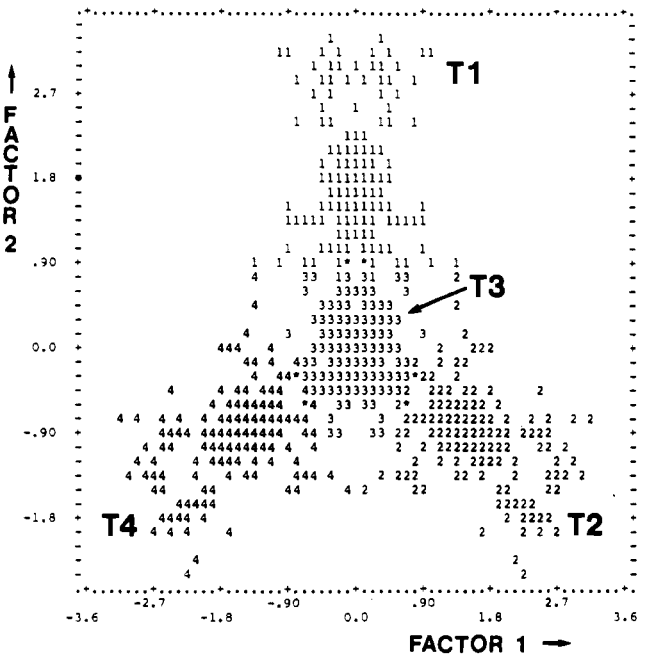

Figure 2. Projection of T-space clusters onto the plane defined by the first two factors. Numbers indicate cluster affiliation, and asterisks indicate superposition of points belonging to different clusters. T1, T2, T3, and T4 are the notation used for the various clusters throughout this study. The 3-fold symmetry is slightly distorted by the line plotter.

Table II. Internal Angles (deg) of Cluster Centrotypes in T-Space[8],[a]

angle	cluster			
	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)	90 (7)	89 (6)
θ_{45}	89 (6)	89 (6)	90 (7)	94 (7)
M	548	548	708	548
N	137	137	59	137

[a] Absolute conformations of T1, T2, and T4 are identical, but the ligand sites are permuted. Estimated standard deviations are given in parentheses. M is the number of cluster members; N is the number of refcodes in the cluster. See Figure 3 for diagrams depicting centrotype conformations and corresponding ligand permutations.

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 center of the triangle.[7] These clusters will from now on be referred to as T1, T2, T4, and T3, respectively. Table I gives some inter- and intracluster statistics, while Figure 2 shows a projection of the four clusters onto the plane defined by the first two factors extracted from the T-space data set.

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 center of the cluster, around which the observed molecular geometries aggregate. The program PKM allows the user to attach a label indicating cluster affiliation to the end of each observation's record. This enables one to subsequently analyze 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 II.

(7) This 3-dimensional description of the arrangement of the clusters relative to each other is meant merely to enable a visualization of the result, since such a description can obviously be of only very limited use in describing a 12-dimensional picture.

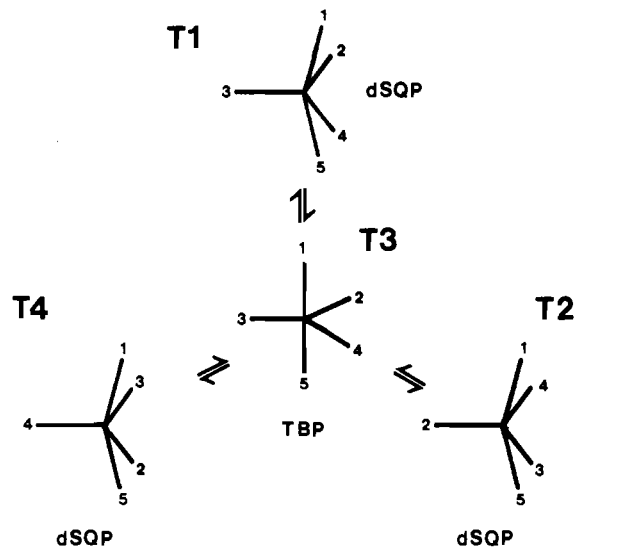

Figure 3. Interpretation of Figure 2, showing three possible distortions of the central TBP (corresponding to the T3 centrotype) into three dSQPs (corresponding to the T1, T2, and T4 centrotypes).

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 TBP of D_{3h} symmetry.[8] Moreover, the three SQP conformers coincide with the three possible SQPs that may be formed via the Berry mechanism from the TBP corresponding to the T3 centrotype, as shown in Figure 3. According to this scheme, starting from the TBP in the center, 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 that are more trigonal bipyramidal in nature and those that tend more toward 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 considering the possible distortions of the archetypal TBP begins to tempt 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.[9] In our case, the square-pyramidal conformer formed has preserved only the C_{2v} symmetry of the distortion coordinate and the D_{3h} TBP, but has 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 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 is meaningful lies in how well this would

(8) Naive averaging of the internal angles by the algorithm does not take into account the geometrical interdependence of the angles, and leads to average values that are geometrically inconsistent. An obvious example is T3, whose angles ought to be 90, 120, and 180°, but where all three equatorial angles are 119°, even though all equatorial–axial angles are 90°.

(9) Of course, the presence of precisely these three isomeric SQPs results from the data expansion, but this fact does not detract from the argument developed subsequently.

Table III. Inter- and Intracluster Statistics for S-Space[a]

cluster		S1	S2	S3	S4
no. of members[b]		276	312	276	704
av size		0.697	0.932	0.697	0.719
dist matrix		S1	S2	S3	
	S2	1.836			
	S3	2.531	1.836		
	S4	1.441	0.833	1.441	

[a] Average size = average distance of cluster members from center of cluster (Å). Distance matrix gives intercluster distances (Å). [b] Total number of members = 1568.

