Volume 242, number 2, 211-214 FEB 06654 January 1989

Discussion Letter

A unique signature identifies a family of zinc-dependent metallopeptidases

C. Victor Jongeneel, Jacques Bouvier* and Amos Bairoch+

*Ludwig Institute for Cancer Research, Lausanne Branch, *Department of Biochemistry, University of Lausanne* and *+Department of Medical Biochemistry, University of Geneva, Switzerland*

Received 27 October 1988

The primary sequence motif HExxH has been found in many zinc-dependent endopeptidases. We show that a larger signature comprising this sequence is common to most of the known zinc-dependent endopeptidases, and that the presence of the signature can be indicative of membership in the family. A search of the protein sequence databases for entries containing the signature retrieved several unexpected potential zinc endopeptidases.

Metalloproteinase; Zinc; Amino acid sequence; Information system

The exponentially growing supply of nucleic acid and derived protein sequences is beginning to make it possible to predict secondary structure, and sometimes biological function, on the basis of recurrent patterns of primary structure [1]. This commentary concerns one family of enzymes for which the availability of sequence information has grown rapidly over the last few years, the zinc-dependent metallopeptidases. We argue that most of the known zinc-dependent metallopeptidases (with the notable exception of the carboxypeptidases) share a common pattern of primary structure, and that the presence of this pattern can have predictive value.

A number of investigators [2–7] have noted that zinc-dependent peptidases contain the sequence HExxH, which constitutes the core of one of the two zinc-binding sites in thermolysin. It is known from the tertiary structure of thermolysin that the two histidines participate in the co-ordination of the zinc atom, while the conserved glutamic acid is part of the active site [8]. All three of the conserved residues have been shown to be essential for the activity of neutral endopeptidase (EC 3.4.24.11) [9]. A third histidine (His-637 in NEP, or His-231 in thermolysin), which was thought to be involved in stabilizing the transition state through hydrogen bonding, is probably not essential [9]. Therefore, the HExxH motif contains all of the residues that have been positively identified as part of the active site, and could define a unique signature for zinc metallopeptidases.

We screened the Swiss-Prot database (version 8, July 1988) for the occurrence of the HExxH motif in known protein sequences: it was found 92 times in 83 different sequences, of which only 15 were known zinc-dependent proteases. Obviously, HExxH does not define a primary structure pattern unique to zinc metallopeptidases. In order to refine our search, we proceeded to align the sequences of all of the known zinc metallopeptidases that contain the motif (fig.1). The alignment showed clearly that the similarity between the

Correspondence address: C.V. Jongeneel, Ludwig Institute for Cancer Research, ch. des Boveresses, CH-1066 Epalinges, Switzerland

Published by Elsevier Science Publishers B.V. (Biomedical Division)

