Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Metal–ligand geometry relevant to proteins and in proteins: sodium and potassium

Marjorie M. Harding

Institute of Cell and Molecular Biology, University of Edinburgh, Michael Swann Building, Edinburgh EH9 3JR, Scotland

Correspondence e-mail: marjorie.harding@ed.ac.uk

Received 14 December 2001
Accepted 26 February 2002

In previous papers [Harding (2001), *Acta Cryst.* D**57**, 401–411, and references therein] the geometry of metal–ligand interactions was examined for six metals (Ca, Mg, Mn, Fe, Cu, Zn) using the Protein Data Bank and compared with information from accurately determined structures of relevant small-molecule crystals in the Cambridge Structural Database. Here, the environments of Na^+ and K^+ ions found in protein crystal structures are examined in an equivalent way. Target $M^+\cdots O$ distances are proposed and the agreement with observed distances is summarized. The commonest interactions are with water molecules and the next commonest with main-chain carbonyl O atoms.

1. Introduction

In previous papers (Harding, 1999, 2000, 2001) the geometry of metal–ligand interaction in proteins containing any of the metals Ca, Mg, Mn, Fe, Cu and Zn was examined using the Protein Data Bank (PDB; Bernstein *et al.*, 1977; Berman *et al.*, 2000) and compared with information from accurately determined structures of relevant small-molecule crystals in the Cambridge Structural Database (CSD; Allen & Kennard, 1993*a*,*b*). The small-molecule structures contained water molecules and ligands which were, as far as possible, analogues of the amino-acid side chains found in proteins. The objectives were to provide information for protein crystallography which might be useful (i) in the interpretation and fitting of models to electron-density maps, (ii) for target distances in restrained refinement, (iii) in the validation of protein structural data and possibly (iv) to contribute to the basic understanding of structure in relation to biological function. Sodium and potassium ions are often found to be present in protein crystal structures, even though the ions may be readily dissociated from the protein in solution – very much more readily than the ions of most of the other metals above. It therefore seemed useful to gather equivalent structural information from the CSD for these two ions and then examine the environments of Na^+ and K^+ in protein crystal structures in the PDB, even though the interactions of Na^+ and K^+ are predominantly electrostatic and it is not really appropriate to talk of ligands, donor groups or donor bonds.

2. Methods and procedures

CSD searches used the October 2000 or October 2001 release of the database. Otherwise, procedures were as those used for Ca and Mg (Harding, 1999), *e.g.* only structures with $R \leq 0.65$ Å were used. The distributions including all observed distances (Fig. 1; Table 1*a*) were obtained using the default CSD radii for the metals, which give Na—O, K—O and Ca—O 'bonds' of 1.65, 2.01 and 1.67 Å, respectively, and searching for all non-bonded contacts 0.4 Å (tolerance) or more longer than these out to the limit given. (Distances as short as these 'bonds' never occur, but their use as defaults in other searches allows sensible chemical connectivity of anions to be established.) For specific $M\cdots O$ distances (Table 1*b*) and for coordination-number estimation, the facility to recalculate connectivity with the MECALC instruction (*QUEST* version 5.17 and later) was used. Recommended target distances are given in Table 2(*a*), but for donor atoms which rarely if ever coordinate to Na^+ or K^+ these are necessarily only estimates based on comparisons with other metals.

Protein structures in the PDB were identified using the search procedure in the Jena Image Library hetero-components database at http://www.imb-jena.de/ImgLibPDB/. The local program *MP* and other smaller associated programs then extracted the coordination-group geometry from the PDB files (July 2001 release from RCSB; see Berman *et al.*, 2000), using the target distances of Table 2(*a*) and other procedures as in Harding (2000), except that in this case the primary coordination sphere included atoms at distances up to (target + 0.75 Å) from the metal and the secondary coordination sphere from here up to (target + 0.95 Å). Coordination-group geometry was examined for structures with resolution ≤ 1.6 Å (there were 30 structures containing 57 Na^+ ions and ten structures containing 22 K^+ ions) and checked against the

Table 1
Interatomic distances from the CSD for Na^+ and K^+ and some comparisons with Ca^{2+}..

