J. Mol. Biol. (1996) **256**, 228–234

JMB

COMMUNICATION

Valence Screening of Water in Protein Crystals Reveals Potential Na^+ Binding Sites

Murad Nayal and Enrico Di Cera*

Department of Biochemistry and Molecular Biophysics Washington University School of Medicine Box 8231, St. Louis MO 63110, USA

**Corresponding author*

Identification of Na^+ binding sites in protein crystals is complicated by the comparable electron density of this monovalent cation and water. Valence calculations can predict the location of metal ion binding sites in proteins with high precision. These calculations were used to screen 332,242 water molecules in 2742 protein structures reported in the Protein Data Bank (PDB), searching for molecules with Na^+-specific valence values $v_{Na^+} \geq 1.0$ v.u., as expected for a bound Na ion. Thirty-three water molecules (<0.01% of the total) were found to have $v_{Na^+} \geq 1.0$ v.u. and to be located within 3.5 Å from at least two protein oxygen atoms. These water molecules, with a high Na^+-specific valence, do not have valences specific for other cations, like Li^+, K^+, Mg^{2+} or Ca^{2+}. They belong to nine different proteins (deoxyribonuclease I, enolase, hen egg-white lysozyme, human lysozyme, phospholipase A2, proteinase A, rubredoxin, thrombin and phage T4 lysozyme) and appear with similar coordination geometry, typically octahedral, in the same place in multiple crystal structure determinations of the same protein. In the case of thrombin, the water molecule singled out by valence calculations is, in fact, a bound Na ion as demonstrated by molecular replacement with Rb^+. Valence calculations provide an accurate screening of water in protein crystals and may help identify Na^+ binding sites of functional importance.

Keywords: computational biology; lysozyme; molecular recognition; monovalent cations; thrombin; water

Specific monovalent cation effects in biological macromolecules are widespread and of utmost physiological relevance (Ayala & Di Cera, 1994; Hohenester *et al*., 1994; Lu & Draper, 1994; Overman & Lohman, 1994; O'Brien & McKay, 1995; Woehl & Dunn, 1995). In over 100 enzymes, Na^+ or K^+ are required for optimal catalysis and act as allosteric effectors that change the structure of the whole protein, or as cofactors binding to specific substrates (Suelter, 1970). Structural information on the molecular origin of these effects has begun to emerge for enzymes activated by K^+ (Toney *et al*., 1993; Larsen *et al*., 1994; Wilbanks & McKay, 1995). Although fewer in number, enzymes activated by Na^+ are involved in key regulatory interactions, as demonstrated by blood clotting factors (Orthner & Kosow, 1978, 1980; Steiner *et al*., 1980; Wells & Di Cera, 1992). Thrombin represents a spectacular example of how Na^+-specific effects may trigger molecular events leading to important physiological functions. In fact, binding of Na^+ to a single site (Di Cera *et al*., 1995), allosterically switches this enzyme from an anticoagulant to a procoagulant factor (Dang *et al*., 1995).

Unlike the case of K^+, Na^+ binding sites in proteins are difficult to locate crystallographically due to the identical number of electrons carried by this metal ion and a water molecule. Hence, an electron density due to a bound Na ion can easily be attributed to a water oxygen atom in the refinement procedure. In view of this difficulty and of the importance of Na^+-specific effects in blood clotting factors and other enzymes, it is important to develop methods that can identify water molecules that might have been misassigned in crystal structures. An algorithm that systematically screens all water molecules in a protein crystal and singles out those molecules that may be, in fact, Na ions would be of great practical utility. Identification of such molecules may prompt further investigation in terms of more direct methods, like molecular replacement, that can identify the bound Na ion unequivocally.

Abbreviations used: PDB, Protein Data Bank; v.u., valence units.

0022–2836/96/070228–07 $12.00/0

A number of computational algorithms to predict metal ion binding sites in proteins have been developed (Yamashita *et al*., 1990; Nayal & Di Cera, 1994; Bagley & Altman, 1995). Valence calculations are particularly accurate (Nayal & Di Cera, 1994) because they exploit an empirical expression between the bond length and the bond strength of a metal ion–oxygen pair parameterized from results obtained in the extensive analysis of structures of metal oxides (Brown & Wu, 1976; Brown, 1992). The expression returns values of the valence for any accessible point in the crystal structure and singles out regions of the protein where a given metal ion is likely to be found. In this study, we show that valence calculations can be used to screen protein crystals to identify bound Na ions that might have been misassigned as solvent molecules.

