Lundblad, R. L., Uhteg, L. C. N., Vogel, C. N., Kingdon, H. S., & Mann, K. G. (1975) *Biochem. Biophys. Res. Commun. 66*, 482–489.
Maizel, J. V. (1971) *Methods Virol. 5*, 179–246.
Marlar, R. A., Kleiss, A. J., & Griffin, J. H. (1982) *Blood 59*, 1067–1072.
McMullen, B. A., Fujikawa, K., Kisiel, W., Sasagawa, T., Howald, W. N., Kwa, E. Y., & Wienstein, B. (1983) *Biochemistry 22*, 2875–2884.
Morita, T., Isaacs, B. S., Esmon, C. T., & Johnson, A. J. (1984) *J. Biol. Chem. 259*, 5698–5704.
Öhlin, A.-K., & Stenflo, J. (1987) *J. Biol. Chem. 262*, 13798–13804.
Öhlin, A.-K., Linse, S., & Stenflo, J. (1988a) *J. Biol. Chem. 263*, 7411–7417.
Öhlin, A.-K., Landes, G., Bourdon, P., Oppenheimer, C., Wydro, R., & Stenflo, J. (1988b) *J. Biol. Chem. 263*, 19240–19248.
Owen, W. G., & Jackson, C. M. (1973) *Thromb. Res. 3*, 705–714.
Plutzky, J., Hoskins, J. A., Long, G. L., & Crabtree, G. R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 546–550.
Pollock, J. S., Shepard, A. J., Weber, D. J., Olson, D. L., Klapper, D. G., Pedersen, L. G., & Hiskey, R. G. (1988) *J. Biol. Chem. 263*, 14216–14223.
Rees, D. J. G., Jones, I. M., Handford, P. A., Walter, S. J., Esnouf, M. P., Smith, K. J., & Brownlee, G. G. (1988) *EMBO J. 7*, 2053–2061.
Ryan, J., Wolitsky, S., Heimer, E., Felix, A., Huang, L., Tam, J., Kisiel, W., Nawroth, P., & Stern, D. (1989) *FASEB J. 3*, A1050 (Abstract 4787).
Skogen, W. F., Bushong, D. S., Johnson, A. E., & Cox, A. C. (1983) *Biochem. Biophys. Res. Commun. 111*, 14–20.
Stearns, D. J., Kurosawa, S., Sims, P. J., Esmon N. L., & Esmon, C. T. (1988) *J. Biol. Chem. 263*, 826–832.
Stenflo, J. (1976) *J. Biol. Chem. 251*, 355–363.
Stenflo, J. (1988) in *Protein C and Related Proteins* (Bertina, R. M., Ed.) pp 21–54, Churchill-Livingstone, London.
Stenflo, J., & Suttie, J. W. (1977) *Annu. Rev. Biochem. 46*, 157–172.
Stenflo, J., & Jönsson, M. (1979) *FEBS Lett. 101*, 377–381.
Stenflo, J., & Fernlund, P. (1982) *J. Biol. Chem. 257*, 12180–12190.
Stenflo, J., Holme, E., Lindstedt, S., Chandramouli, N., Tsai Huang, L. H., Tam, J. P., & Merrifield, R. B. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 444–447.
Sugo, T., Björk, I., Holmgren, A., & Stenflo, J. (1984) *J. Biol. Chem. 259*, 5705–5710.
Sugo, T., Persson, U., & Stenflo, J. (1985) *J. Biol. Chem. 260*, 10453–10457.
Suzuki, K., Stenflo, J., Dahlbäck, B., & Theodorsson, B. (1983) *J. Biol. Chem. 258*, 1914–1920.
Vehar, G. A., & Davie, E. W. (1980) *Biochemistry 19*, 401–410.
Walker, F. J., Sexton, P. W., & Esmon, C. T. (1979) *Biochim. Biophys. Acta 571*, 333–342.

Systematics in the Interaction of Metal Ions with the Main-Chain Carbonyl Group in Protein Structures†

P. Chakrabarti*
Department of Chemistry and Biochemistry, Molecular Biology Institute, University of California, Los Angeles, California 90024

Received June 8, 1989; Revised Manuscript Received September 5, 1989

ABSTRACT: An analysis of the geometry of metal binding by peptide carbonyl groups in proteins is presented. Such metal ions are predominantly calcium in known protein structures. Cations tend to be located in the peptide plane, near the C=O bond direction. This distribution differs from that observed for water molecules bound to carbonyl oxygens. Most metal ions are bound to carbonyl oxygens of peptides in turns or in regions with no regular secondary structure. More infrequent binding interactions occur at the C-terminal end of α-helices or at the edges and sides of β-sheets, where the geometrical preferences of the metal–carbonyl interaction may be satisfied. In many proteins carbonyl groups that are one, two, or three residues apart along the polypeptide chain bind to the same cation; these structures show a limited number of main-chain conformations around the metal center.

Metal ions perform a wide variety of physiological functions, such as structural stabilization, electron transfer, catalysis, transport, and storage (Williams, 1983). The main-chain carbonyl group is an important protein ligand, and the interaction of many model compounds with metal ions has been studied spectroscopically and crystallographically (Chakrabarti et al., 1981; Einspahr & Bugg, 1984). In this paper we extend the work of Einspahr and Bugg (1984) to study the geometry of such interactions in protein structures. The orientation of metal ions with respect to the peptide group, the secondary structure where such a group is located, and the folding of the peptide chain containing more than one carbonyl ligand group are also examined. Such an analysis will help to understand why a metal ion binds to a given site in a protein structure and how it functions.

Materials and Methods

The analysis is based on atomic coordinates from the Brookhaven Protein Data Bank (PDB) (Bernstein et al., 1977). Only the refined structures were included; as a result, taka-

† This work was supported by USPHS Grant GM 31299.
* Address correspondence to the author at the Division of Chemistry, California Institute of Technology, 147-75 CH, Pasadena, CA 91125.

