Analysis of Protein Metal Binding Selectivity in a Cluster Model[†]

M. Krauss*,[‡] **and W. J. Stevens**[§]

Contribution from the Molecular Spectroscopy Division and Center for Advanced Research in Biotechnology, National Institute of Standards and Technology, Gaithersburg, Maryland 20899.

Received March 20, 1989

Abstract: Ab initio molecular orbital calculations of the binding energy of metal cations to octahedral clusters of water, formamide, and formate ligands are used to analyze Ca binding sites in proteins. The intrinsic energetics of the first coordination shell provide a basis for evaluating the conformation behavior and the selectivity of cation binding. The enthalpies of binding are modeled by estimating the environmental polarization energy relative to the model cluster of the first shell. Cluster reaction energies are calculated for transferring Mg, Ca, and Na cations from a water cluster to the protein model cluster as a function of metal–ligand distance which is found to strongly affect cation binding selectivity between cations of the same and different charges. The selectivity is a function of both steric and electrostatic interactions. As is well-known selectivity between cations of the same charge is dependent on steric factors, but there is also an electrostatic component, independent of steric influences, which is selective between cations of different charge. The data are applied to an analysis of two Ca binding sites in the protein, Subtilisin BPN′. The cluster model shows that ion desolvation energies determine the ion selectivity between Mg and Ca in the buried site. Interligand repulsion in the octahedral cluster prevents the larger formamide ligand from approaching the smaller Mg cation as closely as it does for single ligand binding. For the surface binding site, competition between cations of different charge, Ca and Na, can be understood in terms of the rigidity of the site cavity.

1. Introduction

The binding of metal cations to a site consisting of neutral ligands is well understood.[1] If we restrict our attention to Group I and II ions where there is no directed valence interaction, the binding is governed by the size of the cation and the electrostatic, steric, and inductive interactions among the cation and the ligands. In order to rely on the energetics and conformation of a single-shell cluster to emulate the first coordination sphere in water or a protein, the interaction energy of the cation with the first-shell ligands must be significantly larger than water–water or protein ligand–ligand interactions. The energy-optimized structure of the first-shell cluster can then be used without recourse to a statistical analysis, since the interaction of the cluster with its environment will not have a large effect on the energy or conformation of the cluster. A model for calculating the enthalpy of hydration then evaluates the first coordination shell energetics with optimal cation-to-ligand cluster orientations[2,3] and uses the Born model[4] for the polarization interaction of the first-shell cluster with the environment. The intrinsic energetics in the first coordination shell of idealized cluster models of the protein binding sites and the hydrated cations will be obtained with quantum mechanical calculations in order to accurately account for the large quantum and many-body effects. Charge transfer is not calculated to be large by a population analysis, but such effects also have to be treated with ab initio methods. The hydration energies can be obtained as a function of the first-shell coordination number, which provides insight into both the electronic and geometrical structure of the hydrated ions.

In proteins the metal binding site invariably contains an anionic residue. The strong ionic interactions can alter the factors that have been proposed as the basis for the conformational behavior and selectivity among the cations when they are bound to neutral sites. The three most common binding ligands in a protein are the carboxylate moiety, the peptide–carbonyl moiety, and the water molecule.[5] Even though coordination polyhedra with a coordination number of seven or greater are common for Ca, many sites are roughly octahedral with the larger polyhedra containing several water molecules. The carboxylate moiety can bind in either a monodentate or bidentate mode. It is common to count the coordination number for the bidentate binding as two, even though it occupies only one polyhedron apex. The intrinsic binding behavior of model clusters will be systematically analyzed for trends that are typed by the number of anionic residues bound to the cation. In the present study, we restrict the binding site model to have approximate octahedral symmetry, one anionic ligand, and carbonyl oxygen ligands represented by formamide. The binding behavior of the ligands in smaller clusters will also be considered.

The following questions are explored: (1) Does first coordination sphere energetics determine metal selectivity?. (2) What is the relationship between the size of the coordination sphere, the charge of the cation, and the enthalpy of metal binding? (3) What are the intrinsic energetics and conformational behavior of the metal binding to the different ligands? The answers to these questions will permit a quantitative exploration of the energetics of desolvation. The size and polarity of a ligand determines the magnitude of ligand–ligand repulsion and the equilibrium metal–oxygen bond length in an optimum cluster. However, the energetics of a cluster in its equilibrium geometry is not sufficient to understand binding competition in a protein. The study of the energetics of the clusters as a function of the M–O distance for cations of different size and charge is necessary for a deeper insight into binding selectivity. Cluster or cavity size will be varied in a study of the reaction enthalpy for exchange of two cations between the water and protein model clusters.

Although most of the Ca binding proteins have three or four Asp or Glu residues bound to the cation,[5] the Subtilisin protease has only one Asp in each of two binding sites.[6] In both sites, the carboxylate side chain of the Asp residue is bound in a bidentate manner and, therefore, is assumed to be deprotonated. The B, or surface, site has Ca bound to only two residues and four waters. The analysis of binding to either site requires obtaining the energies

---

[†] Presented in part at the OHOLO conference, May 1988, Eilat, Israel.

[‡] Molecular Spectroscopy Division.

[§] Center for Advanced Research in Biotechnology.

(1) Lehn, J.-M. *Structure and Bonding*; Dunitz, J. D., Hemmerich, P., Ibers, J. A., Jorgensen, C. K., Neilands, J. B., Reinin, D., Williams, R. J. P., Eds.; Springer Verlag, Inc.: New York, 1973; Vol. 16, pp 1–69.

(2) (a) Morf, W. E.; Simon, W. *Helv. Chim. Acta* **1971**, *54*, 794–810; (b) *Ibid.* **1971**, *54*, 2683–2704.

(3) Conway, B. E. *Ionic Hydration in Chemistry and Biophysics*; Elsevier: New York, 1981; Chapter 15, pp 312–341.

(4) Born, M. *Z. Phys.* **1920**, *1*, 45.

(5) Einspahr, H.; Bugg, C. E. *Metal Ions in Biological Systems*; Vol. 17, Calcium and Its role in Biology; Sigel, H., Ed.; Marcel Dekker, Inc.: New York, 1984; pp 51–97.

