Volume 315, number 2, 105–108 FEBS 11935 January 1993

Hypothesis

The ovalbumin family of serpin proteins

Eileen Remold-O'Donnell

The Center for Blood Research and the Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA

Received 18 September 1992; revised version received 13 November 1992

A protein family, the 'Ov-serpins' has been identified by comparing amino acid sequence, protein characteristics and gene organization. The Ov-serpins would not be recognized as a family based on sequence identity alone. This example suggests that combinations of characteristics may need to be examined to identify family groupings within the serpin superfamily.

Serpin; Protease inhibitor; Protein family

1. INTRODUCTION

The serpins (serine protease inhibitors) and related proteins, constitute one of the earliest described protein superfamilies recognized by Hunt and Dayhoff in 1981 by computer analysis of amino acid sequence identity [1]. That seminal study revealed that the protease inhibitors α1-antitrypsin (α1-AT) and antithrombin III (AT_{III}) and chicken ovalbumin, which lacks protease inhibitory activity, are distantly related and predicted that the three molecules are members of different protein families within the same superfamily [1]. In the present manuscript, a combination of characteristics are described, which strongly indicate that five serpin proteins, including ovalbumin and several recently cloned human proteins, belong to a single family ('Ov-serpin family'). The family would not be recognized based on amino acid sequence identity alone, and the example suggests that 'combinations of characteristics' may be more appropriate as criteria for recognizing serpin families.

2. SERPIN SUPERFAMILY

Serpins are thought to share a highly ordered tertiary structure defined by the crystal structure for the prototype molecule α1-AT and consisting of nine α-helices and three β-sheets arranged in a stressed configuration with the reactive center, the most variable region, located in an exposed loop where it acts as a bait for target protease [2,3]. Unusual features of serpin protease inhibitors include extreme sequence variability of the reactive center region and, as mentioned above, the stressed configuration, which converts to a relaxed, more ordered configuration when cleaved within the reactive center loop region. Topography is thereby altered, contributing in some cases to new functions. Other serpins including ovalbumin [1] and angiotensinogen [4], which are not protease inhibitors, are susceptible to proteolytic cleavage in the exposed loop region, but fail to undergo the stressed-to-relaxed configurational switch. At the genomic level, exon–intron borders of serpins generally fail to correlate with discernible protein domains, and intron positions vary widely (e.g. [5]).

Correspondence address: E. Remold-O'Donnell, The Center for Blood Research, 800 Huntington Avenue, Boston, MA 02115, USA. Fax: (1) (617) 278 3493 or 278 3330.

3. PROTEINS OF THE Ov-SERPIN FAMILY

Molecules qualifying as Ov-serpins are ovalbumin, a secreted chicken oviduct protein, and gene Y, the product of a linked gene expressed in oviduct cells [6,7]. Gene Y has not been studied at the protein level. Human Ov-serpins include plasminogen activator-2 (PAI-2), formerly called placental-PAI and monocyte-PAI. PAI-2 exists as a cytoplasmic non-glycosylated protein in unstimulated monocytes and as a secreted glycosylated molecule of LPS-treated monocytes [8,9]. Both the cytoplasmic and secreted forms of PAI-2 result from translation of one mRNA species ('facultative polypeptide translocation', [10]). Squamous cell carcinoma antigen (SCCA), a recently sequenced human protein of unknown function, is an Ov-serpin found within normal squamous epithelial cells and secreted at high levels by the corresponding carcinoma cells [11]. Finally, elastase inhibitor (EI) an Ov-serpin found at high levels within human monocytes, macrophages and neutrophils, has been recently sequenced [12], as has an elastase inhibitor from horse leukocytes [13]. These possibly counterpart

molecules have 82% identity of amino acid sequence. The apparent counterpart of EI in guinea pigs is expressed as both intracellular and secreted protein [14].

4. EXTENT OF AMINO ACID IDENTITY

The extent of amino acid identity of Ov-serpins is high, 58% for the pair of chicken proteins, 46–50% for the three human pairs, and 39–46% for the six interspecies pairs (Table I). However, since the Ov-serpins also have considerable amino acid identity with proteins of the larger serpin superfamily, e.g. α1-AT (30–34% identity) and AT_{III} (33–40%) (Table I), this single criterion is inadequate for identifying family members.

5. STRUCTURAL FEATURES

Ov-serpins lack the N-terminal extension regions common to other serpins. They begin relative to α1-AT at position 23, the first buried residue [2]. Ov-serpins also lack C-terminal extensions, ending with proline-391, the last buried residue (Fig. 1). Structural importance is suggested for residues near proline-391 because α1-AT mutants shorter by even one residue are degraded in the cell [17]. The Ov-serpins have serine rather than asparagine as found in the larger superfamily at the penultimate position and a variable residue rather than valine at position 388 (Fig. 1).