0006-2960/90/0429-651$02.50/0

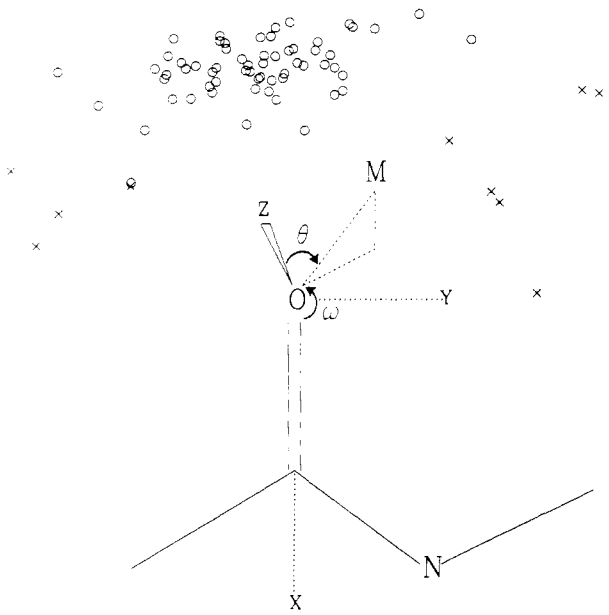

FIGURE 1: Projection of metal ions on the peptide plane for structures given in Table I. Unidentate and bidentate ligands are represented by circles and crosses, respectively. The diagrammatic representation of the spherical polar coordinates used to express metal (M)–carbonyl interactions is also shown.

amylase A containing a calcium ion (Matsuura et al., 1984) was excluded. Only one protein was used to represent a family of homologous molecules. However, for trypsin, structures based on two different crystal forms and the zymogen and, for phospholipase A2, two structures containing different numbers of metal ions have been used to calculate geometrical parameters.

The relative position of the metal ion was expressed in the coordinate system depicted in Figure 1. The ligand O atom is at the origin, the x axis is along the direction from O to C, the peptide plane is coincident with the xy plane, and the z axis is along the normal to the peptide plane. The geometry of metal–ligand bonding is given by the following spherical polar coordinates: the M–O distance; the acute angle θ between the M–O direction and the z axis; the angle ω between the x axis and the projection of the O–M direction on the xy plane.

Usually the peptide group coordination is only through the carbonyl oxygen (unidentate binding). In a few instances, the coordination involves not only the oxygen but another atom from any one of the two residues constituting the peptide bond, and these are classified as bidentate binding. The secondary structural features of all ligand residues were defined according to the method of Kabsch and Sander (1983).

Results and Discussion

Geometry of Metal–Carbonyl Interaction. Values for various parameters are given in Table I. The majority of the structures contain calcium; Ca–O distances are usually distributed in the range 2.2–2.5 Å (shown in Figure 2), with an average of 2.4 (2) Å for 62 values. Given the lesser accuracy of protein structures, these values compare very well with those found in small molecules (Einspahr & Bugg, 1984). The distribution of the M—O=C angle that the metal ion makes with the carbonyl group is shown in Figure 3. For unidentate cases most of the angles are in the range 140–170°, and for bidentate examples the favored range is 110–130°. Although metal–ligand distances and angles are usually not restrained during the refinement, the geometry is found to be similar in various structures.

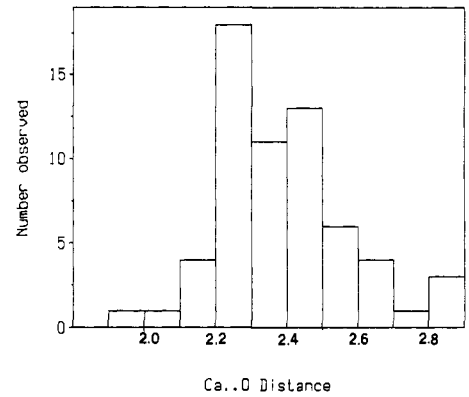

FIGURE 2: Histogram showing the distribution of Ca–O distances (Å).

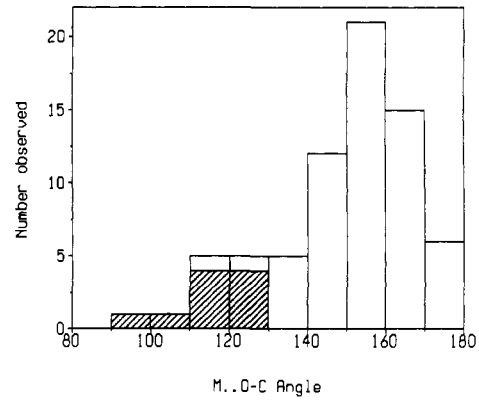

FIGURE 3: Histogram showing the distribution of M—O=C angles (deg) for all structures in Table I. The shaded area corresponds to bidentate cases.

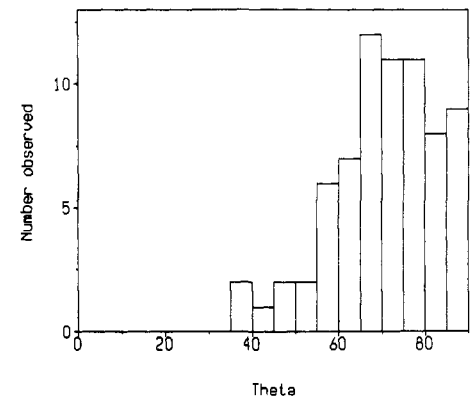

FIGURE 4: Histogram of θ (deg) showing the deviation of metal ions from the peptide plane.

Orientation of the Metal Ion. The deviation of metal ions from the peptide plane, as given by the angle θ, is shown in Figure 4. Most of the θ values are in the range 55–90°, which means that metal ions generally do not deviate by more than 35° from the peptide plane. Larger deviations may occur with a few carbonyl groups that have another carbonyl group one to three residues away binding to the same ion; for example, in thermolysin, the metal ion bound to both Y193 and T194 shows large deviations, as does the one to Y171 (flanked between A169 and V174) in subtilisin (Table I).

The distribution of metal ions in the plane of the peptide group is given by the angle ω and is shown in Figure 1. This angle for bidentate cases is outside the range shown by unidentate examples. In the bidentate mode of binding, residues comprising the peptide group provide, in addition to the carbonyl group, another binding site to the metal ion, and the

metal ion has a restricted region in space from which it can interact with both of the sites simultaneously. The second site is the amino group of residue G1 in 1PPT [a similar coordination is shown by residue W1 in the recently determined structure of phospholipase C (Hough et al., 1989)], the side-chain hydroxyl group of T194 in 3TLN, the side-chain carboxyl group of E60 in 2WRP, and the side-chain nitrogen of H46 in 2AZA.