(6) Finzel, B. C.; Howard, A. J.; Pantoliano, M. W. *J. Cell. Biochem. Suppl.* **1986**, *10A*, 272. Bryan, P. N. *Biotech. Adv.* **1987**, *5*, 221–234. Pantoliano, M. W.; Whitlow, M.; Wood, J. P.; Rollence, M. L.; Finzel, B. C.; Gilliland, G. L.; Poulos, T. L.; Bryan, P. N. *Biochem.* **1988**, *27*, 8311–8317.

This article not subject to U.S. Copyright. Published 1990 by the American Chemical Society

Table I

Compact Effective Potential for Ca[a]

CEP	A_k	n_k	B_k
V_d	−2.16548	1	0.40639
	−5.81970	1	2.38045
V_{s-d}	8.00522	0	2.10543
	8.94751	2	0.66865
V_{p-d}	8.66222	0	1.79413
	5.57218	2	0.49323

Energy-Optimized Shared-Exponent Gaussian Basis Set for Ca

shell	type(*ij*)	α_i	c_i	c_j
1	SP	0.29698	−0.341644	−0.121714
	SP	0.11036	0.249819	0.271719
	SP	0.04919	0.665402	0.541646
2	SP	0.02244	1.0	1.0

[a] The form of the analytic representation is $r^2V_1(r) = \Sigma_k Ar^n \exp(-Br^2)$.

for transfer of the cation from a water cluster to the binding site cluster as a function of the metal–oxygen distance in the cluster. With the binding sites of Subtilisin BPN′ in mind, three metal cations, Mg^{2+}, Ca^{2+}, and Na^+, are considered at this time bound to only one anionic residue. The accuracy with which we calculate the enthalpy of hydration for the cations supports the cluster model as a means of estimating the energetics for the displacement of one cation by the other.

The experimentally determined structure of the first coordination shell in liquid water is very different for Mg and Ca. There is a six-coordinate octahedron for Mg,[7] but Ca is coordinated by nine or ten waters.[8] There are six waters in the first coordination sphere of Na.[3] Our model will examine the energetics of six and nine water clusters. These results will be compared to previous ab initio calculations of binding energies and conformations of cation–water clusters.[9] The binding conformations to the amide ligand has been described by the experimental results of Rao et al.[10] and by various theoretical calculations[11] for a single ligand. Theoretical studies of binding of a carboxylate ligand to these cations have also been published.[12]

2. Method and Models

Table II. Dipole Moment and Polarizability of Ligands

ligand	μ, ea$_0$	α_1, a$_0^3$ xx	yy	zz
water (W)				
DZ	1.059	1.33	7.06	4.72
DZd	0.895	2.65	7.24	5.26
DZd2	0.766	6.81	8.49	7.55
DZP	0.872	2.83	7.40	5.46
DZPd	0.774	6.71	8.57	7.57
HF limit[a]	0.782	7.99	9.04	8.47
exp[b]	0.728	9.26	10.01	9.62
Formamide (F)				
DZ	1.911	9.67	17.75	26.91
DZd	1.777	11.52	19.10	26.77
DZP	1.770	11.75	19.35	27.09
DZPd	1.715	17.03	22.85	30.50
exp	1.468			
formate (F⁻)				
DZ	0.874	9.43	15.37	23.27
DZd	0.703	11.33	17.00	22.75
DZP	0.706	11.46	17.08	22.91
DZP+	0.733	15.49	23.55	30.49
DZPd	0.655	16.70	22.88	27.91
DZPd+	0.672	19.14	26.37	31.21

[a] Werner, H. J.; Meyer, W. *Mol. Phys.* **1976**, *31*, 855. [b] John, I. J.; Bacskay, G. B.; Purvis, G. D. *Chem. Phys.* **1980**, *51*, 49. [c] The DZPd+ basis has one additional d function on each heavy atom and one diffuse sp set of functions on the oxygen atoms which is designated by the +.

Table III. Metal–Water Binding Energies and Geometries

complex[a]	*R*(M–O), au	*R*(O–H)	*A*(H–O–H)	BE/W, kcal/mol
Na				
W_1 opt	2.328	0.953	105.3	23.2
W_1 f	2.3			23.1
exp[b]				24.0
Mg				
W_1 opt	1.985	0.964	105.4	73.8
W_1 f	2.0			73.5
W_2 opt	1.979	0.962	105.5	71.6
Ca				
W_1 opt	2.256	0.961	103.3	53.3
W_1 f	2.3			53.0
W_2 opt	2.271	0.960	103.6	51.8
W frozen		0.950	105.5	

[a] Complexes with one, W_1, or two water ligands, W_2, were considered with all coordinates optimized, opt, or the ligand geometry frozen, f. [b] Enthalpy of binding: Dzidic, L.; Kebarle, P. *J. Phys. Chem.* **1970**, *74*, 1466.

Ab initio molecular orbital (MO) studies of an octahedral cluster with the water (W), formate anion (F⁻), and formamide (F) ligands can be used to model protein binding sites in cases where the geometry of the first shell is either fixed by the protein or determined by the cation to have this symmetry. In this study, valence-electron self-consistent-field (SCF-MO) calculations were performed with the HONDO code.[13] Compact effective potentials (CEP) were used in place of the chemically unimportant core electrons. CEP and their concomitant basis sets have been reported for the first two rows of the periodic table and the CEP and basis set for Ca is given in Table I.

The attractive metal–ligand interaction energy is dominated by charge-multipole and polarization contributions. In MO calculations, the chosen basis set affects the accuracy with which such interactions are predicted. Due to the size of the clusters being considered, only double-ζ (DZ) and double-ζ plus heavy atom d-type polarization functions (DZd) basis sets are used for the octahedrally coordinated complexes. For the largest complexes, only the DZ basis is practical at this time, but not even the DZd basis is sufficient for Hartree–Fock limit accuracy for the moments and polarizabilities. We therefore have to determine if energy differences or reaction energies in the clusters can be calculated with greater accuracy than the individual cluster binding energies.

The dipole moments and polarizabilities for the various ligands are compared for different bases in Table II. The double-ζ plus all-atom polarization (DZP) basis set and one with a second diffuse d function on oxygen (DZPd) were used to test the relative accuracy of the DZd basis. For the polarizability of the formate anion, the addition of a diffuse sp function is also required for a DZPD+ basis. The large overestimate of the dipole moment for all ligands that is found for the DZ basis leads to a substantial overestimate of the charge-dipole interaction. This is offset

---

(7) (a) Dietz, W.; Riede, W. O.; Heinzinger, K. *Z. Naturforsch.* **1982**, *37a*, 1038–1048. (b) Palinkas, G.; Radnai, T.; Dietz, W.; Szasz, Gy. I.; Heinzinger, K. *Ibid.* **1982**, *37a*, 1049–1060.