All protein structures currently in the PDB were included in the analysis. The sample contained 2742 protein structures, with a total of 342,201 assigned water molecules. Only water molecules that could be replaced with a Na ion without significant overlap of van der Waals radii with neighbor atoms were considered. In 9959 cases, the overlap was found to be >10% and the water molecules were disregarded. The Na^+-specific valence of each water molecule was calculated using the equation (Brown & Wu, 1976; Brown, 1992; Nayal & Di Cera, 1994):

$$v_{Na^+} = \sum_{j=1}^{M} v_j = \sum_{j=1}^{M} \left(\frac{R_j}{R_0}\right)^{-N} \qquad (1)$$

where v_j is the valence contributed by the jth ligating oxygen in the coordination shell located at a distance R_j, and M is the total number of oxygen atoms within 4.0 Å. The parameters R_0 and N translate the bond length into the bond strength, or valence, and are specific of a given metal ion-oxygen pair. The values for the Na^+-O pair are $R_0 = 1.622$ and N = 4.29 (Brown & Wu, 1976). At the cutoff distance of 4.0 Å, the valence contributed by a ligating oxygen atom is only 0.02 v.u., or 2% of the value of 1.0 v.u. expected for Na^+. In a coordination shell where all M ligating oxygen atoms are equally disposed around the metal ion, the theoretical distance for the metal ion–oxygen pair, R_{th}, can be derived from equation (1) as:

$$R_{th} = R_0\left(\frac{v_{th}}{M}\right)^{-1/N} \qquad (2)$$

The value of R_{th} expected for a Na^+-O pair in an octahedral coordination shell ($M = 6$) is 2.46 Å. Equation (1) was also used to screen solvent molecules for other cations, and specifically Li^+, K^+, Mg^{2+} and Ca^{2+}. This was done to test the sensitivity of the method and the robustness of the assignments derived from the screening for Na^+. The values of R_0 and N for the cation–O pair in the various cases were derived from the literature (Brown & Wu, 1976). The program WASP (WAter Screening Program) for screening water molecules in a crystal structure for different monovalent and divalent cations using the valence approach is available upon request to enrico@caesar.wustl.edu.

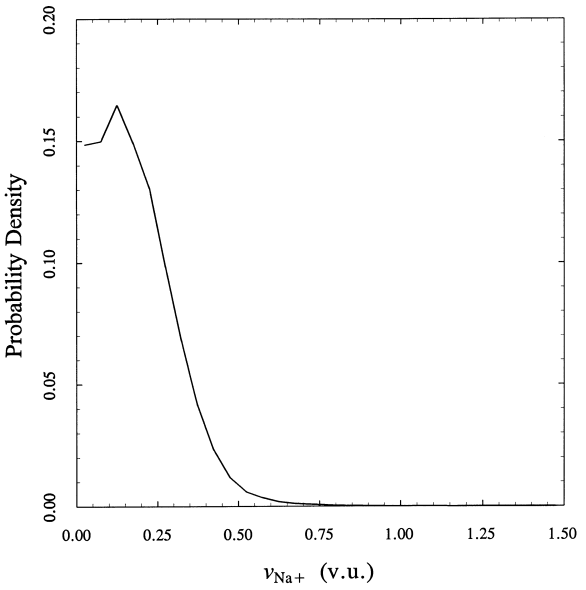

Figure 1. Distribution of v_{Na^+} values for 332,242 water molecules in 2742 protein crystal structures currently in the PDB.

Figure 1 shows the distribution of v_{Na^+} values calculated for 332,242 water molecules in the protein sample studied. The average value of v_{Na^+} is 0.18 v.u., with a standard deviation of 0.13 v.u., and is significantly lower than the value of 1.0 v.u. expected for Na^+†. Sixty-four water molecules (0.02% of the total) have $v_{Na^+} \geqslant 1.0$ v.u., whereas none of the water molecules in the PDB sample has $v_{Li^+} \geqslant 1.0$ v.u., $v_{K^+} \geqslant 1.0$ v.u., $v_{Mg^{2+}} \geqslant 2.0$ v.u., or $v_{Ca^+} \geqslant 2.0$ v.u. This indicates that the valence method is extremely sensitive to the specific ion screened for and provides robust predictions on bound Na ions misassigned as water molecules. The water molecules with $v_{Na^+} \geqslant 1.0$ v.u. show noteworthy differences compared to the rest of the sample (see Figure 2). They are surrounded almost exclusively by oxygen atoms up to 3.4 Å. On the other hand, water molecules with $v_{Na^+} < 1.0$ v.u. tend to be surrounded also by nitrogen atoms peaking at 3.0 Å. In both cases, carbon atoms populate an outer shell around 4.0 Å. These atoms provide support for oxygen and nitrogen atoms, or are directly hydrated by water molecules (Blake *et al*., 1983). A bound Na ion should be surrounded by oxygen atoms only, in view of its nature as a weak acid (Brown & Skowron, 1990; Brown, 1992). A water molecule, on the other hand, can engage in hydrogen bonding interactions with either oxygen or nitrogen atoms.

† This distribution is quite robust and is not biased by the atomic resolution of the crystal structures. Extensive Monte Carlo calculations were carried out introducing a Gaussian pseudo-random error with zero mean and $\sigma = 0.3$ Å in the atomic coordinates of the water molecules. This resulted in perturbations of the v_{Na^+} values of $\leqslant 10\%$.