Most of the ω values for the unidentate mode of binding are in the range 170–208°. Those residues with metal ions outside this range usually have within three residues another carbonyl group binding to the same metal ion (Table I). Even in satellite tobacco necrosis virus, where the carbonyl group (T138) binds to a calcium ion sitting on a 5-fold axis (Jones & Liljas, 1984), the ω angle is within the preferred range. The distribution of metal ions is not symmetrical about the C=O direction (ω = 180°); there are more metal ions with ω greater than 180°. The reason could be the increased steric interaction of metal ions with the C_α-position of the second residue constituting the peptide moiety, as the ω angle is decreased beyond 170°. There is much more freedom on the other side of the C=O direction (ω increasing from 180°) before metal ions start interacting with the C_α-position of the first residue.

Values of θ and ω observed indicate that metal ions do not interact with carbonyl lone-pair orbitals, which are in the peptide plane, and with ω angles of 120 and 240°. In contrast, water molecules bound to the peptide carbonyl group usually approach along the lone-pair direction, and a very few are along the C=O direction (Thanki et al., 1988; Rees et al., 1983; Baker & Hubbard, 1984). This difference in the geometry of interaction can also be used to identify solvent molecules as metal ions during the refinement of the protein structure (Poulos et al., 1987). In azurin, the Cu ion is close to the lone-pair direction. This coordination geometry is similar to other bidentate ligands (Table I). However, being a transition metal, Cu could prefer to interact with the lone-pair orbital (Lewinski & Lebioda, 1986). The slightly longer Cu–O distance observed (Baker, 1988) could be the consequence of meeting the stereoelectronic requirements for the two ligand groups in the bidentate mode of binding.

Because of the limited range of ω angles, a peptide carbonyl cannot act as a bridge between metal ions, as has been observed in some metal complexes of amides (Chakrabarti et al., 1981; Lewinski & Lebioda, 1986). Residues that are close along the sequence usually bind to the same metal ion; however, in thermolysin two carbonyl groups that are four residues apart bind to two different calcium ions.

Table I shows that different crystal forms of trypsin (2PTN, 3PTN, 2TGA) and phospholipase A2 (1BP2, 1P2P) can have some differences in θ and ω values.

Secondary Structural Preference. Secondary structural characteristics for the two residues (i and i + 1) that make up the peptide group binding the metal ion are presented in Table II. The main-chain dihedral angles ψ_i and ϕ_{i+1} for the two residues are also tabulated and plotted in Figure 5.

As metal ions interact with the peptide carbonyl group close to the C=O direction and not in a direction perpendicular to the peptide plane, the only place where a main-chain carbonyl group belonging to a helix can bind a metal ion is at the carboxyl end. This mode of binding is indeed found in the calcium-binding protein (shown in Figure 6), cytochrome P450cam, and *trp* repressor. It is known that anions bind to the N-terminus end of a helix, and this has been explained in terms of the helix dipole (Hol, 1985). However, the binding of metal ions to the C-terminus of a helix seems to be due to

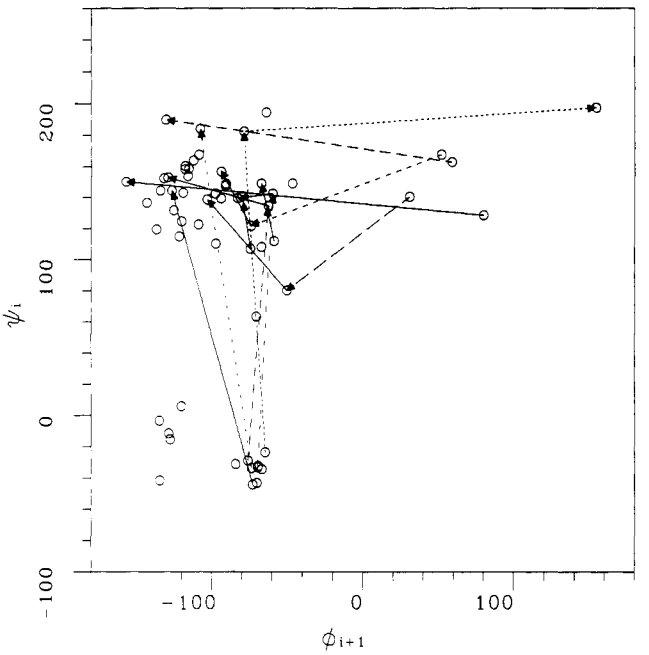

FIGURE 5: Plot of paired values of ψ_i and ϕ_{i+1} torsional angles (deg) for structures given in Table II. (360° has been added to a few ψ_i values in the table, so that these values in the plot are in the range −100° to 260°.) Vectors drawn in solid lines join points for residue n to n + 3, and those drawn in dashed lines connect residue n to n + 1 or n + 2 (only nonhelix residues have been considered).

the spatial requirement of the interaction of a metal ion with the peptide carbonyl group; in the three cases mentioned, the cations are not along the helix axis, and the ligand carbonyl groups point outward away from the helix direction. For the same reason, metal ions can bind only to the peripheral strands and at the sides of a β-sheet. This is found in the structure of thermolysin (depicted in Figure 7), dihydrofolate reductase, and calmodulin. However, in STNV, a residue (Q64) on a strand in the middle of a β-sheet binds a calcium ion, because the residue is toward the edge of the sheet where adjacent strands move away from each other; a turn preceding the strand provides another ligand. A similar arrangement of two ligands coming from a strand and the adjacent turn is also found in 1SGT.

As can be seen from Table II, the majority of the ligand carbonyl groups are from turns or from regions with nonregular structure; these carbonyls are not tied up in regular secondary structure and are free to interact with metals. Peptide carbonyls and mostly carboxylic side chains are found coordinating to calcium ions (Einspahr & Bugg, 1984). However, in 2CPP all protein ligands are peptide carbonyl groups. Similarly, Ca(1) in 3ICB has four peptide carbonyls and just one carboxyl group in its coordination sphere. In both these cases calcium is at the C-terminus of a helix; in this position the ligands can be provided by both the helix and the adjacent loop. The majority of protein ligands in 1BP2 and those for Ca(2) and Ca(3) in 2SEC are carbonyl groups; these are close along the sequence and are usually from turn regions.

