Systematic analysis of metal coordination sphere geometry from crystallographic data: a general method for detecting geometrical preferences, deformations and interconversion pathways

Judith A. K. Howard,[*a] **Royston C. B. Copley,**[a] **Jing Wen Yao**[a] **and Frank H. Allen**[*b]

[a] *Department of Chemistry, University of Durham, South Road, Durham, UK DH1 3LE. E-mail: J.A.K.Howard@durham.ac.uk*
[b] *Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, UK CB2 1EZ. E-mail: allen@ccdc.cam.ac.uk*

A simple and general computational procedure using the L–M–L angles for all crystallographic observations of an ML_n coordination sphere generates mappings which reveal geometrical preferences, deformations from standard polyhedral geometries, and interconversions between polyhedral forms; a systematic analysis of the geometry of seven-coordination is described.

Systematic knowledge about geometries, deformations and interconversion pathways of metal coordination spheres is fundamental to inorganic chemistry.[1] For ML_n coordination, where L is any ligand, much knowledge is implicit in the $N_a = n(n-1)/2$ values of the L–M–L valence angles. This has been demonstrated[2] for the lower coordination numbers, $n = 3$, 4, using crystallographic data retrieved from the Cambridge Structural Database (CSD),[3] and using data analysis methods analogous to those applied to conformational studies of organic systems based on torsional metrics.[4]

However, the high topological symmetry of generalised ML_n systems having $n \geqslant 5$ has so far prevented routine systematic studies of the associated angular data sets. For a given system, the topological symmetry gives rise to $n!$ possible enumerations of the chemically equivalent L atoms, hence to $n!$ permutations of the valence angle sequence. There are two ways of addressing this problem so that we can properly visualise and analyse a data matrix of N_a valence angles for N_f occurrences of an ML_n substructure in the CSD, *viz.*: (*a*) By completely filling the parameter space spanned by the valence angles, *i.e.* by including all $n!N_f$ permutationally equivalent angular sequences in the data matrix. This method has frequently been used in the conformational (torsional) analysis of k-membered cyclic systems where topological symmetry requires a maximum $2k$-fold data expansion.[4] Given that $n! = 120, 720, 5040, 40320$ for $n = 5, 6, 7, 8$-coordination, visualisation and analysis of permutationally complete valence angle matrices for ML_n systems becomes increasingly intractable.

(*b*) By placing each angular sequence into a single asymmetric unit of the valence angle parameter space. This can be effected by comparing all permutations of each angular sequence with some predefined set of standard values, so as to locate the permuted sequence that provides the best numerical fit to the standard sequence. The numerical overlay therefore assigns a unique atomic enumeration of the L-atoms to each example of ML_n coordination retrieved from the CSD with respect to the atomic enumeration chosen for the standard. The resulting data matrix comprises the N_a valence angles for the basic number (N_f) of fragments retrieved from the database: a much more tractable proposition for visual display and numerical analysis.

In this communication, we describe a general computational procedure in which the topological problem is dealt with according to method (*b*), and which also addresses the problems of data analysis inherent in the dimensionality of the angular description, *i.e.* the fact that $n(n-1)/2$ angles are needed to provide a full description of each coordination sphere. We

illustrate the application of the method to a systematic study of the geometrical characteristics of the 372 examples of 7-coordination available in the CSD.

Standard and symmetric n-polyhedra [*e.g.* the tetrahedron (T), octahedron (O), pentagonal bipyramid (PBP), capped trigonal prism (CTP), *etc.*] are archetypal reference frames that are used to describe coordination sphere geometries. In many cases, the N_a valence angles for the standard are fixed by symmetry, irrespective of the M–L (or L–L edge) distances. In others, and notably some of the higher-coordination archetypes, valence angle sequences may vary under the symmetry of the archetype. Here, it is necessary to select a set of angles for use as a standard. This does not obviate our method in any way, since the purpose of the standard is to provide a fixed origin in valence angle space that represents an idealised n-polyhedron appropriate to the problem under investigation.

We denote the standard set of L–M–L angles as $\theta_{k(std)}$ [$k = 1 \rightarrow N_a$], and determine how far an observed geometry, defined by angles $\theta_{k(obs)}$ derived from the CSD, is distorted from an appropriate archetype, by computing the Euclidean dissimilarity:

$$\text{Rc}(x) = 100 \min \left\{ \frac{\Sigma_k [\theta_{k(obs)} - \theta_{k(std)}]^2}{\Sigma_k \theta_{k(std)}^2} \right\}_p^{\frac{1}{2}} \qquad (1)$$

in which (x) identifies the archetypal standard, and the subscript p indicates that $\text{Rc}(x)$ is the minimum value obtained through application of eqn. (1) to all possible sequences of the $\theta_{k(obs)}$ that arise from the $n!$ permutational ligand enumerations. Local FORTRAN77 code derives and applies the permuted L–M–L angle sequences and, for each CSD observation, identifies the specific permutation that is closest to the standard angular description used for archetype (x). Thus, the procedure implicitly identifies the specific labelling of the CSD atoms, L, that best matches the labelling used for the standard.

