



ACADEMIC  
PRESS

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Biochemical and Biophysical Research Communications 305 (2003) 28–36

BBRC

[www.elsevier.com/locate/ybbrc](http://www.elsevier.com/locate/ybbrc)

## Genomic overview of serine proteases<sup>☆</sup>

George M. Yousef,<sup>a,b</sup> Ari D. Kopolovic,<sup>a</sup> Marc B. Elliott,<sup>a</sup>  
and Eleftherios P. Diamandis<sup>a,b,\*</sup>

<sup>a</sup> Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ont., Canada M5G 1X5

<sup>b</sup> Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ont., Canada M5G 1L5

Received 25 March 2003

### Abstract

Serine proteases (SP) are peptidases with a uniquely activated serine residue in the substrate-binding pocket. They represent about 0.6% of all proteins in the human genome. SP are involved in many vital functions such as digestion, blood clotting, fibrinolysis, fertilization, and complement activation and are related to many diseases including cancer, arthritis, and emphysema. In this study, we performed a genomic analysis of human serine proteases utilizing different databases, primarily that of MEROPS. SP are distributed along all human chromosomes except 18 and Y with the highest density (23 genes) on chromosome 19. They are either randomly located within the genome or occur in clusters. We identified a number of SP clusters, the largest being the kallikrein cluster on chromosome 19q13.4 which is formed of 15 adjacent genes. Other clusters are located on chromosomes 19p13, 16p13, 14q11, 13q35, 11q22, and 7q35. Genes of each cluster tend to be of comparable sizes and to be transcribed in the same direction. The members of some clusters are sometimes functionally related, e.g., the involvement of many kallikreins in endocrine-related malignancies and the hematopoietic cluster on chromosome 14. It is hypothesized that members of some clusters are under common regulatory mechanisms and might be involved in cascade enzymatic pathways. Several functional domains are found in SP, which reflect their functional diversity. Membrane-type SP tend to cluster in 3 chromosomes and have some common structural domains. Several databases are available for screening, structural and functional analysis of serine proteases. With the near completion of the Human Genome Project, research will be more focused on the interactions between SP and their involvement in pathophysiological processes.

© 2003 Elsevier Science (USA). All rights reserved.

**Keywords:** Serine proteases; Kallikreins; MEROPS; Evolution; Phylogenetic analysis; Gene clusters; Protein domains; Mapping; Human genome project

Proteases are enzymes that cleave proteins by the catalysis of peptide bond hydrolysis. Based on their catalytic mechanisms, they can be classified into five main classes: proteases that have an activated cysteine residue (cysteine proteases), an aspartate (aspartate proteases), a metal ion (metalloproteases), a threonine (threonine proteases), and proteases with an active serine (serine proteases). According to the widely used and most comprehensive database of proteases (MER-

OPS; <http://www.merops.co.uk>), enzymes of each catalytic type are classified into evolutionarily distinct “clans” and each clan is subdivided into “families” based on sequence homology and order of the catalytic triad [1].

Serine proteases (SP) are a family of enzymes that utilize a uniquely activated serine residue in the substrate-binding pocket to catalytically hydrolyze peptide bonds [2]. SP carry out a diverse array of physiological functions, the best known being digestion, blood clotting, fibrinolysis, fertilization, and complement activation during immune responses [3]. They have also been shown to be associated with many diseases including cancer, arthritis, and emphysema [3–8].

Serine proteases exhibit preference for hydrolysis of peptide bonds adjacent to a particular class of amino

<sup>☆</sup> Abbreviations: *KLK*, human kallikrein (gene); kb, kilobase; SP, serine proteases; NCBI, National Center for Biotechnology Information; UCSC, University of California at Santa Cruz; HGP, Human Genome Project.

\* Corresponding author. Fax: +416-586-8628.

