

Small revisions to predicted distances around metal sites in proteins

Marjorie M. HardingInstitute of Structural and Molecular Biology,
Michael Swann Building, University of
Edinburgh, Edinburgh EH9 3JR, ScotlandCorrespondence e-mail:
marjorie.harding@ed.ac.uk

Received 23 February 2006

Accepted 21 April 2006

A new analysis has been made of distances around metal sites in protein structures in the Protein Data Bank determined with resolution ≤ 1.25 Å and equivalent distances have been extracted from the Cambridge Structural Database. They are for the metals Na, Mg, K, Ca, Mn, Fe, Co, Cu, Zn and the donor atoms O of water, O of Asp and Glu, O of the main-chain carbonyl group, N of His and S of Cys. Some revisions are recommended to the tables of 'target distances' previously given [Harding (2001), *Acta Cryst. D* **57**, 401–411; Harding (2002), *Acta Cryst. D* **58**, 872–874]. As well as small changes in many distances and a large improvement for Mg–O_{carboxylate}, the table includes an indication of how reliable each prediction may be. Special attention was given to carboxylate interactions. When the carboxylate group is monodentate, the M–O_{carboxylate} distance is well defined, but for bidentate carboxylate groups a wide range of distances is allowable; when the metal is Co, Cu or Zn the M–O₁ and M–O₂ distances are clearly inversely correlated; for the more purely electrostatic interactions involving Na, K and Ca there is a wider scatter of distances and little correlation.

1. Introduction

An analysis of metal sites in protein structures in the Protein Data Bank (PDB; Berman *et al.*, 2000; Bernstein *et al.*, 1977) combined with information from analogous metal-coordination compounds in the Cambridge Structural Database (CSD; Allen & Kennard, 1993*a,b*) gave a set of 'target distances' for different combinations of metal and donor group (Harding, 2001, 2002). These target distances are relevant for the interpretation of electron-density maps in new protein structures and for restraints in refinement when data resolution is limited or for validation of the structures. Since then, many more protein structures have been determined at or near atomic resolution and there are also more structures in the CSD; the predictions about distances made in 2001 have been reassessed and small revisions are proposed.

It is also important to consider how precisely these distances can be predicted and to distinguish experimental error in coordinate determination from true flexibility of some kinds of distances. The interactions considered range from almost purely electrostatic for Na and K to those with a substantial covalent contribution to the chemical bonding, Fe, Co, Cu, Zn; the latter have well defined characteristic bond lengths, while the former are more variable. Special attention is also given to the interactions of carboxylate groups, which are potentially bidentate, with metals.

2. Methods

All protein structures determined with resolution ≤ 1.25 Å were selected from the PDB in March 2005. Distances around metal atoms were extracted as described by Harding (2001) and their means and sample standard deviations derived. (A check in a few of the PDB files found no mention of any

restraint on these distances in the structure refinement and it is assumed they are all unrestrained; to restrain them in a refinement at this resolution would not normally be appropriate.) Most of the distributions of these metal to donor atom distances have a standard deviation of <0.10 Å. A small number of observations more than 0.4 Å from each mean were excluded as outliers. Mean distances for the equivalent metal and donor atom combinations were derived from the CSD (November 2005); the search queries were very similar to those in Harding (1999) (when different, they correspond a little more closely to the protein side-chain donor groups than previously). The CSD was used through the UK Chemical Database Service at Daresbury Laboratory (Fletcher *et al.*, 1996). Target distances for each type of bond were derived from the PDB and CSD observations, weighted according to the standard deviations of their means.