For 7-coordination, the three common archetypal polyhedra (Scheme 1) are the pentagonal bipyramid (PBP), capped
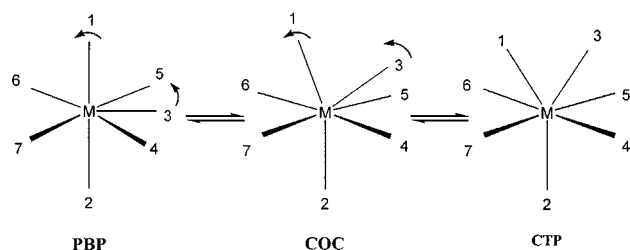
Scheme 1 The PBP ⇔ COC ⇔ CTP interconversion path: atomic movements in 7-coordination. The unique standard angles (°) for CTP: L¹–M–L² 141.8, L¹–M–L³ 76.4, L¹–M–L⁴ 127.3, L¹–M–L⁵ 127.3, L¹–M–L⁶ 76.4, L¹–M–L⁷ 76.4, L²–M–L⁴ 76.4, L²–M–L⁵ 76.4, L⁴–M–L⁵ 87.7, L⁴–M–L⁶ 152.7, L⁴–M–L⁷ 85.9.
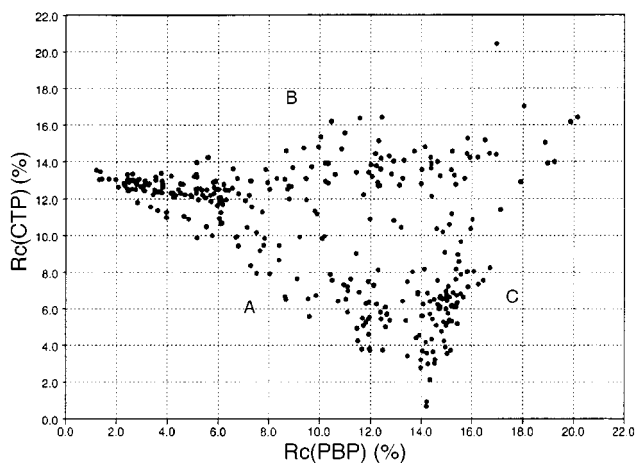
Fig. 1 Rc(CTP) *vs.* Rc(PBP) for all 372 fragments having 7-coordination

Fig. 2 D_{5h} expanded PCA: PC_1 *vs.* PC_2 for all 372 fragments (7440 observations after expansion)

octahedron (COC), and the capped trigonal prism (CTP). The COC can be considered as an intermediate form on the PBP $\Leftrightarrow$ CTP interconversion pathway.[1] Starting from PBP, one of the five equatorial ligands (3 in Scheme 1) moves out of the plane towards apical ligand 1, which then moves away from its apical position. Small relocations of equatorial ligands 4–7 then generate the COC geometry, and further small movements of atoms 3 and 1 generate the CTP polyhedron. The standard CTP angles used in the present report are given in Scheme 1. In practice, users may choose their own standard (origin) angles for the CTP case.

The CSD System program QUEST3D[3] was used to locate 372 examples of ML_7 coordination (L = any non-H atom), and to output the (permutationally random) sequence of 21 L–M–L valence angles for each example. The local code was used to minimise both Rc(PBP) and Rc(CTP) over the 5040 possible permutations using eqn. (1). The dataset was visualised by plotting Rc(PBP) *vs.* Rc(CTP), and Fig. 1 immediately reveals a number of significant features: (*a*) a well populated cluster of PBP geometries having Rc(PBP) up to *ca.* 7%, (*b*) a linear PBP $\Leftrightarrow$ CTP interconversion pathway (A in Fig. 1) that passes through a populated COC area, and (*c*) interconversion pathways B and C that connect the PBP and CTP examples with a loose cluster of observations for which both Rc(*x*) values are *ca.* 15% or higher. Examination of individual structures in these areas shows that they arise almost entirely from ML_7 fragments having small ligand bite angles. Conversely, the PBP area is dominated by ML_7 species having either seven or two unidentate ligands. The COC/CTP area is structurally diverse and requires further detailed study.