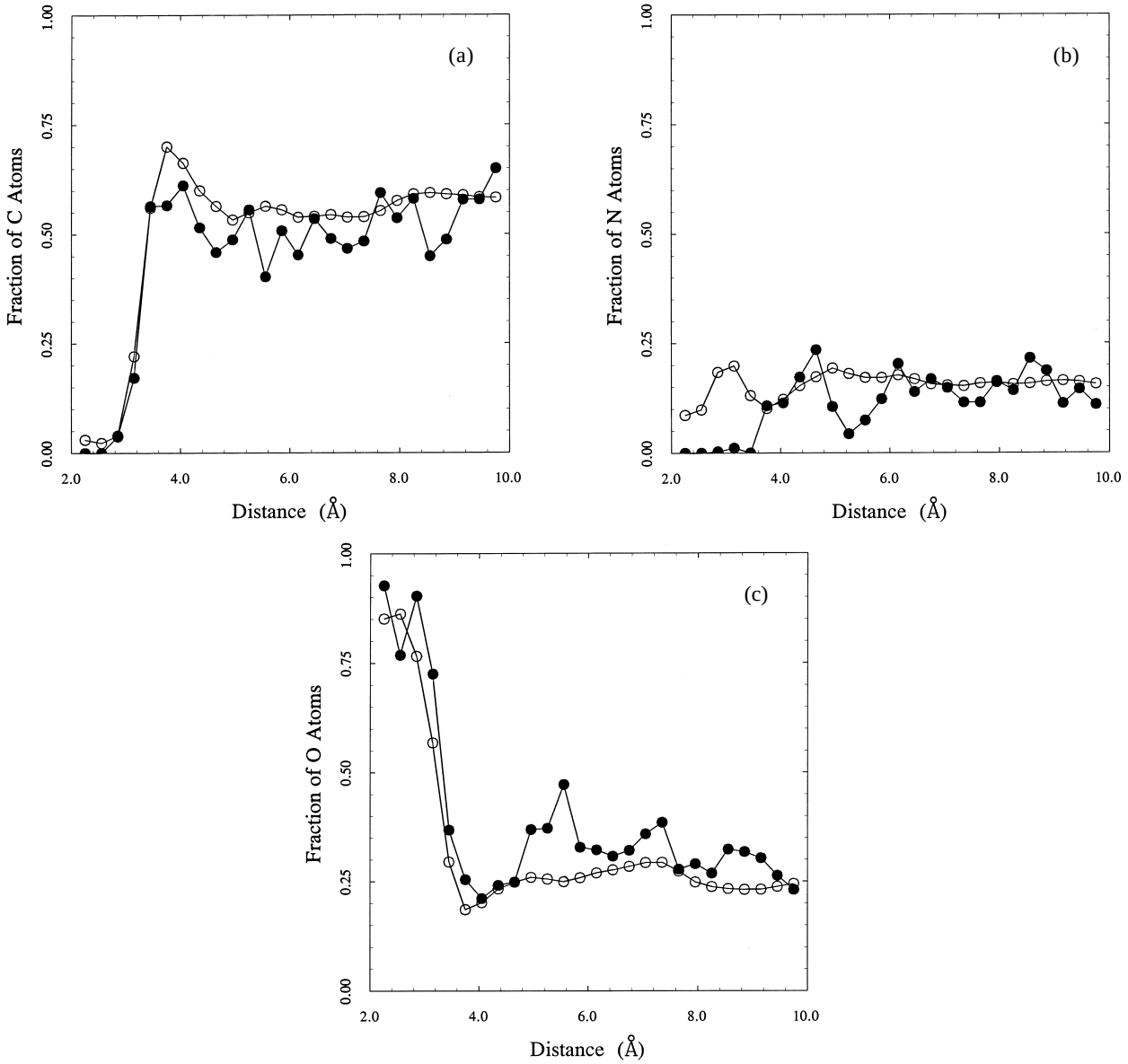

Figure 2. Atom distribution in the environment of water molecules in protein crystals. Water molecules with $v_{Na^+} \geqslant 1.0$ v.u. (●), or $v_{Na^+} < 1.0$ v.u. (○). Atoms are C (a), N (b), O (c). Note the shift in the peak of N atoms around the water molecules with $v_{Na^+} \geqslant 1.0$ v.u.

Hence, the results in Figure 2 may be revealing of Na ions misassigned as water molecules in the refinement procedure.

The 64 water molecules with $v_{Na^+} \geqslant 1.0$ v.u. belong to 11 different proteins and appear with similar coordination and in the same position in multiple determinations of the structure of the same protein. Thirty-one of these water molecules are surrounded by at most one oxygen from the protein within 3.5 Å and are found in the crystal structures of rubredoxin (7rxn), insulin (4ins), apo-liver alcohol dehydrogenase (8adh), and seven structures of mutants of bacteriophage T4 lysozyme (109l, 112l, 113l, 1l87, 1l95, 1l96, 224l). The other 33 water molecules with $v_{Na^+} \geqslant 1.0$ v.u. have at least two protein oxygen atoms within 3.5 Å, which supports the possibility of these molecules being, in fact, Na ions specifically bound to the protein. They are found in nine different proteins (see Table 1) and appear repeatedly in the same position in multiple determinations of the structure of the same protein†.