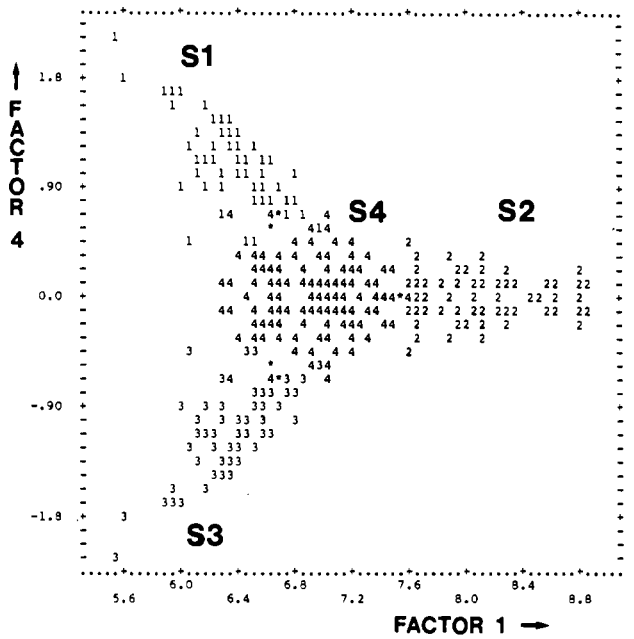

Figure 4. Projection of S-space clusters onto the plane defined by the first and fourth factors. Numbers indicate cluster affiliation, and asterisks indicate superposition of points belonging to different clusters. S1, S2, S3, and S4 are the notation used for the respective clusters throughout this study.

compare with the results of a similar analysis in S-space. 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 bipronged fork, and two unequal clusters (numbers four and two) placed at the center and the bottom tip of the Y, respectively.[7] 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 centers is of the same order as the average size of the cluster. The reason for this will become clear in due course. Table III shows some relevant inter- and intracluster statistics.

The Y-shaped arrangement of the four clusters is best[10] 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. This particular projection shows up only the 2-fold symmetry of the data space, and not the expected 4-fold one. The reason for this is that the 4-fold symmetry is associated with the subspace made up of the second and third factor 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 III.

Table IV gives the internal angles characterizing the various cluster centrotypes of S1, S2, S3, and S4. Although we do not intend at this stage to comment on differences between the bond

(10) 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 sense of the 12-dimensional picture.

Table IV. Internal Aangles (deg) of Cluster Centrotypes in S-Space[a]

	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

[a] Absolute conformations of S1 and S3 are identical, but the ligands are interchanged. Estimated standard deviations are given in parentheses. M is the number of cluster members; N is the number of refcodes in the cluster. See Figure 5 for diagrams depicting centrotype conformations and corresponding ligand permutations.

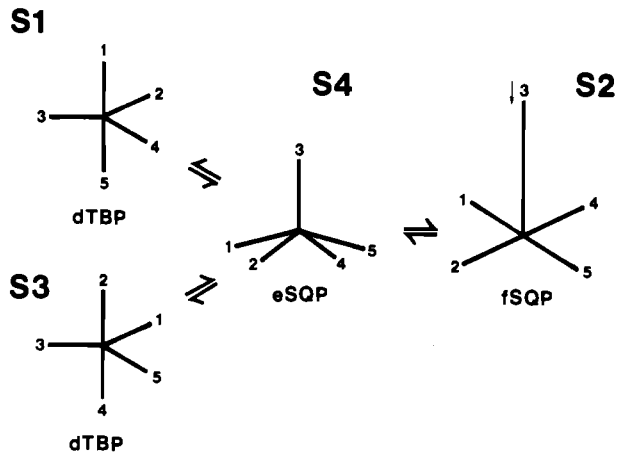

Figure 5. Interpretation of Figure 4, showing possible distortions of the fSQP (corresponding to the S2 centrotype) through an eSQP (S4 centrotype) to either of two dTBPs (S1 and S3 centrotypes).

lengths of the centrotypes, it is important to point out that there is a substantial difference in the apical bond distance between the SQPs characteristic of S2 and S4. (We have attempted to reflect this in the interpretive diagrams of Figure 5.) From Table IV, it can readily be seen that clusters S1 and S3 represent slightly distorted TBPs (dTBP) with C_{2v} symmetry that are marginally displaced along the Berry coordinate toward 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 characterized 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 characterized 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 that is logically consistent with well-established chemical principles emerges when these cluster centrotypes are viewed in the correct sequence, as shown in Figure 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 center. From here, there is a gradual progression to the eSQP (S4 centrotype) as the apical ligand moves closer toward the metal. Finally, the "true" eSQP intermediate has the option of distorting into either of two TBPs, depending on which of the trans-basal angles (θ_{15} or θ_{24}) opens up toward 180°.

Similar to the case with the C_{2v} SQP centrotypes in T-space, the TBP centrotypes of S1 and S3 do not attain the expected D_{3h}

symmetry. Again, this may well be the result of attempting to fit observed conformations approaching D_{3h} symmetry into a C_{4v} framework. The highest symmetry that the two would 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-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 nonidentical 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 data set into two groups—those tending toward a TBP conformation and those tending toward 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, geometries which tend toward 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 that 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, 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. 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 apical 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,[11] we found a smooth transition from the one to the other. We believed not only that this was the result of an absence of criteria differentiating between the fSQP and the eSQP but also that it mirrored the (most likely) rather small energetic differences between the two, leading to an even spread of observed geometries between the 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 of this form of analysis. To enlarge on this point, let us view the observed structures as indeed representing various points on the Born–Oppenheimer energy hypersurface of the square-pyramidal ML_5 fragment.[12] The two points representing the archetypal fSQP and eSQP are then very likely 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 center of the cloud of data points surrounding them.

