Protein Science (1998), 7:2472–2475. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Cation-π (Na^+-Trp) interactions in the crystal structure of tetragonal lysozyme

JOHAN WOUTERS

Facultés Universitaires N.-D. de la Paix, 61 Rue de Bruxelles, B-5000 Namur, Belgium

(RECEIVED May 18, 1998; ACCEPTED July 9, 1998)

Abstract: Experimental evidence of a cation-π interaction between a sodium cation (Na^+) and the indole ring of residue Trp123 in a structure (2.0 Å) of hen egg-white lysozyme is presented. The geometry of the metal ion-π interaction observed in the protein structure (distance between the aromatic plane and the cation ~4 Å) is consistent with geometries observed among small molecules crystal structures and quantum chemistry ab initio calculations. The present crystal structure of lysozyme provides unique structural information about the geometry of binding of cations to π systems in proteins. It shows that the metal ion-π interaction within proteins is not significantly different from similar bindings found in small molecules and that it can be modeled by theoretical methods.

Keywords: ab initio calculations; cation-π interaction; CSD search; lysozyme; X-ray crystallography

Cations can bind to aromatic systems through strong noncovalent forces. Those cation-π interactions were relatively underappreciated in the past, compared to more conventional interactions such as hydrogen bonds, ion pairs (salt bridges), or hydrophobic interactions. In recent years, they have come to be studied in several (bio)chemical systems and should be considered as an important and general binding force (Dougherty, 1996; Ma & Dougherty, 1997). Solid state examples of cation-π interactions are well documented in studies of small molecule crystal structures; stabilizing interactions between ions such as Na^+ or K^+ with simple aromatic rings have been observed (Dougherty & Stauffer, 1990; Ma & Dougherty, 1997). In macromolecules, binding sites for cations through the side chains of aromatic amino acids (phenylalanine (Phe), tyrosine (Tyr), or tryptophane (Trp)) should also exist (Fraenkel et al., 1996). However, to my knowledge, no direct experimental evidence of a metal ion interacting directly with an aromatic residue had ever been reported.

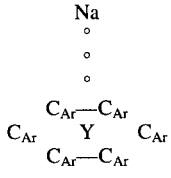
Reprint requests to: Johan Wouters, Lab. Chimie Moleculaire Structurale, Facultes N.-D. de la Paix, 61 Rue de Bruxelles, B-5000 Namur, Belgium; e-mail: wouters@scf.fundp.ac.be.

Results and discussion: During crystallographic analysis of tetragonal hen egg-white (HEW) lysozyme at 2.0 Å resolution, a cation-π interaction (Table 1) between a sodium cation (Na^+) and the six-membered ring of the indole of residue Trp123 was unexpectedly observed (Fig. 1A). In this structure, Trp123 is able to compete with full aqueous solvation for the binding of a sodium cation. The peak corresponding to the sodium cation was clearly present in the first residual electron density maps and persisted, whatever the stage of refinement (addition of water molecules). With a Na occupancy factor set to 1.0, the temperature factor of the ion refined to a value of 42.6 $Å^2$. The position of the cation persists in a difference "omit" Fourier map. Taken together, those arguments strongly suggest the presence of a sodium site facing the aromatic ring of Trp123. To further ascertain this hypothesis, small molecule crystal structures were examined and theoretical calculations performed, using the benzene and indole rings as models of the side chains of Phe and Trp, respectively.

The geometry of the metal ion-π interaction observed in the protein structure (mean distance C_{Ar}-Na^+ = 4.3 Å; distance between the centroid of the six membered ring of the indole and the cation = 4.07 Å) is consistent with geometries observed among small molecules crystal structures deposited at the Cambridge Structure Database (Table 1). This search was performed of precise organic structures ($R_1 < 7\%$). In particular, 44 cation-π interactions involving a sodium ion were retrieved and statistically analyzed; the distance C_{Ar}-Na^+ ranges from 2.83 to 4.95 Å (mean distance C_{Ar}-Na^+ = 3.61 Å).

