Vol. 87, No. 1, 1979 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
March 15, 1979 Pages 32-42

ON THE SUPERSECONDARY STRUCTURE OF ACID PROTEASES

N.S.Andreeva, A.E.Gustchina

Institute of Molecular Biology, Academy of Sciences of USSR

Moscow

Received November 7, 1978

Summary. It was found that polypeptide chains of porcine pepsin molecules and mold acid proteases consist of four topologically equivalent structural units. Each pair of units forms domain. The symmetrical packing of two units within each domain is the important structural feature of acid proteases. Although the primary structures of the four structural units are in general not homologous there are close similarities of the sequences of some topologically equivalent elements. These data are considered as the development of the idea on gene duplications and the subsequent fusion during evolution of acid proteases.

In the last year data on the three-dimensional structure of four acid proteases were published (1-4). While investigating the structural properties of the porcine pepsin molecule and comparing it to the molecules of mold acid proteases we have found some features which are common for all members of this class of enzymes and by our consideration represent an important structural property. The purpose of this publication is the description of this property.

The conformation of a pepsin chain (4,5) is presented in Fig.1. It is very close to those of other acid proteases (6,7) - all elements of the secondary structure of pepsin - β -strands and helices are arranged in the same order within a molecule as for mold acid proteases. The overall dimensions and the shape of a molecule consisting of two domains separated by a large cleft

0006-291X/79/050032-11$01.00/0

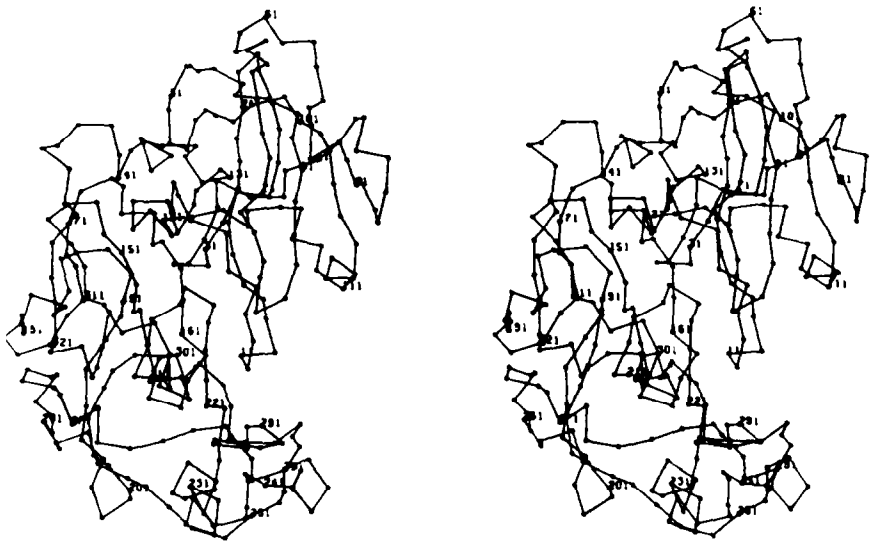

Fig. 1 Stereo-drawing of the conformation of pepsin chain (5).

are also close to those of other enzymes ,which belong to the class of acid proteases.

To describe the structural properties of each domain we tried at first to follow the sequence of the prominent loops along the chain.Two types of loops are presented in acid proteases: β -hairpins with the arms formed from extended segments which are close enough to each other to form hydrogen bonds, and wide loops with the arms which are far enough to incorporate an additional extended chain segment between them with the formation of so called "psi shaped" structure (8) very specific of acid proteases.

The first prominent loop of the N-terminal domain is the β -hairpin which has one arm containing residues with numbers 14–18 (labelled in (8) as b_I^* strand) and the other (c_I) beginning at residue 24. This c_I strand is also the arm of the wide

* We use here the notation of strands and segments proposed in (8)

loop containing the bend at 28-32 segment (c_2) and the second arm (d-strand) with residue numbers 37-42. The next structural element is a small piece of a helix 58-62 (7), then the next β-hairpin follows, containing e_I and f strands with residue numbers 70-83.

Following along the chain one can see that this sequence of the structural elements β-hairpin, wide loop, helix, β-hairpin repeats in the N-terminal domain of acid proteases once more: the β-hairpin 88-96 including f_I and f_2 strands, then after small insertion the wide loop with the bend at 1o8-112 segment and arms 104-107 (g_I) and 117-122 (h_I), the helix 138-143 and the β-hairpin 152-165 containing i and j strands.