Classification of metal-carboxylate interactions as monodentate or bidentate requires an arbitrary definition of the maximum $M-O$ distance that could be regarded as a bond in a bidentate interaction. The distinction between simple bidentate interactions (i) and bridging interactions (ii) or (iii) (see Fig. 2) was made for the PDB results by examining the list of contacts to each metal ion and each carboxylate group involved; a small program was written to perform this. The CSD was searched for simple bidentate interactions, but not for bridging interactions. In CSD searches involving Na, Mg, K and Ca it was always necessary to redefine the $M-O$ distance which would be regarded as a 'bond', as described in more detail by Harding (1999); for other metals this was performed

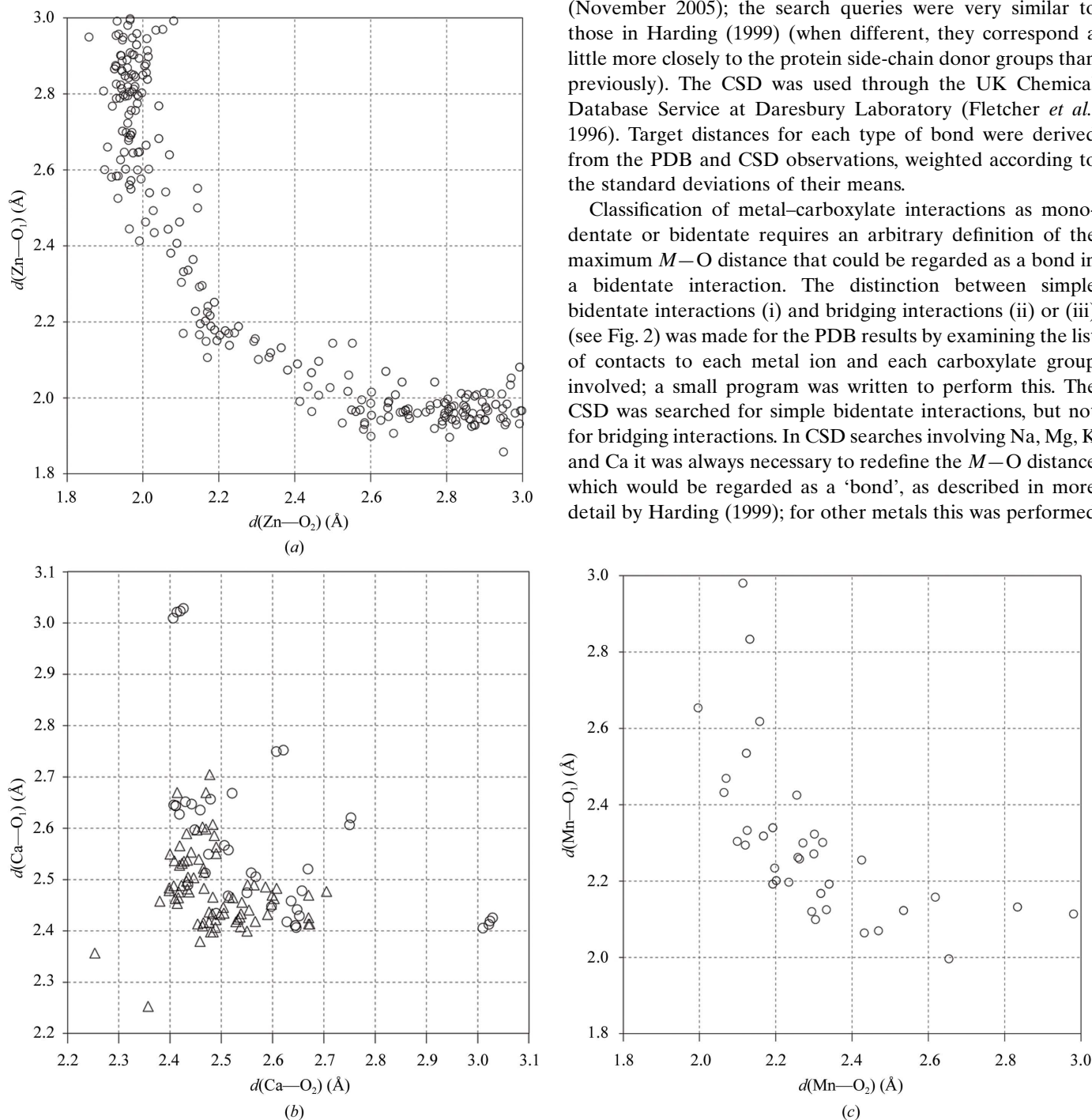


Figure 1

The interaction of metals with carboxylate groups, showing the relationship of the two $M-O$ distances when they are both <3.0 Å (Zn, Mn) or <3.1 Å (Ca). [The labels O_1 and O_2 can be permuted, so each carboxylate group is shown twice and the pattern is symmetrical about the diagonal line, $d(M-O_1) = d(M-O_2)$]. (a) Zn carboxylates in CSD data. (b) Ca bidentate interactions with Asp and Glu in the PDB data. Those of type (i) are shown as triangles and of type (iii) as circles (see Fig. 2). (c) Mn carboxylates in the CSD data.