Because the Rc(*x*) minimisation process does assign standard atomic labels to the seven equivalent ligands in ML_7 systems, we can use this information as a basis for more detailed multivariate data analyses, *e.g.* the principal component analysis (PCA) technique[5] that is commonly applied to conformational problems.[4] In the ML_7 case however, we are now starting with data that defines a single asymmetric unit of parameter space. Thus, we can expand the data set in a controlled manner to reflect full or partial topological symmetries in the resultant PCA plots. Based on previous PCA studies of the conformations of five-membered rings,[4] we may treat the five ligands that most closely map to their coplanar equatorial counterparts in the PBP standard as a cyclic system, which would have exact D_{5h} symmetry in the planar form (a perfect PBP form). To examine the deformations in this system exhibited in real structures, we apply a D_{5h}-expansion to the angular dataset prior to application of PCA. The analysis generates PC_1 and PC_2 as a degenerate pair and Fig. 2, which shows a plot of the PC scores along these orthogonal axes for the expanded dataset, is a readily interpretable visualisation of the dataset: the central cluster of Fig. 2 maps the PBP examples (five equatorial ligands closely
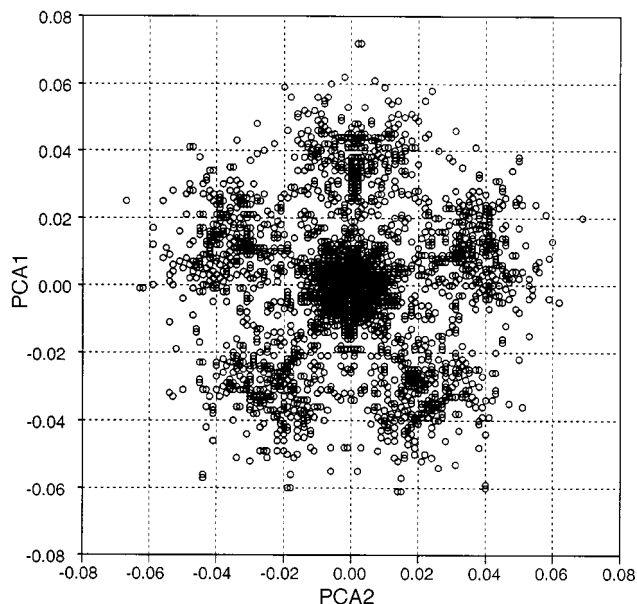
coplanar), while the five symmetric 'spokes' map the PBP $\Leftrightarrow$ COC $\Leftrightarrow$ CTP pathways that arise from each of the equatorial atoms of the PBP standard being designated, in turn and according to the D_{5h} topology, as the mobile ligand 3 in Scheme 1.

An extraordinary richness of information is immediately apparent in the simple two-dimensional visual representation of a 21-dimensional parameter space shown in Fig. 1, while Fig. 2 shows how the analysis can be further enhanced through use of standard multivariate techniques. We stress that any interpretation of structural diversity, in terms of geometrical arrangements, distortions from symmetrical forms, and interconversion pathways, is wholly dependent on the investigator and, as in most structural data mining experiments, a full interpretation (currently in preparation) involves a close examination of individual structures. The technique reported here provides the essential underpinning for such data mining experiments by treating the most difficult problem of atomic permutational symmetry that has, until now, presented a serious disincentive to knowledge discovery in coordination chemistry. Further, the fact that eqn. (1) generates a one-dimensional metric, Rc(*x*), that quantifies how far the shape of an observed coordination sphere deviates from some idealised standard (*x*), raises the possibility of including such metrics in the CSD itself, to form the basis for simple and rapid searches for coordination sphere geometries that are close to (or even far from) a particular reference frame.

Notes and References

1 D. L. Kepert, in *Comprehensive Coordination Chemistry,* Editor-in-Chief G. Wilkinson, 1987, vol. 1, p. 31; M. G. B. Drew, *Prog. Inorg. Chem.,* 1977, **23**, 1.

2 G. Klebe and F. Weber, *Acta Crystallogr., Sect. B,* 1994, **50**, 50; R. Taylor and F. H. Allen, in *Structure Correlation,* ed. H.-B. Bürgi and J. D. Dunitz, VCH, Weinheim, 1994, p. 111.

3 F. H. Allen and O. Kennard, *Chem. Des.-Automat. News,* 1993, **8**, 31.

4 F. H. Allen, M. J. Doyle and T. P. E. Auf der Heyde, *Acta Crystallogr., Sect. B,* 1991, **47**, 412.

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