Table V. Analysis of Cluster Membership of Trigonal-Bipyramidal and Square-Pyramidal Clusters in T-Space and S-Space[a]

T-space	S-space
T3 (TBP) (59)	S1 (dTBP) (69)
	S3 (dTBP)
T1 (dSQP) (137)	
T2 (dSQP)	S2 (fSQP) (39)
T4 (dSQP)	S4 (eSQP) (88)

[a] The only structures that are defined differently in the two data spaces are those indicated by the asterisk: CMBPNI, CMPMNI, CSMRHC, DPPCRH, DPMCRH10, EDCRCN, IPESNI, SALDNI, TMAGEP, and BEZPIF. Numbers in parentheses indicate number of refcodes for each cluster.

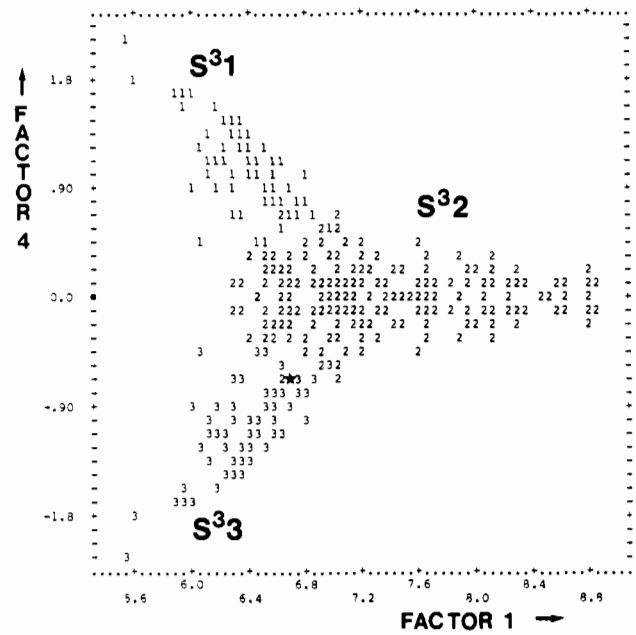

Figure 6. Projection of S-space clusters for $K = 3$ onto the plane defined by factor 1 and factor 4 extracted from S-space for $K = 4$. Numbers indicate cluster affiliation, and the asterisk indicates the superposition of points belonging to different clusters.

Let us continue now to comparing the actual cluster membership in the two data spaces. This is outlined schematically in Table V. As can be seen all CSD reference codes (refcodes)[13] in T3, representing a TBP in T-space, are found again in the identical clusters S1 and S3, representing the TBP in S-space. Similarly, almost all refcodes of the identical clusters T_1, T_2, and T_4 representing the SQP in T-space, are found in either S2 or S4, both representing the SQP in S-space. With only 10 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.

(f) T-Space versus S-Space Continued. Although, as we have seen, the algorithm sorted the square pyramidally disposed 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 that 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 11 iterations are needed to achieve an optimum clustering in S-space when $K =$

(11) Auf der Heyde, T. P. E.; Nassimbeni, L. R. *Inorg. Chem.* **1984**, *23*, 4525–4532.
(12) See the introduction to part 1 of this study.[1]
(13) CSD = Cambridge Structural Database. See ref 1.

Table VI. Internal Angles (deg) of Cluster Centrotypes for $K = 3$ in S-Space[a]

	S^31	S^32	S^33
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

[a] Estimated standard deviations are given in parentheses. M is the number of cluster members; N is the number of refcodes in the cluster.

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 similar, but not quite identical, clusters with 288 and 284 members, respectively. Figure 6 shows a scatter plot of this clustering pattern projected onto the same plane as that in Figure 4, with which it should be compared. Table VI gives the internal angles corresponding to the centrotypes of the three clusters, which we will refer to as S^31, S^32, and S^33.

The very slight dissymmetry of this clustering is evidenced by the unequal number of cluster members for S^31 and S^33. These clusters are represented by distorted TBP centrotypes of C_{2v} symmetry that are almost isometric partners to each other, as are those of S1 and S3 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 IV and VI 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 being taken over into that of the latter except for only three compounds that are lost to S^32 (SQP). The S^32 centrotype corresponds to an eSQP with almost perfect C_{4v} symmetry.[14] Its geometry can be seen to be intermediate between that of S2 (fSQP) and S4 (eSQP) for $K = 4$, as the comparison of Tables IV and VI will also show. Clearly, the reduction in the number of iterations from 21 (for $K = 4$) to 11 (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 identified by the refcode MOPAON[15] in S^31 (dTBP) and its exclusion, on the other hand, 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,[11] contains two six-membered metal–ligand rings that 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 and 0.960 Å the representative point for MOPAON is almost equidistant from the centers of S^31 (size = 0.713 Å) and S^32 (size = 0.851 Å), respectively.

Nørskov-Lauritsen and Bürgi[16] have examined how different hierarchical clustering algorithms deal with a highly symmetric

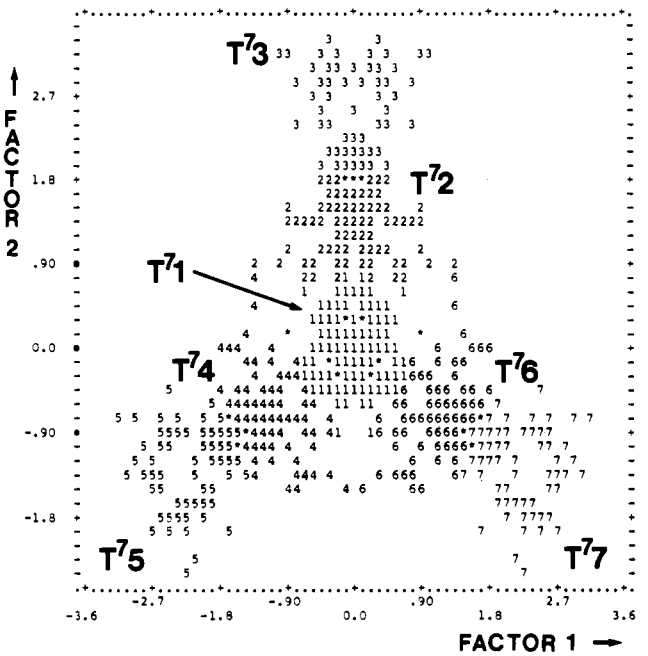

Figure 7. Projection of T-space clusters for $K = 7$ onto the plane defined by factor 1 and factor 2 extracted from T-space for $K = 4$. Numbers indicate cluster affiliation, and asterisks indicate the superposition of points belonging to different clusters.