E-mail address: [ediamandis@mtsinai.on.ca](mailto:ediamandis@mtsinai.on.ca) (E.P. Diamandis).

acids. In the trypsin-like group, the protease cleaves peptide bonds following basic amino acids such as arginine or lysine, since it has an aspartate (or glutamate) in the substrate-binding pocket, which can form a strong electrostatic bond with these residues. The chymotrypsin-like proteases have a non-polar substrate-binding pocket, and thus, require an aromatic or bulky non-polar amino acid such as tryptophan, phenylalanine, tyrosine or leucine. The elastase-like enzymes, on the other hand, have bulky amino acids (valine or threonine) in their binding pockets, thus requiring small hydrophobic residues, such as alanine [9].

Owing to the expanding roles for serine proteases, there has been increasing interest in the identification, structural, and functional characterization of all members of the serine protease family of enzymes in humans and other organisms. The near completion of the Human Genome Project provides a unique opportunity for such efforts. Structural characterization of all serine proteases and extensive analysis of their location are the first steps towards understanding the control of their gene expression and their involvement in various physiological and pathological conditions.

In the present study, we provide an overview of the distribution of human serine proteases along different chromosomes with fine mapping of the serine protease clusters, including genes in each cluster, their direction of transcription, and distances between genes. We also screened all serine proteases for common functional domains and performed a preliminary phylogenetic analysis.

## Materials and methods

We screened several available databases of proteins and proteases for all confirmed and predicted serine proteases. These databases included MEROPS, SwissProt, and the verified and annotated genes of the Human Genome Project, available through NCBI. Details of these databases are described below.

The amino acid sequence of each protein was scanned using the ProfileScan (<http://hits.isb-sib.ch/cgi-bin/PFSCAN?>) and ScanProsite algorithms (<http://ca.expasy.org/tools/scanprosite/>). Domains and secondary structural features were screened by several resources including the PFam database (release 7.5) (<http://pfam.wustl.edu>) and the PROSITE databases of profiles and verified in some cases with data available from the SwissProt and InterPro databases ([www.expasy.org](http://www.expasy.org)).

Mapping of serine proteases in each cluster was done through the University of California at Santa Cruz databases and confirmed by the Human Genome Project database and other resources available at NCBI. Multiple alignments were done using the “Clustal W” software [10]. Phylogenetic analyses were performed with the “Phylip” software package (<http://evolution.genetics.washington.edu/phylip.html>) and the Molecular Evolutionary Genetics Analysis (Mega) program ([www.megasoftware.net](http://www.megasoftware.net)). Different trees were constructed using a range of methods (UPGMA, Neighbor joining, Minimum Evolution, and Maximum Parsimony) with different distance option models (Number of Differences, *p*-Distance, Poisson Correction, and Gamma Distance). Transmembrane prediction was also screened by the

“Classification and Secondary Structure Prediction of Membrane Proteins” Web page “SOSUI” of the Tokyo University of Agriculture and Technology.

## Results

### *Distribution of serine proteases across the human genome*

According to our survey of the human genome, utilizing the databases mentioned under ‘Methods,’ there are about 500 confirmed, non-redundant proteases in the human genome so far, including non-peptidase homologues (see also Appendix 1). This represents about 2% of all gene products in humans [1]. This number increases to about 700 when we include the “predicted” genes and proteins. Our approximate figures while conducting this study (July 2002) indicate that proteases are distributed as follows: 4% aspartate, 26% cysteine, 34% metallo, 5% threonine, and 30% serine proteases. These figures are in accordance with those of other recently published reports [11,12]. Fig. 1 shows the densities of different classes of proteases on different human chromosomes.