Quantum chemistry (ab initio, RHF 6-31G**) calculations were also performed on benzene/Na^+ and indole/Na^+ systems to predict their geometry and electron density. The overall shape of the experimental electron density ($2F_o - F_c$) map obtained in the crystal structure of HEW lysozyme around Trp123 (Fig. 1B) is nicely reproduced by the total electron density calculated on modeled cation-π complexes between monovalent sodium cation and the indole ring (data not shown). The quantitative calculations also confirm that the benzene ring of indole is the preferred cation-π binding site as the difference in relative energies ($\Delta E^{HF/6\text{-}31G^{**}}$) between the six-membered ring/Na^+ optimized geometry and the five-membered ring/Na^+ complex is about 3.6 kcal/mol. The geometry deduced from those computations are a valuable comple-

Table 1. *Geometry of the metal ion-π interaction (in Å)*



	C_{Ar}-Na^+	$[C_{Ar}$-$Na^+]_{mean}$	Y^a-Na^+
Lysozyme structure			
Trp123-Na^+	4.16	4.30	4.07
	4.32		
	4.46		
	4.43		
	4.27		
	4.14		
CSD search			
CAr_6-Na^+ [b]	2.82–4.95[c]	3.86	3.61 (2.84–4.95)[d]
Theoretical calculations			
CAr_6H_6-Na^+ [e]	2.82[f]	2.82	2.45
Indole-$Na_{\|6}{}^+$ [e,g]	2.85	2.80	2.42
	2.87		
	2.82		
	2.76		
	2.72		
	2.76		

[a]Y is the centroid of the aromatic moiety (six aromatic carbons ring).
[b]Forty-four fragments are within the 1.0–4.5 Å distance range defined in the search; C_{Ar} stands for aromatic carbons.
[c]The minimal and maximal distances are given.
[d]The minimal and maximal distances are given in parenthesis.
[e]6-31G** optimized geometry.
[f]All six C_{Ar}-Na^+ are identical.
[g]$Na_{|6}{}^+$ defines the cation along the sixfold axis of the six-membered ring.

ment to the crystal structure geometries (Table 1). However, as a general rule, the computed geometries are more compact than the corresponding experimental structures, as reflected by shorter equilibrium bond lengths (e.g., the distance between the cation and the centroid of the aromatic ring are underestimated by about 30%). As was already pointed out by other groups (Williams, 1993; Dougherty, 1996; Mecozzi et al., 1996a, 1996b; Ma & Dougherty, 1997), the electrostatic potential surfaces calculated around aromatics also provide a reliable qualitative guide to cation-π interactions, the more negative the maximum in electrostatic potential over the center of the aromatic ring, the stronger the cation-π interaction. Both the six-membered ring and the five-membered ring of indole generate an attractive molecular electrostatic potential indicating that the benzene as well as the pyrrole-type ring could be cation-π binding sites. Compared to benzene, indole is clearly a stronger cation-π binder suggesting that Trp may be especially important in cation-π binding.

The present crystal structure of lysozyme provides thus unique structural information about the geometry of binding of metal ions to π systems in proteins. It shows that the cation-π interaction within proteins is not significantly different from similar bindings found in small crystal structures and that it can be partially modeled by theoretical methods.

Identification of Na^+ binding sites in protein crystals is complicated by the comparable electron density of this monovalent cation and water (the scattering powers of H_2O, OH^-, $NH_4{}^+$, and Na^+ are virtually indistinguishable by X-ray protein crystallography). As for structurally ordered water molecules, only as a structure analysis approaches higher (i.e., beyond 2.5 Å) resolution is it possible to begin to identify the positions of individual sodium ions. The main argument to differentiate a sodium ion from a water molecule is the average distances between the ion and the protein atoms.

A distance criterion was used, in some other structures of HEW lysozyme (Vaney et al., 1996), to locate a Na^+ in the middle of the loop Ser60–Leu75: the ion makes six short bonds (between 2.18 and 2.62 Å) with the surrounding atoms. In the present structure, a water molecule occupies this site and makes a H bond with OG of residue Ser72 (2.99 Å) and with a second water molecule (2.32 Å). The absence of sodium in this octahedral site can be explained by the fact that, in the present study, the carbonyl group of the peptide bond Arg73–Asn74 points outside the Pro70–Leu75 loop (toward solvent), while in other structures of lysozyme it points inside the loop, coordinating a Na^+ (Turner & Howell, 1995; Vaney et al. 1996). The new cation-π binding site of Na^+ near Trp123 is clearly different from the "classical" distorted octahedral binding site. A distance criteria-based methodology, however, can be developed to detect it.