(*a*) *M*···O distances: mean and sample standard deviation found from the distributions in Fig. 1 and the range over which the mean was evaluated.

	No. of observations	Mean *M*···O (Å)	Range for evaluation (Å)
Na···O	4315	2.42 (9)	2.01–2.70
K···O	4277	2.84 (12)	2.42–3.15
Ca···O	782	2.43 (9)	2.20–2.80

(*b*) Mean *M*···O distances for water molecules and carboxylate O atoms, with sample standard deviations.

	No. of observations	Mean (Å)	Observed range (Å)
$Na^+\cdots OH_2$	1149	2.41 (9)	2.13–2.76
$K^+\cdots OH_2$	446	2.85 (13)	2.52–3.20
$Na^+\cdots O_{carboxylate}$	178	2.40 (10)	2.23–2.74†
$K^+\cdots O_{carboxylate}$	268	2.77 (8)	2.59–2.97†

† Nearly all monodentate.

Table 2
Na^+, K^+ and the PDB.

(*a*) Recommended target distances (Å) for Na^+ and K^+ in proteins, based on distances in the CSD. As in Harding (2001), values in square brackets are estimates based on comparisons with other metal ions, since there are insufficient observations.

	Na^+	K^+
M···O in water, carboxylate, Ser, Thr, main-chain carbonyl groups	2.42	2.84
M···O in phenolate	[2.20]	[2.62]
M···N in imidazole	[2.38]	[2.80]
M···S in cysteine	[2.56]	[2.98]
M···O, N not in protein	2.42	2.84

(*b*) Number of observations (N_{obs}) in the PDB for structures with resolution ≤ 1.6 Å and agreement with targets (where the number of observations is sufficient). dif (Å) is (observed distance − target distance); the sample standard deviation is given.

	Na^+			K^+		
	N_{obs}	Mean dif	Dif range	N_{obs}	Mean dif	Dif range
M···OH_2	148	0.16 (29)	−0.5 to 0.9	49	0.16 (27)	−0.3 to 0.8
M···$O_{carboxylate}$†	28	0.27 (39)	−0.8 to 0.9	5	0.05 (32)	−0.4 to 0.5
M···$O_{main\text{-}chain\ carbonyl}$	70	0.04 (22)	−0.4 to 0.6	47	−0.07 (19)	−0.3 to 0.6
M···$O_{Ser/Thr}$	6	0.05 (24)	−0.2 to 0.5	2	−0.07 (18)	−0.2 to 0.1
M···O, all	261	0.15 (31)	−0.8 to 0.9	113	0.05 (25)	−0.4 to 0.8

† 24 monodentate Na^+ carboxylates and two bidentate; three monodentate K^+ carboxylates and one bidentate.

results for all structures with resolution ≤ 2.8 Å (328 structures including 565 Na^+ ions; 107 structures with 427 K^+ ions).

3. Results and discussion

3.1. Na^+ and K^+ in the CSD

The interactions found in the CSD are predominantly (∼90% in each case) with O atoms; these O atoms may be in anions or in water molecules or in other molecules where they retain some negative charge. The main other participants in interactions are N, F^- and Cl^-. Observed distributions of *M*···O distances are shown in Fig. 1 and are compared with those for Ca^{2+}; Table 1(*a*) gives details for these and Table 1(*b*) gives details for a small number of specific types of interactions. The distributions of Fig. 1 are for all *M*···O distances out to the limit given, with no attempt made to distinguish between a 'bond' and a 'non-bonded contact'. Ca^{2+} and Na^+ are very similar in size; K^+ is ∼0.4 Å larger. The peak in the K···O distribution is distinctly broader than that in the Na···O distribution, which is very similar to that in the Ca···O distribution. Even more importantly, substantial numbers of K···O distances are found in the range 3.2–3.8 Å and a modest number of Na···O distances in the equivalent range 2.8–3.4 Å, whereas there are very very few Ca···O distances in this range. This difference in behaviour is related to the even more electrostatic nature of the K^+ and Na^+ interactions, which results in greater flexibility in the cation environment. While for Ca^{2+} (and most of the metals studied earlier such as Mg^{2+} and Mn^{2+}) coordination number or coordination group can be clearly defined, for Na^+ and K^+ it is not so easy because an arbitrary decision must be made on the limit of the coordination sphere.

