Volume 249, number 1, 5-7 FEB 07158 May 1989

Commentary

Comparative analysis of viral cysteine protease structural models

J. Fernando Bazan*+ and Robert J. Fletterick+

*Department of Biophysics, University of California, Berkeley, CA 94720 and +Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA

Received 2 March 1989

Picornavirus; Trypsin fold; Sequence pattern; Protein homology

A recent Discussion Letter in this journal by the Moscow group of Gorbalenya et al. [1] presents similar conclusions to an earlier study by Bazan and Fletterick [2]. Both papers propose an evolutionary relationship between viral Cys proteases (encoded by related positive-stranded RNA viruses of animals and plants; for review see [3]) and the family of cellular, trypsin-like Ser proteases. This association is suggested by alignment of viral protease sequences and secondary structural patterns that reveal common features with the bilobal β-barrel fold of trypsin-like enzymes. The spatial identity of the active site Ser of trypsin-like proteases (the nucleophilic residue of the catalytic triad His-57/Asp-102/Ser-195, chymotrypsin numbering scheme [4]) and a conserved Cys residue found in the C-termini of viral Cys protease domains [5] is a striking measure of the divergence of the viral and cellular protease classes. Both papers predict the location of twelve component β-strands of the trypsin fold from the viral protease sequences, placing two other viral equivalents of the trypsin catalytic triad at the active site. We describe below how our structural model [2] of the viral Cys protease differs in approach, detail and functional deductions from the construct of Gorbalenya et al. [1].

Correspondence address: J.F. Bazan, Dept of Biochemistry and Biophysics, Univ. of California, San Francisco, CA 94143-0448, USA

We sought to generate a robust structural model that was consistent with all the observed viral Cys protease homologs. The rhino- and enterovirus genera of the picornaviruses encode an additional, smaller (average 150 amino acid) 2A Cys protease that is similar in sequence to the larger (~185 amino acid long) 3C proteases in the vicinity of the putative active site Cys, but possesses a different cleavage specificity [3]. Proteolytic enzymes of the plant como-, nepo- and potyviruses appear to belong the the larger 3C viral class [3]. We showed that the merging of the 2A and 3C-like protease alignments strongly suggested that both subgroups of viral Cys proteases were structurally related. Allowing for the placement of gaps between elements of predicted secondary structure, two absolutely conserved residues in the 2A/3C proteases were located in the N-terminal half of the protease domains: residues His-20/40 and Asp-38/85 (picornavirus 2A/3C protease consensus numbering [2]). These amino acids, in concert with the conserved Cys-109/147, were proposed as the viral Cys protease catalytic residues (fig.1) [2,6].

Gorbalenya et al. [1] do not address the smaller 2A Cys proteases and arrive at a markedly different alignment of the N-terminal segments of the 3C-like (animal and plant virus) protease domains. This results in a choice of Glu/Asp at position 74 (our 3C numbering scheme) as a picornaviral 3C catalytic residue that has uncertain equivalents in the plans virus proteases [1].