Standard automatic "water divining" procedures are not able to locate cations bound to π systems as they usually only retain residual electron density implied in H-bonds (this condition is translated into a distance criteria). However, using proper search geometries, novel binding sites for cations can be constructed through the side chains of aromatic amino acids. As a test, I found that it was possible, a posteriori, to identify the sodium binding site at Trp123 by using the wARP program (Perrakis et al., 1997), in a nonconventional manner, imposing search distances compatible with the geometry of the cation-π bond deduced in the present study (Table 1). If this procedure could be generalized to all new structures determined at high resolution, more examples of cation-π interactions could be found in the future. Development of methods able to predict metal ion binding sites in proteins could be used during this process (Yamashita et al., 1990; Nayal & Di Cera, 1996).

The identification of the sodium cation in the present structure emphasizes the need for careful identification and refinement of solvent that should never be neglected as a tedious aspect during the "end game" of structure analysis. Furthermore, this study clearly demonstrates the possibility of stabilization of metal ions by proteins via aromatic residues and provides new distance criteria to search for those binding sites. It opens new possibilities to study the biological relevance of cation bound to π systems for molecular recognition in receptors, active transfer of ions, and stabilization of folded structures.

Material and methods: *X-ray crystallography:* Data collection and refinement statistics are provided in Table 2. Tetragonal crystals (space group $P4_32_12$) of hen egg-white (HEW) lysozyme were obtained in 2% $NaNO_3$, in a 0.1 M sodium acetate buffer pH 4.6 using a protein stock solution containing 20 mg mL^{-1}. A colored small molecule, bromoisatine, was added in the solution with the initial hope to cocrystallized this molecule with the protein. Intensity data were collected at 278 K at the EMBL BW7B beamline at

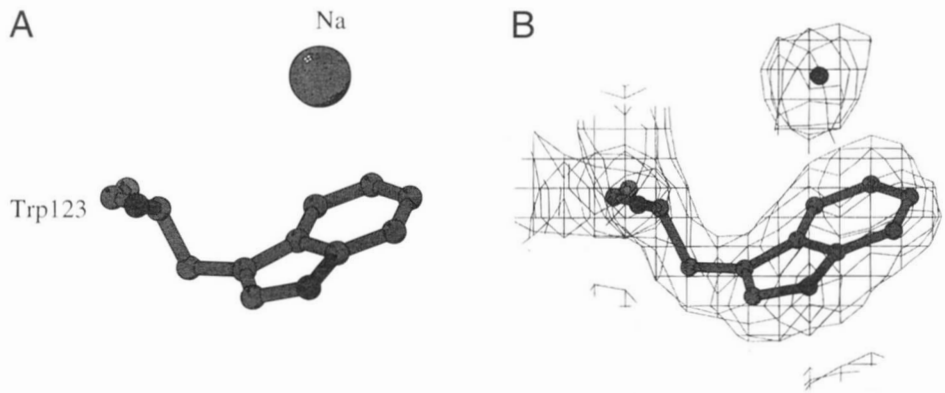


Fig. 1. A: Sodium cation (Na^+) and the indole ring of residue Trp123 in the structure of HEW lysozyme forming the proposed cation-π complex. **B**: The experimental electron density ($2F_o - F_c$) map generated around Trp123 is presented. Figure generated with Bobscript (Esnouf, 1997).

DESY (Hamburg, Germany). Data were processed with Mosflm (CCP4, 1994). The initial model corresponds to the structure deposited in the Brookhaven Protein Data Bank under accession code 1HEL. Starting from this structure, minor conformational changes had to be applied to lateral chains, mainly for residues exposed to the solvent. The loop Thr69–Leu75 was also rebuilt as its rather flexible nature had been emphasized in previous studies (Vaney et al., 1996). The peptide bond of this loop is well ordered and defined into the $3F_o - 2F_c$ or $2F_o - F_c$ maps. In particular the carbonyl group between Arg73 and Asn74 points toward the solvent, and excludes the formation of a octahedric binding site for Na^+. All water molecules were deleted in the initial model. Refinement was performed with Shelxl97 (Sheldrick, 1997) and was monitored by cross-validation using a 5%-test set of reflections. Special attention was paid to the refinement of the solvent. No density corresponding to the dye (bromoisatine) could be detected. The atomic model was finally refined to an *R*-factor of 0.170 (R_{free} = 0.240) for 6,156 reflections between 18.0 and 2.0 Å resolution. Analysis of the stereochemistry with Shelxpro (Sheldrick, 1997) showed that all of the main-chain atoms fall within the allowed regions of the Ramachandran plot.