(8) (a) Probst, M. M.; Radnai, T.; Heinzinger, K.; Bopp, P.; Rode, B. M. *J. Phys. Chem.* **1985**, *89*, 753–759. (b) Palinkas, G.; Heinzinger, K. *Chem. Phys. Lett.* **1986**, *126*, 251–254.

(9) (a) Kollman, P. A.; Kuntz, I. D. *J. Am. Chem. Soc.* **1972**, *94*, 9236–9237. (b) Pullman, A.; Berthod, H.; Gresh, N. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1976**, *10*, 59–76. (c) Ortega-Blake, I.; Barthelat, J. C.; Costes-Puech, E.; Oliveros, E.; Daudey, J. P. *J. Chem. Phys.* **1982**, *76*, 4130–4134. (d) Ortega-Blake, I.; Novaro, O.; Les, A.; Rybak, S. *J. Chem. Phys.* **1982**, *76*, 5405–5413. (e) Sano, M.; Yamatera, H. *Ions and Molecules in Solution, Studies in Physical and Theoretical Chemistry*; Tanaka, H., Ohtaki, H., Tamamushi, R., Eds.; Elsevier: New York, 1983; Vol. 27, pp 109–116. (f) Probst, M. M. *Chem. Phys. Lett.* **1987**, *137*, 229–233. (g) Davy, R. D.; Hall, M. B. *Inorg. Chem.* **1988**, *27*, 1417–1421.

(10) Chakrabarti, P.; Venkatesan, K.; Rao, C. N. R. *Proc. R. Soc. London* **1981**, *A375*, 127–153.

(11) (a) Rode, B. M.; Preuss, H. *Theor. Chim. Acta* **1974**, *35*, 369–378. (b) Fussenegger, R.; Rode, B. M. *Chem. Phys. Lett.* **1976**, *44*, 95–99. (c) Fuchs, D. N.; Rode, B. M. *Chem. Phys. Lett.* **1981**, *82*, 517–519. (d) Welti, M.; Pretsch, E.; Clementi, E.; Simon, W. *Helv. Chim. Acta* **1982**, *65*, 1996–2007.

(12) (a) Gottschalk, K. E.; Hiskey, R. G.; Pedersen, L. G.; Koehler, K. A. *J. Mol. Struct.* **1982**, *90*, 265–269. (b) Maynard, A. T.; Hiskey, R. G.; Pedersen, L. G.; Koehler, K. A. *Ibid.* **1985**, *124*, 213–221.

(13) Dupuis, M.; King, H. F. *Int. J. Quantum Chem.* **1977**, *11*, 613; *J. Chem. Phys.* **1978**, *68*, 3998–4004.

(14) Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1984**, *81*, 6026–6033.

Table IV. Ca^{2+} + Formamide (F); DZP Optimized Geometry and Binding Energy (kcal/mol)

	CaF(frozen)			CaF	CaF_2
R(C-O)	1.234			1.266	1.258
R(C-H)	1.087			1.085	1.088
R(C-N)	1.370			1.306	1.311
$R(N-H)_c$	0.998			1.004	1.004
$R(N-H)_t$	0.996			1.005	1.005
A(HC-C-O)	121.6			119.5	119.1
A(N-C-O)	124.3			124.1	123.1
A(HNC-N-C)	119.3			121.9	120.0
A(HNT-N-C)	121.4			120.1	122.5
R(Ca-O)	2.1			2.067	2.079
A(C-O-Ca)[a]	170.0 c	150.0 c	150.0 t	173.7	179.4
BE	87.4	83.7	78.3	90.9	170.2

[a] The angle is measured with Ca cis, c, or trans, t, to the HC atom.

Table V. Ca^{2+} + Formate Anion (F^-); Optimized Geometry and Binding Energy (kcal/mol)

	CaF^-(frozen)	CaF^-
R(C-H)	1.124	1.088
R(C-O)	1.244	1.266
A(O-C-H)	114.8	119.8
R(Ca-O)	2.191	2.108
A(Ca-C-H)	180.0	180.0
BE	301.3	307.8

somewhat by the underestimate of the ligand polarizability. For Ca binding to one water, the DZ binding energy is 62.5 kcal/mol compared to the binding energy of 53.3 kcal/mol for the optimized structure described in Table III. Although it is evident that electron correlation is needed to improve the dipole moments of the isolated ligands, the inclusion of correlation has been found to have a minor effect in the binding of both Li^+ and F^- to water[15] and correlation would be even smaller for cations with small dipole polarizabilities. In this study we will explore only the SCF cluster binding energies.

Choosing the geometry of the ligands in a cluster model presents a problem. The X-ray data[5,10,11] are too variable and dependent upon the environment to be used as a guide for an ab initio calculation of a cluster. Geometry optimization by energy gradient methods is only practical for small clusters, which can be suggestive of the electronic and geometric changes that are induced. Such studies show that the relative energetics of metal binding are not very dependent on the small variations in the ligand geometries. We have calculated the binding energies and optimized geometries for MW (M = Na, Mg, Ca), CaF, CaF_2, and CaF^-. The results for the metal-water systems with the DZPd+ basis are given in Table III. As has been observed in earlier work,[9] the change in ligand geometry due to interaction with the metal is relatively small considering the substantial binding energy and polarization of the electronic wave function. The geometry changes are even smaller in the larger clusters. This is evident in the MW2 (M = Mg, Ca) data in Table III and supported by calculations for MW4 systems.[16] In the equilibrium conformation, the symmetry axis of the water is coincident with a vector drawn from the metal to oxygen. The average angle of the water relative to the axis is found to deviate from zero in an analysis of both experiment and molecular dynamics.[7] However, the energy required to bend the water from the equilibrium position by 30° is calculated to be only 2 kcal/mol and for 20° it is less than 1 kcal/mol.

The intrinsic energetically favorable orientation of formamide (F) with respect to calcium was determined by the energy gradient optimization of CaF by using the DZP basis for F. A. comparison in Table IV of the binding energy of frozen versus optimized formamide in CaF or CaF_2 finds a change of about 3 kcal/mol per formamide relative to a binding energy of about 90 kcal/mol per formamide.