In terms of absolute numbers, we identified approximately 150 serine proteases within the human genome, which are distributed in all chromosomes except 18 and Y; in these two chromosomes, no serine protease genes could be localized so far. Data obtained from the MEROPS, NCBI, Human Genome Project, and UCSC databases allowed us to locate all serine proteases on different chromosomes as shown in Fig. 2. The density of serine proteases varies from just 1–3 genes in chromosomes 8, 9, 10, 12, 17, 20, 22, and X to up to 23 genes on chromosome 19. Most genes are sporadic and relatively few clusters exist. The largest cluster of serine proteases is the kallikrein family, located on the long arm of chromosome 19 (19q13.4). This cluster is formed of 15 genes and covers a genomic area of about 300 kb. A second cluster is found on the short arm of the same chromosome and is formed of four genes, in addition to a non-peptidase homologue (azurocidin 1). Approximately ~500 kb centromerically there is another serine protease gene (PRSS15) which is separated by about 600 kb from another serine protease (endopeptidase Clp; CLPP). We examined all genes that localize within these two gaps and found no genes with the characteristic serine protease domain.

Another serine protease cluster is the tryptase cluster located on the short arm of chromosome 16 and is formed of four tryptases and their alleles (e.g., tryptase  $\beta$ 1 and 2), which are followed (1400 kb centromerically) by another cluster of three genes on chromosomal band 6p13.3 (Fig. 2). Two other serine proteases (PRSS8 or prostatic and rhomboid-like protease 1) are located further centromerically. A third cluster includes four hematopoietic

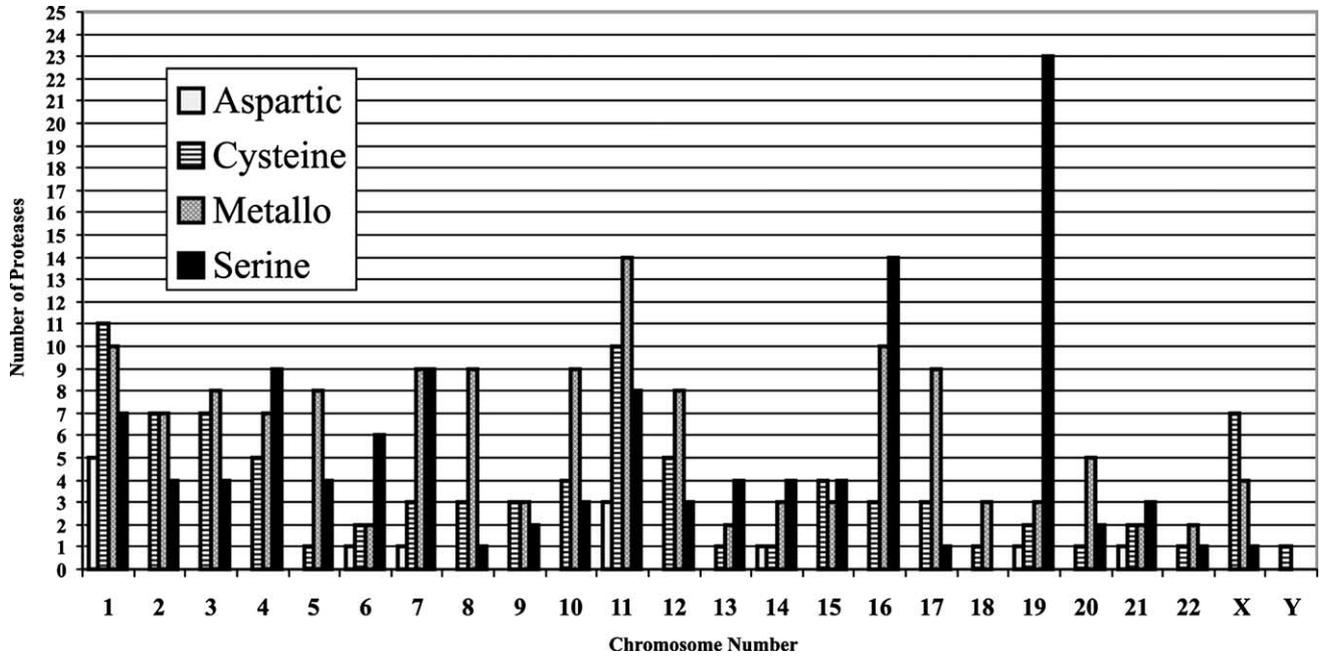


Fig. 1. Density of distribution of different proteases in chromosomes of the human genome. For details, see text.

