Volume 312, number 2,3, 110–114 FEBS 11715 November 1992

Minireview

Families of metalloendopeptidases and their relationships

Weiping Jiang[a] and Judith S. Bond[b]

[a]*Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA* and [b]*Department of Biological Chemistry, Pennsylvania State University, College of Medicine, Hershey, PA 17033, USA*

Received 2 September 1992; revised version received 21 September 1992

Crystal structures available for four metalloendopeptidases have revealed zinc ligands for these enzymes. New sequence information has made it possible to compare the primary structures of the zinc-binding site in metalloendopeptidases. A scheme based on the zinc-binding site is proposed to classify metalloendopeptidases into five distinct families: thermolysin, astacin, serratia, matrixin, and snake venom metalloproteinases. Two histidines and one glutamate are zinc-ligands in the thermolysin family. Three histidines and one tyrosine are zinc ligands in the other four families, which are further distinguished by the identity of the residue following the third histidine and by the environment surrounding the tyrosine.

Metalloendopeptidase; Zinc ligand; Astacin family; Bacterial neutral protease; Collagenase; Snake venom metalloproteinase

1. INTRODUCTION

Most metalloendopeptidases contain zinc as an essential metal ion for their catalytic activity. The coordination of the zinc atom in the active site has been resolved in four metalloproteases so far, including thermolysin from *Bacillus thermoproteolytics*, a neutral protease from *B. cereus*, an elastase from *Pseudomonas aeruginosa*, and astacin from the crayfish, *Astacus fluviatilis* [1–4]. The three bacterial enzymes are closely related and contain three amino acid residues and one water molecule which bind zinc. The first two ligands are two histidine residues located within the HEXXH motif, in which the glutamic acid acts as a catalytic base. The third ligand is a remote glutamic acid, which is located 20 amino acids toward the C-terminus from the second histidine residue. Astacin contains three histidine residues, a remote tyrosine residue, and one water molecule which bind zinc. The three histidine residues are located within the signature sequence for the astacin family, **HE**(L,I)X**H**XX**G**F**X**H**E**(Q,H)X**R**XD**R**DX(Y,H)(V,I)-X(I,V) [5]. By analogy to thermolysin, the first glutamic acid is probably a catalytic base. The remote tyrosine residue is 50 amino acids toward the C-terminus from the third histidine residue in the signature.

The HEXXH motif has been used to identify zinc-binding sites in metalloendopeptidases when new amino acid sequences are obtained. Although this motif is well conserved in these enzymes, the third zinc ligand (the remote glutamic acid) is present only in a few closely related bacterial metalloendopeptidases [6]. However, high homologies in the sequences following the HEXXH motif have been noticed in several groups of metalloproteases, including collagenases, astacin family proteins, and snake venom metalloproteinases [7–9]. The information on the crystal structure of astacin has greatly facilitated identification of additional zinc ligands in these enzymes [4]. A comparison is presented herein for the sequences of metalloendopeptidases surrounding the zinc-binding site.

Correspondence address: W. Jiang, Department of Biological Chemistry, Pennsylvania State University, College of Medicine, Hershey, PA 17033, USA. Fax: (1) (717) 531-7072.

2. FAMILIES

Thermolysin and several related bacterial metalloproteases [6,10] are well conserved in the two regions containing the zinc-binding site (Fig. 1). The first region is the HEXXH motif, and the second region is **GAXNEAFSD** (bold letters herein represent strictly conserved residues in consensus sequences), which contains the third zinc ligand (E25). The numbering scheme used in this and following figures is arbitrary, and #1 is placed on the first histidine residue in the HEXXH motif. In addition to active site residues (two histidines and two glutamic acids), seven amino acids in the 29 amino acid sequence are strictly conserved, which may be important for either structure or function. For example, the aspartic acid (D29) forms a salt bridge in thermolysin [11].

The astacin family consists of several proteins from diverse sources, including *Drosophila*, *Xenopus*, crayfish, sea urchin, mouse, rat, and human [8,12–18]. There are three regions in the proteins of this family with high

 Published by Elsevier Science Publishers B.V.

