Genealogy of mammalian cysteine proteinase inhibitors

Common evolutionary origin of stefins, cystatins and kininogens

Werner Müller-Esterl, Hans Fritz, Josef Kellermann*, Friedrich Lottspeich*, Werner Machleidt+ and Vito Turk°

*Abteilung für Klinische Chemie und Klinische Biochemie in der Chirurgischen Klinik Innenstadt der Universität München, Nussbaumstr. 20, D-8000 München 2, *Max-Planck-Institut für Biochemie, Am Klopferspitz, D-8033 Martinsried, +Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie, Universität München, Goethestr. 33, D-8000 München 2, FRG* and *°Department of Biochemistry, J. Stefan Institute, Jamova 39, Yu-61000 Ljubljana, Yugoslavia*

Received 22 August 1985

A model for the evolution of mammalian cysteine proteinase inhibitors has been constructed on the basis of sequence homology. This model suggests that the diversity of cysteine proteinase inhibitors has evolved from two ancestral units forming the building blocks of stefin and cystatin. Gene triplication of the archetypal inhibitor generated the kininogen heavy chain which contains three cystatin-like copies. Hence, the superfamily of mammalian cysteine proteinase inhibitors is constituted by at least three distinct families, with stefin, cystatin and kininogen as their prototypes.

Cysteine proteinase inhibitor *Stefin* *Cystatin* *Kininogen* *Evolution* *Sequence homology*

1. INTRODUCTION

Inhibitors of cysteine proteinases are present in tissues and body fluids of mammalian species [1,2]. By the criterion of their M_r, cysteine proteinase inhibitors (CPIs) form 2 classes [3]: low-M_r CPIs including stefin and cystatin, and high-M_r CPIs represented by the kininogens (also termed α-CPIs). The peptapeptide Gln-Val-Val-Ala-Gly present in most of the known mammalian CPIs is believed to form part of the reactive site of the inhibitors [1,4]. Two copies of this consensus peptide are harbored by human low-M_r and high-M_r kininogen thus suggesting that their common heavy chain [5] might be composed of 2 inhibitor domains [4] capable of interacting with cysteine proteinases [6,7].

To date, the evolutionary relationship among the various CPIs has remained largely obscure. Here, we demonstrate that the superfamily of mammalian CPIs encompasses at least 3 distinct families, i.e. the stefins, cystatins and kininogens, which have evolved from common ancestors.

2. MATERIALS AND METHODS

2.1. *Computer program RELATE*

This program compares all segments of a length of 25 amino acids from one sequence with all segments of identical length from another [8]. A segment score is accumulated by adding the pair scores of the corresponding segments. The statistical significance of the relationship is evaluated by determining the segment scores for 100 randomly permuted sequences and expressing the comparison scores in standard deviation (SD) units. Scores $\geqslant 3.0$ SD are considered to indicate significant sequence homology.

2.2. *Hydropathic index*

The hydropathy profiles were calculated using the moving segment approach which continuously