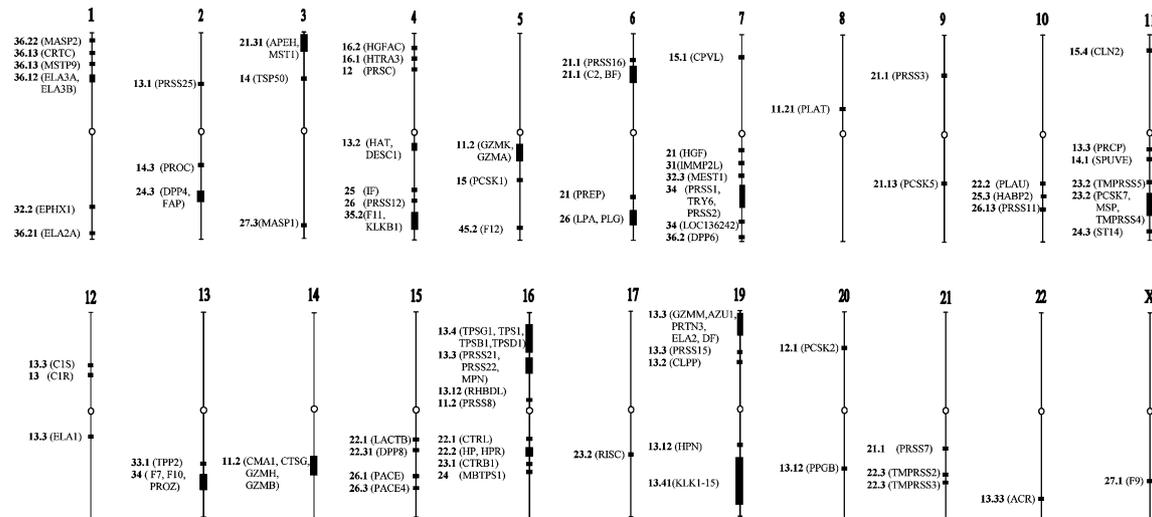


Fig. 2. A schematic presentation of the location of most serine proteases in the human genome. Chromosomes are presented by vertical lines with the number of each chromosome indicated above. The centromere is presented by a circle in the middle of each chromosome which separates the p-arm (above) from the q-arm (below). Protease positions are indicated by solid bars with the gene name, followed by its chromosomal band shown beside each bar. Gene clusters are indicated by wider rectangles. Note that no serine protease genes were found so far on chromosomes 18 and Y. For full gene names, see Appendix 1. Note that some genes are not as yet precisely mapped due to the presence of gaps in the human genome. Figure is not drawn to scale.

serine proteases (granzymes and cathepsin G) and is located on chromosome 14q11.2. A cluster of three genes is located on the long arm of chromosome 11, separated ~4000 kb centromerically from the transmembrane serine protease 5, and ~11,000 kb telomerically from the matriptase gene (ST14). Interestingly, many serine protease members of this region have transmembrane

domains. The serine protease cluster on chromosome 7 is formed of three trypsin genes.

*Detailed mapping of serine protease clusters*

We utilized the UCSC and HGP databases to perform fine mapping of the major serine protease clusters

on different chromosomes. The largest cluster of serine proteases is the human kallikrein gene cluster. We previously characterized the human kallikrein gene locus with a single base accuracy and indicated the direction of transcription and lengths of all genes, in addition to the distances between them [13,14]. This locus spans a region of 261 kb and is formed of 15 kallikrein genes with no intervening non-kallikreins in-between. Distances between adjacent genes range from