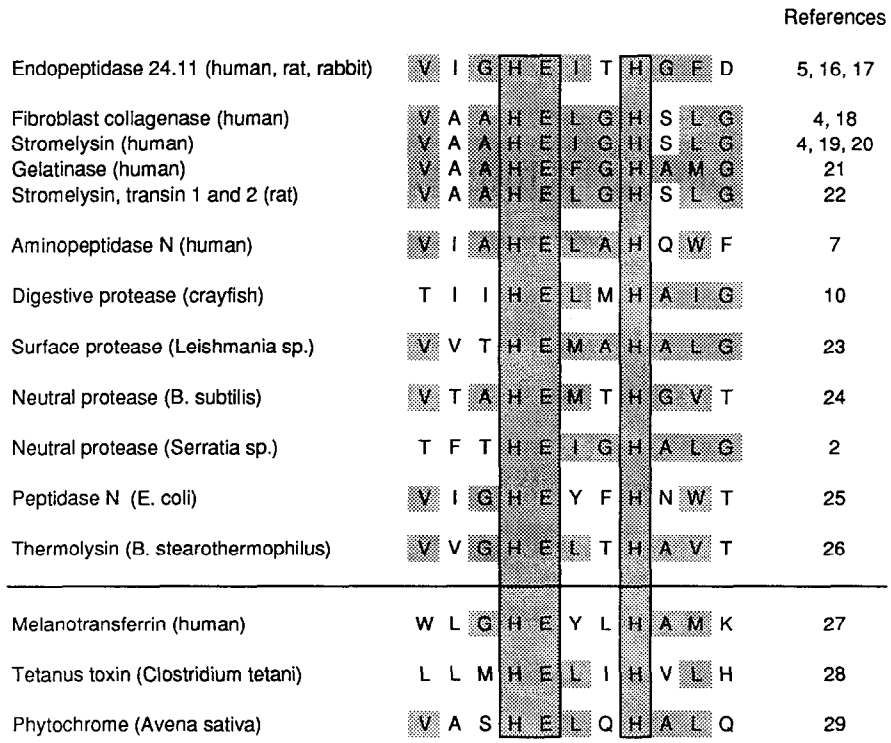

Fig.1. Comparison of the primary structures of known and putative zinc-dependent peptidases. Proteins with known proteolytic activities are shown above the line, and potential members of the family below the line. Amino acids conserved between family members are shaded, and the three residues known to be required for hydrolytic activity are boxed. The alignment from which we generated the 'zinc signature' used to scan the databases included the following sequences: endopeptidase 24.11, fibroblast collagenase, *Leishmania* surface protease, the neutral proteases of *B. subtilis* and *Serratia*, and thermolysin.

metallopeptidases extended beyond the HExxH motif. In particular, none of the enzymes contained any charged amino acids (other than the two His and the Glu) within a 10 amino acid stretch extending from 3 amino acids before the first His to 2 amino acids after the second His: in addition, all of them contained a hydrophobic residue 2 amino acids after the second His. Based on these observations, we redefined the putative zinc signature as (uncharged)-(uncharged)-H-E-(uncharged)-(uncharged) - H - (uncharged) - (hydrophobic), and screened the Swiss-Prot databank again for proteins containing this pattern of primary sequence. In addition to the proteins used in the original alignment (see the legend to fig.1), the search retrieved five new entries: (i) a protease from the digestive juice of the crayfish, *Astacus fluviatilis*; (ii) peptidase N of *E. coli*; (iii) phytochromes 3 and 4 from *Avena sativa*; (iv) melanotransferrin, a surface protein belonging to the transferrin family; (v) the tetanus toxin, from *Clostridium tetani* (see fig.1 for the primary sequences).

The sequence of the crayfish protease was originally described as having 'no homologous relationship to any of the known protein sequences' [10]. A recent report [6], of which we became aware after the initial database search had been done, has shown that this protease is indeed zinc-dependent, confirming the predictive value of the sequence pattern. The authors also noted the presence in the *Astacus* protease of the HExxH motif.

E. coli peptidase N has been described as an aminopeptidase [11], but data on its specificity and ionic requirements are still scant. The presence in this enzyme of the zinc endopeptidase signature makes it a very likely candidate for membership in the family. It should be noted that the aminopep-

tidase N of higher eukaryotes, which was thought to share nothing but a name with the *E. coli* enzyme, has recently been shown to have a surprisingly high degree of homology with its prokaryotic namesake [7]. Aminopeptidase N has long been known to be a zinc metallopeptidase, and does indeed contain the signature [7].