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

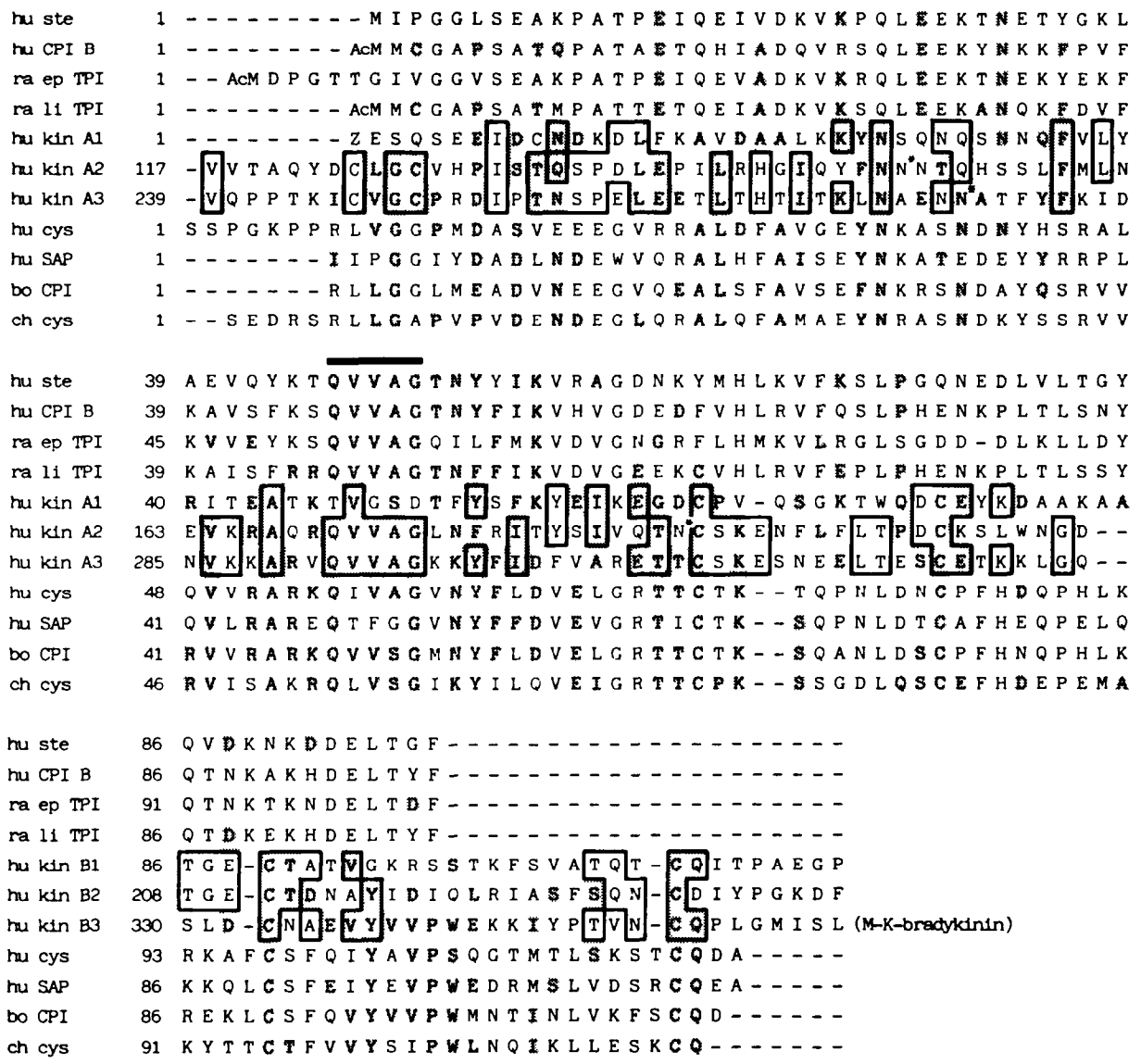

Fig.1. Sequence homologies between stefin-like inhibitors, the human kininogen heavy chain repeats and cystatin-like inhibitors. Stefin-like inhibitors: hu ste, stefin from human granulocytes [12]; hu CPI B, cysteine proteinase inhibitor B from human liver (also named cystatin B) [23]; ra ep TPI, thiol proteinase inhibitor from rat epidermis [13]; ra li TPI, thiol proteinase inhibitor from rat liver [14]. Kininogen repeats: A1B1, A2B2, A3B3, internal repeats of the human kininogen heavy chain (Kellermann et al., submitted). Cystatin-like inhibitors: hu cys, cystatin (γ-trace) from human plasma [3,15]; hu SAP, acidic protein from human saliva [16]; bo CPI, cysteine proteinase inhibitor from bovine colostrum [17]; ch cys, cystatin from chicken egg white [18,19]. Allowance for gaps (indicated by dashes) has been made to improve the alignment. Residues identical among the internal repeats of the kininogen heavy chain are boxed. Sequence identities among the kininogen repeats and at least one of the other inhibitors are shaded. The solid bar marks the potential reactive sites. Numbers indicate the relative positions of the residues in the amino acid sequences. Asterisks locate carbohydrate attachment sites. Ac, acetyl; Z, pyroglutamic acid.

determines the average hydropathy within a segment length of 9 amino acids as it advances through the sequence [9].