Main-Chain Conformation. Table II shows that in many structures, when the carbonyl group of a residue at position n binds to a metal ion, the carbonyl group at position n + 1 or n + 2 or n + 3 also binds to the same ion. Residues 93 and 94 in 2CPP, being at the C-terminus end of a helix, are in a unique position to bind to the same metal ion. In Figure 5 vectors have been drawn connecting ($\psi_i\phi_{i+1}$) for the first ligand peptide group to that of the second ligand peptide group. It can be seen that the end points of the vectors between n and n + 3 residues are very close. Those connecting n to n + 1

Table I: Geometric Parameters for Metal (M)–Carbonyl Interactions

PDB code[a]	protein name	metal[b]	residue[b]	M–O distance (Å)	M—O=C angle (deg)	angle (deg)[c] ω	angle (deg)[c] θ
			(A) Unidentate Binding				
2APR	aspartic proteinase	Ca	G220	2.34	150.8	207.1	78.3
3BCL	bacteriochlorophyll *a*	Mg(5)	L234	2.04	155.1	204.6	86.4
1BP2	phospholipase A2, bovine	Ca	Y28	2.28	158.1	201.0	84.5
			G30	2.47	164.4	194.6	84.5
			G32	2.32	156.3	190.4	68.4
1P2P	phospholipase A2, porcine	Ca(1)	Y28	2.56	156.6	202.9	84.9
			G30	2.50	169.0	169.3	87.1
			G32	2.68	132.0	221.6	63.5
		Ca(2)	S72	2.37	148.3	207.8	74.1
3CNA	concanavalin A	Ca	Y12	1.96	170.3	189.7	88.8
2CPP	cytochrome P450cam	Ca[d]	E84	2.69	157.1	170.0	67.6
			G93	2.75	148.8	148.8	88.5
			E94	2.88	113.6	228.9	39.3
			Y96	2.64	142.5	208.5	64.5
2CPV	paravalbumin B	Ca(1)	K96	2.28	135.6	193.5	58.7
		Ca(2)	F57	2.25	156.8	190.2	76.3
4DFR	dihydrofolate reductase	Ca	S135(B)	2.45	151.0	203.4	71.8
3EST	elastase	Ca	N72	2.42	153.8	191.4	64.8
			Q75	2.41	156.0	184.4	69.9
3ICB	Ca-binding protein	Ca(1)	A14	2.16	137.3	217.4	67.4
			E17	2.39	160.5	176.0	69.6
			D19	2.38	161.7	180.5	69.8
			Q22	2.21	164.6	173.7	77.0
		Ca(2)	E60	2.26	155.2	182.6	65.7
2PRK	proteinase K	Ca(1)	P175	2.41	143.8	199.7	60.0
			V177	2.46	173.1	183.9	84.9
		Ca(2)	T16	2.44	157.9	194.0	73.8
2PTN	trypsin	Ca	N72	2.28	158.5	178.3	66.2
			V75	2.26	148.1	168.7	57.9
3PTN	trypsin	Ca	N72	2.26	147.3	186.6	57.2
			V75	2.29	165.5	172.8	76.1
2TGA	trypsinogen	Ca	N72	2.24	154.3	183.0	62.9
			V75	2.25	146.9	170.5	57.4
1SGT	trypsin *S. griseus*	Ca	A177	2.24	159.5	197.1	78.0
			E180	2.26	159.1	171.5	68.5
2SEC	subtilisin	Ca(1)	L75	2.23	163.5	187.1	73.5
			T79	2.41	164.7	175.6	75.1
			V81	2.28	157.0	203.0	88.4
		Ca(2)	A169	2.57	161.4	165.3	77.7
			Y171	2.57	128.1	177.3	39.1
			V174	2.54	156.5	194.9	71.0
		Ca(3)	A37	2.81	155.9	158.7	79.6
			H39	2.64	139.1	185.6	49.6
			L42	2.58	164.9	181.2	72.8
4TNC	troponin C	Ca(1)	F112	2.38	174.9	184.9	88.2
		Ca(2)	R148	2.27	156.1	191.1	69.2
3CLN	calmodulin	Ca(1)	T26	2.46	164.0	187.0	74.5
		Ca(2)	T62	2.17	170.9	179.8	80.2
		Ca(3)	Y99	2.06	170.2	188.1	85.2
		Ca(4)	Q135	2.38	147.6	188.6	58.3
4SBV	SBMV	Ca(A)	F199(B)	2.38	140.7	208.4	62.1
		Ca(B)	F199(C)	2.42	163.2	194.0	80.9
		Ca(C)	F199(A)	2.45	153.9	205.1	82.1
2STV	STNV	Ca(1)	S61	2.22	172.4	186.7	86.4
			Q64	2.33	148.4	198.1	63.3
		Ca(3)	T138	2.13	163.5	169.4	73.2
2SNS	staphylococcal nuclease	Ca	T41	2.86	136.0	223.1	79.8
3TLN	thermolysin	Ca(1)	E187	2.29	143.2	184.9	53.2
		Ca(2)	N183	2.37	161.1	181.5	70.3
		Ca(3)	Q61	2.18	161.9	198.1	88.0
		Ca(4)	I197	2.31	145.4	199.6	60.5
			(B) Bidentate Binding[e]				
2AZA	azurin	Cu(A)	G45(A)	3.16	127.0	127.8	75.3
		Cu(B)	G45(B)	3.09	128.6	129.7	75.5
1PPT	pancreatic polypeptide	Zn	G1	2.13	114.5	245.1	82.6
3TLN	thermolysin	Ca(4)	Y193	2.43	128.2	140.2	52.7
			T194	2.54	116.2	229.7	43.3
2WRP	*trp* repressor	Na	E60	2.74	117.1	241.6	74.3
3CPA[f]	carboxypeptidase A complex	Zn	G1	2.34	105.4	253.4	68.3
				(2.78)	(106.8)	(244.0)	(42.0)
4TLN	thermolysin complex	Zn	O	2.00	119.7	120.9	74.0
5TLN	thermolysin complex	Zn	O1	2.10	121.0	124.2	65.2
7TLN	thermolysin complex	Zn	O1	2.85	94.9	97.0	47.0