1.5 to 32.5 kb. All genes, except *KLK2* and *KLK3*, are transcribed in the same direction from telomere to centromere (Fig. 3A). This locus is bounded telomerically by the siglec family of genes [15] and centromerically by the testicular acid phosphatase gene (ACPT) [16]. In this study, we characterized another smaller serine protease cluster on the short arm of the same chromosome (Fig. 3A). The cluster is formed of four serine proteases (complement factor D, elastase 2,

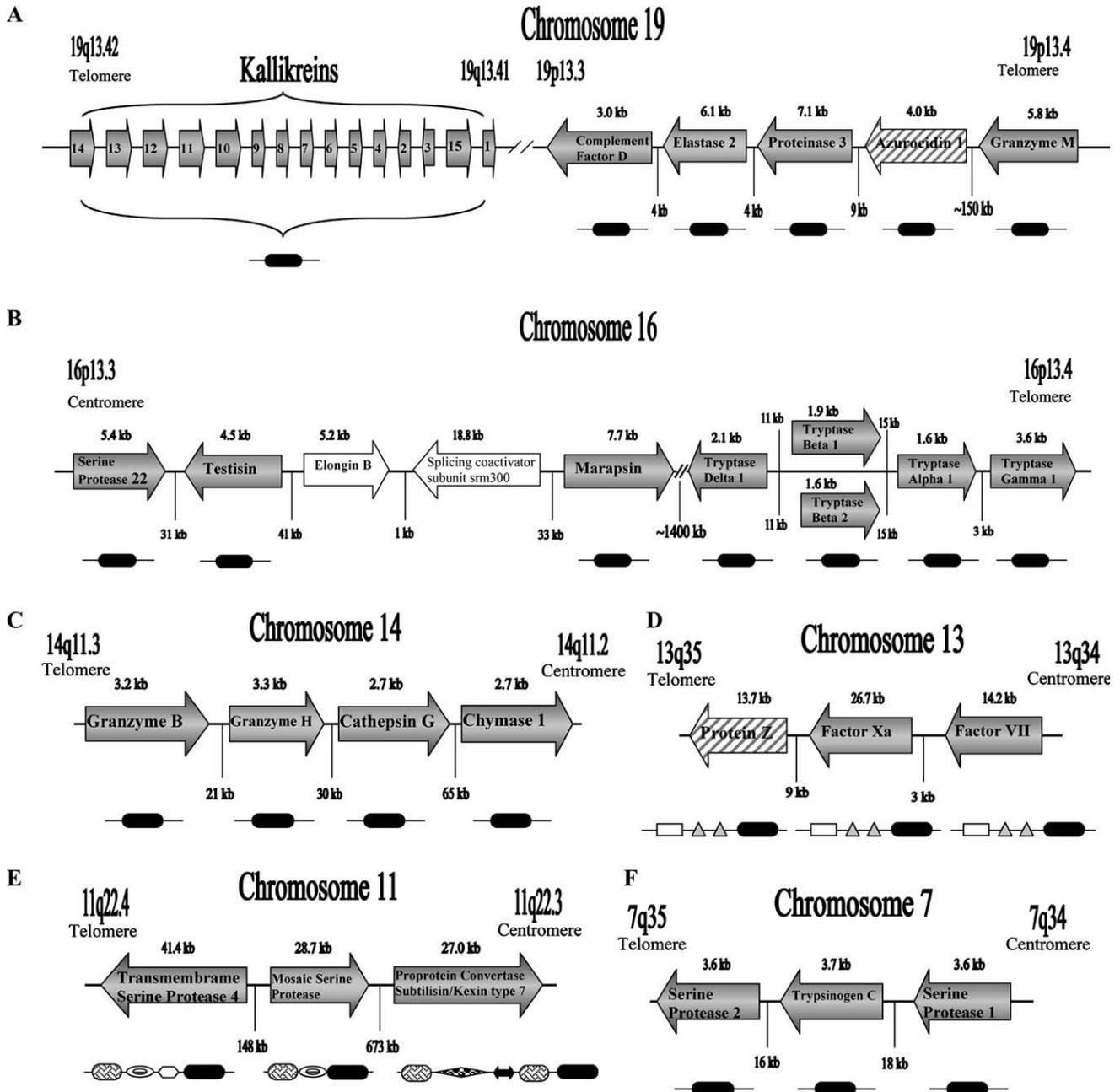


Fig. 3. Mapping of serine protease gene clusters. Genes are represented with arrows, indicating the direction of transcription. Gene lengths are indicated above each gene and distances between genes, rounded to the nearest kilobase, are indicated below. Characteristic functional domains are shown below each gene. Tryptase  $\beta 1$  and 2 are two alleles of the same gene. For domain symbols see Appendix 2. Figure is not drawn to scale. kb, kilobases;  $\square$ , proteases;  $\square$ , non-peptidase homologues;  $\square$ , not proteases.

proteinase 3, and granzyme M in this same order, from centromere to telomere) in addition to a non-peptidase homologue (azurocidin 1). All genes are transcribed from telomere to centromere and again, the distances between genes range from 4 to 9 kb (except between granzyme M and azurocidin 1 which is about 150 kb) (Fig. 3A). In order to study the extent of each locus, we examined all identified genes on either side of each cluster, for sequence similarity with, or for the presence of, the characteristic trypsin domain of serine proteases. No other related genes could be found. The possibility still exists, however, for the presence of other, not as yet annotated genes on either site of this locus.

The tryptase cluster of serine proteases is located on the short arm of chromosome 16 (16p13.4) with all genes transcribed from centromere to telomere (except tryptase delta 1 which runs on the opposite direction) (Fig. 3B). Genes in this locus are tightly clustered (distances between genes are as short as 3 kb). This cluster is followed centromerically by another cluster of three proteases on 16p13.3. This cluster is, however, interrupted by two non-protease genes. Analysis of genes on both ends of each of these clusters did not reveal any sequence similarity to serine proteases.