Table 1

Numbers of observations and mean distances in structures in the PDB determined at near atomic resolution and in the CSD with *R* factor < 0.065, with sample standard deviations, treating all metal-coordination numbers together.

	Ca	Mg	Mn	Fe	Co	Cu [†]	Zn	Na	K
<i>M</i> –H ₂ O									
PDB									
<i>N</i> _{obs}	302	269	17	8	2	4	31	133	24
Mean distance (Å)	2.40 (10)	2.09 (8)	2.22 (6)	2.17 (8)	—	—	2.06 (13)	2.42 (19)	2.82 (14)
CSD									
<i>N</i> _{obs}	169	326	289	121	552	379	270	334	104
Mean distance (Å)	2.39 (5)	2.07 (3)	2.19 (4)	2.09 (5)	2.09 (3)	2.13 (22)†	2.09 (5)	2.41 (10)	2.80 (19)
<i>M</i> –O monodentate carboxylate									
PDB									
<i>N</i> _{obs}	105	43	19	12	1	1	16	4	1
Mean distance (Å)	2.33 (7)	2.08 (8)	2.12 (5)	2.10 (6)	2.01	1.96	2.01 (9)	2.3 (3)	2.8
CSD									
<i>N</i> _{obs}	170	4	5	8	33	95	84	931	1049
Mean distance (Å)	2.38 (7)	2.05 (5)	2.15 (1)	2.03 (2)	2.05 (6)	1.96 (4)	1.99 (5)	2.41 (11)	2.82 (13)
<i>M</i> –O main-chain carbonyl									
PDB									
<i>N</i> _{obs}	130	12	—	—	—	—	5	44	25
Mean distance (Å)	2.36 (10)	2.26 (23)	—	—	—	—	—	2.46 (24)	2.80 (15)
CSD									
<i>N</i> _{obs}	6	4	8	26	30	137	12	15	11
Mean distance (Å)	2.39 (11)	—	2.19 (5)	2.04 (6)	2.08 (5)	2.04 (14)†	2.07 (5)‡	2.37 (6)	2.67 (10)
<i>M</i> –N of imidazole (for His)									
PDB									
<i>N</i> _{obs}	2	3	22	24	7	19	62	3	—
Mean distance (Å)	—	—	2.16 (5)	2.03 (8)	2.04 (9)	2.02 (4)	2.04 (4)	—	—
CSD									
<i>N</i> _{obs}	1	—	10	7	47	110	34	—	—
Mean distance (Å)	—	—	2.25 (3)	2.17 (1)	2.14 (5)§	2.02 (9)†	2.01 (4)	—	—
<i>M</i> –S of thiolate (for Cys)									
PDB									
<i>N</i> _{obs}	—	—	—	239	—	10	59	—	—
Mean distance (Å)	—	—	—	2.30 (3)	—	2.15 (9)	2.34 (5)	—	—
CSD									
<i>N</i> _{obs}	—	—	43	47	46	3	28	10	—
Mean distance (Å)	—	—	2.35 (4)	2.28 (4)	2.25 (4)§	—	2.28 (4)	2.88 (8)	—