Table I footnotes

[a] 2APR, Suguna et al., 1987. 3BCL, Tronrud et al., 1986. 1BP2, Dijkstra et al., 1981. 1P2P, Dijkstra et al., 1983. 3CNA, Hardman & Ainsworth, 1972. 2CPP, Poulos et al., 1987. 2CPV, Moews & Kretsinger, 1975. 4DFR, Bolin et al., 1982. 3EST, Meyer et al., 1988. 3ICB, Szebenyi & Moffat, 1986. 2PRK, Betzel et al., 1988. 2PTN, 3PTN, and 2TGA, Marquart et al., 1983. 1SGT, Read & James, 1988. 2SEC, McPhalen et al., 1985. 4TNC, Satyshur et al., 1988. 3CLN, Babu et al., 1988. 4SBV, Silva & Rossmann, 1985. 2STV, Jones & Liljas, 1984. 2SNS, Cotton et al., 1979. 3TLN, Holmes & Matthews, 1982. 2AZA, Baker, 1988. 1PPT, Blundell et al., 1981. 2WRP, Lawson et al., 1988. 3CPA, Christianson & Lipscomb, 1986. 4TLN and 5TLN, Holmes & Matthews, 1981. 7TLN, Holmes et al., 1983. [b] 1, 2, or A, B etc. are used to distinguish different metal atoms or subunits in the same protein molecule. Amino acids are represented by one-letter codes. In the last three entries involving peptide-like substrates, ligand atoms are indicated by names as given in the original papers. [c] As defined in the text. [d] A solvent molecule that has been identified as Ca. [e] The carbonyl group belongs to the protein in the first four entries and to substrates in the rest. [f] The first set of values is from the low-temperature study and has been used in Figure 1; the lower set is based on the analysis at room temperature.

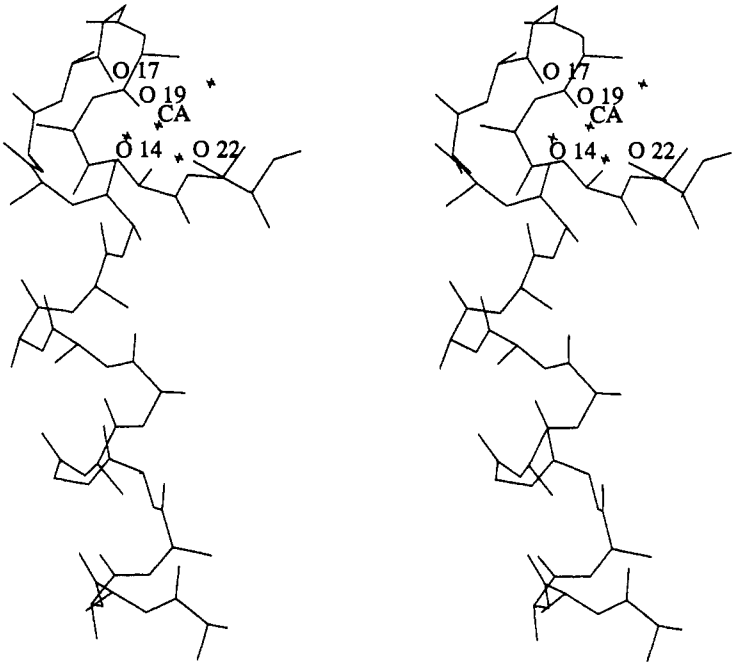

FIGURE 6: Stereoview showing the binding of a calcium ion at the C-terminus of a helix (side-chain atoms beyond C_β are not shown) in a calcium-binding protein (3ICB). Peptide carbonyl ligands are indicated; Ca and three other oxygen ligands are shown as crosses.

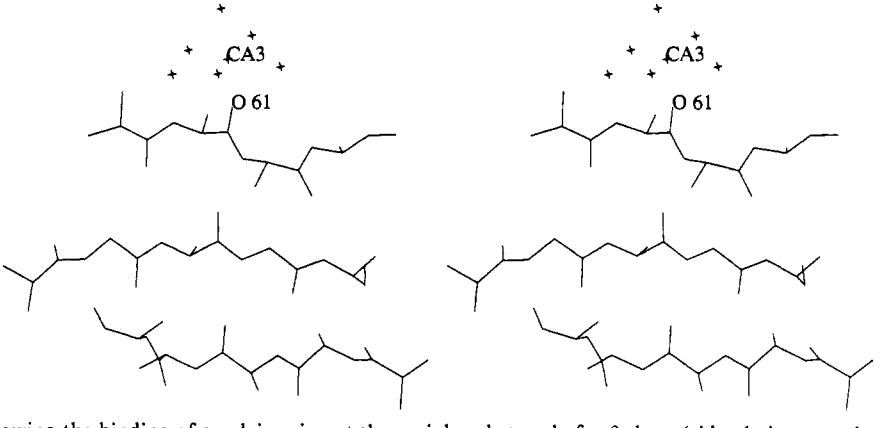

FIGURE 7: Stereoview showing the binding of a calcium ion at the peripheral strand of a β-sheet (side-chain atoms beyond C_β are not shown) in thermolysin. The ligand carbonyl group and the Ca ion are indicated; positions of other oxygen ligands are shown by crosses.

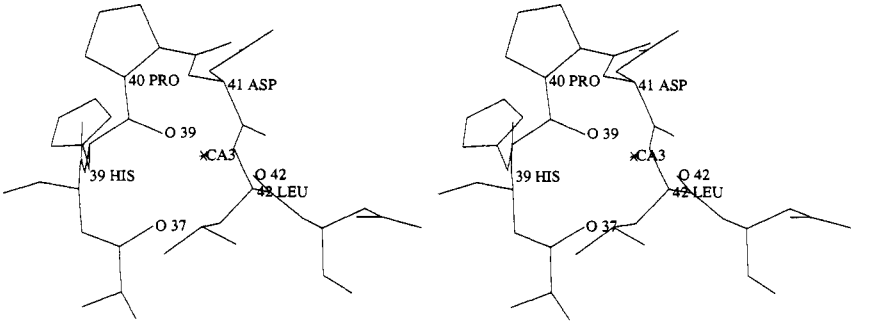

FIGURE 8: Stereoview showing the binding of a calcium ion at the type I β-turn (residues 39–42) in subtilisin Carlsberg.