For Na^+, six is by far the commonest coordination number. When the limit of the coordination sphere is set at 2.72 Å (*i.e.* including the whole of the main peak in Fig. 1*a*), the numbers of Na^+ ions found with coordination numbers 4, 5, 6, 7 and 8 are 218, 460, 807, 115 and 37, respectively; coordination numbers <4 and >8 are also found. When the limit of the coordination sphere is increased to 3.01 Å, a substantial

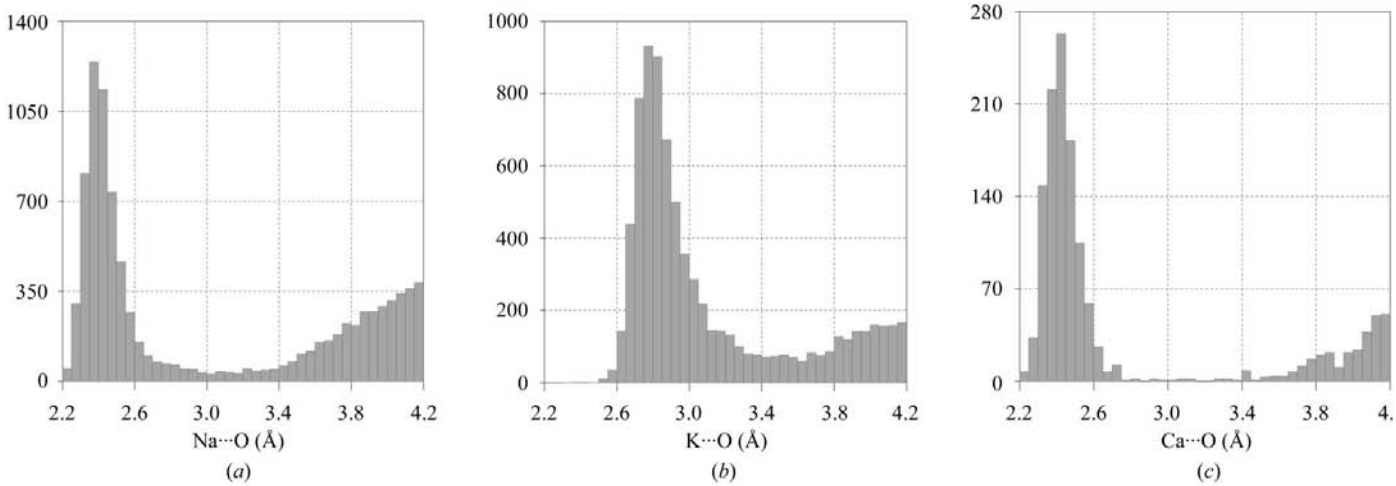

Figure 1
Distributions of $M^{n+}\cdots$O distances in crystals of small-molecule compounds (CSD) for O in all types of chemical combination. (*a*) Na···O distances, (*b*) K···O distances, (*c*) Ca···O distances

number of additional contacts are included and many coordination numbers are increased, but coordination number 6 is still the most favoured and is more than three times as common as any other. For K^+, eight is the commonest coordination number, with seven nearly as favourable. When the coordination sphere limit is set at 3.20 Å, the numbers of K^+ ions with coordination numbers 5, 6, 7, 8 and 9 are 90, 280, 373, 481 and 136, respectively; when the limit of the coordination sphere is increased to 3.43 Å, eight is still the commonest coordination number.