The energy minimum in the case of calcium binding to a single formate anion (F^-) occurs in the bidentate conformation.[12] The formate geometry in the complete gradient optimized structure for CaF^-(bi) with the DZPd+ basis is shown in Table V. It does not differ substantially from the optimized geometry of the isolated ligand. The metal-ligand binding energy varies by only 2% for a complete optimization versus one in which formate is kept frozen. Freezing the ligands in their isolated conformations in subsequent cluster calculations is justified by these results. Reducing the size of the basis to the DZd or the smallest polarized one yields binding energies close (1 to 2%) to the results in Tables IV and V. The binding energies and equilibrium metal-to-oxygen distance are summarized in Table VI for the frozen F and F^- ligands with the DZd basis set.

Table VI. Metal-Single Ligand Binding Energy (kcal/mol)

	Na	Mg	Ca
F			
energy	35.0	115.4	89.3
R(M-O)[a]	2.2	1.8	2.0
F^- uni			
energy	124.7	309.0	266.1
R(M-O)	2.0	1.7	1.9
F^- bi			
BE	141.1	351.9	298.2
R(M-O)	2.4	2.15	2.4

[a] Equilibrium distance in Å.

Table VII. Binding Energies for Water Clusters

cluster		energies, kcal/mol			Mg-Ca diff
		Na	Ca	Mg	
W6	DZ		301.0	361.5	60.5
	DZd		255.0	318.7	63.7
	DZP		255.2	317.2	62.0
	DZPd	99.5	259.9	315.0	55.1
W9	DZ		331.8	361.2	29.4
	DZPd		292.0	320.2	28.2

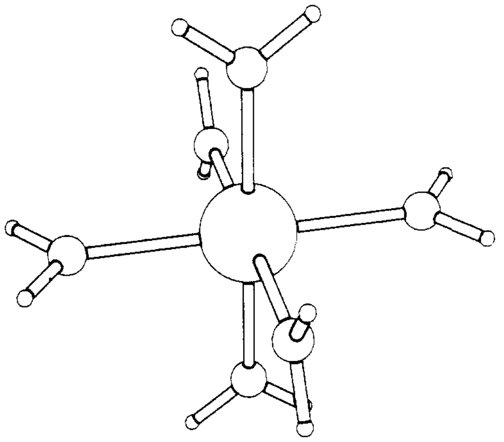

Figure 1. Schematic of the MW6 cluster generated by the T point group.

To understand the competitive binding of the metal ions to protein sites, the cluster reaction energetics for the following reaction need to be determined

$$M_1W_1 + M_2P \rightarrow M_1P + M_2W_2 + W_1 - W_2, \Delta E_1 \quad (1)$$

where M_iW_i represents the metal and its first hydration shell, M_iP represents the metal bound to the model protein cluster, and W_i is the first-shell water cluster. The difference in the number of first-shell waters for the two metals, $W_1 - W_2$, is required to balance the total number of waters, assuming no change in water content between M_1P and M_2P. The reaction energy is equal to the difference in binding energies of the products and reactants. We are therefore concerned primarily with the variation of metal binding energies with different types and numbers of ligands.

If the reaction energy is reasonably constant with variation in the basis set, then the energies from smaller basis set calculations can be used to analyze the larger clusters. For the single ligand clusters, the binding energies in Tables IV-VI show that there is only a small difference in the binding energies with the polarized basis sets. The variation of binding energy with basis set was then tested for both the six- and nine-coordinate water clusters. The six-coordinate water cluster has a T symmetry that reduces ligand-ligand repulsive interactions as seen in Figure 1. The DZd basis gives binding energies in reasonable agreement with the DZP and DZPd+ values listed in Table VII. The binding energies obtained for the DZ basis set are much too large, but the binding energy differences are not unreasonable. Agreement between the DZd and DZP is to be expected, since there is very little difference in the electronic description of the ligand molecule with the two basis sets. However, the agreement with the DZPd+ basis, which gives much

---

[15] Diercksen,, G. H. F.; Kramer, W. P.; Roos, B. O. *Theor. Chim. Acta* **1975**, *36*, 249-274.

[16] Unpublished work on MWn frequencies.

Table VIII. Comparison of Cluster Binding Energies and Cluster Binding Energies for Octahedral Complexes

(a) Comparison of Cluster Binding Energies

complex[a]		energies, kcal/mol		
		Ca	Mg	diff
DZd	W6	255.0	318.7	63.7
	F^-W5(bi)	435.8	487.6	51.8
	F^-W5(uni)	425.5	485.7	60.2
	F^-FW4(bi)	443.9	492.7	48.8
	F^-FW4(uni)	438.5	496.4	57.9
DZ	W6	301.0	361.5	60.5
	F^-W5(bi)	460.7	516.3	55.6
	F^-W5(uni)	461.4	522.7	61.3
	F^-FW4(bi)	470.6	519.7	49.1
	F^-FW4(uni)	474.5	532.7	58.2

(b) Cluster Binding Energies for Octahedral Complexes

	DZ energies, kcal/mol		
complex[a]	Ca	Mg	Na
F^-FW4 (2.8 Å)	420.0	434.1	
F^-F3W2	492.1	543.3	
F^-F4W	498.5	545.4	176.4
F^-F4W (2.7 Å)	438.7		171.2
F^-F4W (2.8 Å)	422.7		166.7
F^-F5(bi)	490.7	530.9	
F^-F5(uni)	504.5	547.2	

[a] F^- = formate, F = formamide, W = water.

smaller ligand dipole moments and much larger polarizabilities, suggests a compensation between the electrostatic and inductive interactions. As an example of the agreement between the energy for the exchange of Mg and Ca cations in the reaction

$$\mathrm{CaW6 + MgF^-FW4 \rightarrow CaF^-FW4 + MgW6}$$

we find 14.9 and 11.4 kcal/mol respectively for the DZ and DZd basis sets.

Another reason for the success of the smaller basis sets in these calculations comes from the minimization of basis set superposition effects that can lead to an overestimate of binding energies. Superposition effects are not found to be as significant for the CEP calculations, because the core electrons are not present. For the W6 cluster, the basis set superposition energy (BSSE) is less than 0.0005 au per ligand for both the DZ and DZP bases. The small difference in binding energy between the present results and an all-electron DZP calculation[9c] is probably due to the larger superposition error that occurs for the K shells in the all-electron case. The most important result in Table VII is the relative constancy with basis set of the predicted difference in the water cluster binding energies between the Mg and Ca clusters.