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, first, to the subdivision of the data space into K sections and, second, 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 program is first (in some nonhierarchical sense of the word) to have its approximate center fixed and be allocated its members, followed by that named two and so on. In this case, MOPAON is classified together with cluster one and then with cluster two, but not with cluster three. Short of analyzing the PKM flow chart, we can only speculate that MOPAON's close proximity to both TBP and SQP coupled to the order in which the PKM algorithm allocates cluster centers and members gives rise to the dissymmetry that emerges for $K = 3$.[17] MOPAON is represented by the asterisk in Figure 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 scatter plot of the cluster pattern onto the factor 1/factor 2 plane, as in Figure 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 centers and then let PKM optimize them. As "seeds"

(14) Although Table VI seems to indicate perfect C_{4v} symmetry for S^32, it does so only because the θ_{ij}'s and esd's are rounded off. In fact the symmetry is only very nearly C_{4v}.

(15) Orioli, P. L., Di Vaira, M. *J. Chem. Soc. A* **1968**, 2078–2083.
(16) Nørskov-Lauritsen, L.; Bürgi, H.-B. *J. Comput. Chem.* **1985**, *6*, 216–228.

(17) Willet 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. See: Willet, P. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 29–32.

Table VII. Internal Angles (deg) and Refined Cluster Centrotypes for $K = 7$ in T-Space[a]

	$T^7$1	$T^7$2 (≈$T^7$4, $T^7$6)	$T^7$3 (≈$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

[a] Estimated standard deviations are given in parentheses. Angles for $T^7$4 and $T^7$6, and $T^7$5 and $T^7$7, may be obtained by permuting $T^7$2 and $T^7$3, respectively, according to T2 and T4 in Figure 3. M is the number of cluster members; N is the number of refcodes in the cluster.

we chose (i) the D_{3h} TBP centrotype of T3 for cluster one, (ii) the C_{4v} SQP centrotype of S4 (eSQP) for clusters two, four, and six, with the angles permuted so as to correspond to the three isometric SQPs shown in Figure 3, and (iii) the C_{4v} centrotype of S2 (fSQP) for clusters three, five, and seven, also permuted according to Figure 3. Input simply consisted of the values of the 12 symmetry coordinates for the D_{3h} TBP evaluated for the various conformations listed in parts i–iii above.

After just 10 iterations (only one more than for $K = 4$), the algorithm had managed to produce a symmetrical clustering pattern, of which a scatter plot onto the factor 1/factor 2 plane of Figure 2 is shown in Figure 7. The refined centrotypes for the seven clusters $T^7$1, $T^7$2, $T^7$3, $T^7$4, $T^7$5, $T^7$6, and $T^7$7 are given in Table VII.

An analysis of the cluster membership indicates that $T^7$2 (eSQP) and $T^7$3 (fSQP) together contain the membership of T1 (dSQP) with an additional four reference codes absorbed from T3. Obviously then, T1 has been split in two, and a comparison of Tables II and VII 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 also it has been greatly facilitated by seeding the data set with approximate cluster centrotypes.

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 10 compounds identified as divergent in Table V, 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 in Table V represent observed molecular geometries that are classified as SQPs of C_{2v} symmetry in T-space, and as TBPs of C_{2v} symmetry in S-space. Accordingly one might expect these compounds to have a geometry that is intermediate between the TBP and the SQP. Indeed, when the papers[18] in which these structures were originally reported are examined, it becomes clear that the authors had great difficulty 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[19] and Holmes,[20] 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 that 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 toward a SQP, as outlined in our earlier study[11] of nickel complexes, we find that this ranges from 36% in the case of IPESNI[18g] to 59% in that of CMBPNI.[18a] 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 favored 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[18f] 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*" (our emphasis). This statement is all the more impressive when one considers it in the context of its date of formulation: 1968. At this stage the Berry mechanism was still largely unrecognized in coordination 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 that lie at greater than average distances from the center of the cluster with which they have been classified. Part of the output from PKM is a histogram displaying the distance from the cluster center to each observation. From these histograms, one can estimate the density or homogeneity of a given cluster and thereby identify cluster outliers.

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 centers in parentheses behind the refcode. In 11 cases the outliers lie more than twice the average size[21] of the cluster away from its center.