3.2. Na^+ and K^+ in the PDB

Table 2(*a*) gives the proposed target distances used in the analysis of sodium- and potassium-containing structures in the PDB. In Table 2(*b*) the ranges of metal–donor atom distances found in structures with resolution $\leq$ 1.6 Å are given and the distances are compared with the target values. The ranges are large, consistent with the findings in small-molecule structures. The mean observed distances are a little larger than the targets, except for main-chain carbonyl O as donor, but the differences are probably not significant. (The accurate location of Na^+ in protein structures is not easy, since its electron density is scarcely greater than that of an O atom; the number of observations for K^+ is fairly small.) For restraints and validation, therefore, these targets can be used, but as a result of the more ionic and more variable nature of the $M^+\cdots$O interaction discussed above, they will need to be used with a fairly large standard deviation such as 0.3 Å. Better still, the target should be asymmetric: a distance 0.5 Å less than the target is not acceptable, but a distance 0.5 Å greater than the target is certainly possible, though not common.

Coordination numbers found here for Na^+ and K^+ in proteins tend to be lower than those in small-molecule compounds. In the Na^+ structures with resolution $\leq$ 1.6 Å and with a coordination sphere limit of 3.17 Å, the numbers with coordination number 4, 5, 6 and 7 are 6, 19, 21 and 1, respectively; for K^+ with a coordination sphere limit of 3.59 Å, the numbers with coordination number 5, 6, 7 and 8 are 5, 6, 5 and 4, respectively. There are several possible reasons for the apparently low coordination numbers of Na^+ and K^+ in these protein structures: it could be that not all coordinating water molecules have been found or it could be because the program *MP* still does not search symmetry-related asymmetric units for contacts, but it also seems plausible that it is because it is difficult to fold the protein molecule, the multidentate ligand, in a way that achieves the high number of contacts commonly found in small-molecule structures.

For ionic interactions like these, directional covalency, such as in the regular tetrahedron for coordination number 4, is not to be expected. Ions or negatively charged groups will pack around the cation as closely as they can without becoming too close to each other.

The most common donor group when Na^+ or K^+ is associated with protein is the main-chain carbonyl group; the commonest participating amino acids are Ala, Gly, Leu, Ile, Val, Ser, Thr, Asp, and Asn, but there is at least one example for every other amino acid except Trp. It is common to find two or three different main-chain carbonyl O atoms interacting with K^+ or Na^+, often in addition to one or more carboxylate or amide O atoms and water molecules. Often the interacting amino-acid residues are close in the polypeptide chain (*e.g.* one, two or three residues from each other). Listings of all the 'coordination' groups occurring in the PDB can be found at http://tanna.bch.ed.ac.uk/.

4. Conclusions

Target distances for $K^+\cdots$O and $Na^+\cdots$O can be set (usually 2.42 and 2.84 Å, respectively; see Table 2*a*) for use as restraints or in validation, but it is important to recognize that because these interactions are predominantly electrostatic a wider range of distances is acceptable than in many of the $M\cdots$O interactions studied previously (Harding, 2001). Nevertheless, in protein crystal structures, K^+ and Na^+ ions are often found surrounded by a cluster of O atoms belonging to the protein molecule as well as water molecules.

At the University of Edinburgh, I am very grateful to Professor Malcolm Walkinshaw for access to computing facilities, to Dr Paul Taylor for computational support and to both for helpful discussions. I am also grateful to the Cambridge Crystallographic Data Centre, the Protein Data Bank and the European Bioinformatics Institute (at Hinxton, Cambridgeshire) for the databases used.

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. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Harding, M. M. (1999). *Acta Cryst.* D**55**, 1432–1443.
Harding, M. M. (2000). *Acta Cryst.* D**56**, 857–867.
Harding, M. M. (2001). *Acta Cryst.* D**57**, 401–411.