Table II: Secondary Structural Features for Residues i and $i + 1$ of the Peptide Group Binding Metal Ions

PDB code	residue number (i)	residue		torsion angle (deg)[a]		structure[b]	
		i	$i + 1$	ψ_i	ϕ_{i+1}	i	$i + 1$
2APR	220	Gly	Thr	5.9	−120.1	T	C
3BCL	234	Leu	Phe	−11.5	−128.5	T	C
1P2P	28	Tyr	Cys[c]	−23.5	−64.5	T	B
	30	Gly	Leu	−177.6	−78.9	S	S
	32	Gly	Gly	−162.6	154.4	C	C
	72	Ser	Tyr	136.5	−143.0	C	C
3CNA	12	Tyr	Pro	108.1	−67.1	S	C
2CPP	84	Gly	Cys	−41.5	−134.4	T	S
	93	Gly	Glu	−32.3	−69.3	H	H
	94	Glu	Ala	−43.1	−70.1	H	H
	96	Tyr	Asp	119.4	−136.7	C	C
2CPV	96	Lys	Ile	142.9	−118.9	E	E
	57	Phe	Ile	142.4	−97.4	S	E
4DFR	135	Ser	Val	122.3	−108.5	E	E
3EST	72	Asn	Leu	112.0	−58.9	B	T
	75	Gln	Asn	148.9	−67.2	S	C
3ICB	14	Ala	Ala	−31.0	−84.2	H	T
	17	Glu	Gly	167.9	51.8	S	S
	19	Asp	Pro	121.5	−74.0	S	T
	22	Gln	Leu	156.7	−93.8	S	B
	60	Glu	Val	153.7	−115.7	C	B
2PRK	175	Pro	Ser	−33.5	−73.7	T	T
	177	Val	Cys[c]	−176.1	−107.7	S	E
	16	Thr	Ser	−15.3	−127.7	S	S
3PTN	72	Asn	Ile	107.0	−74.4	C	T
	75	Val	Val	140.8	−79.9	S	C
1SGT	177	Ala	Asn	−44.1	−72.8	T	T
	180	Glu	Ile	144.3	−126.4	E	E
2SEC	75	Leu	Asp	147.5	−90.7	C	S
	79	Thr	Gly	163.1	58.8	S	S
	81	Val	Leu	−170.3	−130.6	C	C
	169	Ala	Lys	−33.1	−69.6	T	T
	171	Tyr	Asp	142.4	−59.7	S	T
	174	Val	Ile	139.4	−82.9	S	E
	37	Ala	Ser	−28.8	−76.0	T	T
	39	His	Pro	134.4	−62.6	C	T
	42	Leu	Asn	152.9	−129.1	C	C
4TNC	112	Phe	Ile	148.7	−90.5	S	B
	148	Arg	Ile	167.5	−108.5	S	B
3CLN	26	Thr	Ile	158.5	−115.2	E	E
	62	Thr	Ile	163.9	−112.1	S	E
	99	Tyr	Ile	139.2	−93.7	E	E
	135	Gln	Val	160.3	−117.7	S	E
4SBV[d]	199	Phe	Lys	110.4	−97.2	C	C
				131.6	−125.1	C	C
				114.9	−121.6	C	C
2STV	61	Ser	Gly	128.6	80.2	S	T
	64	Gln	Val	150.1	−156.8	E	E
	138	Thr	Gly	−3.4	−134.7	S	S
2SNS	41	Thr	Pro	148.9	−46.8	C	C
3TLN	187	Glu	Ile	158.3	−117.8	E	E
	183	Asn	Pro	63.3	−70.5	S	C
	61	Gln	Phe	124.5	−119.8	E	E
	193	Tyr	Thr	140.6	30.7	B	C
	194	Thr	Pro	80.2	−50.3	C	T
	197	Ile	Ser	138.5	−102.9	S	S
2AZA[d]	45	Gly	His	144.3	−134.0	C	B
				152.4	−131.7	C	B
1PPT	1	Gly	Pro	−165.7	−64.4	C	C
2WRP	60	Glu	Leu	−34.3	−66.9	H	H

[a] $\psi_i = N_i$–$C_{(\alpha)i}$–C_i–N_{i+1}; $\phi_{i+1} = C_i$–N_{i+1}–$C_{(\alpha)i+1}$–C_{i+1}. [b] As defined by Kabsch and Sander (1983): B, residue in isolated β-bridge; C, nonregular structure; E, extended strand; G, 3_{10}-helix; H, α-helix; S, bend; T, H-bonded turn. [c] Involved in disulfide bond. [d] Values for various subunits or independent molecules in crystal structures are given.

Table III: Amino Acid Preferences at Positions i and $i + 1$ of the Peptide Group Binding Metal Ions

position	amino acid[a]																			
	G	A	V	L	I	F	P	M	S	T	Y	W	N	Q	C	D	E	K	R	H
i	7	4	4	3	1	3	1	0	3	7	6	0	3	5	0	1	6	1	1	1
$i + 1$	5	2	5	5	11	2	7	0	4	2	1	0	3	0	3	3	1	2	0	1

[a] Represented by one-letter code and taken from Table II.

or $n + 2$ form two groups with starting points in two different regions, but they broadly point to the same final position. Only the second vector from 1P2P behaves anomalously. This shows that there is a restriction in conformation when two carbonyl groups closely placed along the sequence bind to the same metal ion, as has been observed for cysteine ligands that are three residues apart (Chakrabarti, 1989). There may be a limited number of ways the polypeptide chain can fold around a metal ion.

Sequence. The composition of the first, i, and the second, $i + 1$, residues constituting the ligand peptide group has been analyzed in Table III. Hydrophobic residues are favored in both positions, as can be expected from the average composition of such residues in protein structures (Janin et al., 1988). Glycine seems to be favored in both positions, and many of these are from cases where two carbonyl groups one or two residues apart bind to the same metal ion, as in phospholipase A2. The occurrence of proline at the second position has been noted before (Einspahr & Bugg, 1984), and for a few cases, this seems to be due to the preference of proline to be at the second position of a type I β-turn (Wilmot & Thornton, 1988). The binding of a metal ion at a turn conformation is presented in Figure 8, which also shows that this conformation makes it possible for the first and the last residues in a type I β-turn to bind to the same metal ion. Residues 19–22 in 3ICB and 194–197 in 3TLN show similar conformations.