No.	1					5																25			
Thermolysin	**H**	**E**	L	T	**H**	A	V	T	**D**	Y	T	A	G	L	I	Y	Q	N	E	S	G	A	I	N	E	A	I	S	D
B.st.NP	**H**	**E**	L	T	**H**	A	V	T	**D**	Y	T	A	G	L	V	Y	Q	N	E	S	G	A	I	N	E	A	M	S	D
B.c.NP	**H**	**E**	L	T	**H**	A	V	T	E	N	S	S	N	L	I	Y	Q	N	E	S	G	A	L	N	E	A	I	S	D
B.su.NP	**H**	**E**	M	T	**H**	G	V	T	Q	E	T	A	N	L	I	Y	E	N	Q	P	G	A	L	N	E	S	F	S	D
B.a.NP	**H**	**E**	M	T	**H**	G	V	T	Q	E	T	A	N	L	N	Y	E	N	Q	P	G	A	L	N	E	S	F	S	D
P.a.Elastase	**H**	**E**	V	S	**H**	G	F	T	E	Q	N	S	G	L	I	Y	R	G	Q	S	G	G	M	N	E	A	F	S	D
L.p.Protease	**H**	**E**	V	S	**H**	G	F	T	E	Q	H	S	G	L	E	Y	F	G	Q	S	G	G	M	N	E	S	F	S	D
Consensus	**H**	**E**	X	T	**H**	G	V	T	X	X	T	A	G	L	I	Y	X	N	Q	S	G	A	X	N	E	A	F	S	D

Fig. 1. Comparison of the zinc-binding site in members of the thermolysin family. Sequences are aligned starting from the HEXXH motif. The zinc-binding residues are numbered. Absolutely conserved residues are in bold. A consensus sequence is generated, which includes absolutely conserved residues and common residues present in at least half of the sequences. X represents a variable amino acid. B., *Bacillus*; st., *stearothermophilus*; c., *cereus*; su., *subtilis*; a., *amyloliquefaciens*; NP, neutral protease; P.a., *Pseudomonas aeruginosa*; and L.p., *Legionella pneumophila*.

homology in a 60 amino acid sequence (Fig. 2). The first region, **HEIGHAIGFXHE**, contains three histidines for zinc binding, and one glutamic acid (E2) for catalysis. In astacin, Gly-8 is important for the secondary structure and the second glutamic acid (E12) forms a salt bridge with the N-terminus of the mature enzyme [4]. The second region, **RXDRD**, is extremely hydrophilic. The third region, **YDYXSIMHY**, contains the fourth amino acid zinc ligand, Y61. Great variability is present in the more than 30 amino acids between the second and third conserved regions.

Several bacterial extracellular metalloproteases, including a protease from *Serratia* sp., and protease B and C from *Erwinia chrysanthemi* [19–22], show a closer relationship to each other than to members of the thermolysin family (Fig. 3). Besides the HEXXH motif, there are two glutamic acids (E19 and E32) conserved in these enzymes. It seems unlikely that either of these two glutamic acids is involved in zinc binding because the regions containing these two glutamic acids show no homology with that containing E25 in thermolysin. There is, however, a region similar to that in thermolysin that has been recognized previously [19]. This region is located 120 amino acids away from the second histidine in the motif (Fig. 3). It was suggested that the glutamic acid (or aspartic acid) in this region functions as the third zinc ligand because of sequence similarity to that in thermolysin.

Members of the serratia family show a strong similarity to members of the astacin family. The three histidine residues in the consensus sequence have exactly the same spacing as those in astacin. Furthermore, the glycine residue (G8), that is important for astacin secondary structure, is also conserved. Therefore, by homology with astacin, it is more likely that the third histidine (rather than the glutamic or aspartic acid as discussed above) is the third zinc ligand in members of the serratia family. Instead of the glutamic acid (E12) in the astacin family, a proline (P12) is present and distinguishes this family from the astacin family. In addition, there is a region, **QFSIMSY**, which is very similar to the region including the fourth zinc ligand (Y61) in astacin. Tyr-42 in this region is located 31 amino acids from His-11 in the serratia family. Great variability in the longer spacer region (50 amino acids) between H11 and Y61 in the astacin family may explain the shorter spacer region (31 amino acids) present in the serratia family. Therefore, Y42 in the serratia family is likely to bind zinc.

The matrixin family consists of collagenases, gelatinases, and stromelysins [7]. They all contain the **HELGHSLGLXHS** motif, in which the first two histidine residues are proposed to bind zinc from comparisons with thermolysin (Fig. 4). By analogy to astacin, the third histidine residue may be the third ligand for zinc. This prediction is further supported by the conservation of Gly-8 in the motif. A serine residue at position 12 distinguishes members of this family from members of the astacin family, which contain a glutamic acid residue at this position. There is no region containing either the glutamic acid similar to that in thermolysin or the tyrosine similar to that in astacin. However, there is a tyrosine at position 46, which is conserved in all the sequences; this tyrosine could be involved in binding zinc.