(i) S1 (dTBP) = S3; Size = 0.697 Å. DPPCRH (1.568)[18d] and CSMRHC (1.836)[18c] are both compounds that have previously (see Table V) been identified as having a geometry intermediate between a TBP and a SQP, thereby explaining why they

(18) (a) CMBPNI: Powell, H. M.; Watkin, D. J.; Wilford, J. B. *J. Chem. Soc. A* **1971**, 1803–1810. (b) CMPMNI; Jansen, J. C.; van Koningsveld, H.; van Ooijen, J. A. C.; van Reedijk, J. *Inorg. Chem.* **1980**, *19*, 170–174. (c) CSMRHC: Mague, J. T. *Inorg. Chem.* **1970**, *9*, 1610–1618. (d) DPPCRH: Pignolet, L. H.; Doughty, D. H.; Nowicki, S. C.; Casalnuovo, A. L. *Inorg. Chem.* **1980**, *19*, 2172–2177. (e) DPMCRH10: Cowie, M. *Inorg. Chem.* **1979**, *18*, 286–292. (f) EDCRCN: Raymond, K. N.; Corfield, P. W. R.; Ibers, J. A. *Inorg. Chem.* **1968**, *7*, 1362–1372. (g) IPESNI: Fälth, L. *Chem. Scr.* **1976**, *9*, 167–170. (h) SALDNI: Seleborg, M.; Holt, S. L.; Post, B. *Inorg. Chem.* **1971**, *10*, 1501–1504. (i) TMAGEP: Estes, E. D.; Hodgson, D. J. *Inorg. Chem.* **1973**, *12*, 2932–2935. (j) BEZPIF: Kalck, P.; Bonnet, J.-J.; Poilblanc, R. *J. Am. Chem. Soc.* **1982**, *104*, 3069–3077.
(19) Muetterties, E. L.; Guggenberger, L. J. *J. Am. Chem. Soc.* **1974**, *96*, 1748–1756.
(20) Holmes, R. R. *Acc. Chem. Res.* **1979**, *12*, 257–265.
(21) The average size is defined as the average distance of all cluster members to its center.

are found on the fringes of this cluster. PEANIC (2.606)[22a] and PEANNI (2.741)[22b] are nickel carbonyl and nitrosyl complexes, respectively, with the tripodal ligand tris((2-diphenylphosphino)ethyl)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 1 Å 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 that forces these compounds onto the outer edge of this cluster whose centrotype, after all, exhibits C_{2v} symmetry.

(ii) S2 (fSQP); Size = 0.932 Å. CHESNI (1.481)[22c] is a four-plus-one coordinate compound[1] 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 than it does a square-pyramidal geometry, thereby possibly accounting for its distance from the cluster center. CNPLPT (1.398)[22d] 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)[22e] contains a four-membered nickel–dithiophosphinato ring and the large and bulky 2,9-dimethyl-1,10-phenanthroline ligand, both of which contribute toward 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,[11] although in that case it was done merely on the basis of a graphical analysis. MPNBNI (1.451)[22f] is a dimer that similarly contains a four-membered ring leading to a geometry which is quite "asymmetric". FPHPRH (1.367)[22g] is also distorted toward "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 that lie at large distances from the center 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); Size = 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.

3. Hierarchical Cluster Analysis.[23] The nature of the output from the hierarchical cluster analysis program of BMDP makes the processing of large data sets by this program impractical. As a result, we chose not to use this package but to follow, instead, the approach outlined by Nørskov-Lauritsen and Bürgi[16] in employing the procedure CLUSTER incorporated in the Statistical Analysis System (SAS).[24] Nørskov-Lauritsen and Bürgi[16] outline the application of *Ward's method* as implemented in the program 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 that 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.

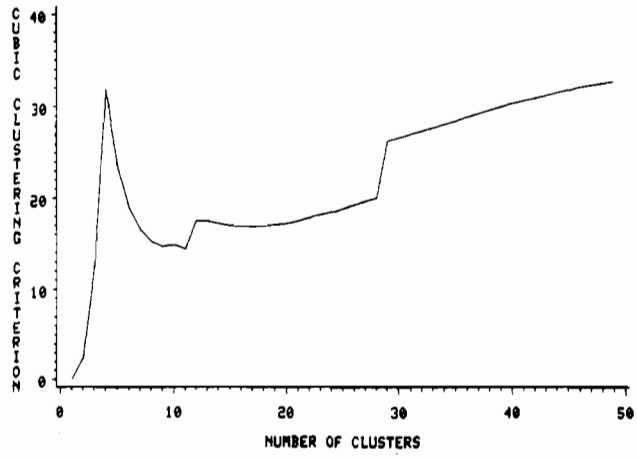

Figure 8. Cubic clustering criterion plotted against number of clusters in T-space.

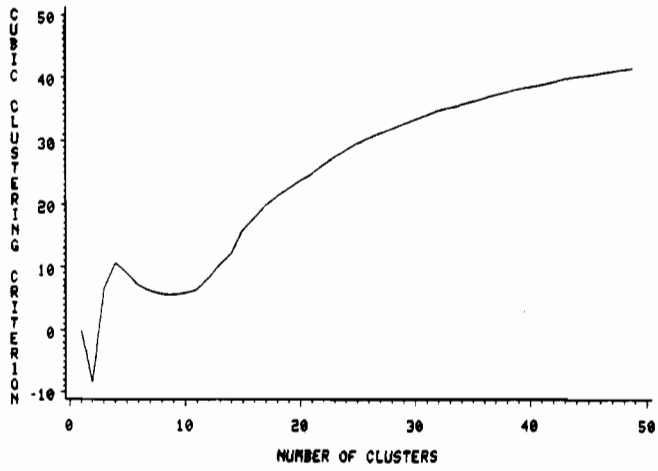

Figure 9. Cubic clustering criterion plotted against number of clusters in S-space.

The descriptor that they found "particularly useful for estimating the optimal number of clusters" is the cubic clustering criterion (CCC).[24b,25] This criterion rests on an approximation to the expected value of the within-cluster sum of squares, which in turn assumes that "a uniform distribution on a hyperrectangle will be divided into clusters shaped roughly like hypercubes". It is pointed out in ref 24 that for certain large samples this assumption gives very accurate results. The exact definition of the CCC is "very complicated",[16] but essentially it is a statistical measure that when plotted against the number of clusters formed, can indicate an optimum number of clusters by a peak in the graph.