† The cutoff of $v_{Na^+} \geqslant 1.0$ v.u. effectively singles out most of the proteins with a water molecule scoring high Na^+-specific valence. Using a cutoff of $v_{Na^+} \geqslant 0.90$ v.u., and retaining the criterion of the water being surrounded by at least two oxygen atoms from the protein within 3.5 Å, singles out a total of 99 water molecules in structures that belong to 14 different proteins. In addition to the nine proteins listed in Table 1, these water molecules are found in β-lactamase (water 432 in 1blc.pdb), proteinase K (water 412 in 1pek.pdb), subtilisin (water 300 in 1s01.pdb), transthyretin (water 683 in 1ttc.pdb) and D-xylose isomerase (water 641 in 1xib.pdb; water 690 in 1xin.pdb).

When screened for other monovalent and divalent cations, in no case is the valence of these water molecules found close to the expected value (see Table 1). K^+ is excluded in all cases because of the overlap of the van der Waals radii with the oxygen atoms in the coordination shell. Li^+ yields $\nu_{Li^+} < 1.0$ v.u. in all cases. Likewise, Mg^{2+} and Ca^{2+} score well below 2.0 v.u. and Ca^{2+} is excluded in many cases because of the overlap of van der Waals radii. The temperature factor of these water molecules is a poor predictor of the presence of a Na ion and does not correlate with the Na^+-specific valence. In most cases the value is actually above the average temperature factor of the protein. On the other hand, both the number of oxygen atoms surrounding these water molecules and the average distance in the coordination shell signal the presence of a bound Na ion.

Some key examples of these water molecules are given in Figure 3. In deoxyribonuclease I (2dnj), the water with $\nu_{Na^+} \geqslant 1.0$ v.u. has six nearest-neighbors, two oxygen atoms from the protein and four water molecules, arranged in an octahedral geometry. The average distance from the central water molecule is 2.48 Å, a value practically identical to R_{th} for Na^+. A similar architecture is found in the case of thrombin (1thr), where the central water molecule is octahedrally coordinated by two oxygen atoms from the protein and four water molecules, with an average distance of 2.35 Å. The water molecule

Table 1. Coordination shell and properties of water molecules with $\nu_{Na^+} \geqslant 1.0$ v.u.

Protein	r[a]	wat[b]	B[c]	ν_{Na^+}	ν_{Li^+}	ν_{K^+}	$\nu_{Mg^{2+}}$	$\nu_{Ca^{2+}}$	N[d]	R[e]	Atoms in the coordination shell[f]
A. *Deoxyribonuclease I*											
2dnj	2.0	539	7.0 (20.0)	1.01	0.58	n.a.[g]	1.01	1.51	6	2.48	D172 $O^{\delta 2}$ (2.51), D198 O (2.37), 523 O (2.56), 546 O (2.47), 547 O (2.45), 556 O (2.51)
B. *Enolase*											
1nel	2.6	755	7.4 (20.9)	1.01	0.57	n.a.[g]	1.01	1.63	5	2.45	E377 $O^{\epsilon 1}$ (2.83), E377 $O^{\epsilon 2}$ (2.24), 453 O (2.24), 493 O (2.78), 495 O (2.15)