It was more surprising to find the signature in phytochromes. These proteins can undergo a reversible transition between an active (P_r) and an inactive (P_{fr}) form, as a result of exposure to red ($P_{fr} \longrightarrow P_r$) or far red ($P_r \longrightarrow P_{fr}$) light. In their active form, the phytochromes induce the transcription of many genes involved in plant growth. Basing his argument on the effects of pH and chelators on phytochrome action, Hoober [12] recently proposed that the difference between the P_r and P_{fr} forms lies in the co-ordination of a zinc ion. The presence of the metallopeptidase signature certainly re-inforces this notion. Whether binding of zinc induces the appearance of a phytochrome-specific proteolytic activity (e.g. for the inactivation of a repressor or the conversion of a transcription factor precursor to an active form) remains to be investigated. A note of caution should be added, since a phytochrome from *Cucurbita pepo* (cucumber) [13], while being highly homologous to the *Avena* proteins, is missing the second histidine of the signature. Though it is possible that a histidine from a different part of the polypeptide chain could provide the necessary co-ordinating activity, it could also mean that the presence of the signature in phytochromes is spurious.

As to the tetanus toxin and melanotransferrin, there is (to our knowledge) no indication that they may be zinc-binding proteins or have an endopeptidase activity. Our observation may warrant a search for such properties.

Although the pattern we sought for in the protein databases did retrieve a number of interesting proteins, we cannot say whether it really defines the zinc signature accurately. The alignment shown in fig.1 suggests that the use of a pattern taking into account the conservation of specific residues at positions other than those of His and Glu could increase the selectivity of database searches. For the moment, the constraining factors are the limited availability of protein sequence and tertiary structure information [1], and our poor understanding of the structural requirements for peptide hydrolysis by zinc metallopeptidases.

Zinc-dependent endopeptidases form a very heterogeneous family, with widely differing specificities and sensitivities to inhibitors. The algorithms used to detect evolutionary relationships between proteins fail to detect any homology between the proteases listed in fig.1 (except within the collagenase family), because the region of homology is too short to produce statistically significant sets of overlapping alignments; hence the initial oversight in assigning the crayfish protease or *E. coli* peptidase N to the family. Only more sophisticated methods, such as clustering analysis [14], reveal patches of similar secondary structures. Yet, a simple signature seems to define the family in a unique fashion. Whether this reflects a distant evolutionary relationship remains a subject for speculation. Carboxypeptidase A, which uses a similar reaction mechanism and contains zinc at its active center, is not evolutionarily related to thermolysin, and does not contain the signature. Neither, for that matter, do any of the many other zinc-binding proteins that have been sequenced.

Acknowledgements: We thank Drs Clément Bordier, John Kenny and Michelle Letarte for stimulating discussions, and Dr Ove Norén for communicating the sequence of aminopeptidase N before publication.

REFERENCES

[1] Rooman, M.J. and Woda, S.J. (1988) Nature 335, 45–49.
[2] Nakahama, K., Yoshimura, K., Marumoto, R., Kikuchi, M., Lee, I.S., Hase, T. and Matsubara, H. (1986) Nucleic Acids Res. 14, 5843–5855.
[3] McKerrow, J.H. (1987) J. Biol. Chem. 262, 5943.
[4] Whitham, S.E., Murphy, G., Angel, P., Rahmsdorf, H.J., Smith, B.J., Lyons, A., Harris, T.J.R., Reynolds, J.J., Herrlich, P. and Docherty, A.J.P. (1986) Biochem. J. 240, 913–916.
[5] Devault, A., Lazure, C., Nault, C., Le Moual, H., Seidah, N.G., Chrétien, M., Kahn, P., Powell, J., Mallet, J., Beaumont, A., Roques, B.P., Crine, P. and Boileau, P. (1987) EMBO J. 6, 1317–1322.
[6] Stöcker, W., Wolz, R.L., Zwilling, R., Strydom, D.J. and Auld, D. (1988) Biochemistry 27, 5026–5032.
[7] Olsen, J., Cowell, G.M., Königshöfer, E., Danielsen, E.M., Möller, J., Laustsen, L., Hansen, O.C., Welinder, K.G., Engberg, J., Hunziker, W., Spiess, M., Sjöström, H. and Norén, O. (1988) FEBS Lett. 238, 307–314.