There are only a few cases where positively charged residues are adjacent to the carbonyl involved in metal binding. In all these cases a short stretch of the polypeptide chain provides multiple ligands to the metal; only in 4SBV does the ligand peptide group involve a lysine, and there is no other ligand nearby. Of the three cysteine residues at the second position in Table II, two are involved in disulfide bonds, which impart rigidity to the protein fold close to the metal center.

Argos et al. (1978) have shown that a few zinc-containing enzymes have similar active center geometries. In this analysis we have examined metal–ligand interactions from the perspective of the ligand group and have found that such interactions have a preferred geometry that is independent of the type of metal ion found in protein structures. A knowledge of this geometry for the carbonyl and other protein ligands (Chakrabarti, 1989) and the protein secondary structure where a ligand can be found will help in engineering metal-binding sites in protein structures (Pantoliano et al., 1988) and in designing coordination sites in model compounds (Ibers & Holm, 1980).

ACKNOWLEDGMENTS

I am grateful to Professor D. Rees for discussion and to Professors C. Bugg and W. Lipscomb and Drs. Y. Babu and D. Christianson for providing coordinates based on their work.

Registry No. Ca, 7440-70-2.

REFERENCES

Argos, P., Garavito, R. M., Eventoff, W., Rossmann, M. G., & Brändén, C. I. (1978) *J. Mol. Biol. 126*, 141–158.

Babu, Y. S., Bugg, C. E., & Cook, W. J. (1988) *J. Mol. Biol. 204*, 191–204.

Baker, E. N. (1988) *J. Mol. Biol. 203*, 1071–1095.

Baker, E. N., & Hubbard, R. E. (1984) *Prog. Biophys. Mol. Biol. 44*, 97–179.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.

Betzel, C., Pal, G. P., & Saenger, W. (1988) *Acta Crystallogr. Sect. B 44*, 163–172.

Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P., & Wu, C.-W. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4175–4179.

Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., & Kraut, J. (1982) *J. Biol. Chem. 257*, 13650–13662.

Chakrabarti, P. (1989) *Biochemistry 28*, 6081–6085.

Chakrabarti, P., Venkatesan, K., & Rao, C. N. R. (1981) *Proc. R. Soc. London, Sect. A 375*, 127–153.

Christianson, D. W., & Lipscomb, W. N. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7568–7572.

Cotton, F. A., Hazen, E. E., Jr., & Legg, M. J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2551–2555.

Dijkstra, B. W., Kalk, K. H., Hol, W. G. J., & Drenth, J. (1981) *J. Mol. Biol. 147*, 97–123.

Dijkstra, B. W., Renetseder, R., Kalk, K. H., Hol, W. G. J., & Drenth, J. (1983) *J. Mol. Biol. 168*, 163–179.

Einspahr, H., & Bugg, C. E. (1984) *Met. Ions Biol. Syst. 17*, 51–97.

Hardman, K. D., & Ainsworth, C. F. (1972) *Biochemistry 11*, 4910–4919.

Hol, W. G. J. (1985) *Prog. Biophys. Mol. Biol. 45*, 149–195.

Holmes, M. A., & Matthews, B. W. (1981) *Biochemistry 20*, 6912–6920.

Holmes, M. A., & Matthews, B. W. (1982) *J. Mol. Biol. 160*, 623–639.

Holmes, M. A., Tronrud, D. E., & Matthews, B. W. (1983) *Biochemistry 22*, 236–240.

Hough, E., Hansen, L. K., Birknes, B., Jynge, K., Hansen, S., Hordvik, A., Little, C., Dodson, E., & Derewenda, Z. (1989) *Nature 338*, 357–360.

Ibers, J. A., & Holm, R. H. (1980) *Science 209*, 223–235.

Janin, J., Miller, S., & Chothia, C. (1988) *J. Mol. Biol. 204*, 155–164.

Jones, T. A., & Liljas, L. (1984) *J. Mol. Biol. 177*, 735–767.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.

Lawson, C. L., Zhang, R., Schevitz, R. W., Otwinowski, Z., Joachimiak, A., & Sigler, P. B. (1988) *Proteins 3*, 18–31.

Lewinski, K., & Lebioda, L. (1986) *J. Am. Chem. Soc. 108*, 3693–3696.

Matsuura, Y., Kusunoki, M., Harada, W., & Kakudo, M. (1984) *J. Biochem. 95*, 697–702.

McPhalen, C. A., Schnebli, H. P., & James, M. N. G. (1985) *FEBS Lett. 188*, 55–58.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983) *Acta Crystallogr., Sect. B 39*, 480–490.

Meyer, E., Cole, G., Radhakrishnan, R., & Epp, O. (1988) *Acta Crystallogr., Sect. B 44*, 26–38.

Moews, P. C., & Kretsinger, R. H. (1975) *J. Mol. Biol. 91*, 201–228.

Pantoliano, M. W., Whitlow, M., Wood, J. F., Rollence, M. L., Finzel, B. C., Gilliland, G. L., Poulos, T. L., & Bryan, P. N. (1988) *Biochemistry 27*, 8311–8317.

Poulos, T. L., Finzel, B. C., & Howard, A. J. (1987) *J. Mol. Biol. 195*, 687–700.

Read, R. J., & James, M. N. G. (1988) *J. Mol. Biol. 200*, 523–551.

Rees, D. C., Lewis, M., & Lipscomb, W. N. (1983) *J. Mol. Biol. 168*, 367–387.

Satyshur, K. A., Rao, S. T., Pyzalska, D., Drendel, W., Greaser, M., & Sundaralingam, M. (1988) *J. Biol. Chem. 263*, 1628–1647.

Silva, A. M., & Rossmann, M. G. (1985) *Acta Crystallogr., Sect. B 41*, 147-157.
Suguna, K., Bott, R. R., Padlan, E. A., Subramanian, E., Sheriff, S., Cohen, G. H., & Davies, D. R. (1987) *J. Mol. Biol. 196*, 877-900.
Szebenyi, D. M. E., & Moffat, K. (1986) *J. Biol. Chem. 261*, 8761-8777.
Thanki, N., Thornton, J. M., & Goodfellow, J. M. (1988) *J. Mol. Biol. 202*, 637-657.
Tronrud, D. E., Schmid, M. F., & Matthews, B. W. (1986) *J. Mol. Biol. 188*, 443-454.
Williams, R. J. P. (1983) *Pure Appl. Chem. 55*, 35-46.
Wilmot, C. M., & Thornton, J. M. (1988) *J. Mol. Biol. 203*, 221-232.