† In the CSD the distributions for Cu–O and Cu–N are obviously composite. For Cu^{II} with coordination number 5 or 6 there are very substantial Jahn–Teller distortions; as a result there are usually a large group of observations clustered around a value near 2.0 Å, corresponding to equatorial ligands, together with a wide distribution of longer axial distances up to at least 2.5 Å; there are also a small number of Cu^I compounds, most notably the thiolates, in all of which the coordination number is 3. ‡ Zn–O_{carbonyl}: here, the number of ‘outliers’ is more significant, six in four different molecules; the distances are in the range 2.3–2.5 Å. They appear to be further illustrations of the ability of Zn to make one or more additional bonds (see bidentate carboxylates) which are abnormally long when there are already four or more normal bond lengths [an early example of this was noted in bis(histidinato)zinc, where there are four normal Zn–N bonds and two Zn–O contacts at ~2.8 Å; Harding & Cole, 1963; Kretsinger *et al.*, 1963]. § Co–N: there are obviously three components with different oxidation states and/or coordination numbers.

in the exploration of bidentate carboxylates and the production of Figs. 1(a) and 1(c).

The 248 protein structures which were used in this study are indicated in the supplementary information¹, which also contains some details of the metal sites.

3. Results and discussion

3.1. Target distances

The observations that are now available from the PDB and CSD are summarized in Table 1. Sample standard deviations of the various mean distances are given; these indicate the spread of the observations and so show how well the distance should be predicted. Some of the means have much larger standard deviations than others, even when adequate numbers

of observations are available. Many effects contribute to this scatter of observed distances: experimental coordinate errors in the structure determinations as well as real differences due to different oxidation states or coordination numbers of the metal or other factors affecting the nature of the bond. In all but two cases the means from the PDB and the CSD agree within about one standard deviation and most agree rather better. Coordinate errors in PDB structures, resolution ≤1.25 Å, are likely to be greater than those in the CSD structures, with *R* < 0.065. The small standard deviations, ≤0.05 Å, in favourable cases such as Mn–N, Co–N and Zn–N(His) provide an upper limit for the coordinate errors in the PDB at this resolution.

For each metal, the values in Table 1 include all coordination numbers. For some types of complex, mainly those of Co, Cu and Zn, several different coordination numbers are found and the distances represent the mixture present (which may be different in the PDB and CSD). A large proportion of the complexes with water and carboxylate donors have metal-ion

¹ Supplementary material has been deposited in the IUCr electronic archive (Reference: BE5055). Services for accessing these data are described at the back of the journal.

Table 2

CSD results showing variation of distances with coordination number.

Mean distances are in Å with sample standard deviations.

CSD	Co		Cu ^{II†}		Zn	
	<i>N</i> _{obs}	<i>d</i> _{mean}	<i>N</i> _{obs}	<i>d</i> _{mean}	<i>N</i> _{obs}	<i>d</i> _{mean}
<i>M</i> –H ₂ O						
CN = 4			48	1.96 (3)	42	2.01 (4)
CN = 5			69	1.96 (3)	13	2.09 (7)
CN = 6			110	1.98 (4)	215	2.10 (4)
<i>M</i> –N imidazole						
CN = 4	6	2.02 (1)	28	2.00 (3)	25	2.00 (2)
CN = 5			31	2.01 (2)		
CN = 6	41	2.15 (3)	42	2.01 (2)	5	2.18 (4)
<i>M</i> –O monodentate carboxylate						
CN = 4	12	1.98 (2)	53	1.96 (3)	70	1.97 (2)
CN = 6	21	2.09 (2)	10	1.96 (2)	14	2.09 (3)

† The values given here are for the very close clusters of bond distances, presumed to be all equatorial; other longer distances are also found when CN = 5 or 6 and are presumably axial bonds. For Cu–OH₂ there are 151 observations of longer distances in the range 2.10–2.91 Å, for Cu–O_{carboxylate} there are six observations in the range 2.12–2.52 Å and for Cu–N_{imidazole} two observations, 2.23 and 2.59 Å.

coordination number six; some Zn and Cu complexes are four- or five-coordinate and Ca, Na and K may also be seven- or eight-coordinate. Where imidazole is present, most Zn complexes are four-coordinate and Cu has approximately equal numbers of four-, five- and six-coordinate examples. In thiolate complexes the common coordination numbers are Mn 5, Fe 4 and 5, Co 4 and 6 and Zn mostly 4, while all the Cu complexes are three-coordinate. For some of the CSD results there are clear differences in metal–donor atom distance for different coordination numbers and these are given in Table 2. The variations of *M*–S distance with coordination number are much less significant than those of *M*–N or *M*–O distances.