6. SIGNAL SEQUENCE

Another feature distinguishing the Ov-serpins is the lack of a cleavable hydrophobic signal sequence commonly found in the larger serpin superfamily [11,12,18,19]. The apparent reliance of Ov-serpins on non-cleavable internal signal sequences is of special interest since, as mentioned above, several Ov-serpins appear to exist as dualistic molecules that can be either secreted or cytoplasmic. In the case of ovalbumin, a fragment was isolated corresponding to residues 239–282 of Fig. 1, which inhibited functional signal activity in an in vitro assay [20]. Examination of hybrid proteins, on the other hand, indicated that the ovalbumin signal sequence is located within residues 22–41 (positions 44–64 in Fig. 1) [21].

Table I

Percent identities of amino acid sequences

	EI	PAI-2	SCCA	gene Y	Oval
(human) EI	–	50.2 (1)	49.9 (4)	43.0 (2)	39.9 (3)
(human) PAI-2	50.2 (1)	–	45.5 (5)	42.3 (3)	39.2 (4)
(human) SCCA	49.9 (4)	45.5 (5)	–	46.0 (2)	42.6 (1)
(chicken) gene Y	43.0 (2)	42.3 (3)	46.0 (2)	–	57.9 (1)
(chicken) Oval	39.9 (3)	39.2 (4)	42.6 (1)	57.9 (1)	–
(human) AT_{III}	39.6 (6)	35.7 (7)	40.1 (6)	35.8 (6)	32.5 (6)
(human) α1-AT	30.3 (7)	31.2 (6)	33.7 (7)	29.7 (6)	29.8 (5)

Shown are percent identical amino acids in pairs of aligned sequences with the number of gaps indicated in brackets. The comparison is limited to regions common to the Ov-serpin family (379 residues of EI, 378 of PAI-2, 374 in SCCA and gene Y and 372 in ovalbumin). Interhelical regions of PAI-2, SCCA, gene Y, ovalbumin were omitted, as were N- and C-terminal extensions of α1-AT and AT_{III}. α1-AT [15] and AT_{III} [16], shown for comparison, are members of the larger serpin superfamily, but are not Ov-serpins.

7. INTERHELICAL REGION

The region between helices C and D is variable in length and non-homologous in the superfamily and is accommodated in structural models as a separate loop that does not affect overall folding [3]. This region is exceptionally long (37 residues) in PAI-2, 16 residues in SCCA, 14 residues in ovalbumin and gene Y and is absent in EI (Fig. 1, enclosed box), as well as in several molecules of the larger serpin superfamily [3]. There is thus no common feature of this region that distinguishes the Ov-serpins as a family.

8. GENE ORGANIZATION

At present, at least five very different patterns of gene organization have been defined for serpin molecules [22,23]. The three Ov-serpins characterized at the genomic level, ovalbumin [24], gene Y [7] and PAI-2, have virtually identical gene organization consisting of seven exons and eight introns, and including identical splice junctions [22,25]. Since PAI-2 and gene Y/ovalbumin have different functions and are from different species, their nearly identical exon-intron organization in the face of diverse patterns in the larger superfamily, is strong evidence that they belong to the same family. Another serpin family, consisting of the four molecules α1-AT, α1-antichymotrypsin, angiotensinogen, and heparin co-factor II/hLS2 (leuserpin 2) has also been identified based on a shared gene organization of five exons and four introns [26]. The criterion of gene organization pattern may yet prove to be most important in defining serpin families.

9. DISCUSSION

Thus, based on extent of sequence identity, absence of a cleavable signal sequence, common structural features, and common exon-intron organization, the proteins ovalbumin, gene Y, PAI-2, EI and SCCA constitute a separate branch, or family, of the serpin superfamily as predicted by Hunt and Dayhoff.* The human

*Recently sequenced protease inhibitors encoded by cowpox virus and vaccinia virus have been shown to have serpin-like sequence [27]; they share with Ov-serpins the properties of lacking cleavable signal sequences and N-terminal extensions.