Isolation and Characterization of a Tyrosyl Phosphatase Activator from Rabbit Skeletal Muscle and *Xenopus laevis* Oocytes†

Xavier Cayla,‡ Jozef Goris,*,§ Jacques Hermann,‡ Peter Hendrix,§ René Ozon,‡ and Wilfried Merlevede§
Afdeling Biochemie, Faculteit Geneeskunde, Katholieke Universiteit te Leuven, B-3000 Leuven, Belgium, and Laboratoire de Physiologie de la Reproduction, Université Pierre et Marie Curie, 75005 Paris, France
Received March 7, 1989; Revised Manuscript Received August 29, 1989

ABSTRACT: PTPA, a specific phosphotyrosyl phosphatase activator of the PCS_{H2} and PCS_L protein phosphatases, was purified up to apparent homogeneity from *Xenopus laevis* ovaries and rabbit skeletal muscle and highly purified from dog liver. PTPA appears as a 40-kDa protein in gel filtration, as well as in sucrose gradient centrifugation, and as a 37-39-kDa protein doublet in SDS-PAGE. Its estimated cellular concentration of 0.75 μM in oocytes or 0.25 μM in rabbit skeletal muscle is suggestive of an important role in the regulation of the cellular PTPase activity. The PTPase activation reaction of the PCS_L phosphatase is time-dependent, ATP and Mg^{2+} being essential cofactors [A_{50}(ATP) = 0.12 mM in the presence of 5 mM $MgCl_2$]. With RCM lysozyme as substrate, the specific activity of the PTPA-activated PCS_L phosphatase is 700 nmol of P_i/(min·mg). The pH optimum of the PTPase shifts from 8.5-9 in basal conditions to a neutral pH (7-7.5), and the A_{50} for the essential metal ion Mg^{2+} is decreased (3 mM). The activation is rapidly reversed in the presence of the substrate, and more slowly after removal of ATP·Mg. The PTPA-activated PCS_L phosphatase represents a major PTPase activity in the cytosol of *X. laevis* oocytes (at least 50% of the measurable PTPase with RCM lysozyme phosphorylated on tyrosyl residues). The PTPA activation is specific for the PTPase activity of the PCS_L and PCS_{H2} phosphatases, without affecting their phosphoseryl/threonyl phosphatase activity. However, effectors of the phosphorylase phosphatase activity, such as polycations and okadaic acid, also influence the PTPase activity. Phosphorylase *a* inhibits the activated PTPase activity (I_{50} = 5 μM). The PTPase activity of the other oligomeric PCS phosphatases (PCS_{H1} and PCS_M) is not influenced, suggesting an inhibitory role for some of their subunits. This activation is compared with the recently described PTPase stimulation of the PCS phosphatases by ATP/PP_i [Goris, J., Pallen, C. J., Parker, P. J., Hermann, J., Waterfield, M. D., & Merlevede, W. (1988) *Biochem. J. 256*, 1029-1034] and by tubulin [Jessus, C., Goris, J., Cayla, X., Hermann, J., Hendrix, P., Ozon, R., & Merlevede, W. (1989) *Eur. J. Biochem. 180*, 15-22].

A role for the phosphorylation of proteins on tyrosine residues in the control of cell proliferation is implied by the observation that protein tyrosine kinase activities are intrinsic to a number of growth factor receptors (Ushiro & Cohen, 1980; Kasuga et al., 1982; Petruzzelli et al., 1982; Ek et al., 1982; Jacobs et al., 1983) and that they are also subverted in the generation of a number of oncogenes (Downward et al., 1984; Sherr et al., 1985; Stern et al., 1986). The importance of these activities in signal transfer is evident from studies demonstrating that abolition of kinase activity through site-directed mutagenesis abolishes the signaling/transforming capabilities (Honegger et al., 1987; Chou et al., 1987; Chen et al., 1987). In the model systems that have been used to elucidate the control of cellular functions by phosphorylation, it has become increasingly clear that the role played by protein phosphatases is far from a passive one and that complex regulatory mechanisms exist to affect steady-state protein phosphorylation through alterations in protein phosphatase activity (Cohen, 1982; Merlevede et al., 1984; Ballou & Fischer, 1986; Goris et al., 1989a).

Evidence was presented by several laboratories for the existence of specific PTPases,[1] distinct from the mammalian acid

† This work was supported by the Onderzoeksfonds K.U. Leuven and the Fonds voor Geneeskundig Wetenschappelijk Onderzoek (Belgium) and by the Institut National de la Santé et de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Institut National de la Recherche Agronomique, and the Ministère de la Recherche et de l'Enseignement Supérieur (France).
* To whom correspondence should be addressed.
‡ Université Pierre et Marie Curie.
§ Katholieke Universiteit te Leuven.

[1] Abbreviations: PTPA, phosphotyrosyl phosphatase activator protein; PTPase, phosphotyrosyl phosphatase; *p*NPPase, *p*-nitrophenylphosphate phosphatase; PCS_{H1}, PCS_{H2}, PCS_M, PCS_L, and PCS_C phosphatases, polycation-stimulated, high (H), medium (M), and low (L) molecular weight phosphatases and the catalytic subunit of the same species; AMD phosphatase; ATP·Mg-dependent phosphatase; MAP_2, microtubule-associated protein; RCM lysozyme, reduced carboxamidomethylated and maleylated lysozyme; MLC, myosin light chains; kDa, kilodalton(s); SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; Bis-Tris, 2,2-bis(hydroxymethyl)-2,2',2''-nitrilotriethanol; TLCK, L-1-chloro-3-(4-tosylamino)-7-amino-2-heptanone hydrochloride; EGTA, ethylene glycol bis(β-aminoethyl ether)-*N,N,N',N'*-tetraacetic acid; PMSF, phenylmethanesulfonyl fluoride.

0006-2960/90/0429-658$02.50/0