The same sequence of the structural elements repeats also two times in C terminal domain, the β-hairpin 194-206 containing l and m strands, the wide loop with the bend at 213-217 segment or m_2 and arms 202-207*, 220-224 or n strand, helix 225-235 and the β-hairpin 237-248 containing n_1 and n_2 strands; the next structural unit is beginning at 275 residue and has the first -hairpin at the segment with residues 275-284, then the wide loop 286-303 follows with the strands o and p, helix 304-308 and the β-hairpin with the bend at 316-317 residues containing the strands r and q. The mutual arrangement of hairpins, wide loops and helices is the same in different structural units. Root mean square deviations of coordinates for topologically equivalent α-carbons are the following: for the first and the second structural units - 2.48Å, for the second and the third units - 2.20Å, for the second and the fourth - 2.07Å (Fig. 2).

The inspection of these data suggests that the polypeptide chain of acid proteases contains a small structural unit of about

*There is a small insertion in the primary structure of pepsin - cystine loop 206-210, which was not taken into account.

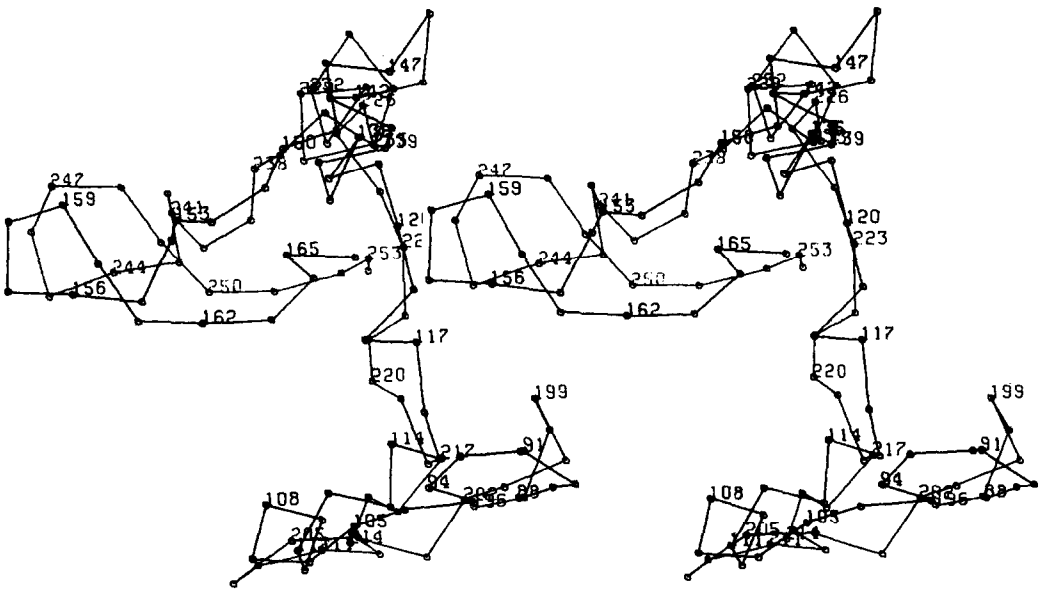

Fig. 2 Superposition of the second and the third structural units of pepsin.

70 amino acid residues. The unit includes four elements: A- the first β-hairpin, B - wide loop, C - helix and D - the second β-hairpin. The repeating of the unit leads to the formation of the N-terminal domain, followed by the connecting region between domains (165-195), and then this structural unit repeats two times once more providing the formation of the C-terminal lobe. This consideration permits to describe the structure of acid proteases in very simple terms. The first unit contains elements A_1, B_1 C_1, D_1 , the second - A_2, B_2, C_2, D_2 and so on. The connecting regions between elements can be denoted as for example $A_I B_I$, $B_I C_I$, $C_I D_I$*. Then 225-235 helix is C_3 element, active aspartic acid 215 is in B_3 loop, active aspartic acid 32 belongs to B_I loop, D_I is the "flap" above active site region containing Tyr-75, D_4 is C-terminal hairpin and so on.

* The structural similarity of units does not concern some regions connecting structural elements, and the length of helices.