As shown in Table VIIIa, basis set comparisons were also carried out for clusters containing at least one formate and zero or one formamide. Since the differences in metal binding energies are the same to within 10% for the DZ and DZd bases, the DZ basis set is considered sufficiently accurate for the analysis of the binding energy differences.

For clusters with several formamides, it is also necessary to vary the cluster geometry to find the optimum ligand orientations. Geometries of the clusters with additional formamides were compared only for selected points, since gradient-driven energy surface searches are impractical because of the large size of the basis sets, the large number of atoms, and the slow variation of the energy with angular variation of the ligand relative to the metal atom. The clusters were eventually arranged to maximize the internal hydrogen binding. The calculated binding energies for the most stable clusters considered are summarized in Table VIIIa,b. A representative structure for MF^-F5 is shown in Figure 2.

The major components of the binding enthalpy of the cations are the energy of the first coordination shell cluster, E, the Born polarization enthalpy for this cluster, ΔH_B, and the energy required to free $2N$ first-shell hydrogen bonds from the bulk, ΔH_W.

$$\Delta H_{\mathrm{solv}} = E + \Delta H_B + \Delta H_W \qquad (2)$$

The last term consists of the energy required to free N water from the bulk, ΔH_{vap}, the intermolecular binding of the first- and second-shell waters, E_{12}, and the energy required to form the cavity which accommodates the metal cation.

$$\Delta H_W = \Delta H_{\mathrm{vap}} + E_{12} + \Delta H_{\mathrm{cav}} \qquad (3)$$

The first-shell cluster energy is evaluated by ab initio quantum chemical methods as described earlier. We have included the SCF binding energy of the cation to the first-shell waters, the dispersion binding between the first-shell waters, and an ab initio estimate of the zero-point energy. The Born polarization enthalpy differs only slightly from the Born free energy because the derivative of the dielectric constant of water with respect to the temperature is small.[17] The heat of vaporization of water is 10.5 kcal/mol. We neglect the increase in the first-shell hydrogen bond energies due to the presence of the cation, so the intermolecular binding of the first and second shells cancels half of the energy required to free the water. The energy required to form a cavity to accommodate the cations is evaluated from the surface tension of water by using the ion radii of Rashin and Honig.[17] No distinction was made between Ca in the six or nine coordinate complex. The ΔH_{cav} are found to be 7.4, 5.6, and 9.0 kcal/mol respectively for Na, Mg, and Ca. The cavity energy is estimated to be about 10 kcal/mol for the larger Ca ion. Since we are interested in metal transfer reaction enthalpies, small thermodynamic terms that are similar for the different ions are neglected. The ΔH_W determined by Morf and Simon[2] is found to be comparable and either estimate can be used.

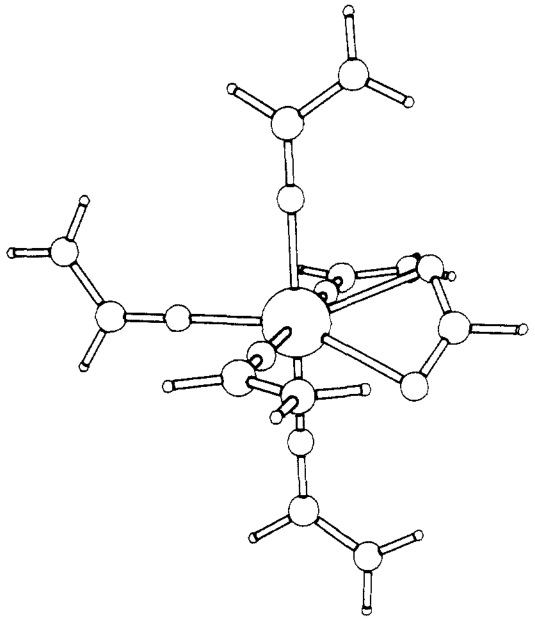

Figure 2. Schematic of the MF^-F5 cluster with the formate ligand, F^-, attached to the metal in a bidentate bond.

3. Metal–Water Clusters and Hydrogen Enthalpies

Both six- and nine-water clusters have been studied, because the first coordination shell of Ca in water is not observed to be six coordinate.[8] The model six-coordinate structure is an octahedron, while the nine-coordinate structure is the idealized tricapped trigonal pyramide that is inferred as an important feature in the hydrated ion.[8] In the octahedral site, the rotational orientation of the waters is fixed by generating the complex with the T point group operations. The only free parameter is the radial distance between the metal and the oxygen atom. This was varied for all the metals, Na, Mg, and Ca, with the DZPd+ basis. The calculated minima for Na, Mg, and Ca are 2.38, 2.12, and 2.24 Å, respectively. The radial minima for both Mg and Ca are found to be somewhat longer for the smaller basis sets. The nine-coordinate cluster was generated with the D_{3h} point group. A large number of parameters can be varied, with three important coordinates being the metal–oxygen distance in the two independent planes and the distance between the planes or R(Ca–O1), R(Ca–O2), and R(O1–O2) in Figure 3. The rotamer orientations of the waters were fixed to minimize the repulsive interactions between the waters. With the water orientations fixed, variation of the free parameters to maximize the binding energy results in $R(\mathrm{Ca{-}O_1})$, $R(\mathrm{Ca{-}O_2})$, and R(O1–O2) equal to 2.5, 2.6, and 2.90 Å. Replacing Ca with Mg yields a slightly smaller complex with the three parameters equal to 2.4, 2.4, and 2.73 Å. In the Mg complex, however, the three additional waters in the nine-coor-

---

(17) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588–5593.

Table IX. Components of Enthalpies of Hydration (kcal/mol)

M	SCF	disp	ZPE	Born	free W	total	exp[a]
NaW6	-99.5	-4.3	5.7	-42.3	38.0	-103	-98.5
MgW6	-315.0	-7.0	10.2	-181.3	38.0	-455	-462.4
CaW6	-259.9	-5.0	6.7	-176.8	38.0	-397	
CaW9	-292.0	-17.7	11.8	-150.2	56.9	-391	-383.6

[a] See ref 17.

