Homology within the *N*-terminal extension of cysteine proteinases

It is now well established that the cysteine proteinases of plants (e.g. papain, actinidin, bromelain) and animals (e.g. cathepsins B, H and L) are homologous (see Takio *et al.*, 1983) and almost certainly evolved from a common ancestral protein. For all but one of the cysteine proteinases sequenced to date the *N*-terminal residues can be exactly aligned with one another at a distance of 25–29 residues from the active site cysteine: the exception is bovine cathepsin H which has two additional *N*-terminal residues (Turk *et al.*, 1985). However there have been a number of reports which indicate that, at least for the cathepsins, cysteine proteinases can exist in larger forms, and that these have functions that differ from those of the well characterized forms. High- M_r proteinases related to the lysosomal enzyme cathepsin B have been found extracellularly (see Gordon *et al.*, 1985; Mort & Recklies, 1986) and intracellularly in insulin secretory granules (Docherty *et al.*, 1984). A high- M_r extracellular form of lysosomal cathepsin L has also been reported (Recklies & Mort, 1985). Some of the high- M_r enzymes display differences in substrate specificity which may relate to a role in processes involving limited proteolysis such as proprotein processing (Docherty *et al.*, 1984; Gordon *et al.*, 1985). The high- M_r forms may also represent precursors of the lysosomal proteinases and so, by analogy with other lysosomal enzymes, their larger size would be due to an extension at the *N*-terminus.

It would be of considerable interest to know the nature of these different forms of cysteine proteinase. As yet the complete purification of a high- M_r enzyme has not been reported, and so details of the primary structure can only be obtained indirectly by DNA cloning and sequencing. cDNA clones for rat cathepsin B have been identified, but do not encode the complete translation product (San Segundo *et al.*, 1985). Nevertheless, the DNA sequence does show that the latter must extend at least 11 amino acids beyond the *N*-terminus of cathepsin B.

In the last year there have been three reports of DNA sequences encoding complete proteins closely related to known cysteine proteinases. Two of the sequences are for developmentally-regulated proteins of the cellular slime mould *Dictyostelium discoideum*, namely CP1 (Williams *et al.*, 1985) and CP2 (Pears *et al.*, 1985), the third is for aleurain, a gibberellic acid-regulated protein of barley aleurone (Rogers *et al.*, 1985). All three have an *N*-terminal extension, and the availability of these sequences has provided the first opportunity to assess the extent to which this region is conserved in proteinases of unrelated species.

The predicted sequences for CP1, CP2 and aleurain are shown in Fig. 1. The homologies between these proteins and other cysteine proteinases have already been described (Pears *et al.*, 1985; Rogers *et al.*, 1985; Williams *et al.*, 1985): 38%, 37% and 39% of papain residues and 41%, 46% and 62% of rat cathepsin H residues appear in CP1, CP2 and aleurain respectively. A typical signal sequence occupies the extreme *N*-terminus of all three proteins. This is not unexpected because CP1 and CP2 are probably lysosomal enzymes, and aleurain

is secreted. The degree of homology between the three signal sequences is low. Within the remainder of the extension region a number of residues are shared by two or all three proteins. The region of greatest homology lies between aleurain residues 70 and 123: 25 of these residues (46.3%) are present in one or both of the slime mould sequences, and 14 residues (25.9%) are shared by all three proteins. Within this region, 35.1% of CP1 residues and 35.2% of CP2 residues appear in the aleurain sequence; the overall homologies with aleurain are 33.5% and 31.6% for CP1 and CP2 respectively. The degree of sequence conservation in this part of the *N*-terminal extension implies that a common function has been retained in cysteine proteinases of the two species.

Since other cysteine proteinases are as closely related to CP1, CP2 and aleurain as the latter three are to one another, it seems likely that at least some will have similar *N*-terminal regions. At present the functions of the extension region remain a matter for speculation. The three sequences in Fig. 1 show that it is more hydrophilic in character than the remainder of the protein. While there is no evidence that the CP1, CP2 or aleurain sequences represent proenzymes, each of their extension sequences has one or more putative processing sites containing adjacent basic amino acid residues. Clearly, if an *N*-terminal extension were part of any of the previously sequenced cysteine proteinases it must be removed at some stage during biosynthesis. The extension may have a role in targetting the protein and in determining its specificity. One conclusion which may be drawn from the properties of high- M_r forms of cathepsin B is that an *N*-terminal extension might impose a restricted specificity on the proteinase (Docherty *et al.*, 1984; Gordon *et al.*, 1985; Mort & Recklies, 1986). Hopefully, sequences of this region will soon be available for cysteine proteinases for which a high- M_r form has been identified. These, together with studies on the properties of CP1, CP2 and aleurain, will allow the relationship between the structure of the *N*-terminal extension and its function to be established.

Michael J. NORTH

Department of Biological Science, University of Stirling, Stirling FK9 4LA, Scotland, U.K.

- Docherty, K., Hutton, J. C. & Steiner, D. F. (1984) *J. Biol. Chem.* **259**, 6041–6044
- Gordon, J. I., Sims, H. F., Edelstein, C., Scanu, A. K. & Strauss, A. W. (1985) *J. Biol. Chem.* **260**, 14824–14831
- Mort, J. S. & Recklies, A. D. (1986) *Biochem. J.* **233**, 57–63
- Pears, C. J., Mahbubani, H. M. & Williams, J. G. (1985) *Nucleic Acids Res.* **13**, 8853–8866
- Recklies, A. D. & Mort, J. S. (1985) *Biochem. Biophys. Res. Commun.* **131**, 402–407
- Rogers, J. C., Dean, D. & Heck, G. R. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6512–6516
- San Segundo, B., Chan, S. J. & Steiner, D. F. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2320–2324
- Takio, K., Towatari, T., Katunuma, N., Teller, D. C. & Titani, K. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3666–3670
- Turk, V., Brzin, J., Lanarcic, B., Locnikar, P., Popovic, T., Ritonja, A., Babnik, J., Bode, W. & Machleidt, W. (1985) in *Intracellular Protein Catabolism* (Khairallah, E. A., Bond, J. S. & Bird, J. W. C., eds.), pp. 91–103, Alan R. Liss, New York
- Williams, J. G., North, M. J. & Mahbubani, H. (1985) *EMBO J.* **4**, 999–1006

1986