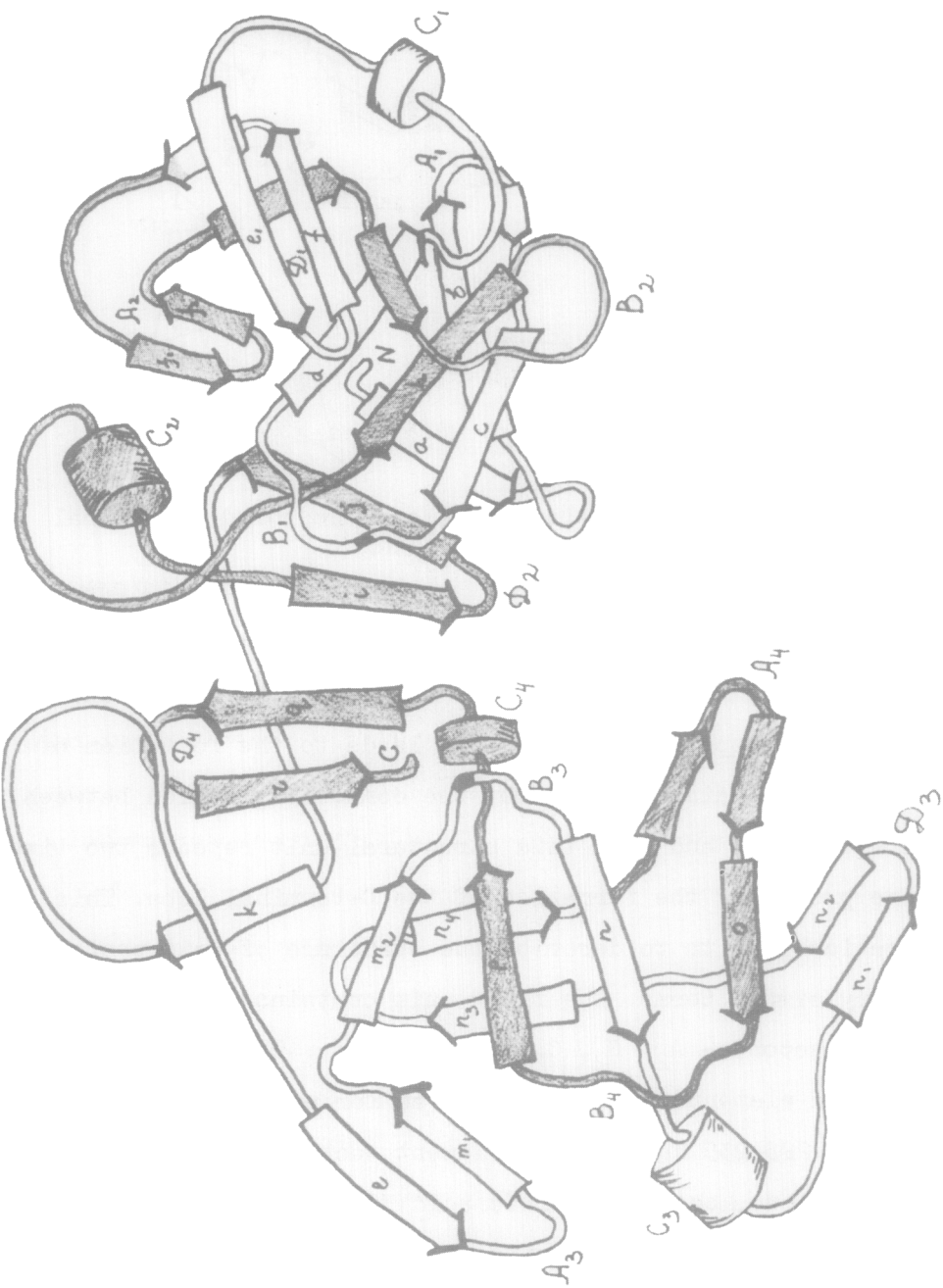

Fig. 3 Supersecondary structure of the N and C terminal domains of acid proteases.

One of the most interesting structural features of acid proteases is the symmetrical arrangement of different units in each domain (Fig.3). Two wide B-loops penetrate into each other forming two "psi-shaped" regions,containing two opposite pairs of parallel strands,followed by helices. This kind of arrangement of β -strands is the specific feature of acid proteases, the examples of the supersecondary structure of this type for β -proteins are not yet known (9). We call this type of β -strands arrangement "pepsin fold". One can see that every structural unit has only antiparallel β -structure,however, the packing of units in each domain provides the formation of the extensive β -sheet with the antiparallel and parallel strands described in previous publication (7).