3. RESULTS

3.1. *Internal homologies in the human kininogen heavy chain*

Search for internal homologies in the heavy chain shared by low-M_r and high-M_r kininogen (Kellermann et al., submitted) has led to the identification of 3 internal repeats spanning from residues 1 to 116, 117 to 238 and 239 to 360 (fig.1). Significant sequence homologies (scores from 3.87 to 12.27) were found among these repeats (table 1). Thus, evolvement of the heavy chain of the human kininogens by gene triplication can be envisaged [10]. The lengths of the internal repeats vary from 116 to 122 amino acids. Considering the underlying pattern of sequence homology, each of these long repeats can be subdivided into an N-terminally located portion of 85–91 amino acids (referred to as segment 'A') and a stretch of 31 amino acids ('B') positioned at their C-terminus. The entire structure of the human kininogen heavy chain is then schematically represented by A1B1-A2B2-A3B3 or $(AB)_3$. Similar patterns of tandemly repeated AB units are found in the rat and bovine kininogens (not shown). Among the A/B segments, the pairs A2/A3 and B1/B2 had the highest scores (12.91 and 22.23) indicating that they have diverged more recently in evolution.

3.2. *Sequence homologies among cysteine proteinase inhibitors*

The sequences of mammalian low-M_r CPIs and of human kininogen repeats A1, A2, A3, B1, B2, and B3 are aligned in fig.1. By the criterion of sequence homology, the low-M_r CPIs fall into 2 categories: the stefin group (fig.1, upper) with a moderate sequence homology to the kininogen AB repeats, and the cystatin group (fig.1, lower) with a considerable homology to the AB repeats. The stefin-type inhibitors share a single region of extended sequence homology ('homology box') with the kininogen A segment centered around the proposed reactive site. Only few additional sequence identities are present in the peripheral portions of these inhibitors. Unlike the stefin group, the cystatin-type inhibitors share 2 homology boxes with the kininogen repeats. The first box holds the reactive site residues embedded in kininogen segment A. The second homology box is located near the C-terminus of the inhibitors and covers portion B of the kininogen repeats. This latter homology box is not found in the stefin-like inhibitors; rather, their sequences stop 13 amino acids after the start of segment B (fig.1, bottom).

In a quantitative approach, we have evaluated the sequence similarities between the kininogen repeats and the various low-M_r CPIs by the RELATE program [8]. Table 1 lists the resulting scores. A significant homology to at least one of the internal repeats was found for each of the low-M_r CPIs except for rat epidermal TPI. Mean scores

Table 1

Sequence homologies between the internal repeats A1B1, A2B2, and A3B3 of the human kininogen heavy chain, stefin-like inhibitors and cystatin-like inhibitors

	A1B1	A2B2	A3B3	hu ste	hu CPI B	ra ep TPI	ra li TPI	hu cys	hu SAP	bo CPI	ch cys
A1B1	–	3.89	3.87	0.39	2 64	1.96	3.14	1.56	0.70	3.15	4.89
A2B2	3.89	–	12.27	1.38	1.04	0.83	2.52	4.32	1.52	5.14	5.02
A3B3	3.87	12.27	–	6.01	3.19	1.80	4.00	6.26	8.37	10.25	5.99

Segment comparison scores ⩾ 3.0 SD units indicating significant homology among 2 sequences are shaded. The acronyms from fig.1 are used

of 4.76 and 2.41 were calculated for the cystatin-like and stefin-like CPIs. This indicates that the cystatin-type inhibitors are more closely related to the heavy chain repeats than the stefin-like inhibitors. Along this line, 2 of the 4 cystatin-like inhibitors, but not a single stefin-like inhibitor had significant sequence homologies to each of the 3 AB repeats. From these results we conclude that the kininogen heavy chain contains 3 copies of a cystatin-like inhibitor rather than of a stefin-like inhibitor.