We also mapped the four serine proteases that form a cluster on chromosome 14q11.2–q11.3 and found that all genes are transcribed from telomere to centromere. Genes of this cluster are generally smaller (2.7–3.3 kb) and more spaced, with distances of 21–65 kb between adjacent genes. No other serine proteases were identified on either end (Fig. 3C). Mapping of the three serine proteases that form a cluster on chromosome 13 is shown in Fig. 3D, which also shows the direction of transcription and distances between genes. Genes of this cluster are relatively larger (13–27 kb). Chromosome 11 contains another cluster of widely spaced serine proteases (Fig. 3E), two of them run centromerically and the third is in the opposite strand. The three trypsin cluster members on chromosome 7 are shown in Fig. 3F, with all genes being of comparable sizes and transcribed in the same direction (centromere to telomere).

#### *Functional domains of the human serine proteases*

We scanned the amino acid sequence of human serine proteases using the “ProfileScan” algorithm. In addition to the characteristic trypsin domain, we identified 32 other domains. A full list of all domains is included in Appendix 2 and the domains identified (verified and/or predicted) are shown besides each protein in the phylogenetic tree presented in Fig. 4 and under each gene of the map in Fig. 3. The presence of a wide variety of domains with different functions reflects the functional diversity among serine proteases.

Fig. 3 shows that members of each cluster do not usually have the same panel of domains apart from the common trypsin domain, and most of the domains were randomly distributed. An exception is the three members of the chromosome 13 cluster which all have two EGF domains and a GLA region (Fig. 3D). Interestingly, some functionally (as indicated by the presence of conserved domains) and evolutionarily (as indicated by clustering in same sister groups in phylogenesis) related serine proteases, e.g., the proprotein convertases, are distributed along four different chromosomes, suggesting the involvement of a different evolutionary mechanism.

#### *Evolutionary history*

A representative phylogenetic tree of selected serine proteases is presented in Fig. 4. The characteristic functional domains for each enzyme are shown. As previously reported, phylogenetic trees appear to be governed by surface-exposed residues that control substrate and modulatory ligand recognition [17]. The tree presented here agrees with a previous attempt to segregate proteases into functional groups based on phylogenetic analysis [18].