The amino acid sequences of several snake venom metalloproteinases, including hemorrhagic toxin (Ht) and non-hemorrhagic proteins, have been determined. The primary structures indicated that the venom enzymes belong to a distinct metalloproteinase family, all of which contain the HEXXH motif, but which have no significant sequence similarity with any other known metalloproteinases except for this region [9]. However, as can be seen from Fig. 5, members of this family show a significant homology to the astacin family. They contain a region of **HELGHNLGMEHD**, which is similar to the region of **HEIGHAIGFXHE** in the astacin family. Therefore, the third histidine is probably a zinc ligand. An aspartic acid at position 12 distinguishes this family from the astacin family. The tyrosine residue at

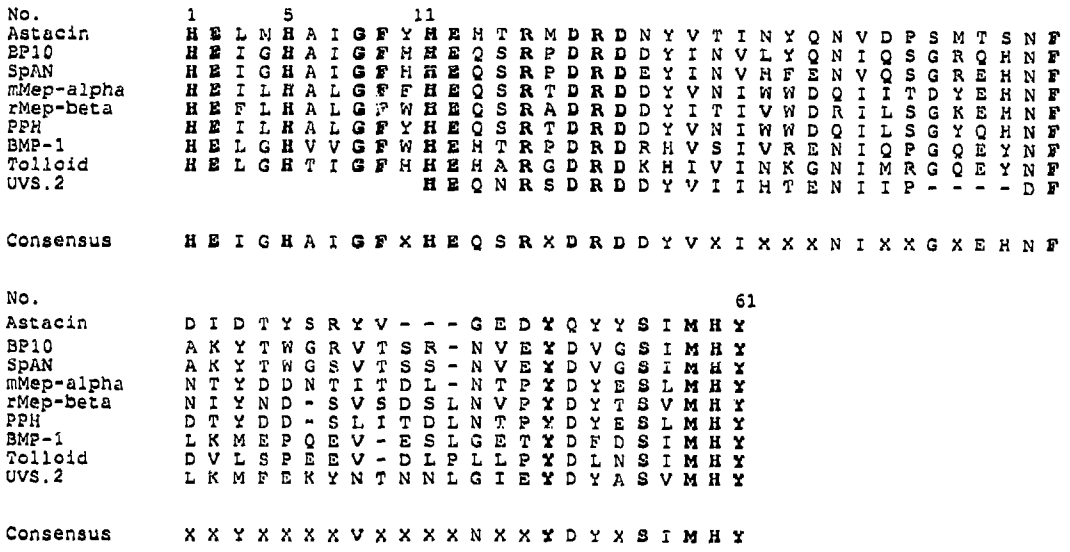

Fig. 2. Comparison of the zinc-binding site in members of the astacin family. Gaps are created for optimal alignment. mMEP, mouse meprin; rMEP, rat meprin; PPH, *N*-benzoyl-L-tyrosyl-*p*-amino-benzoic acid hydrolase; BMP-1, bone morphogenetic protein-1. BP 10 and SpAN are sea urchin proteins. Tolloid and UVS.2 are proteins from *Drosophila* and *Xenopus*, respectively.

position of 46, located within **FSDCSKXYY**, is probably another zinc ligand, as compared to that in the astacin family.

3. RELATIONSHIPS BETWEEN FAMILIES

From the comparisons presented above, it seems apparent that each family has its own identity and can be related to one another based on the sequences surrounding the zinc-binding site. A scheme to classify metalloendopeptidases is proposed in Fig. 6. On the basis of the first two zinc ligands, metalloendopeptidases can be divided into two major categories, one containing the HEXXH motif and the other containing the HXXEH. The HXXEH endopeptidases include bacterial protease III and the human and *Drosophila* insulin-degrading enzymes [26]. Based on the third zinc ligand, the HEXXH endopeptidases can be divided into