Among the various AB repeats, A3B3 had the most pronounced homology to low-M_r CPIs (mean score 5.73), while A2B2 and A1B1 showed significantly lower relatedness (mean scores 2.72 and 2.30). This suggests that the A3B3 domain which precedes the kinin segment in mammalian kininogens has been best conserved during the evolution of mammalian CPIs (vide infra). Our conclusion is corroborated by the fact that the hydropathy profile of human cystatin matches almost perfectly with that of A3B3, while the corresponding profile of stefin fits only moderately (fig.2).

3.3. *Evolution of mammalian cysteine proteinase inhibitors*

Comparison of the known sequences of 12 CPIs revealed (table 2) that the superfamily of mammalian CPIs is constituted by 3 well-defined families, one of which containing the kininogen heavy chains ('kininogen family'), another comprising the stefin-type inhibitors ('stefin family'), and a third holding the cystatin-related inhibitors ('cystatin family'). Typically, sequence homologies among the members of each family were high (mean scores from 24.56 to 69.12), while those among the members of different families were low though mostly significant (mean scores from 2.59 to 7.73). The data from table 2 rank the pairs of CPI families according to their mean comparison scores in the order kininogens/cystatins (7.73), cystatins/stefins (4.28) and stefins/kininogens (2.59).

On the basis of these findings we have constructed a scheme for the evolution of mammalian CPIs (fig.3). This model predicts that the diversity of mammalian CPIs has evolved from 2 ancestral units A and B. The stefin progenitor represents the archetype of the whole superfamily in that it comprises a single A unit. The cystatins have acquired a second element (B), possibly by gene fusion thus forming the prototype of an AB unit. Then, in a gene duplication event, intermediate $(AB)_2$ molecules were formed. Eventually, another gene duplication gave rise to the present day $(AB)_3$ form of the kininogen heavy chains. At present, the mode and time point of the acquirement of the

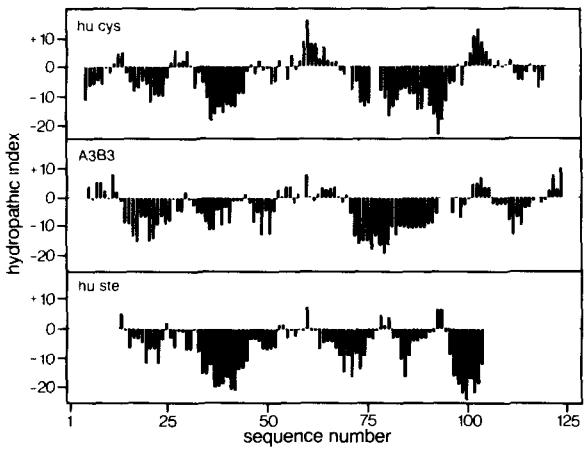

Fig.2. Hydropathy profiles of human cystatin (hu cys), internal repeat A3B3 of the human kininogen heavy chain (A3B3), and human stefin (hu ste). Numbers indicate relative positions of the residues in the amino acid sequences. Gaps are introduced as in fig.1.