The tight clustering, size homogeneity, sequence similarity, and uniform direction of transcription support the idea that genes of each cluster arose from a common ancestor by gene duplication and exon shuffling [19,20]. The proposed mechanism involves unequal crossing-over of sister chromatids during meiotic recombination. The ongoing genome projects in different animal species will be very helpful in establishing the phylogenetic relatedness of serine proteases and their model of evolution.

In the near future, the full sequencing of the genomes of other species will allow for a more precise analysis of the evolutionary history of serine proteases among different species. Recently, Olsson and Lundwall [21] reported the presence of mouse orthologues to human kallikreins hK1 and hK4–hK15. Interestingly, these genes also occur in a cluster and are arranged in the same order as their human counterparts.

An interesting observation is that some members of certain clusters were sometimes evolutionarily unrelated. An example is the kallikrein gene locus which represents the largest cluster of serine proteases within the human genome. Many trees grouped human kallikrein 6 (in addition to kallikrein 14 in some cases) in distinct groups, away from the kallikrein group (data not shown). The main purpose of the phylogenetic analysis in this study was to examine the possible evolutionary relationship within and in-between the different clusters. Several other evolutionary trees for individual clans and families of serine proteases can be found elsewhere [22].

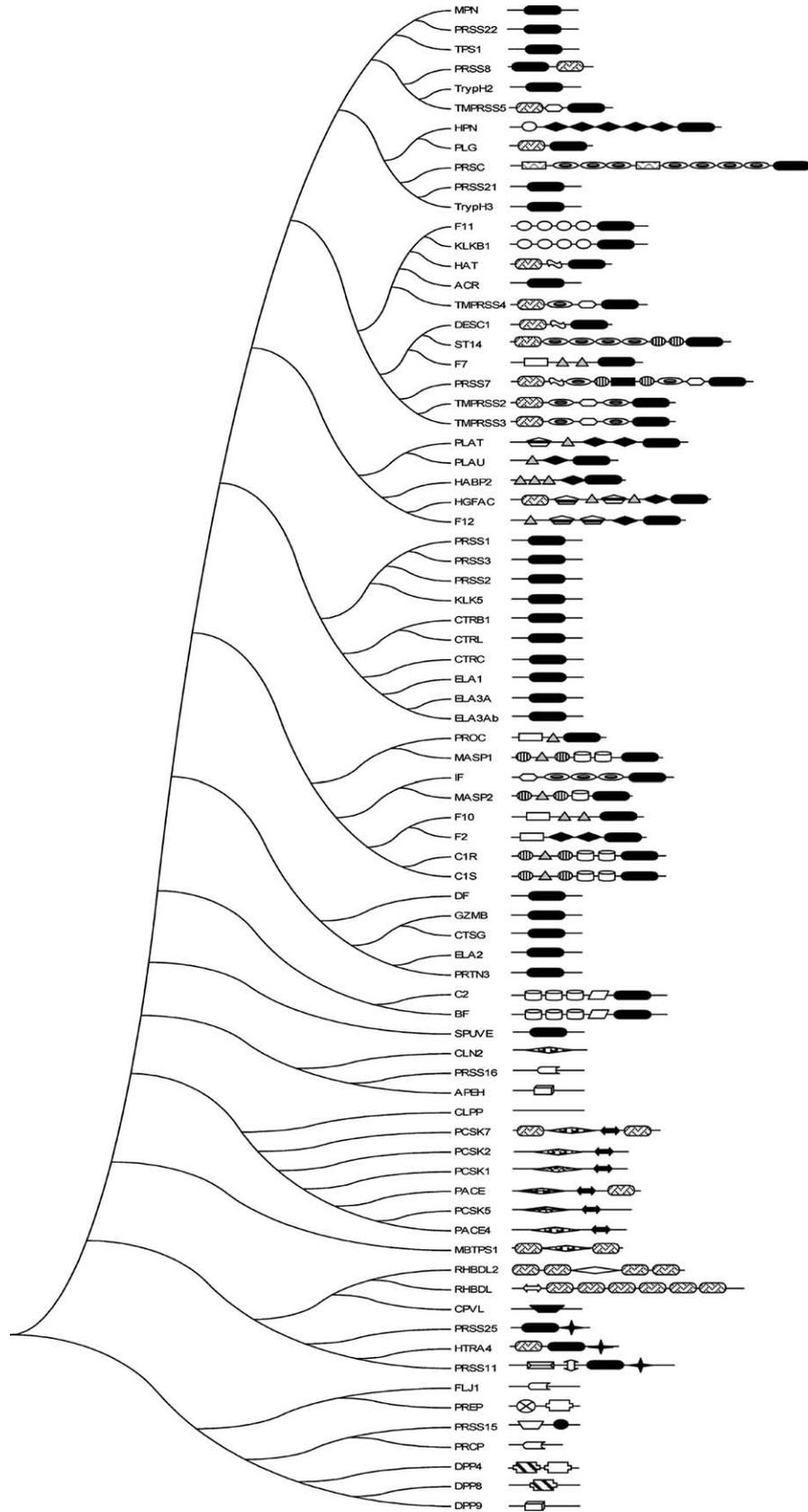


Fig. 4. An example of an evolutionary tree for selected serine proteases. The tree was constructed by the “Minimum Evolution” method using the ‘Number of Differences’ distance. Characteristic domains are shown beside each gene. For full gene and domain names, see Appendices 1 and 2. For simplicity, only representative members of some families (e.g., kallikreins and granzymes) were included in the analysis.