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

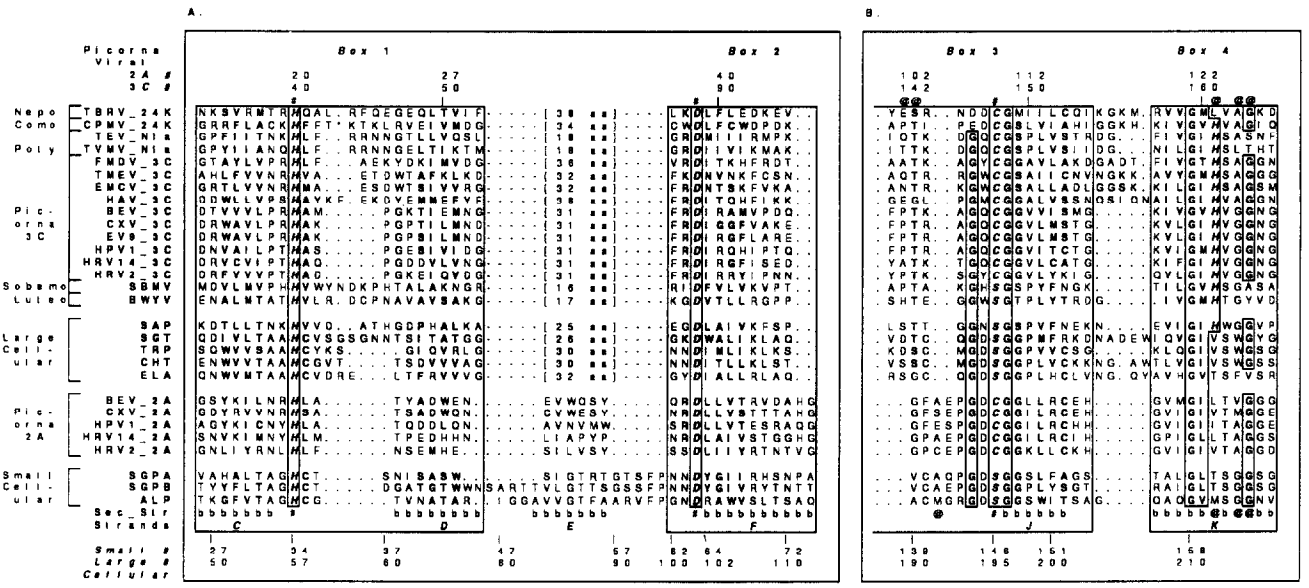

Fig.1. Alignment of viral Cys/Ser proteases with representative trypsin-like Ser proteases from mammals and bacteria (see fig.2 in [2], fig.1 in [7]). The consensus numbering schemes of [2] are followed. The similarity boxes 1–4 encase the 'conserved' catalytic triads of His, Asp and Ser/Cys residues (#) as well as the few other identical and conserved matches (boxed, in bold). β-Strands are marked by letters under the alignments. (A) Box 1,2 alignment [2,7]. Note the length variability of the intervening region (which includes β-strand E). Revisions in this segment for the plant virus proteases is discussed in [7]. (B) Box 3,4 alignment [2,7]. Note residues that are important in forming the substrate-binding pocket in cellular enzymes (@). Abbreviations: HRV2-14, human rhinovirus, strains 2-14; HPV1, human poliovirus; EV9, echovirus, strain 9; CXV, coxsackievirus, strain B3; BEV, bovine enterovirus; HAV, hepatitis-A virus; EMCV, encephalomyocarditis virus; TMEV, Theiler's murine encephalomyelitis virus; FMDV, foot-and-mouth disease virus; TVMV, tobacco-vein-mottling virus; TEV, tobacco-etch virus; CPMV, cowpea mosaic virus; TBRV, tomato black ring virus; SBMV, southern bean mosaic virus; BWYV, beet western yellows virus; SAP, *S. aureus* protease, strain V8; SGT, *S. griseus* trypsin; TRP, trypsin; CHT, chymotrypsin; ELA, elastase; SGPA-B, *S. griseus* proteases A-B; ALP, *L. enzymogenes* α-lytic protease. For references to these sequences see [2,4,7,16,17].

The alignment of 2A and 3C viral proteases is important for several reasons, not the least of which is the unambiguous location of conserved, putative catalytic residues in these viral enzymes. First of all, sequence profiles [7] of the 2A and 3C viral proteases respectively matched the small and large cellular trypsin-like Ser proteases in databank [8] searches. Secondly, the putative 2A/3C active site residues His-20/40, Asp-38/85 and Cys-109/147 were proposed to be spatially and mechanistically equivalent to the small/large trypsin-like Ser protease catalytic triads of His-34/57, Asp-64/102 and Ser-146/195 (fig.1) [2,6]. Finally, the closer relationship of 2A-'small trypsins' (of bacterial origin: *S. griseus* proteases A-B and *L. enzymogenes* α-lytic protease [4]) and 3C-'large trypsins' suggested a parallel classification of viral protease structural subclasses. The problematic sequence alignment of small and large cellular trypsins [4] is mirrored in the alignments of the more economical 2A-3C proteases. An accurate alignment (as pertaining to structure) of viral enzymes, an important prelude to molecular modeling of the tertiary folds, is bolstered by the known superposition of crystollographically determined structures of small and large trypsin molecules [9].

The structural model for the picornavirus 3C protease by Gorbalenya et al. [1] differs from ours principally in the placement of β-strands E-F and the inter-barrel loop. Importantly, the Gorbalenya alignment places the conserved Asp-85 of the proposed catalytic triad in a loop distal from the active site according to our model. The role of the homologous Asp-102 in trypsin catalysis has recently been explored by X-ray crystallography [10]; its interaction with the base, His-57, should not be contingent on the presence of either a Ser or Cys nucleophile as Gorbalenya et al. [1] speculate.

The structural classification of viral 2A/3C subclasses had implications for the observed, distinctive substrate specificities of the viral proteases. The conserved His-161 of the 3C proteases (save

for the recent nepovirus enzyme [11]) is predicted to influence the choice of Gln/Glu-Gly/Ser as the preferred 3C processing site because of its placement in the substrate-binding pocket [2,6]. We raised a strong parallel with the (large) bacterial enzyme *S. aureus* protease, strain V8, that has a similar affinity for Glu residues at the P_1 position and has a His residue at the equivalent 213 position [2]. Indeed, this bacterial protease had the greatest sequence similarity with 3C viral proteases of any cellular trypsin-like enzyme [2]. Interestingly, the 2A viral proteases have instead an equivalent aliphatic residue (Ile/Val/Leu-123) and share an affinity for an aromatic P_1 residue with the small bacterial *S. griseus* enzymes [2]. Molecular models of the active site and binding-pocket regions of 2A/3C viral enzymes built on the known frameworks of small/large trypsin structures were consistent with the above predictions [2,6]. We further suggest that the observed 'inactivation' of picornaviral 3C enzymes upon mutation of the His-161 residue is instead a reflection of altered specificity or loss of activity arising from a changed P_1 pocket.