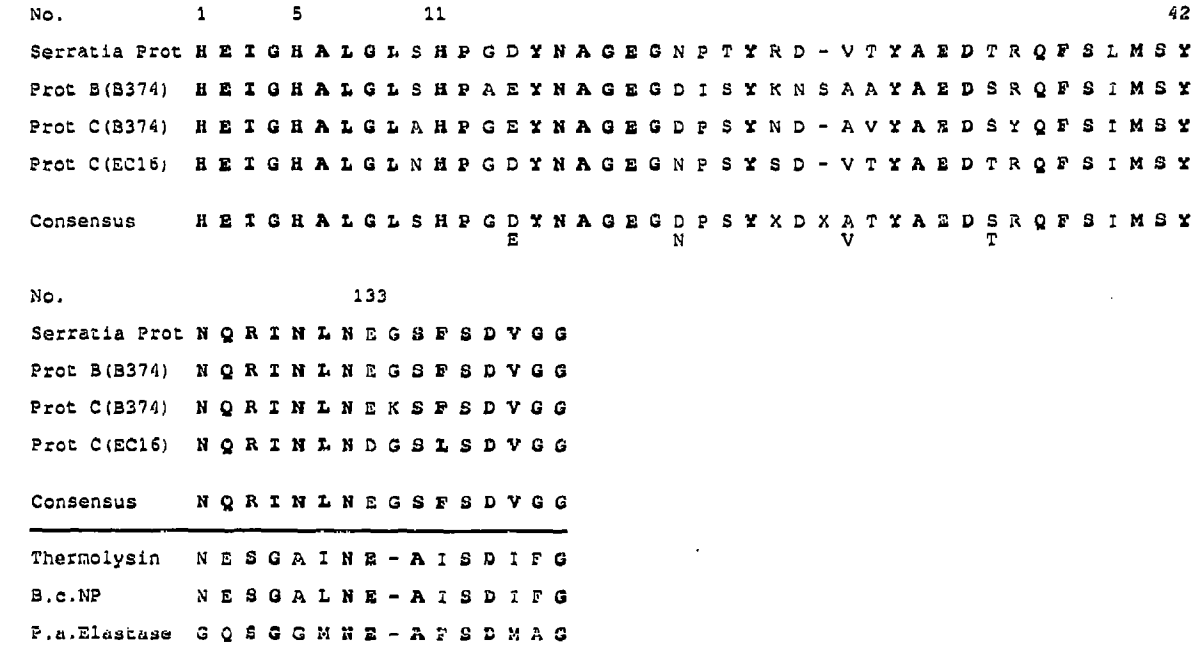

Fig. 3. Comparison of the potential zinc-binding site in members of the serratia family. In addition, the bottom three lines show homologous regions in thermolysin, B.c. NP, and P.a. elastase (their crystal structures are known) in which the conserved residues are highlighted. Prot, protease; other abbreviations as in Fig. 1.

Fig. 4. Comparison of the potential zinc-binding site in members of the matrixin family. For names and references of the matrix metalloproteinases (MMP), see [7]. Sources are from human (a), rabbit (b), rat (c), and mouse (d). In addition, MMP-3d encodes for the mouse transin [23]. MME, macrophage metalloelastase [24]; Consen, consensus. HE-6 is a cDNA clone coding for a hatching enzyme from sea urchin [25].

two groups which contain either a glutamic acid or a histidine. The thermolysin family, as described above, is a typical example of those enzymes containing the gluatmic acid as the third zinc ligand. Neprilysin (NEP, or endopeptidase 24.11), a mammalian neutral peptidase, is another example of this group. The glutamic acid of neprilysin, located 59 amino acids away toward the C-terminus, has been shown by site-directed mutagenesis to be involved in binding of zinc [27,28]. Thimet oligopeptidase (endopeptidase 24.15), another mammalian neutral peptidase [29], contains two glutamic acids in regions of GTHVERDFV and RTGGEAPED, that show weak homology to both thermolysin (GAINEAISD) and neprilysin (NTLGENIAD). The HEXXH metalloendopeptidases of the second group contain a histidine (H11) as the third zinc ligand, which is located in the **HEXXHXXGXXH** motif. These enzymes contain an additional zinc ligand, tyrosine. On the basis of the residue present immediately after the third histidine in the motif and, to a lesser extent, homology present in the sequences containing the tyrosine, these enzymes can be further divided into several families. The astacin family contains the glutamic acid (E12) right after the histidine (H11) and the sequence, **SIMHY**, which includes the tyrosine. The families of serratia, matrixin, and snake venom metalloproteinases contain the proline (P12) and the sequence, **SIMSY**, the serine (S12) and the sequence, **IQSLY**, and the aspartic acid (D12) and the sequence, **SKXYY**, respectively.