C. *Hen egg-white lysozyme*											
6lyt	1.9	136	13.6 (16.8)	1.09	0.61	n.a.[g]	1.09	n.a.[g]	6	2.49	S60 O (2.13), C64 O (2.18), S72 O^{γ} (2.74), R73 O (2.40), 141 O (2.70), 148 O (2.77)
D. *Human lysozyme*											
1lhh	1.8	140	9.0 (12.7)	1.01	0.57	n.a.[g]	1.01	1.61	5	2.36	S61 O (2.26), C65 O (2.34), V74 O (2.48), 162 O (2.43), 225 O (2.30)
1lhi	1.8	104	19.2 (17.2)	1.04	0.58	n.a.[g]	1.04	1.67	6	2.64	S61 O (2.30), R62 O (3.95), C65 O (2.26), V74 O (2.23), 100 O (2.76), 110 O (2.37)
1lhm	1.8	87	19.0 (17.0)	1.00	0.56	n.a.[g]	1.00	1.58	6	2.65	S61 O (2.27), R62 O (3.89), C65 O (2.44), V74 O (2.23), 84 O (2.71), 94 O (2.36)
1lz5	1.8	231	8.0 (14.6)	1.07	0.60	n.a.[g]	1.07	1.72	6	2.59	S61 O (2.27), R62 O (3.79), C65 O (2.28), S74 O (2.26), 228 O (2.63), 239 O (2.33)
2lhm	1.8	117	7.2 (13.9)	1.06	0.59	n.a.[g]	1.06	1.72	5	2.33	S61 O (2.24), C65 O (2.35), V74 O (2.34), 115 O (2.36), 125 O (2.36)
E. *Phospholipase A2*											
1pod	2.1	236	4.6 (30.6)	1.06	0.59	n.a.[g]	1.06	n.a.[g]	5	2.39	H27 O (2.22), G29 O (2.24), G31 O (2.34), D48 $O^{\delta 1}$ (2.92), D48 $O^{\delta 2}$ (2.22)
F. *Proteinase A*											
1sgc	1.8	45	4.8 (14.4)	1.06	0.61	n.a.[g]	1.06	1.57	7	2.57	G116 O (2.63), D123 $O^{\delta 1}$ (2.77), I124 O (2.31), L242 O (2.87), L242 O^{ter} (2.40), 39 O (2.42), 53 O (2.59)
3sga	1.8	229	1.8 (14.2)	1.12	0.63	n.a.[g]	1.12	n.a.[g]	7	2.55	G116 O (2.47), D123 $O^{\delta 1}$ (2.79), I124 O (2.14), L242 O (2.79), L242 O^{ter} (2.43), 224 O (2.38), 237 O (2.88)
4sga	1.8	229	6.0 (14.1)	1.09	0.62	n.a.[g]	1.09	1.66	7	2.54	G116 O (2.41), D123 $O^{\delta 1}$ (2.81), I124 O (2.21), L242 O (2.83), L242 O^{ter} (2.45), 224 O (2.51), 237 O (2.59)
5sga	1.8	229	4.5 (14.4)	1.11	0.63	n.a.[g]	1.11	1.70	7	2.54	G116 O (2.43), D123 $O^{\delta 1}$ (2.81), I124 0 (2.19), L242 O (2.83), L242 O^{ter} (2.37), 224 O (2.54), 237 O (2.58)
2sga	1.5	37	25.9 (15.9)	1.07	0.61	n.a.[g]	1.07	n.a.[g]	7	2.75	S195 O^{γ} (3.07), S214 O (3.40), 32 O (2.95), 72 O (2.14), 82 O (2.36), 94 O (2.38), 148 O (2.95)
G. *Rubredoxin*											
7rxn	1.5	315	28.6 (16.1)	1.13	0.66	n.a.[g]	1.13	1.54	7	2.71	E12 O (2.76), E17 $O^{\epsilon 1}$ (3.49), 324 O (2.70), 406 O (2.63), 415 O (2.37), 417 O (2.51), 432 O (2.51)
7rxn	1.5	407	32.3 (16.1)	1.20	0.69	n.a.[g]	1.20	1.70	8	2.74	E12 $O^{\epsilon 1}$ (2.51), E12 $O^{\epsilon 2}$ (3.31), 221 O (2.48), 414 O (2.67), 426 O (2.67), 438 O (2.43), 441 O (2.49), 465 O (2.63)
H. *Thrombin*											
1hgt	2.2	540	27.9 (28.6)	1.10	0.62	n.a.[g]	1.10	n.a.[g]	5	2.43	R221a O (2.37), K224 O (2.21), 708 O (2.96), 744 O (2.48), 974 O (2.12)
1thr	2.3	467	13.0 (28.2)	1.33	0.74	n.a.[g]	1.33	n.a.[g]	6	2.35	R221a O (2.26), K224 O (2.24), 469 O (2.81), 487 O (2.40), 545 O (2.14), 662 O (2.22)