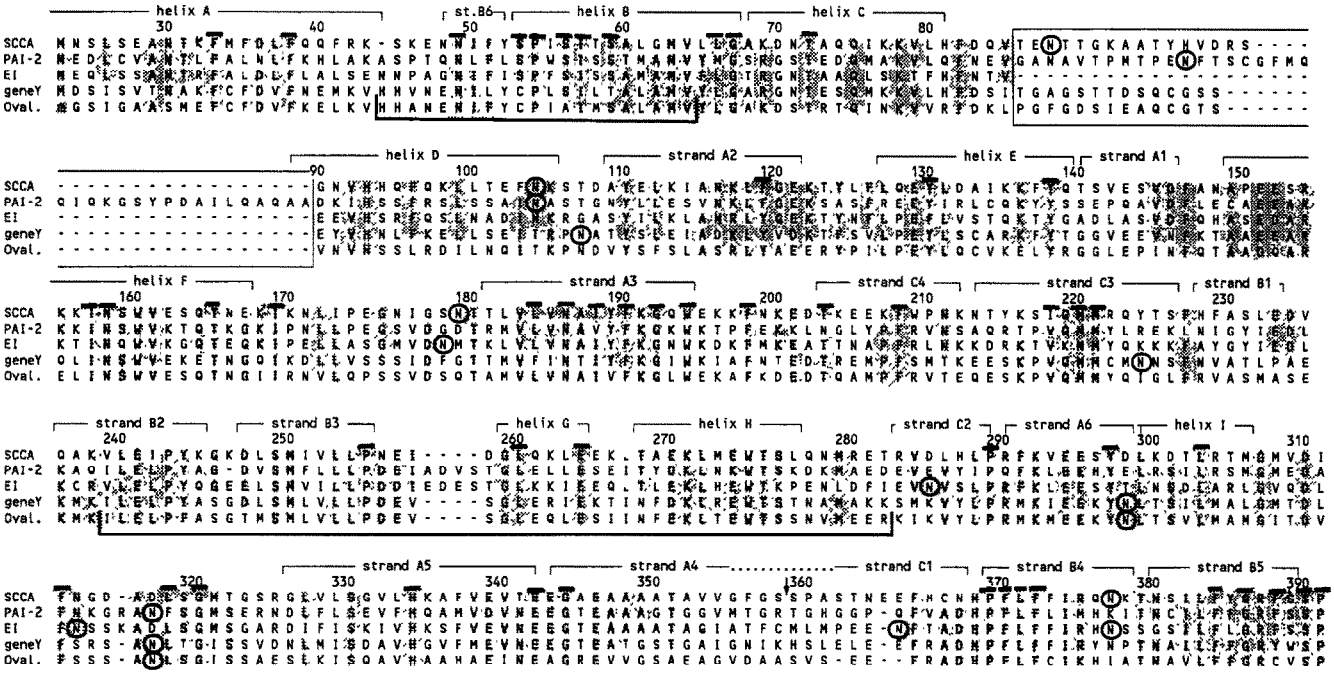

Fig. 1. Comparison of proteins of the ovalbumin family of the serpin superfamily ('Ov-serpins'). Shown are human squamous cell carcinoma antigen (SCCA), human plasminogen activator-2 (PAI-2), human monocyte/neutrophil elastase inhibitor (EI), chicken ovalbumin (Oval) and chicken gene Y (references in text). Manual alignment and position numbers are based on the crystal structure for the prototype serpin α1-AT (indicated above the aligned sequences) [2,3]. Shading indicates residues that are identical in 4 of 5 proteins or 3 of 3 human proteins. Also shaded are the homologous residues R/K or F/Y or D/E in 4 or 5 proteins. A box encloses the variable length region with minimal identities corresponding to α1-AT residues 86–89, which forms a loop between helices C and D; it is present in four Ov-serpins, but is absent in EI. The 51 dark bars indicate residues categorized as conserved in the larger serpin superfamily [3] and for the most part conserved in the five Ov-serpins. The exceptions are position 388 where the superfamily V is replaced in Ov-serpins by Y, F or C and position 390 where the superfamily N is replaced by S. Double underlines indicate ovalbumin fragments with 'signal activity' in vitro, positions 44–64 [21] and positions 239–282 [20]. The bond corresponding to the reactive site is indicated by an arrow at positions 358–359. Potential *N*-glycosylation sites, four in SCCA, three in PAI-2, five in EI, four in gene Y and two in ovalbumin, are encircled.

Ov-serpins, PAI-2, EI, and SCCA merit study because of their likely functions in fibrinolysis, inflammation and neoplasia, respectively. In addition, integrated study of two or three of these molecules might reveal mechanisms that are shared, and possibly specific, to the Ov-serpins, such as the regulated synthesis of dualistic molecules, which can be either cytoplasmic or secreted. Finally, extensive effort is currently directed toward identifying structural features underlying the 'stressed-to-relaxed' physiological switch mechanism unique to serpins. Since the Ov-serpins include both active protease inhibitor molecules (PAI-2 and EI) capable of the configurational switch and also molecules lacking this capability (ovalbumin), physicochemical comparisons within the family might provide a uniform background against which to identify the structural features responsible for the switch mechanism.