The classification scheme presented here is further supported by the overall sequence homologies between and within the proposed families. There are significant homologies present in members from the same family described in this paper. In the thermolysin family, *B. cereus* neutral protease and *P. aeruginosa* elastase, for example, are 73% and 49% identical to thermolysin, respectively [6]. In the astacin family, all members are

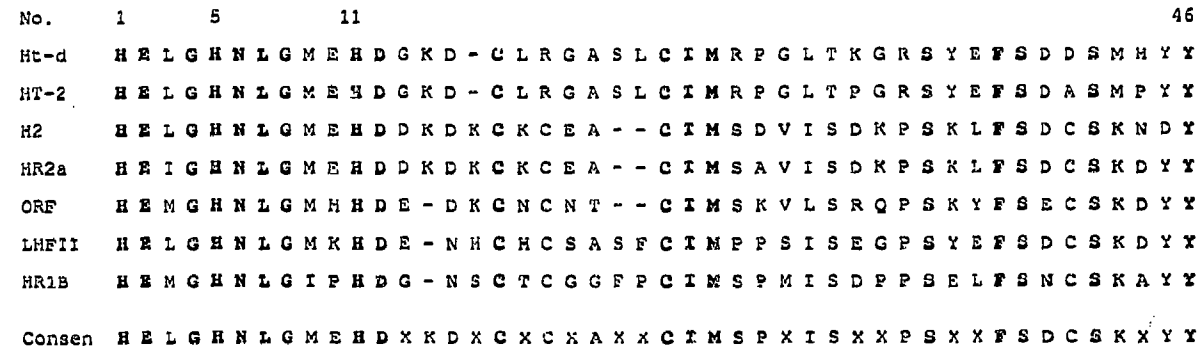

Fig. 5. Comparison of the potential zinc-binding site in members of the snake venom family. For more information, see [9].

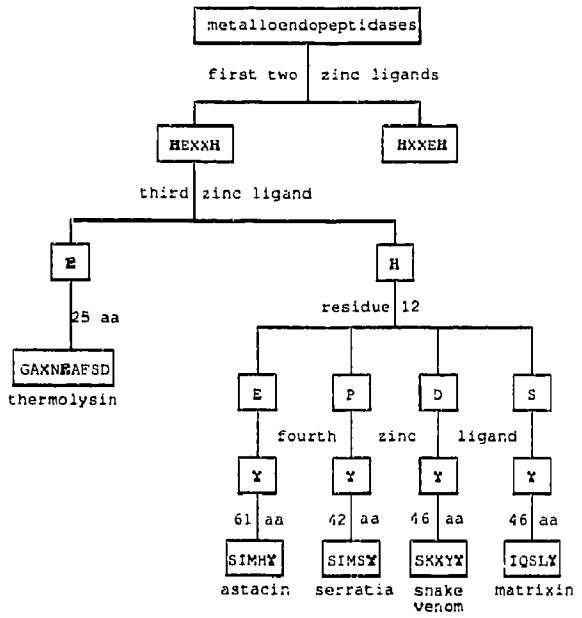

Fig. 6. Families of metalloendopeptidases based on the zinc-binding site. The residues in bold are zinc-ligands. The numbers of amino acids (aa) indicated below either E or Y represent the distances between these residues and the first H residue in the HEXXH motif.

approximately 30% identical to astacin. The identities in the serratia family are 80% between proteases B and C, and 57% between protease B and the serratia protease [20,21]. The homology of MME-d with other members in the matrixin family ranges from 33% (MMP-7) to 48% (MMP-3a and MMP-3d) [24]. Snake venom metalloproteinases all share approximate 50% identity [30]. By contrast, there is no significant homology present in the overall sequences for the members from different families.

This classification scheme will be useful for identifying zinc ligands in metalloendopeptidases and the structural relationships of these enzymes. Site-directed mutagenesis and X-ray analysis of the crystal structures for some of these enzymes will provide further experimental tests for this scheme. In addition, this classification scheme may be expanded to include metalloexopeptidases such as carboxypeptidases and aminopeptidases [6], and angiotensin-converting enzyme, a carboxyl terminal dipeptidase [31]. It will be of interest to see whether the two fundamental modes of binding zinc are conserved in all metallopeptidases.

Acknowledgements: We thank Dr. Russell Wolz for his critical reading of this manuscript. This work was supported by National Institutes of Health Grant DK 19691.