Table 2

Mean segment comparison scores (SD units) for the members of the superfamily of mammalian cysteine proteinase inhibitors

	kin	cys	ste
kin	69.12	7.73	2.59
cys	7.73	32.57	4.28
ste	2.59	4.28	24.56

Highest scores identifying the kininogen family (kin), the stefin family (ste) and the cystatin family (cys) are shaded. Sequence data sources for the heavy chains of human kininogens (Kellermann et al., submitted), bovine kininogens [20] and rat kininogens [21,22] are as indicated

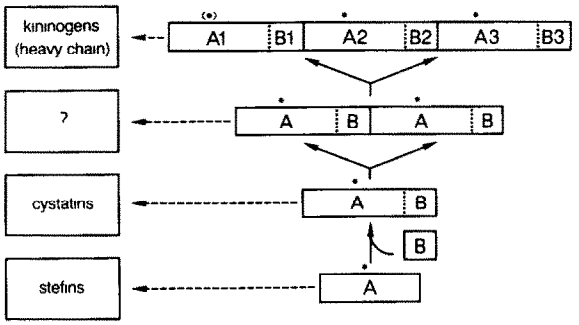

Fig.3. Genealogy of mammalian cysteine proteinase inhibitors. 'A' denotes the N-terminal portion of the long internal repeats of human kininogens heavy chain, and 'B' the corresponding C-terminal segment. Asterisks mark the potential reactive sites.

other constituents of the kininogens, i.e. the kinin region and the light chain(s), are still obscure.

4. DISCUSSION

The model for the evolution of mammalian cysteine proteinase inhibitors proposed here has several intriguing implications. First, it strongly suggests that the kininogen heavy chain encompasses 3 copies of an archetypal cystatin and might thus represent a triple-headed inhibitor of cysteine proteinases. Second, it provides a rationale for the nomenclature of mammalian cysteine proteinase inhibitors. Third, it predicts that the kininogens have descended from a primordial $(AB)_2$ inhibitor presently unknown in mammalian or non-mammalian species.

4.1. *Kininogens as multi-headed inhibitors*

In view of the 3 cystatin blocks establishing the kininogen heavy chain one might expect that the kininogens are trivalent inhibitors. However, inspection of the sequence portions holding a potential reactive site indicates that only repeats A2B2 and A3B3 have conserved the consensus sequence Q-V-V-A-G, while the A1B1 domain has almost entirely lost this sequence (cf. fig.1). It is tempting to hypothesize that at most 2 of the 3 internal repeats expose functionally active inhibitor sites, i.e. A2B2 and A3B3. Only recently, this conclusion has been confirmed in an elegant study by Salvesen et al. [11].

4.2. *Nomenclature of mammalian cysteine proteinase inhibitors*

In light of the newly established evolutionary model for mammalian cysteine proteinase inhibitors, consequences as to the nomenclature of these inhibitors have to be drawn. Clearly, by the criterion of sequence homology the CPI superfamily comprises 3 distinct families which we have named according to their prominent members: the stefin family, the cystatin family and the kininogen family. Using this classification system, all CPIs for which sequence data are presently available can be assigned to one of the 3 families without ambiguity or overlapping. Other criteria such as M_r, disulfide content and prevalent occurrence (intra- vs extracellular) might also be considered for the classification of CPIs [1,2]. Unlike the former, these latter criteria per se do not meet the requirement of an unequivocal assignment unless they are combined.

4.3. *Double-headed cysteine proteinase inhibitors*

Our model for the evolution of the mammalian CPIs predicts binary precursor(s) composed of 2 cystatin-like AB units. At present, no such inhibitors are known. It seems worthwhile to look for these evolutionary intermediates, as $(AB)_2$ inhibitors would provide the missing link between the cystatins and the kininogens. It is also anticipated that the predicted binary inhibitors should give us an idea as to the time point of the fusion of the various gene elements coding for the kininogen heavy chain, the kinin and the light chain portions [10]. Thus, the finding of such primordial inhibitors might provide important clues to the deeper understanding of the pedigree of the kininogens and hence of the genealogy of mammalian cysteine proteinase inhibitors.

ACKNOWLEDGEMENTS

We like to thank Drs G. Salvesen and S. Nakanishi for making preprints of their papers available to us prior to publication. This work was supported by grants from the Deutsche Forschungsgemeinschaft (SFB 207), the Research Council of Slovenia, Yugoslavia, and the International Bureau of the Kernforschungsanlage Jülich.