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 8 and 9 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 ref 24b are as follows: (i) peaks on the plot with CCC greater than 2

(22) (a) PEANIC: Ghilardi, C. A.; Sabatini, A.; Sacconi, L. *Inorg. Chem.* **1976**, *15*, 2763–2767. (b) PEANNI: Di Vaira, M.; Ghilardi, C. A.; Sacconi, L. *Inorg. Chem.* **1976**, *15*, 1555–1561. (c) CHESNI: Gould, R. O.; Harding, M. M. *J. Chem. Soc. A* **1970**, 875–881. (d) CNPLPT: Wernberg, O.; Hazell, A. *J. Chem. Soc., Dalton Trans.* **1980**, 973–978. (e) DMPADP10: Shetty, P. S.; Fernando, Q. *J. Am. Chem. Soc.* **1970**, *92*, 3964–3969. (f) MPNBNI: Butcher, R. J.; Sinn, E. *Inorg. Chem.* **1977**, *16*, 2334–2343. (g) FPHPRH: Hughes, R. P.; Krishnamachari, N.; Lock, C. J. L.; Powell, J.; Turner, G. *Inorg. Chem.* **1977**, *16*, 314–319.

(23) This part of the analysis was performed during a brief stay by T.A.d.H. 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.

(24) (a) *SAS User's Guide: Statistics*, 1982 ed.; SAS Institute Inc.: Cary, NC, 1982. (b) *SAS User's Guide: Basics*, 1982 ed.; SAS Institute Inc.: Cary, NC, 1982.

(25) Searle, W. S. *SAS Technical Report: A-108*; SAS Institute Inc.: Cary, NC, 1983.

Table VIII. Cluster Membership Broken Up According to Metals[a]

	Ni	Pd	Pt	Rh	Ir	totl.
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

[a] Recall that T1 = T2 = T4 and S1 = S3.

or 3 indicate good clusterings; (ii) very distinct nonhierarchical clusters show a sharp rise before the peak followed by a gradual decline; (iii) peaks with the CCC between 0 and 2 indicate possible clusters.

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 nonhierarchical nature, at least from the level of 12 clusters upward. Far less significant "peaks" occur at 12 and 29 clusters—these presumably result from the clustering together of clouds of isometric conformations that 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 that 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 9, is similarly encouraging. The 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 8) 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. 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.

4. Analysis of Structural Results. We will here extract further structural information from the results of *K-Means* relocation 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.[26] 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 database 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,[27] he compares the observed structures to ones predicted by Rossi and Hoffmann on the basis of extended Hückel studies on the PtL_5^{3-} system[28] and also offers interpretations of the observed conformations in terms of nonbonded 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 favor a more SQP conformation, and his data appear to agree with this assertion. Table VIII 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 the individual metals are regarded, it becomes obvious that this tendency is more pronounced for the second- and third-row elements, which are likely to be low-spin, than it is for the nickel compounds. Without more knowledge about the spin states of the complexes, however, little can be definitely said about the conformational preferences of different spin states.

Table VIII also suggests 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 more dramatically toward an eSQP. Whether this difference in behavior 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.

There has been much discussion concerning the value of the trans-basal angles (θ_{15} and θ_{24}) in an "ideal" SQP. It is generally agreed, however, that these will depend on the *d*-electron configuration.[26-29] Clearly, no such thing as an "ideal" SQP exists; instead, there exists a range of square pyramidally disposed conformations ranging from a fSQP to 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 the eSQP) with the empirical results listed by Holmes[26,27] and the predictions of Rossi and Hoffmann.[28] For SQP nickel complexes, Holmes observed that the trans-basal angle for high-spin complexes is 161°, while that for low-spin complexes is 173°. Rossi and Hoffmann predicted that this angle would be 164°. Both results fall in the range of values given in Tables II, IV, VI, and VII.

Finally, Table IX lists the average bond distance increments d_1–d_5 for the various clusters (and it should be seen as an extension of Tables II and IV). The e.s.d.'s are large, and the results should thus be treated cautiously, but they nevertheless suggest interesting trends. For example, axial bonds in a TBP appear to be longer than equatorial ones, while apical bonds in both SQPs are longer than the basal ones. The latter result is in line with Holmes' observations and Rossi and Hoffmann's predictions; the former is not. Most interesting, however, is the difference in the length of the apical bonds in the fSQP and eSQP, with that in the fSQP being ~23 times longer (tending toward four-plus-one coordinate).

5. Intracluster Statistics. Now that cluster analysis has shown the essential distribution of the data in 12-dimensional space, it might be instructive to investigate *intracluster* univariate and bivariate statistics in a manner similar to that used for T- and S-spaces overall in the previous paper of this series.[1]

(a) Univariate Statistics. T-Space. Table X gives the standard deviations, variances, and percentage variance 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 (linear and angular) 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 a graphic representation of symmetry coordinates.[30] 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},

(26) Holmes, R. R. *Prog. Inorg. Chem.* **1984**, *32*, 119–235.
(27) Holmes, R. R. *J. Am. Chem. Soc.* **1984**, *106*, 3745–3750.
(28) Rossi, A. R.; Hoffmann, R. *Inorg. Chem.* **1975**, *14*, 365–374.
(29) (a) Burdett, J. K. *Adv. Inorg. Chem. Radiochem.* **1978**, *21*, 113–134. (b) Gillespie, R. J. *J. Chem. Soc.* **1963**, 4679–4685.
(30) The previous paper in this series[1] details the symmetry coordinates for the TBP and SQP and gives a graphic interpretation of these.