Table 4 gives the revised set of target distances and an indication of the reliability of each. (It assumes that the prediction of distance may have to be made without a knowledge of the oxidation state or coordination number or, in the case of Cu, whether the bonds are axial or equatorial; with this knowledge, it is obviously possible to do better.) Where the covalent contribution to the bond between metal and donor atom is significant, as in Zn–N and Zn–S, there is very good agreement between the values in different proteins because there is a characteristic bond length; the standard deviation (Table 1) can be quite small, ~0.04 Å, and the most reliable predictions can be made. At the other extreme, the interactions between Na or K and oxygen donors are almost entirely electrostatic with no characteristic ‘bond’ length and a wider scatter of observed distances; the standard deviations rise to 0.1–0.2 Å.

Note that Co was not previously included, but is in the new table. Only five of the new target distances differ by more than 0.05 Å from those given by Harding (2001, 2002); in only one of these is the difference greater than one standard deviation. For Mg–O_{carboxylate} the new value is shorter by 0.19 Å; the old value was based on very few observations and errors in these may have arisen from the difficulty of locating Mg, with its small atomic number, accurately. The new value for Cu–OH₂, which is longer by 0.16 Å than the old, takes account of the

Table 3

Numbers of interactions of metal with O of a simple bidentate carboxylate group (see text for details).

	Ca	Mg	Mn	Fe	Co	Cu	Zn	Na	K
Maximum <i>M</i> –O distance accepted (Å)	2.8	2.5	2.5	2.5	2.5	2.5	2.5	2.8	3.2
PDB <i>N</i> _{obs}	120	18	14	—	6	—	38	6	2
CSD <i>N</i> _{obs}	10†	4†	16†	14	36	46	38	6†	14†

† Also many bridging interactions and more complicated networks.

longer (axial) bonds in five- and six-coordinate complexes (but where the coordination arrangement is clear, a better value may be found from Table 2).

3.2. Bidentate carboxylate groups

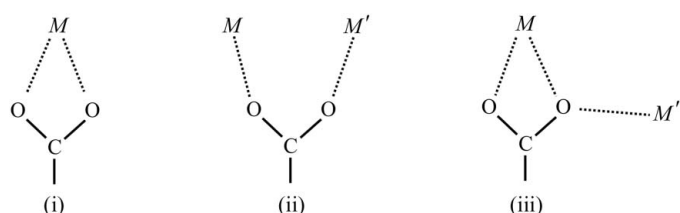
Several kinds of bidentate interactions are possible, simple (i), bridging (ii) or a combination (iii) (Fig. 2), and the PDB analysis allowed these to be distinguished. Ca participates in about equal numbers of type (i) and (iii) and very few of type (ii), whereas nearly all the Zn interactions are of type (i); the numbers for Mg and Mn are small and for these bridging (ii) is favoured.

In the CSD analysis only the simple type (i) interactions are included. Types (ii) and (iii) also occur, quite frequently for some metals (*e.g.* Mn, Ca), and there are many more complicated networks, especially for Na, K and Ca. The numbers of observations are given in Table 3.

The *M*–O distances in these bidentate carboxylates are quite variable and the patterns shown in the CSD and PDB are consistent where there are reasonable numbers of observations. For Co, Cu and Zn in simple bidentate coordination (i), both *M*–O distances may be ~2.2 Å or one may be shorter, down to the characteristic length in a monodentate carboxylate, and the other longer. The distances are inversely correlated as shown in Fig. 1(a); distances (in Å) conform to the relationship

$$[d(M-O_1) - 2.0] = 0.04/[d(M-O_2) - 2.0].$$

Fig. 1(a) also shows that *M*–O distances in the range 2.6–3.0 Å are observed. This is beyond the range that was counted as bidentate coordination in Table 3, but shorter than would normally occur in a van der Waals contact. These should correspond to very weakly bonding interactions (see, for example, Brown, 1992), while the other shorter *M*–O bond is indistinguishable in length from that in a monodentate

**Figure 2**

Bidentate interactions.