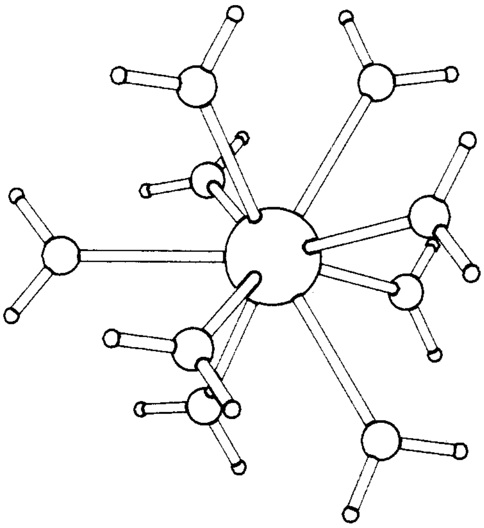

Figure 3. Schematic of the MW9 cluster generated by the D_{3h} point group.

dinate complex are not substantially bound relative to the six-coordinate complex. They would be more stable as second-shell waters bound to the six-coordinate cluster.

From the data in Table VII, it is seen that the binding energies are reasonably stable once polarization functions are included for the heavy atoms. The DZ binding energies are substantially larger than those obtained with polarized bases. Since the sizes of these clusters are comparable, it is not surprising that the basis set errors in the ligand polarization and electrostatic energies are comparable for the different metals, yielding relatively constant binding energy differences. To predict accurate binding energies, polarization is obviously very important in these systems and a polarization basis is required on the heavy atoms.

The number of ligand–ligand interactions is large, so the dispersion interaction between the water ligands can be an important component of the total cluster energy. The water–water dispersion was estimated by a second-order perturbation calculation (MP2)[18] using the DZPd+ basis and correcting for BSSE by evaluating the MP2 energy for each water with the total basis set. The dispersion energy per pair in W6 was 0.58, 0.42, and 0.36 kcal/mol for an inter-water O–H distance of 2.1, 2.3, and 2.4 Å, which represent the Mg, Na, and Ca clusters, respectively.

The zero-point energy can also be important, but it is difficult to obtain accurately since there are many modes involving the metal that are not measured experimentally. Ab initio frequencies[16,19] suggest that the contribution from the metal–oxygen stretch vibrations is not dominant. Evaluation of all frequencies for an octahedral cluster is difficult for a good basis set, so we have carried out ab initio calculations of the frequencies for smaller MW*i* clusters for $i = 1$–4, using both DZ and DZPd+ bases. The sum of the torsion, bending, and stretching frequencies has been extrapolated to $i = 6$ and 9. The shift in the water frequencies due to the cluster interactions is not large, and this contribution to the ΔZPE is ignored.

In order to calculate the Born polarization energy, the radius of the first shell is required. The cavity radius is set equal to the distance between the metal and the midpoint of the hydrogen bond that would be formed between a first-shell and second-shell ligand. This procedure yields radii of 3.87, 3.61, 3.71, and 4.36 Å respectively for the Na, Mg, CaW6, and CaW9 clusters. The components of the solvation enthalpy are given in Table IX. The calculated energies are compared to the values deduced by Rashin and Honig.[17]

The calculation of the enthalpy of hydration of Ca using the six-coordinate cluster yields 397 kcal/mol, which is 6 kcal/mol larger than the value calculated with the nine-coordinate cluster. This difference in the two estimates is small on a percentage basis and comparable to terms neglected in this calculation. If independent experimental evidence did not exist for the size of the first shell, then the energy calculation would incorrectly predict the six-coordinate cluster as most stable. Energetically there is little to distinguish different sized clusters of water bound to Ca. The enthalpy of hydration for Mg agrees within 3% of experiment but the difference between the experimental and calculated values is an appreciable 7 kcal/mol. These two measures reflect both the very approximate nature of the calculations and their qualitative correctness.

4. Replacement of Water by Protein Ligands

The water ligands in the Mg and Ca clusters are replaced systematically by first one F^- ligand and then successively by F ligands. The F^-W5 and F^-FW4 clusters were studied with both the DZ and the DZd bases. This permits a comparison of the predicted binding energy differences. In both of these clusters, the F^- ligand was considered in both uni- and bidentate modes. Again, the conclusion is reached that the binding energy differences between the Mg and Ca are sufficiently accurate with the DZ basis.

Analysis of the components of the binding in the complete first-shell cluster requires values for the smaller clusters, where ligand–ligand interactions are minimized and the perturbation of the ligands is dominated by the metal cation. The binding energies for single-ligand clusters, ML, are summarized in Table VI for formamide and formate. The binding energy for a single ligand is substantially higher than the binding per ligand in the octahedral cluster. Basis set superposition energies (BSSE) are calculated to be less than 0.5 kcal/mol per ligand for the DZd basis set. The internal ligand geometry was not found to change significantly upon bonding to the metal ion.[16]

The binding energy of a cation to a single formamide exceeds the binding to water. The minimum energy structure is very close to a linear M⋯O–C bond (angle M–O–C = 173°), reflecting the strong local interaction with, and polarization of, the CO bond rather than overall interaction with the molecular dipole. This agrees with the theoretical literature on small clusters with alkali and alkaline earth cations.[11] Tilting the amide group away from, rather than toward, the cation yields a flatter, more attractive energy surface as a function of M–O–C angle as seen in Table IV. The conformations observed experimentally are considerably distributed over angle,[5,10] but they probably reflect environmental perturbations.

For small clusters the bidentate bond to the formate anion is always more stable than the unidentate binding as seen in Table VI. Experimentally, the unidentate bond seems to predominate in proteins.[5] Although it is possible that the Asp or Glu side chains are not always deprotonated when a cation is bound to a number of anionic residues, deprotonation is likely for one such residue bound to a divalent cation. The unidentate complex is more stable

---

(18) (a) Moller, C.; Plesset, M. S. *Phys Rev.* **1934**, *46*, 618. (b) Pople, J. A.; Binkley, J. S.; Seeger, R. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1976**, *10*, 1–19.

(19) (a) Probst, M. M.; Limtrakul, J. P.; Rode, B. M. *Chem. Phys. Lett.* **1986**, *132*, 370–376. (b) Hashimoto, K.; Yoda, N.; Iwata, S. *Chem. Phys.* **1987**, *116*, 193–202.

in all DZ octahedral complexes studied, with the relative stability increasing with the number of waters that are replaced by formamide. However, for the DZd basis, the bidentate complexes are still more stable for the MgF$^-$W5, CaF$^-$W5, and CaF$^-$FW4 complexes, but the unidentate systems are again increasingly stable as the waters are replaced. The Mg complexes are more likely to be in the unidentate form, reflecting the shorter radial metal-to-oxygen distances and the increased ligand crowding. In addition to the steric crowding of a bidentate bond, unidentate binding of F$^-$ to the metal ion may also be favored in proteins due to additional strong interactions between the second carboxylate oxygen and other fragments.