The resemblance of the N and C-terminal domain conformations of acid proteases and its relationship to the internal homology of pepsin primary structure was the subject of publication by J.Tang et.al. (8). The resemblance of the three-dimensional structure of the N and C terminal domains is obviously the consequence of the presence of topologically equivalent structural units in each domain.

In spite of the large number of repeating sequences in pepsin (10) , the alignment of the primary structures of the N and C terminal domains of pepsin and penicillopepsin performed in (8) shows lack of similarity except the two active site loops (B_1 and B_3 by our notation). The alignement of amino acid sequences of the four structural units includes some additional homologous elements of pepsin primary structure (Table 1). For example, the repeating segments containing some tryptophans are in equivalent positions, all A hairpins have one or two glycines at turns and their sequences look very much similar,especially,those of A_1 and A_2 hairpins.

Table 1

Amino acids sequences of four structural units of porcine pepsin and penicillopepsin*

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	15	15	16	17	18	19	20	21	22	23	*	*	*	*	*	*	*	*	24	25	26	27
I	G	D	E	P	L	E	N	Y	L	D	T	E	Y	F	-	-	G	T	I	G	I	G	T	P									A	Q	D	F
V	A	T	N	T	P	T	A	N	-	D	E	E	Y	I	T	P	V	T	I	G	-	G	-	-									-	T	T	L
*	*	*	*	*	*	*	*	*	*	*	*	*	*	86	87	88	89	90	91	92	*	93	94	95	96	97	98	99[1]	00	01	02	03	04	05	06	07
														Y	D	T	V	Q	V	G		G	I	S	D	T	N	Q	I	F	G	L	S	E	T	E
														T	D	S	V	T	V	G		G	V	I	A	H	G	Q	A	V	Q	A	A	Q	Q	I
*	*	*	*	*	*[1]	87	88	89	90		*	91	92	93	94	95	96	97	98	99[2]	00	*	01	*	*	*	*	*	*	*	*	*	*	*	02	03	04
						E	G	Y	W			Q	I	T	L	D	S	I	T	M	D		G											E	T	I
						Q	G	F	W			S	F	N	V	D	S	Y	T	A	G		S											Q	S	G
*	*	*	*	*	*	*	*	*	*	*	*	*	*[2]	73	74	75	76	77	78	79	80	81	82		83	84	85	86	*	*	*	*	87	88	89	90
														S	A	Y	I	L	Q	D	D	D	S	-	C	T	S	G					F	E	G	M
														I	N	Y	G	P	S	G	N	G	S	T	C	L	G	G					I	Q	S	N

A

Table 1 (continued)

Table 1 (continued)