We note that the role of the Cys residue as a nucleophile in the viral protease catalytic mechanism is presaged by an evolutionary analogy with the cellular (papain-like) Cys proteases. The active site geometry of the catalytic His-159/Asn-175/Cys-25 triad of papain is surprisingly similar to that of trypsin but displays no further structural similarity in the rest of the protein [12]. Crystallographic analysis of a Ser-195 → Cys trypsin variant ([13] and Wilke, M., personal communication) reveals that the active site geometry is not greatly disturbed. Minute readjustments of the Cys sidechain (presumably because of the larger sulfur atom) may explain the much lowered catalytic efficiency of the mutant enzyme [13]. The conversely analogous variant of the tobacco-etch virus NIa protease, Cys-147 (3C numbering) → Ser results in a marginally active enzyme [14] in contrast to the previous negative findings of Ivanoff et al. [15]. (Dougherty et al. [14] additionally tested His-40 and Asp-85 for a catalytic role; his findings of loss of activity on altering these residues support our conclusions about the viral catalytic triad.) Furthermore, Ser active-center viral proteases may be found in other branches of RNA viruses; we and others (Hellen, C., personal communication) have independently noted the similarity of trypsin-like protein segments of southern-bean mosaic virus [16] and beet western yellows virus [17]. Our alignment of these putative plant Ser proteases with the viral Cys and cellular Ser proteases (fig.1) differs from the alignment of the sobemoviral protein region by Gorbalenya et al. [1]. The evolution of this growing superfamily of trypsin-like enzymes is fully explored elsewhere [6]. We look forward to experiments stimulated by these virus protease models that may resolve the structural ambiguities that have been discussed.

Acknowledgements: J.F.B. is jointly supported by a taxable NIH predoctoral training grant (no. NS07219) and the efforts of Tina S. Bazan.

REFERENCES

[1] Gorbalenya, A.E., Donchenko, A.P., Blinov, V.M. and Koonin, E.V. (1989) FEBS Lett. 243, 103–114.
[2] Bazan, J.F. and Fletterick, R.J. (1988) Proc. Natl. Acad. Sci. USA 85, 7872–7876.
[3] Wellink, J. and Van Kammen, A. (1988) Arch. Virol. 98, 1–26.
[4] Craik, S.C., Rutter, W.J. and Fletterick, R.J. (1983) Science 220, 1125–1129.
[5] Argos, P., Kamer, P., Nicklin, M.J.H. and Wimmer, E. (1984) Nucleic Acids Res. 12, 7251–7267.
[6] Bazan, J.F. and Fletterick, R.J. (1989) Virology, submitted.
[7] Gribskov, M., Homyak, M., Edenfield, J. and Eisenberg, D. (1988) Comput. Appl. Biol. Sci. 4, 61–66.
[8] Release 15.0 of the National Biomedical Research Foundation protein database; Sidman, K.E., George, D.G., Barker, W.C. and Hunt, L.T. (1988) Nucleic Acids Res. 16, 1869–1871.
[9] James, M.N.G. (1976) in: Proteolysis and Physiological Regulation (Ribbons, D.W. and Brew, K. eds) pp. 125–142, Academic Press, New York.
[10] Sprang, S., Standing, T., Fletterick, R.J., Stroud, R.M., Finer-Moore, J., Xuong, N.-H., Hamlin, R., Rutter, W.J. and Craik, S.C. (1987) Science 237, 905–909.
[11] Greif, C., Hemmer, O. and Fritsch, C. (1988) J. Gen. Virol. 69, 1517–1529.
[12] Argos, P., Garavito, R.M., Eventoff, W., Rossmann, M.G. and Branden, C.I. (1978) J. Mol. Biol. 126, 141–158.
[13] Higaki, J.N., Gibson, B.W. and Craik, C.S. (1987) Cold Spring Harbor Symp. Quant. Biol. 52, 615–621.
[14] Dougherty, W.G., Parks, T.D., Cary, S., Bazan, J.F. and Fletterick, R.J. (1989) Virology, submitted.
[15] Ivanoff, L.A., Towatari, T., Ray, J., Korant, B.D. and Petteway, S.R. (1986) Proc. Natl. Acad. Sci. USA 83, 5392–5396.
[16] Wu, S., Rinehart, C.A. and Kaesberg, P. (1987) Virology 161, 73–80.
[17] Veidt, J., Lot, H., Leiser, M., Schneidecker, D., Guilley, H., Richards, K. and Jonard, G. (1988) Nucleic Acids Res. 16, 9917–9932.