Replacing one water in DZd CaF$^-$W5 by a formamide increases the total binding energy by 13 and 8 kcal/mol for the uni- and bidentate cases, respectively. For Mg the binding energy increases by 11 and 5 kcal/mol, respectively. Crowding of the ligands leads to smaller increases for Mg and the bidentate case. For the DZ basis, the binding energy increases are 13 and 10 kcal/mol for Ca and 10 and 3 for Mg in the uni- and bidentate cases, respectively. Thus, the relative effects of ligand replacement are well represented in the DZ basis even though the absolute binding energies differ substantially from the DZd results. The MF$^-$F5 complex was examined in both the uni- and bidentate forms. The unidentate complex is energetically favored over the bidentate by 14 kcal/mol for Ca and 16 kcal/mol for Mg.

The complex, MF$^-$F3W2, is made by replacing two more waters in the unidentate MF$^-$FW4, resulting in an increase in stability per formamide of 9 kcal/mol for Ca and 5 kcal/mol for Mg. Subsequent replacement of the two remaining waters results in a constant binding energy increase per substitution of 6 kcal/mol for Ca and 2 kcal/mol for Mg. Even with the large ligand-ligand repulsive interaction, the addition of formamide is always exothermic, but the crowding of the ligands in the case of Mg yields a smaller relative increase. This is to be contrasted with the large difference of 37 kcal/mol for binding one water versus formamide to the Ca cation.

Several tests of the complex stability was made as a function of the metal-to-formate bond angle, M-O-C, in both systems. For CaF$^-$W5, the energy surface is found to be flat to within 1 kcal/mol as the angle is varied down to 140° with the unbound oxygen rotated toward the hydrogen in a neighboring ligand. On the other hand, for MgF$^-$W5 the same rotation yields an increase in binding of about 4 kcal/mol, which reflects incipient interligand hydrogen bonding. The Mg complex is more compact, so the interligand O⋯H distance is only 1.90 Å, while for Ca it is 2.16 Å. The interligand H bond is made at the expense of the strong electrostatic bonding of the ligand with the cation which prefers a linear orientation.

In summary, for compact Mg complexes, both the attractive and repulsive interligand interactions are stronger than for Ca complexes. The angular distribution of bond angles in real cases would reflect the balance of direct cation bond and ligand-ligand interactions and not the intrinsic angular dependence of the bond energy itself. Further details of interligand hydrogen bonding will not be explored here since they will depend on the environmental influences present in real cases.

Varying the radial size of the MF$^-$FW4 cluster illustrates the dramatic narrowing of the energy difference in binding between Mg and Ca expected as the cations with the same charge interact electrostatically with a fixed octahedron of ligands. The difference in binding energies is reduced to only 14 kcal/mol, when the M-O distance is increased to 2.6 Å for the F$^-$ and F ligands and 2.8 Å for W. For Na and Ca in the MF$^-$F4W cluster, the dependence on the size of the cluster is also dramatic. As the cluster size increases, the energy of the Ca cluster decreases much more rapidly than that of the Na cluster as can be seen in Table VIIIb. This is due to the coulombic attraction of cations with different charges interacting with the anionic formate ligand.

5. Application to the Binding Sties in Subtilisin

The model octahedral metal-protein complex is defined differently for the buried A and the surface B sites in Subtilisin. The A site is modeled by the MF$^-$F$_5$ cluster while the octahedral versions of the B site are the MF$^-$FW4 and MF$^-$F$_4$W clusters depending on the metal. The model protein binding energies are estimated from only the first coordination shell cluster and the Born polarization energy. The energy required to modify the protein in order to optimize metal binding is not considered. The protein environment is considered to be immobile and the dielectric constant is assumed to be 4. The B site is not buried in the protein, and water interactions are present and could increase the dielectric constant. Although the absolute Born energy for a cluster is sensitive to the value of the dielectric constant, the difference in the Born energies for similar sized clusters in any reaction scheme is very small. If we vary the dielectric constant from 2 to 8 for any exchange of Mg and Ca, for example, the reaction energy will increase by only 0.2 kcal/mol. In reality the environmental interaction is certainly dominated by hydrogen bonding and the Born polarization energy must be considered a crude approximation. The ion cluster also has a large dipole moment formed from the well-separated cation and anionic charge distributions. The estimated contribution of the dipole polarization energy[20] is only of the order of 5 kcal/mol for the model protein clusters and so is neglected here since the calculated dipole energy is comparable in all the cation clusters including Na and will cancel in the reactions considered here.

Table X. Binding Energies of Model Clusters

	$-\Delta H$, kcal/mol		
cluster	Mg	Ca	Na
W6	455	397	103
W9		391	
F$^-$F5	570	527	
F$^-$F4W (opt)	568	521	176
F$^-$F4W (2.7 Å)		460	171
F$^-$F4W (2.8 Å)		443	167
F$^-$F4W2 (2.8 Å)		453–463	
F$^-$FW4 (opt) uni	553	495	175
F$^-$FW4 (opt) bi	540	491	175

Within these approximations enthalpies are presented at relevant radial distances for both A and B site models in Table X. It is evident that they are only useful to calculate reaction energy differences. The cluster energetics are used to analyze the following: (1) the relative binding energies of Mg and Ca in the buried A site and (2) the energetics and conformation of Ca and Na binding to the B site and the inherent rigidity of the B site cavity.

For the A site, the reaction energy for replacing Ca by Mg is 20 kcal/mol as shown in Table XI. The model cluster energies show that desolvation energies determine the ion selectivity between Mg and Ca in site A. This is because the differences in binding energies to the protein, which are dominated by the first shell, are not as large as the differences in water. The water ligands can adapt to the smaller size of Mg while the bulkier protein ligands cannot. Increasing the size of the A model cavity would increase the preference for Ca. However, the ligand-ligand repulsion prevents the collapse around the smaller Mg cation. Rigidity of the binding site need not be invoked.[21] Even for optimal radial metal-oxygen distances this site will be very selective between Ca and Mg.