continued

Table 1. *continued*

Protein	r[a]	wat[b]	B[c]	v_{Na^+}	v_{Li^+}	v_{K^+}	$v_{Mg^{2+}}$	$v_{Ca^{2+}}$	N[d]	R[e]	Atoms in the coordination shell[f]
I. *T4 lysozyme*											
110l	1.7	267	31.6 (21.6)	1.01	0.58	n.a.[g]	1.01	1.54	6	2.58	G30 O (2.81), F104 O (2.80), 224 O (2.27), 225 O (2.29), 266 O (2.54), 299 O (2.75)
112l	1.8	267	29.7 (24.6)	1.01	0.58	n.a.[g]	1.01	1.48	6	2.59	G30 O (2.70), F104 O (2.78), 224 O (2.28), 225 O (2.28), 266 O (2.54), 299 O (2.98)
1dyc	2.1	267	50.1 (39.3)	1.02	0.58	n.a.[g]	1.02	1.51	6	2.58	G30 O (2.74), F104 O (2.86), 224 O (2.30), 225 O (2.28), 266 O (2.44), 299 O (2.85)
1l22	1.7	353	41.9 (31.6)	1.03	0.59	n.a.[g]	1.03	1.54	7	2.62	G30 O (2.78), F104 O (2.97), 242 O (2.75), 243 O (2.29), 351 O (2.30), 352 O (2.38), 533 O (2.89)
1l54	1.9	353	34.0 (29.5)	1.02	0.58	n.a.[g]	1.02	1.53	7	2.62	G30 O (2.85), F104 O (2.97), 242 O (2.90), 243 O (2.30), 351 O (2.28), 352 O (2.40), 533 O (2.62)
1l62	1.7	267	33.6 (28.3)	1.01	0.58	n.a.[g]	1.01	1.49	7	2.65	G30 O (2.90), F104 O (2.78), 224 O (2.34), 225 O (2.28), 265 O (2.88), 266 O (2.40), 299 O (2.97)
1l63	1.75	267	37.4 (27.7)	1.00	0.57	n.a.[g]	1.00	1.47	6	2.60	G30 O (2.71), F104 O (2.86), 224 O (2.30), 225 O (2.28), 265 O (2.91), 266 O (2.53)
1l75	1.9	267	31.9 (32.9)	1.00	0.57	n.a.[g]	1.00	1.48	6	2.54	G30 O (2.65), F104 O (2.80), 224 O (2.55), 225 O (2.35), 256 O (2.29), 265 O (2.62)
1l87	1.8	267	41.2 (31.2)	1.04	0.59	n.a.[g]	1.04	1.54	5	2.52	G30 O (2.68), F104 O (2.97), 224 O (2.29), 225 O (2.26), 266 O (2.40)
1l96	2.0	267	24.4 (25.3)	1.04	0.59	n.a.[g]	1.04	1.51	7	2.61	G30 O (2.49), F104 O (3.06), 224 O (2.31), 225 O (2.41), 265 O (2.61), 266 O (2.67), 299 O (2.69)
1lyg	1.8	267	38.6 (28.5)	1.03	0.59	n.a.[g]	1.03	1.52	7	2.64	G30 O (2.72), F104 O (2.94), 224 O (2.36), 225 O (2.30), 265 O (2.85), 266 O (2.35), 299 O (2.97)
1lyh	1.7	267	33.3 (28.3)	1.05	0.60	n.a.[g]	1.05	1.56	7	2.61	G30 O (2.77), F104 O (2.83), 224 O (2.30), 225 O (2.30), 265 O (2.73), 266 O (2.40), 299 O (2.94)
1lyj	1.8	267	40.8 (26.4)	1.01	0.58	n.a.[g]	1.01	1.47	7	2.61	G30 O (2.72), F104 O (2.87), 224 O (2.40), 225 O (2.35), 265 O (2.58), 266 O (2.45), 299 O (2.93)
1lyi	2.0	267	33.9 (24.5)	1.00	0.57	n.a.[g]	1.00	1.48	6	2.60	G30 O (2.83), F104 0 (2.86), 224 0 (2.37), 225 0 (2.31), 265 O (2.95), 266 O (2.30)
7lzm	1.8	267	49.5 (25.8)	1.01	0.57	n.a.[g]	1.01	1.49	7	2.66	G30 O (2.38), F104 O (3.25), 224 O (2.71), 225 O (2.24), 265 O (2.80), 266 O (2.43), 299 O (2.80)

[a] Resolution (in Å) of the crystal structure.
[b] Number of the water molecule with $v_{Na^+} \geqslant 1.0$ v.u. in the PDB file.
[c] Temperature factor of the water molecule with $v_{Na^+} \geqslant 1.0$ v.u.; the average temperature factor for the protein is given in parenthesis.
[d] Number of oxygen ligands in the coordination shell.
[e] Average distance (in Å) in the coordination shell.
[f] Distance (in Å) from the water molecule with $v_{Na^+} \geqslant 1.0$ v.u. is given in parenthesis.
[g] Not allowed because of overlap of van der Waals radii. The values of v_{Na^+} and $v_{Mg^{2+}}$ are identical because of the identical values of R_0 and N in equation (1) for these two cations (Brown & Wu, 1976).

singled out by valence calculations is, in this case, a Na ion as recently demonstrated by molecular replacement with Rb^+ (Di Cera *et al*., 1995). In the case of hen egg-white lysozyme (6lyt), the water molecule with $v_{Na^+} \geqslant 1.0$ v.u. has six nearest-neighbors, four protein oxygen atoms and two water molecules, arranged again in an octahedral geometry. A similar architecture is observed in human lysozyme (see Table 1). The average distance from the central water molecule is 2.49 Å, with S60 O and C64 O less than 2.2 Å away. Phospholipase A2 (1pod) is unusual in so far as the water molecule with $v_{Na^+} \geqslant 1.0$ v.u. has five nearest-neighbors, all from the protein, arranged in an incomplete octahedron. The coordination shell also contains the charged carboxylate oxygen atoms of D48, one of which is in very close contact with the central water molecule. In the case of proteinase A (5sga), there are two water molecules with $v_{Na^+} \geqslant 1.0$ v.u. One is in the active site region, close to S195 O^γ and S214 O, and is surrounded by five water molecules. The other site (shown in Figure 3) has seven nearest-neighbors, five from the protein and two water molecules, arranged in a distorted octahedral geometry reminiscent of canonical EF-hands (Glusker, 1991). The coordination shell contains three charged carboxylate oxygen atoms. Although assigned as a water molecule, the possibility of this molecule being a Na ion instead was mentioned in the original refinement procedure (Sielecki *et al*., 1979).