Acknowledgements: I thank Philip Auron and Andrew Webb for suggestions and critical reading of the manuscript and Abigail Allen for preparation of the figure. This study was supported by the National Institutes of Health Grant HL41579.

REFERENCES

[1] Hunt, L.T. and Dayhoff, M.O. (1980) Biochem. Biophys. Res. Commun. 95, 864–871.
[2] Loebermann, H., Tokuoka, R., Deisenhofer, J. and Huber, R. (1984) J. Mol. Biol. 177, 531–556.
[3] Huber, R. and Carrell, R.W. (1989) Biochemistry 28, 8951–8966.
[4] Doolittle, R.F. (1983) Nature 222, 417–419.
[5] Prochownik, E.V., Bock, S.C. and Orkin, S.H. (1985) J. Biol. Chem. 260, 9608–9612.
[6] McReynolds, L., O'Malley, B.W., Nisbet, A.D., Fothergill, J.E., Givol, D., Fields, S., Robertson, M. and Brownlee, G.G. (1978) Nature 273, 723–728.
[7] Heilig, R., Muraskowsky, R., Kloepfer, C. and Mandel, J.L. (1982) Nucleic Acids Res. 10, 4363–4382.
[8] Genton, C., Kruithof, E.K.O. and Schleuning, W.-D. (1987) J. Cell Biol. 104, 705–712.
[9] Wohlwend, A., Belin, D. and Vassalli, J.-D. (1987) J. Exp. Med. 165, 320–339.
[10] Belin, D., Wohlwend, A., Schleuning, W.D., Kruithof, E.K.O. and Vassalli, J.D. (1989) EMBO J. 8, 3287–3294.
[11] Suminami, Y., Kishi, F., Sekiguchi, K. and Kato, H. (1991) Biochem. Biophys. Res. Commun. 181, 51–58.
[12] Remold-O'Donnell, E., Chin, J. and Alberts, M. (1992) Proc. Natl. Acad. Sci. USA 89, 5635–5639.
[13] Dubin, A., Travis, J., Enghild, J.J. and Potempa, J. (1992) J. Biol. Chem. 267, 6576–6583.

[14] Remold-O'Donnell, E. and Lewandrowski, K. (1983) J. Biol. Chem. 258, 3251–3257.
[15] Carrell, R.W., Jeppsson, J.O., Laurell, C.B., Brennan, S.O., Owen, M.C., Vaughan, L. and Boswell, D.R. (1982) Nature 298, 329.
[16] Bock, S.C., Wion, K.L., Vchar, G.A. and Lawn, R.M. (1982) Nucleic Acids Res. 10, 8113–8125.
[17] Brodbeck, R.M. and Brown, T.L. (1992) J. Biol. Chem. 267, 294–297.
[18] Palmiter, R.D., Gagnon, J. and Walsh, K.A. (1978) Proc. Natl. Acad. Sci. USA 75, 94–98.
[19] Ye, R.D., Wun, T.-C. and Sadler, J.E. (1988) J. Biol. Chem. 263, 4869–4875.
[20] Lingappa, V.R., Lingappa, J.R. and Blobel, G. (1979) Nature 281, 117–121.
[21] Tabe, L., Krieg, P., Strachan, R., Jackson, D., Wallis, E. and Colman, A. (1984) J. Mol. Biol. 180, 645–666.
[22] Ye, R.D., Aahern, S.M., Le Beau, M.M., Lebo, R.V. and Sadler, J.E. (1989) J. Biol. Chem. 264, 5495–5502.
[23] Hirosawa, S., Nakamura, Y., Miura, O., Sumi, Y. and Aoki, N. (1988) Proc. Natl. Acad. Sci. USA 85, 6836–6840.
[24] Woo, S.L.C., Beattie, W.G., Catterall, J.F., Dugaiczyk, A., Staden, R., Brownlee, G.G. and O'Malley, B.W. (1981) Biochemistry 20, 6437–6446.
[25] Samia, J.A., Alexander, S.J., Horton, K.W., Auron, P.E., Byers, M.G., Shows, T.B. and Webb, A.C. (1990) Genomics 6, 159–167.
[26] Ragg, H. and Preibisch, G. (1988) J. Biol. Chem. 263, 12129–12134.
[27] Ray, C.A., Black, R.A., Kronheim, S.R., Greenstreet, T.A., Sleath, P.R., Salvesen, G.S. and Pickup, D.J. (1992) Cell 69, 597–604.