[8] Devault, A., Sales, V., Nault, C., Beaumont, A., Roques, B., Crine, P. and Boileau, G. (1988) FEBS Lett. 231, 54–58.
[9] Matthews, B.W., Colman, P.M., Jansonius, J.N., Titani, K., Walsh, K.A. and Neurath, H. (1972) Nature New Biol. 238, 41–43.
[10] Titani, K., Torff, H.J., Hormel, S., Kumar, S., Walsh, K.A., Rödl, J., Neurath, H. and Zwilling, R. (1987) Biochemistry 26, 222–226.
[11] McCaman, M.T. and Villarejo, M.R. (1982) Arch. Biochem. Biophys. 213, 384–394.
[12] Hoober, J.K. (1988) Carlsberg Res. Commun. 53, 27–41.
[13] Sharrock, R.A., Lissemore, J.L. and Quail, P.H. (1986) Gene 47, 287–295.
[14] Benchetrit, T., Bissery, V., Mornon, J.P., Devault, A., Crine, P. and Roques, B.P. (1988) Biochemistry 27, 592–596.
[15] Kester, W.R. and Matthews, B.W. (1977) J. Biol. Chem. 252, 7704–7710.
[16] Malfroy, B., Kuang, W.J., Seeburg, P.H., Mason, A.J. and Schofield, P.R. (1988) FEBS Lett. 229, 206–210.
[17] Malfroy, B., Schofield, P.R., Kuang, W.J., Seeburg, P.H., Mason, A.J. and Henzel, W.J. (1987) Biochem. Biophys. Res. Commun. 144, 59–66.
[18] Goldberg, G.I., Wilhelm, S.M., Kronberger, A., Bauer, E.A., Grant, G.A. and Eisen, A.Z. (1986) J. Biol. Chem. 261, 6600–6605.
[19] Wilhelm, S.M., Collier, I.E., Kronberger, A., Eisen, A.Z., Marmer, B.L., Grant, G.A., Bauer, E.A. and Goldberg, G.I. (1987) Proc. Natl. Acad. Sci. USA 84, 6725–6729.
[20] Saus, J., Quinones, S., Otani, Y., Nagase, H., Harris, E.D. and Kurkinen, M. (1988) J. Biol. Chem. 263, 6742–6745.
[21] Collier, I.E., Wilhelm, S.M., Eisen, A.Z., Marmer, B.L., Grant, G.A., Seltzer, J.L., Kronberger, A., He, C., Bauer, E.A. and Goldberg, G.I. (1988) J. Biol. Chem. 263, 6579–6587.
[22] Breathnach, R., Matrisian, L.M., Gesnel, M.C., Staub, A. and Leroy, P. (1987) Nucleic Acids Res. 15, 1139–1151.
[23] Button, L.L. and McMaster, W.R. (1988) J. Exp. Med., 724–729.
[24] Yang, M.Y., Ferrari, E. and Henner, D.J. (1984) J. Bacteriol. 160, 15–21.
[25] Foglino, M., Gharbi, S. and Lazdunski, A. (1986) Gene 49, 303–309.
[26] Titani, K., Hermodson, M.A., Ericsson, L.H., Walsh, K.A. and Neurath, H. (1972) Nature New Biol. 238, 35–37.
[27] Rose, T.M., Plowman, G.D., Teplow, D.B., Dreyer, W.J., Hellström, K.E. and Brown, J.P. (1986) Proc. Natl. Acad. Sci. USA 83, 1261–1265.
[28] Eisel, U., Jarausch, W., Goretzki, K., Henschen, A., Engels, J., Weller, U., Hudel, M., Habermann, E. and Niemann, H. (1986) EMBO J. 5, 2495–2502.
[29] Hershey, H.P., Barker, R.F., Idler, K.B., Lissemore, J.L. and Quail, P.H. (1985) Nucleic Acids Res. 13, 8543–8559.