REFERENCES

[1] Turk, V., Brzin, J., Lenarcic, B., Locnikar, P., Popovic, T., Ritonja, A., Babnik, J., Bode, W. and Machleidt, W. (1985) in: Intracellular Protein Catabolism (Khairallah, E. and Bond, J.S. eds) pp.91–103, A.R. Liss, New York.

[2] Barrett, A. (1985) in: Intracellular Protein Catabolism (Khairallah, E. and Bond, J.S. eds) pp.105–116, A.R. Liss, New York.

[3] Brzin, J., Popovic, T., Turk, V., Borchart, U. and Machleidt, W. (1984) Biochem. Biophys. Res. Commun. 118, 103–109.

[4] Ohkubo, I., Kurachi, K., Takasawa, T., Shiokawa, H. and Sasaki, M. (1984) Biochemistry 23, 5691–5697.

[5] Tagaki, Y., Kitamura, N. and Nakànishi, S. (1985) J. Biol. Chem. 260, 8601–8609.

[6] Müller-Esterl, W., Fritz, H., Machleidt, W., Ritonja, A., Brzin, J., Kotnik, M., Turk, V., Kellermann, J. and Lottspeich, F. (1985) FEBS Lett. 182, 310–314.

[7] Sueyoshi, T., Enjyoji, K., Shimada, T., Kato, H., Iwanaga, S., Bando, Y., Kominami, E. and Katunuma, N. (1985) FEBS Lett. 182, 193–195.

[8] George, D.G., Orcutt, B.C., Dayhoff, M.O. and Barker, W.C. (1985) PIR Report REL-0185, Version 2.0, Natl. Biomed. Res. Found., Georgetown Univ. Med. Center, Washington, DC.

[9] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.

[10] Kitamura, N., Kitagawa, H., Fukushima, D., Takagaki, Y., Miyata, T. and Nakanishi, S. (1985) J. Biol. Chem. 260, 8610–8617.

[11] Salvesen, G., Parkes, C., Abrahamson, M., Grubb, A. and Barrett, A.J. (1985) Biochem. J., in press.

[12] Machleidt, W., Borchart, U., Fritz, H., Brzin, J., Ritonja, A. and Turk, V. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 1481–1486.

[13] Takio, K., Kominami, E., Bando, Y., Katunuma, N. and Titani, K. (1984) Biochem. Biophys. Res. Commun. 121, 149–154.

[14] Takio, K., Kominami, E., Wakamatsu, N., Katunuma, N. and Titani, K. (1983) Biochem. Biophys. Res. Commun. 115, 902–908.

[15] Grubb, A. and Löfberg, H. (1982) Proc. Natl. Acad. Sci. USA 79, 3024–3027.

[16] Isemura, S., Saitoh, E. and Sanada, K. (1984) J. Biochem. 96, 489–498.

[17] Hirado, M., Tsunasawa, S., Sakiyama, F., Niinobe, M. and Fujii, S. (1985) FEBS Lett. 186, 41–45.

[18] Turk, V., Brzin, J., Longer, M., Ritonja, A., Eropkin, M., Borchart, U. and Machleidt, W. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 1487–1496.

[19] Schwabe, C., Anastasi, A., Crow, H., McDonald, J.K. and Barrett, A.J. (1984) Biochem. J. 217, 813–817.

[20] Kitamura, N., Takagaki, Y., Furuto, S., Tanaka, T., Nawa, H. and Nakanishi, S. (1983) Nature 305, 545–549.

[21] Cole, T., Inglis, A.S., Roxburgh, C.M., Howlett, G.J. and Schreiber, G. (1985) FEBS Lett. 182, 57–61.

[22] Furuto-Kato, S., Matsumoto, A., Kitamura, N. and Nakanishi, S. (1985) J. Biol. Chem., in press.

[23] Ritonja, A., Machleidt, W. and Barrett, A. (1985) Biochem. Biophys. Res. Commun., in press.