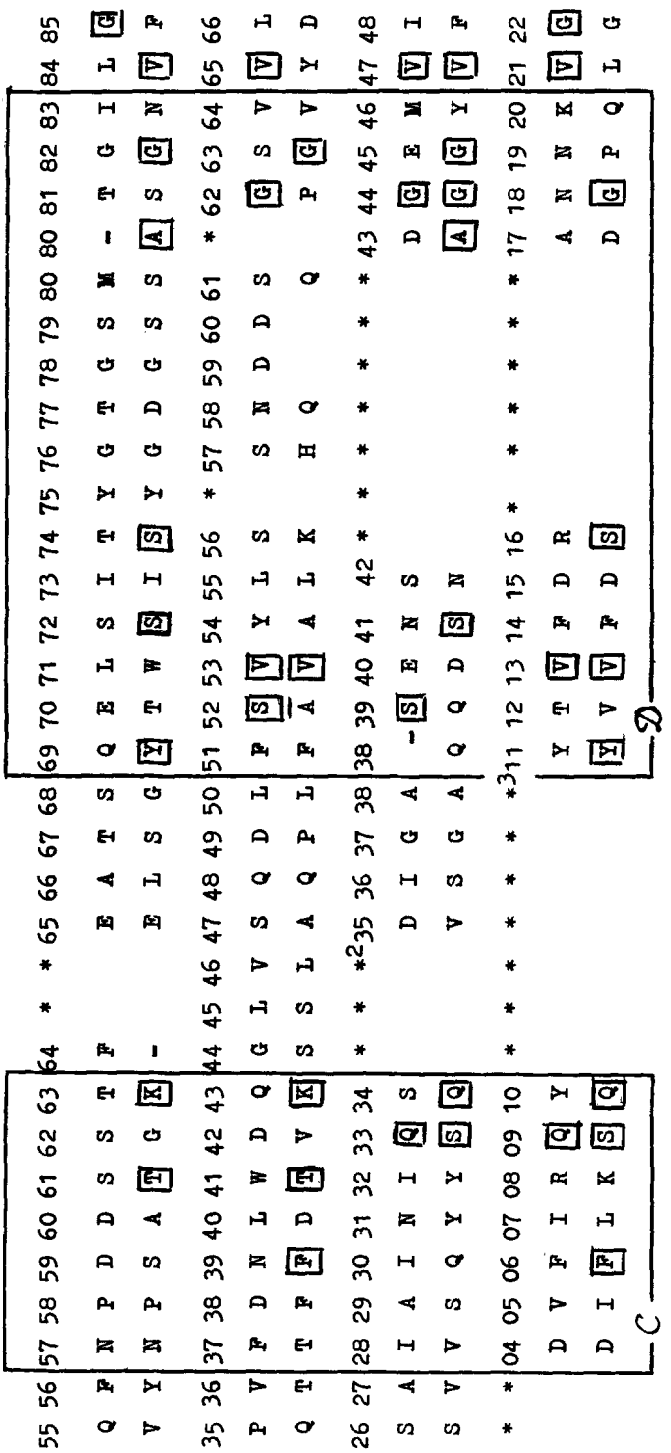

Table 1 (continued)

67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	84
[L]	G	G	I	D	[S]	[S]	Y	Y	T	G	S	L	N	W	V	P	V	-
F	G	F	I	D	S	S	K	Y	T	G	S	L	T	Y	T	[G]	V	D

49	50	*	*	*	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
S	C				[S]	[S]	I	D	S	L	P	D	I	V	F	T	I	D	G	V	Q	Y	P	L	S	P
T	C				[S]	B	V	T	B	L	P	V	S	I	S	[G]	Y	-	T	A	T	V	P	G	S	L

23	24	25	26	27
[L]	A	P	V	A
F	A	P	Q	A

* As the three-dimensional structures of pepsin and penicillopepsin are homologous, the sequence of penicillopepsin is used for statistic. The sequences are aligned in accordance with (7). Exceptions are the displacements of some residue positions in the regions of deletions.
- deletions in the structure of porcine pepsin and penicillopepsin in accordance with (7).
* identify the positions where there are not topologically equivalent residues.

This similarity ,however, concerns only several regions and in general the homology is not obvious.

Neverthless the idea on gene duplication and subsequent fusion proposed for the explanation of structural similarity of the domains of acid proteases may be equally applicable for the explanation of the structural resemblance of the four structural units of these enzymes. The symmetrical packing of the units in each domain seems to be an important feature of acid proteases.

We are grateful to Mr.Gdanov for the preparing of stereodrawings and calculations.

REFERENCES

1. Subramanian,E.,Liu,M., Swan,I.D.A.,Davies,D.R. (1977) Adv.Exp.Med. and Biol., 95, 33-41.
2. Jenkins, J., Tickle,I., Sewell, T., Ungaretti,L., Wollmer,A., Blundell, TT (1977) Adv.Exp.Med. and Biol., 95, 43-60
3. Hsu, I-N., Delbaere,L.T.J., James, M.N.G., Hofmann,T. (1977) Adv.Exp.Med. and Biol., 95, 61-81.
4. Andreeva, N.S., Gustchina, A.E.,Fedorov, A.A., Shutzkever N.E., Volnova, T.V. (1977) Adv.Exp.Med. and Biol.,95, 23-32.
5. Andreeva, N.S., Fedorov, A.A., Gustchina, A.E., Riskulov, R.R., Safro,M.G., Shutzkever,N.E. (1978) J.Mol.Biol. (Russian) 12, 922-935.
6. Subramanian,E.,Swan, I.D.A., Liu,M., Davies, D.R., Jenkins, J.A. Tickle,I.J.,Blundell,T.L. (1977) Proc.Natl.Acad.Sci. USA, 74, 556-559.
7. Hsu, I-N.,Delbaere,L.T.J.,James M.N.G.,Hofmann, T. (1977) Nature (London), 266, 140-145.
8. Tang,J.,James,M.N.G.,Hsu,I-N.,Jenkins,J.,Blundell T. (1978) Nature (London), 271, 618-621.
9. Levitt,M.,Chotia,C., (197) Nature (London), 261,552-557
10. Tang,J.,Sepulveda,P.,Marciniszyn,J.,Chen,K.C.S.,Huang,W-Y.,Tao,N., Liu,D.,Lanier,J.P. (1973) Proc.Natl.Acad.Sci.USA,70,3437-3439.