Our analysis predicts that in at least nine proteins, listed in Table 1, one of the assigned water molecules is a specifically bound Na ion. Our predictions should be tested directly by molecular replacement. In the case of thrombin, molecular

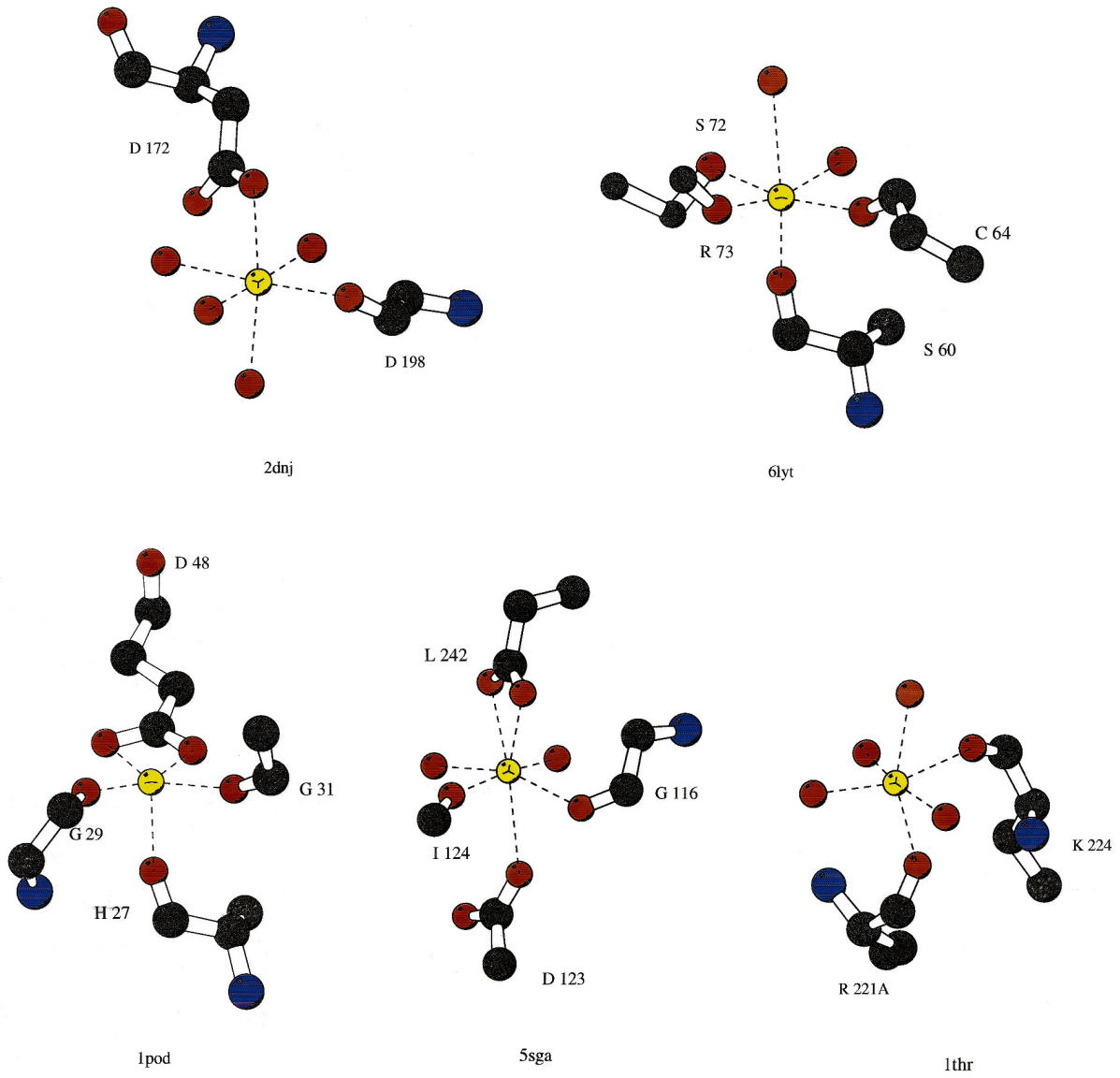

Figure 3. Potential Na^+ binding sites of: deoxyribonuclease I (2dnj), hen egg-white lysozyme (6lyt), phospholipase A2 (1pod), proteinase A (5sga), and thrombin (1thr). Valence calculations suggest that the central water molecule (yellow), assigned as such in the refinement procedure, is instead a specifically bound Na ion. The number of the central water molecule is given in Table 1 (see column under wat). Atoms are color-coded as follows: O (red); N (blue); C (black).

replacement with Rb^+ has recently shown that the water molecule with $v_{NA^+} \geqslant 1.0$ v.u. (see Table 1) is indeed a Na ion. This is the bound Na ion responsible for the slow → fast transition that is crucial for the physiological functions of the enzyme (Dang *et al*., 1995). Whether similarly important Na^+-specific effects exist for lysozyme, rubredoxin and the other proteins singled out by our valence calculations remains to be established by experimental studies. As the number of crystal structures deposited in the PDB increases, so does the need for accurate screening of water molecules. Valence calculations provide a powerful and sensitive tool for the identification of proteins where Na^+ binding sites of functional importance may have been missed in the original refinement.

Acknowledgements

This work was supported in part by NIH research grant HL49413, NSF research grant MCB94-06103 and by a grant from the American Heart Association. E.D.C. is an established investigator of the American Heart Association and Genentech.

References

Ayala, Y. & Di Cera, E. (1994). Molecular recognition by thrombin: role of the slow → fast transition, site-specific ion binding energetics and thermodynamic mapping of structural components. *J. Mol. Biol.* **235**, 733–746.