Table 2. *Data collection and refinement statistics*

Diffraction data	
Resolution range (Å)	18.0–2.0
Source/wavelength	Synchrotron/0.859 Å
Unique data	7,717
R-merge (%)	9.6
Multiplicity	4.6
Completeness (%)	98.5
Refinement	
Space group	P43212
R-factor	0.170
R-free	0.240 (5%-test set)
Number of reflections	6,156
Number of atoms	1,092 (1,002 protein/90 solvent)
RMS deviation of bond lengths (Å)	0.005
RMS deviation of bond angles (°)	0.020

Quantum chemistry: Single determinant ab initio Hartree-Fock-Roothaan methods with an extended 6-31G** basis are adequate to study cation-π complexes of simple ions (Ma & Dougherty, 1997; Mecozzi et al., 1996a, 1996b). Within this formalism, implemented in the Gaussian 94 suite of programs (Frisch et al., 1995), optimizations of cation-π complexes were performed on the following systems: benzene/Na^+, indole/$Na_{|5}{}^+$ and indole/$Na_{|6}{}^+$ (where $Na_{|x}{}^+$ defines the cation along the fivefold (x = 5) or sixfold (x = 6) axis of the five-membered (pyrrole-type) or six-membered rings, respectively).

CSD search: Structures corresponding to metal ion-π complexes involving Na^+ were retrieved from Version 5.14 (containing over 175,000 entries) of the Cambridge Structure Database using the Quest3D program and analyzed with the Vista interface (Allen & Kennard, 1993). Only fragments presenting a distance sodium . . . centroid of a six-membered aromatic ring between 1.0 and 4.5 Å were retained and analyzed.

Acknowledgments: The help of Guy Baudoux was very much appreciated. I also would like to thank Profs. F. Durant and G. Evrard for continuous interest in my work and the staff at EMBL for helpful discussions during data collection.

References

Allen F, Kennard O. 1993. 3D search and research using the Cambridge Structure Database. *Chem Design Auto News 8*:31–37.

Collaborative Computational Project, Number 4. 1994. *Acta Crystallogr D50*:760–763.

Dougherty D. 1996. Cation-π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science 271*:1163–1168.

Dougherty D, Stauffer D. 1990. *Science 250*:1558–1560.

Esnouf RM. 1997. *J Mol Graphics 15*:132–134.

Fraenkel Y, Shalev DE, Gershoni JM, Navon G. 1996. *Crit Rev Biochem Mol Biol 31*:272–301.

Frisch MJ, Trucks GW, Schlegel HB, Gill PMW, Johnson BG, Robb MA, Cheeseman RR, Keith T, Petersson GA, Montgomery JA, Raghavachari K, Al-Laham MA, Zakrzewski VG, Ortiz JV, Foresman JB, Cioslowski J,

Stefanov BB, Nanayakkara A, Challacombe M, Peng CY, Ayala PY, Chen W, Wong MW, Andres JL, Replogle ES, Gomperts R, Martin RL, Fox DJ, Binkley JS, Defrees DJ, Baker J, Stewart JP, Head-Gordon M, Gonzalez C, Pople JA. 1995. *Gaussian 94*, Revision E.2. Pittsburgh, PA: Gaussian, Inc.

Ma J, Dougherty D. 1997. The cation-π interaction. *Chem Rev 97*:1303–1324.

Mecozzi S, West A, Dougherty D. 1996a. Cation-π interactions in simple aromatics: Electrostatics provide a predictive tool. *J Am Chem Soc 118*:2307–2308.

Mecozzi S, West A, Dougherty D. 1996b. Cation-π interactions in aromatics of biological and medicinal interest: Electrostatic potential surfaces as a useful qualitative guide. *Proc Natl Acad Sci USA 93*:10566–10571.

Nayal M, Di Cera E. 1996. Valence screening of water in protein crystals reveals potential Na+ binding sites. *J Mol Biol 256*:228–234.

Perrakis A, Sixma T, Wilson K, Lamzin V. 1997. WARP: Improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. *Acta Crystallogr D53*:448–455.

Sheldrick G. 1997. *Shelx97, program for the refinement of crystal structures.* Göttingen, Germany: University of Göttingen.

Turner M, Howell P. 1995. Structures of partridge egg-white lysozyme with and without tri-N-acetylchitotriose inhibitor at 1.9 Å resolution. *Protein Sci 4*:442–449.

Vaney M, Maignan S, Ries-Kautt M, Ducruix A. 1996. High resolution structure (1.33 Å) of HEW lysozyme tetragonal crystal grown in the APCF apparatus. *Acta Cryst D52*:505–517.

Williams J. 1993. The molecular electric quadrupole moment and solid state architecture. *Acc Chem Res 26*:593–598.

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