Table IX. Average Bond Distance Increments (Å) for the Clusters Indicated[a]

	T1 (dSQP)	T3 (TBP)	S1 (dTBP)	S2 (fSQP)	S4 (eSQP)
d_1	−0.049 (116)	0.030 (205)	0.024 (192)	−0.165 (82)	−0.008 (88)
d_2	−0.057 (106)	−0.025 (83)	−0.032 (83)	−0.165 (82)	−0.008 (88)
d_3	0.219 (353)	−0.025 (83)	0.001 (88)	0.692 (298)	0.030 (125)
d_4	−0.057 (106)	−0.025 (83)	−0.032 (83)	−0.165 (82)	−0.008 (88)
d_5	−0.049 (116)	0.030 (205)	0.024 (192)	−0.165 (82)	−0.008 (88)

[a] T1 = T2 = T4 and S1 = S3. Estimated standard deviations are given in parentheses.

Table X. Standard Deviations (σ), Variances (σ^2), Percentage of Variance in Linear Coordinates (% LV), 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	% LV	% AV	σ, Å	σ^2, Å^2	% LV	% AV
S_1	0.150	0.022	12.8		0.191	0.036	34.6	
S_2	0.140	0.020	11.2		0.084	0.007	6.8	
S_3	0.067	0.004	2.6		0.219	0.048	45.8	
S_4	0.245	0.060		8.2	0.534	0.286		47.2
S_{5a}	0.348	0.121	69.5		0.082	0.007	6.4	
S_{5b}	0.082	0.007	3.9		0.082	0.007	6.4	
S_{6a}	0.432	0.187		25.6	0.331	0.110		18.1
S_{6b}	0.440	0.193		26.5	0.331	0.110		18.1
S_{7a}	0.236	0.056		7.6	0.175	0.031		5.0
S_{7b}	0.196	0.039		5.3	0.175	0.031		5.0
S_{8a}	0.316	0.100		13.7	0.142	0.020		3.3
S_{8b}	0.309	0.096		13.1	0.142	0.020		3.3

Table XI. Standard Deviations (σ), Variances (σ^2), Percentage of Variance in Linear Coordinates (% LV) and Percentage of Variance in Angular Coordinates (% AV) for Clusters S1, S2, and S4

	S1 (dTBP)				S2 (fSQP)				S4 (eSQP)			
	σ, Å	σ^2, Å^2	% LV	% AV	σ, Å	σ^2, Å^2	% LV	% AV	σ, Å	σ^2, Å^2	% LV	% AV
S_1	0.087	0.008	7.98		0.298	0.089	76.8		0.125	0.016	33.3	
S_2	0.133	0.018	18.6		0.116	0.014	11.77		0.118	0.014	29.8	
S_3	0.248	0.061		10.4	0.216	0.046		5.7	0.296	0.087		16.4
S_4	0.148	0.022	22.8		0.054	0.003	2.5		0.078	0.006	12.9	
S_5	0.317	0.101		17.0	0.209	0.044		5.3	0.278	0.077		14.5
S_6	0.151	0.023		3.9	0.411	0.169		20.6	0.260	0.068		12.7
S_{7a}	0.203	0.041	43.2		0.072	0.005	4.5		0.075	0.006	12.0	
S_{7b}	0.084	0.007	7.5		0.072	0.005	4.5		0.075	0.006	12.0	
S_{8a}	0.345	0.119		20.1	0.489	0.239		29.1	0.319	0.102		19.1
S_{8b}	0.306	0.094		15.8	0.489	0.239		29.1	0.319	0.102		19.1
S_{9a}	0.311	0.097		16.4	0.205	0.042		5.1	0.219	0.048		9.1
S_{9b}	0.311	0.097		16.4	0.205	0.042		5.1	0.219	0.048		9.1

S_{6b}). This is the interpretation when the D_{3h} symmetry coordinates are referred to or "fitted" to a SQP.[31] Note that the symmetry of T-space is reflected in the variances of the degenerate coordinates (S_5–S_8) for cluster T3; it is not reflected in the variances of T1 since this cluster does not conform to the full symmetry of D_{3h} data space, but only to C_{2v}.

S-Space. Table XI 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 to the symmetry (C_{4v}) of the 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 of S_{8a} suggests a large degree of freedom in the angular orientation of the pivot atom. These interpretations can be recognized from the graphic representation of the C_{4v} symmetry coordinates referred to a TBP.[30] For the flattened SQP a large variation in the apical bond distance becomes apparent from S_1, while the variances 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 center 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 while S_1 comprises only the apical bond. All symmetry coordinates contribute approximately equally to the overall variance.

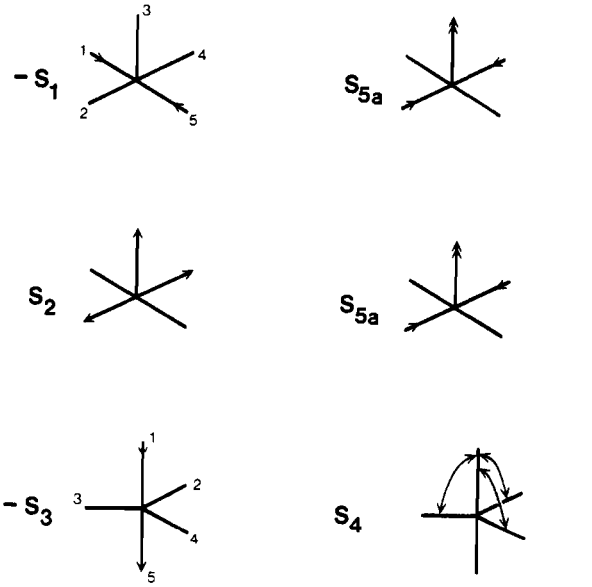

Figure 10. Graphical representation of distortions corresponding to correlated pairs of symmetry coordinates. The first two are from cluster T1 (dSQP) and the third is from cluster T3 (TBP). Where the correlation is negative, the inverse of one coordinate has been drawn, e.g. $-S_1$.