REFERENCES

[1] Holmes, M. A. and Matthews, B.W. (1982) J. Mol. Biol. 160, 623–639.
[2] Stark, W., Pauptit, R.A., Wilson, K.S. and Jansonius, J.N. (1992) Eur. J. Biochem. 207, 781–791.
[3] Thayer, M.M., Flaherty, K.M. and McKay, D.B. (1991) J. Biol. Chem. 266, 2864–2871.
[4] Bode, W., Gomis-Rüth, F.X., Huber, R., Zwilling, R. and Stöcker, W. (1992) Nature 358, 164–167.
[5] Jiang, W., Gorbea, C.M., Flannery, A.V., Beynon, R.J., Grant, G.A. and Bond, J.S. (1992) J. Biol. Chem. 267, 9185–9193.
[6] Vallee, B.L. and Auld, D.S. (1990) Biochemistry 29, 5647–5659.
[7] Woessner, J.F. (1991) FASEB J. 5, 2145–2154.
[8] Dumermuth, E., Sterchi, E.E., Jiang, W., Wolz, R.L., Bond, J.S., Flannery, A.V. and Beynon, R.J. (1991) J. Biol. Chem. 266, 21381–21385.
[9] Hite, L.A., Fox, J.W. and Bjarnason, J.B. (1992) Biol. Chem. Hoppe-Seyler 373, 381–385.
[10] Black, W.J., Quinn, F.D. and Tompkins, L.S. (1990) J. Bacteriol. 172, 2608–2613.
[11] Colman, P.M., Jansonius, J.N. and Matthews, B.W. (1972) J. Mol. Biol. 70, 701–724.
[12] Shimell, M.J., Ferguson, E.L, Childs, S.R. and O'Connor, M.B. (1991) Cell 67, 469–481.
[13] Sato, S.M. and Sargent, T.D. (1989) Dev. Biol. 137, 135–141.
[14] Titani, K., Torff, H.-J., Hormel, S., Kumar, S., Walsh, K.A., Rödl, J., Neurath, H. and Zwilling, R. (1987) Biochemistry 26, 222–226.
[15] Lepage, T., Ghiglione, C. and Gache, C. (1992) Development 114, 147–164.
[16] Reynolds, S.D., Angerer, L.M., Palis, J., Nasir, A. and Angerer, R.C. (1992) Development 114, 769–786.
[17] Johnson, G.D. and Hersh, L.B. (1992) J. Biol. Chem. 267, 13505–13512.
[18] Wozney, J.M., Rosen, V., Celeste, A.J., Mitsock, L.M., Whitters, M.J., Kriz, R.W., Hewick, R.M. and Wang, E.A. (1988) Science 242, 1528–1534.
[19] Nakahama, K., Yoshimura, K., Marumoto, R., Kikuchi, M., Lee, I.S., Hase, T. and Matsubara, H. (1986) Nucleic Acids Res. 14, 5843–5855.
[20] Delepelaire, P. and Wandersman, C. (1989) J. Biol. Chem. 264, 9083–9089.
[21] Delepelaire, P. and Wandersman, C. (1990) J. Biol. Chem. 265, 17118–17125.
[22] Dahler, G.S., Barras, F. and Keen, N.T. (1990) J. Bacteriol. 172, 5803–5815.
[23] Ostrowski, L.E., Finch, J., Krieg, P., Matresian, L., Patskan, G., O'Connell, J.F., Phillips, J., Slaga, T.J., Breathnach, R. and Bowden, G.T. (1988) Mol. Carcinogenesis 1, 13–19.
[24] Shapiro, S.D., Griffin, G.L., Gilbert, D.J., Jenkins, N.A., Copeland, N.G., Welgus, H.G., Senior, R.M. and Ley, T.J. (1992) J. Biol. Chem. 267, 4664–4671.
[25] Lepage, T. and Gache, C. (1990) EMBO J. 9, 3003–3012.
[26] Becker, A.B. and Roth, R.A. (1992) Proc. Natl. Acad. Sci. USA 89, 3835–3839.
[27] Devault, A., Lazure, C., Nault, C., LeMoual, H., Seidah, N.G., Chrétien, M., Kahn, P., Powell, J., Mallet, J., Beaumont, A., Roques, B.P., Crine, P. and Boileau, G. (1987) EMBO J. 6, 1317–1322.
[28] LeMoual, H., Devault, A., Roques, B.P., Crine, P. and Boileau, G. (1991) J. Biol. Chem. 266, 15670–15674.
[29] Pierotti, A., Dong, K.-W., Glucksman, M.J., Orlowski, M. and Roberts, J.L. (1990) Biochemistry 29, 10323–10329.
[30] Takeya, H., Onikura, A., Nikai, T., Sugihara, H. and Iwanaga, S. (1990) J. Biochem. 108, 711–719.
[31] Ehlers, M.R.W. and Riordan, J.F. (1991) Biochemistry 30, 7118–7126.