Bagley, S. C. & Altman, R. B. (1995). Characterizing the microenvironment surrounding protein sites. *Protein Sci.* **4**, 622–635.

Blake, C. C. F., Pulford, W. C. A. & Artymiuk, P. J. (1983). X-ray studies of water in crystals of lysozyme. *J. Mol. Biol.* **167**, 693–723.

Brown, I. D. (1992). Chemical and steric constraints in inorganic solids. *Acta Crystallog. sect. B*, **48**, 553–572.

Brown, I. D. & Skowron, A. (1990). Electronegativity and Lewis acid strength. *J. Am. Chem. Soc.* **112**, 3401–3403.

Brown, I. D. & Wu, K. K. (1976). Empirical parameters for calculating cation-oxygen bond valence. *Acta Crystallog. sect. B*, **32**, 1957–1959.

Dang, Q. D., Vindigni, A. & Di Cera, E. (1995). An allosteric switch controls the procoagulant and anticoagulant activities of thrombin. *Proc. Natl Acad. Sci. USA*, **92**, 5977–5981.

Di Cera, E., Guinto, E. R., Vindigni, A., Dang, Q. D., Ayala, Y. M., Wuyi, M. & Tulinsky, A. (1995). The Na^+ binding site of thrombin. *J. Biol. Chem.* **270**, 22089–22092.

Glusker, J. P. (1991). Structural aspects of metal liganding to functional groups in proteins. *Advan. Protein Chem.* **42**, 1–75.

Hohenester, E., Keller, J. W. & Jansonius, J. N. (1994). An alkali metal ion size-dependent switch in the active site structure of dialkylglycine decarboxylase. *Biochemistry*, **33**, 13561–13570.

Larsen, T. M., Laughlin, L. T., Holden, H. M., Rayment, I. & Reed, G. H. (1994). Structure of rabbit muscle pyruvate kinase complexed with Mn^{2+}, K^+, and pyruvate. *Biochemistry*, **33**, 6301–6309.

Lu, M. & Draper, D. E. (1994). Bases defining an ammonium and magnesium ion-dependent tertiary structure within the large subunit ribosomal RNA. *J. Mol. Biol.* **244**, 572–585.

Nayal, M. & Di Cera, E. (1994). Predicting Ca^{2+}-binding sites in proteins. *Proc. Natl Acad. Sci. USA*, **91**, 817–821.

O'Brien, M. C. & McKay, D. B. (1995). How potassium affects the activity of the molecular chaperone Hsc70. I. Potassium is required for optimal ATPase activity. *J. Biol. Chem.* **270**, 2247–2250.

Orthner, C. L. & Kosow, D. P. (1978). The effect of metal ions on the amidolytic activity of human factor Xa (activated Stuart-Prower factor). *Arch. Biochem. Biophys.* **185**, 400–406.

Orthner, C. L. & Kosow, D. P. (1980). Evidence that human α-thrombin is a monovalent cation-activated enzyme. *Arch. Biochem. Biophys.* **202**, 63–75.

Overman, L. B. & Lohman, T. M. (1994). Linkage of pH, anion and cation effects in protein-nucleic acid equilibria. *Escherichia coli* SSB protein-single stranded nucleic acid interactions. *J. Mol. Biol.* **236**, 165–178.

Sielecki, A. R., Hendrickson, W. A., Broughton, C. G., Beldaere, L. T. J., Brayer, G. D. & James, M. N. G. (1979). Protein structure refinement: Streptomyces griseus serine protease A at 1.8 Å resolution. *J. Mol. Biol.* **134**, 781–804.

Steiner, S. A., Amphlett, G. N. & Castellino, F. J. (1980). Stimulation of the amidase and esterase activity of activated bovine plasma protein C by monovalent cations. *Biochem. Biophys. Res. Commun.* **94**, 340–347.

Suelter, C. H. (1970). Enzymes activated by monovalent cations. *Science*, **168**, 789–795.

Toney, M. D., Hohenester, E., Cowan, S. W. & Jansonius, J. N. (1993). Dialkylglycine decarboxylase structure: bifunctional active site and alkali metal sites. *Science*, **261**, 756–759.

Wells, C. M. & Di Cera, E. (1992). Thrombin is a Na^+-activated enzyme. *Biochemistry*, **31**, 11721–11730.

Wilbanks, S. M. & McKay, D. B. (1995). How potassium affects the activity of the molecular chaperone Hsc70. II. Potassium binds specifically in the ATPase active site. *J. Biol. Chem.* **270**, 2251–2257.

Woehl, E. U. & Dunn, M. F. (1995). Monovalent metal ions play an essential role in catalysis and intersubunit communication in the tryptophan synthase bienzyme complex. *Biochemistry*, **34**, 9466–9476.

Yamashita, M. M., Wesson, L., Eisenman, G. & Eisenberg, D. (1990). Where metal ions bind in proteins. *Proc. Natl Acad. Sci. USA*, **87**, 5648–5652.

Edited by I. A. Wilson

(Received 16 August 1995; accepted in revised form 20 November 1995)