For the B site, only two protein ligands are observed to be bound to the Ca with waters completing the first shell.[6] The ASP residue is bound as a bidentate anion. For the MF$^-$FW4 reaction model, bidentate bound Ca is favored over bidentate Mg in the B site by 14 kcal/mol for optimum M-O bond lengths. If unidentate binding is possible for Mg and optimal metal-to-oxygen distances can be achieved for both Ca and Mg binding, then the two cations may be competitive. The DZd energetics in Table VIIIa show that CaF$^-$FW4 is more stable in the bidentate form in agreement

---

[20] (a) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486. (b) Beveridge, D. L.; Schnuelle, G. W. *J. Phys. Chem.* **1975**, *79*, 2562-2573.

[21] Stuart, D. I.; Acharya, K. R.; Walker, N. P. C.; Smith, S. G.; Lewis, M.; Phillips, D. C. *Nature* **1986**, *324*, 84-87.

Table XI. Description of Subtilisin Sites and Reaction Enthalpies

Site A		
Buried site with metal bound to six protein ligands. one aspartate (ASP41)		
MgW6 + CaF$^-$F5 ---> CaW9 + MgF$^-$F5.		ΔH = 20 kcal/mol
Site B		
Surface site geometry is constant but metal position varies:		
Ca bound to surface with optimal R(Ca-O) to F$^-$(ASP197), F, W4.		
Na bound in center of cavity with R>2.8Å to F$^-$(ASP197), F4, W.		
a) Binding of metal to surface of cavity (C-center, S-surface): Vary radius of cavity, optimal surface bonds		
CaF$^-$F4W(C) + W3 ---> CaF$^-$FW4(S) + F3,	opt	65 kcal/mol
	2.7Å	4
	2.8Å	-13
NaF$^-$F4W(C) + W3 ---> NaF$^-$FW4(S) + F3.	opt	34
	2.8Å	25
b) Competition between cations for binding site: Same charge but different size cation - Mg vs Ca		
MgW6 + CaF$^-$FW4 ---> CaW9 + MgF$^-$FW4.	opt bi-bi	13
	opt bi-uni	1
	2.4Å bi-uni	30
Different charge and approximately same size - Na vs Ca optimal cluster size: Na in center, Ca on surface		
NaW6 + CaF$^-$FW4 + F3 ---> CaW9 + NaF$^-$F4W.	opt	18
'rigid' protein cavity: both Na and Ca in center		
NaW6 + CaF$^-$F4W ---> CaW9 + NaF$^-$F4W,	opt	57
	2.7Å	1
	2.8Å	-12
expand Ca first shell > 6 by adding water:		
NaW6 + CaF$^-$F4W2 ---> CaW9 + NaF$^-$FW4 + W.	2.8Å	-2 to 8

with the experiment,[6] but MgF$^-$FW4 is more stable in the unidentate form.

The experimental observation of Na bound in the B site[6] provides additional insight into the structure of the site. The Na is displaced from the Ca position but it is still bound to the two residues, Asp 197 and Glu 195, that are in the first shell of the Ca ion. In addition, there are bonds to the oxygen sites on Gly 169, Tyr 171, and Val 174. Two waters round out the complement of seven oxygen ligands to sodium. The Asp and Glu residues that are near the protein surface in effect intercept the Ca before it penetrates into the cavity. The Na does enter the cavity and interacts with all the residues, but the average Na–O distance of 2.9 Å is far from optimum. The binding to the cation is not sufficiently strong to reduce the binding distance to the optimal values which are about 0.5 Å smaller. The Ca binding also does not attract the Gly, Tyr, and Val residues to short binding distances. The protein backbone is reported to be similar for both Ca and Na binding,[22] suggesting a relatively stiff binding cavity is formed by the protein, and this cavity is not qualitatively altered by the metal binding. The Ca binds to the ASP and forms an optimal bond but is prevented by the rigid cavity from forming optimal bonds with three of the protein residues. An optimal bond to water is stronger than the long bonds to the neutral amide ligands. The rigidity of the site is maintained by a strong ion pair interaction of Lys 170 and the side chain of Glu 195, which both bind to the side chain of Ser 163. Since the Lys and Glu residues are from opposite sides of the B site, the "salt bridge" counteracts the collapse of the cavity about the divalent cation.

In order to illustrate the effect of cation binding to a cavity that is constrained, the binding of Ca is modeled with the MF$^-$F4W cluster. The reaction enthalpy is compared in Table XI for three sets of metal-to-oxygen distances. For the optimal distances, the Ca is very strongly bound but the binding between the ionic components of the cluster falls off rapidly for the divalent cation. At 2.8 Å the Ca will bind to the B site as a CaF$^-$FW4 cluster as observed.[6]

Competition between cations of different charge can be understood in terms of the rigidity of the B site cavity. As the radius of the cavity increases, the Na cation interaction with the carboxylate moiety falls off less rapidly than the interaction of carboxylate with divalent Ca. From Table XI, we see that Na binding is then favored by more than 11 kcal/mol over Ca at a cavity radius greater than 2.7 Å. We note that Ca can bind additional waters in its first shell and becomes more competitive with Na when both are embedded in the cavity. An upper bound to the binding energy of the seventh coordinated water in the first shell of the B site is obtained from the dissociation energy for water from the CaF$^-$W5 cluster at the larger distance of the B site. The lower bound is obtained by calculating the second-shell hydrogen-bonded water.

The CaF$^-$FW4 model cluster, nontheless, is far more stable (>15 kcal/mol) than Na embedded in the cavity. Experimental observation of Na binding occurs with a large excess of Na.[6] This is one case where the environment is different for the Na and Ca protein binding and the use of different dielectric constants may be justified. However, if we use a value larger than 4 in the calculation of the polarization energy for the Ca cluster, Ca binding will increase over Na. There is also the significant repulsive energy between the three formamide ligands that are no longer bound to the Ca cation. Within this model, no quantitative estimate is possible but this term reduces the reaction energy for replacing a Ca ion bound to the surface site with an embedded Na ion. The 2.8 Å NaF$^-$F4W cluster models the embedded Na cation. It is far more stable (>30 kcal/mol) than the NaF$^-$FW4 cluster with optimal bond distances. The larger metal–oxygen distances for the embedded cation does not penalize the monovalent cation as much as the divalent. The Ca stays near the surface and Na embeds in the cavity.

Registry No. W, 7732-18-5; F$^-$, 71-47-6; F, 75-12-7; Na, 7440-23-5; Ca, 7440-70-2; Mg, 7439-95-4; subtilisin, 9014-01-1.

---

(22) Private communication from the Protein Engineering Group at GENEX of unpublished data and a discussion on the structure of subtilisin.