(31) This can be done by referring to the graphic interpretation of the D_{3h} symmetry coordinates shown in the previous paper[1] and regarding them as projected onto a SQP framework, rather than a TBP framework.

Table XII. Correlation Matrices for Symmetry Coordinates in Clusters T1 (dSQP) and T3 (TBP)[a]

	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}
Cluster T1												
S_1	1.00											
S_2	−0.61	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
Cluster T3												
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

[a] Symmetry species to which the coordinates belong are also shown.

(b) Bivariate Statistics. T-Space. The correlation matrices for clusters T1 (dSQP) and T3 (TBP) are shown in Table XII. The method, introduced in the previous paper,[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 intracluster correlation data. For T1 (dSQP) two major correlations appear: between S_1 and S_{5a} ($r = -0.812$) and between S_2 and S_{5a} ($r = 0.819$). The first corresponds to the departure of the apical ligand and the concomitant shrinking of the basal bonds in the dSQP (or the reverse process). This is represented in Figure 10 where the D_{3h} symmetry coordinates are referred to a SQP[31] to facilitate the interpretation of the data. In other words, this represents the "constant-amount-of-glue" coordinate already identified for T-space overall.[1] The second major correlation, also graphically illustrated in Figure 10, essentially indicates that there is a large degree of freedom in the position of the apical ligand, a characteristic of square pyramidally disposed conformations that we have repeatedly encountered. These two correlations are each able to account for ~80% of the variance in T1. Note, also, that symmetry coordinates are correlated *across different symmetry species*; i.e., irreducible representations are mixed. This arises out of the inability of T1 to conform to the full 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 contribution ($r = -0.84$), accounting for 70% of the sample variance in T3, characterizes the distribution of data in this cluster. This correlation (between S_3 and S_4) is graphically represented in Figure 10, and it mirrors the S_N2 coordinate.

Three major correlated distortions, described as an S_N2 coordinate, a Berry 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 be associated with T1 (dSQP), while the S_N2 is associated with T3 (TBP). However, in the case of the Berry coordinate, both clusters have retained a small component of it, thereby reducing its importance from that for T-space overall. T1 and T3 show correlations with coefficients of only 0.53 and 0.25, respectively, between coordinates S_{5a} and S_{6a}. This is a consequence of the clustering that has separated the data into those close to D_{3h} symmetry and those distorted along the Berry coordinate toward C_{4v} symmetry.

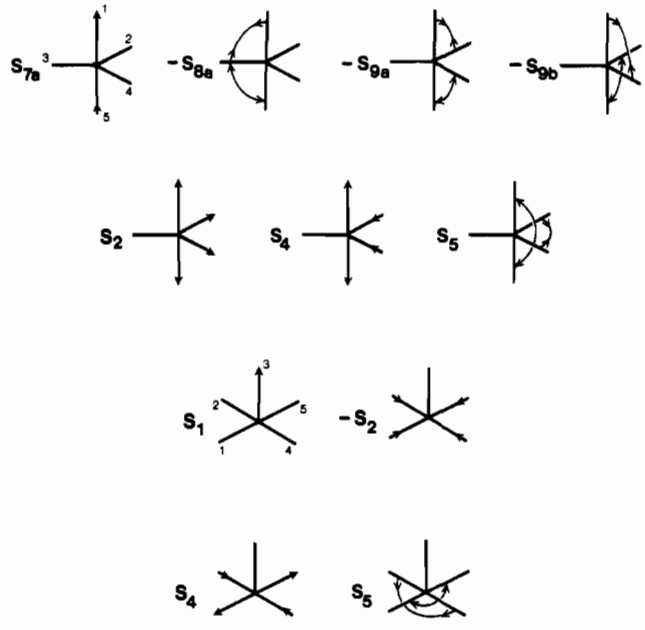

Figure 11. Graphical representation of distortions corresponding to correlated pairs of symmetry coordinates. Where the correlation is negative, the inverse of one coordinate has been drawn, e.g. $-S_{8a}$. Correlations are from clusters S1, S2, and S4 and are discussed in the text.

S-Space. Table XIII 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_{9b} with coefficients of −0.70, −0.76, and −0.76, respectively. A graphic interpretation of these correlations, illustrated in Figure 11, 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 co-

Table XIII. Correlation Matrices for Symmetry Coordinates in Clusters S1 (dTBP), S2 (fSQP), and S4 (eSQP).[a]

	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}
						Cluster S1						
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
						Cluster S2						
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
						Cluster S4						
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

[a] Symmetry species to which the coordinates belong are also shown.

herent 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} and S_{9b}. Another characteristic, the variance in the axial bonds, is mirrored in the correlation between S_2 and S_4 (r = 0.647). This correlated distortion is also graphically illustrated in Figure 11. 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.

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 truism 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 is associated with S1 (dTBP) and S4 (eSQP) but not with S2 (fSQP).

6. Summary. The clustering techniques employed in this study—nonhierarchical relocation clustering and hierarchical clustering using Ward's method—have been shown to yield essentially identical, robust clusters in the 12-dimensional deformation space of ML_5 molecular fragments. They have revealed that the ML_5 fragment is characterized by three archetypal geometries: the TBP, the eSQP, and the fSQP. Cluster statistics indicate distortions suggesting a constant amount of bonding at the central metal atom in all three geometries, as well as distortions mirroring ligand substitution reactions. The Berry distortion is associated with both TBP and SQP conformations, and its relative importance in these clusters is therefore reduced as a consequence of the clustering, which separates the data into those close to TBP symmetry and those distoring away from it along a C_{2v} coordinate.

The next paper examines these distortions in greater detail by using factor analysis.

Acknowledgment. Through a travel grant Rustenburg Platinum Mines Ltd. generously enabled T.A.d.H. to perform the hierarchical cluster analysis in Berne, while the University of the Western Cape helped with a study leave.