Table 4

Metal–donor atom target distances.

A revised table based on the distances in the CSD and PDB given in Table 1. Values marked *** are the most reliable, with good agreement between the CSD and PDB, and for these the standard deviation is ~ 0.05 Å or less. For values marked ** a standard deviation of 0.10 is appropriate and for values marked * 0.15–0.20 Å; values with no star marking are the least reliable. Distances may be less precisely predictable because the interactions are less covalent/more electrostatic (e.g. Na, K) or because there are very few observations or because of variations in coordination number or Jahn–Teller distortions (e.g. for Cu). For asparagine and glutamine, expect $M-O$ distances similar to monodentate carboxylates or perhaps very slightly longer. For serine and threonine, expect $M-O$ distances between those for water and for monodentate carboxylate. For tyrosine, expect $M-O$ distances that are significantly shorter (by ~ 0.1 Å) than for monodentate carboxylate.

	O, water	O, Asp or Glu monodentate	O, main-chain carbonyl	N, histidine	S, cysteine
Na	2.41**	2.41**	2.38**	—	—
Mg	2.07***	2.07**	2.26	—	—
K	2.81*	2.82*	2.74*	—	—
Ca	2.39**	2.36**	2.36**	—	—
Mn	2.19***	2.15***	2.19	2.21**	2.35
Fe	2.09**	2.04**	2.04	2.16*	2.30***
Co	2.09**	2.05***	2.08	2.14**	2.25*
Cu	2.13	1.99*	2.04	2.02**	2.15
Zn	2.09***	1.99***	2.07	2.03***	2.31**

carboxylate. The figure shows that there is a continuous range of allowable states between monodentate and bidentate coordination to metal.

For Ca, Mg, Na and K the pattern looks different. For Ca, there is certainly variability of the $Ca-O$ distance, but little evidence of correlation of $Ca-O_1$ and $Ca-O_2$ (Fig. 1*b*) and for Na and K there is more scatter and even less suggestion of correlation. The scatter can be attributed to the greater flexibility of the more electrostatic interactions. The two Mg bidentate carboxylates in the CSD are very nearly symmetrical, with all $Mg-O$ distances between 2.09 and 2.14 Å. For Mn (Fig. 1*c*) and Fe the patterns of behaviour are probably intermediate between Ca and Zn, but there are rather few observations.

4. Conclusions

A revised table of distances at metal sites in proteins is presented (Table 4). As well as small revisions for many distances and a large improvement for $Mg-O_{\text{carboxylate}}$, there is an indication of how reliable each prediction may be. The table includes mean distances for monodentate carboxylate interacting with metals and these are well defined. For bidentate carboxylate groups there are wide ranges of allowable distances; for Co, Cu and Zn, where the binding to metal is a little more covalent than in the others, the $M-O_1$ and $M-O_2$ distances are clearly inversely correlated in a way which might form the basis for a reaction pathway.

I am very grateful to Professor Malcolm Walkinshaw and to the University of Edinburgh for computing facilities and to Dr Paul Taylor for computational support. I acknowledge the use of the EPSRC's Chemical Database Service at Daresbury.

References

- Allen, F. H. & Kennard, O. (1993*a*). *Chem. Des. Autom. News*, **8**, 1.
 Allen, F. H. & Kennard, O. (1993*b*). *Chem. Des. Autom. News*, **8**, 31–37.
 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
 Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
 Brown, I. D. (1992). *Acta Cryst.* **B48**, 553–572.
 Fletcher, D. A., McMeeking, R. F. & Parkin, D. (1996) *J. Chem. Inf. Comput. Sci.* **36**, 746–749.
 Harding, M. M. (1999). *Acta Cryst.* **D55**, 1432–1443.
 Harding, M. M. (2001). *Acta Cryst.* **D57**, 401–411.
 Harding, M. M. (2002). *Acta Cryst.* **D58**, 872–874.
 Harding, M. M. & Cole, S. J. (1963) *Acta Cryst.* **16**, 643–650.
 Kretsinger, R. H., Cotton, F. A. & Bryan, R. F. (1963). *Acta Cryst.* **16**, 651–657.