### Transmembrane proteases

Most serine proteases are secreted proteins. Seven members of a small class of membrane-bound serine proteases, called Type II Transmembrane Serine Proteases (TTSPs), were recently characterized [23]. In addition to the proteolytic domain, transmembrane proteases have a transmembrane domain and stem cytoplasmic domains that serve protein/protein or protein/ligand interactions. Our sequence analysis of all human serine proteases indicated the presence of putative transmembrane domains in other SP. Most of the seven characterized transmembrane proteases were found to cluster on two chromosomes. TMPRSS4, TMPRSS5, and ST14 are located on the long arm of chromosome 11. Of note is the presence of the MSP gene adjacent to TMPRSS4, a splice variant of which contains a transmembrane domain [24]. Distal to MSP is the PCSK7 gene, a type I transmembrane protease. The second cluster of transmembrane proteases is located on the long arm of chromosome 21, and is composed of PRSS7, TMPRSS2, and TMPRSS3, in order, from centromere to telomere.

### Discussion

In the current study, we provide an overview of serine proteases in the human genome, with detailed mapping of the main serine protease clusters. With the near-completion of the human genome, our data should be fairly accurate. It should be noted, however, that our figures did not take into account the presence of splice variants; several of them had been reported for serine proteases [12,25–27] and pseudogenes [12].

An important observation is that most of the genes in each cluster are transcribed in the same direction (e.g., 13 of the kallikrein genes on chromosome 19, three tryptase genes on chromosome 16, and the four-gene cluster on chromosome 14). This points to the possibility of the existence of a common mechanism that controls expression of all genes in a cluster, e.g., a “locus control region.” Also, the presence of tightly clustered genes draws attention to the possibility of involvement of these proteins in a cascade-like enzymatic pathway.

Interactions between serine proteases are common, and substrates of serine proteases are sometimes other serine proteases that are activated from an inactive precursor (zymogen) [2]. The involvement of serine proteases in cascade pathways is well documented. One important example is the blood coagulation cascade. Blood clots are formed by a series of zymogen activations. In this enzymatic cascade, the activated form of one factor catalyzes the activation of the next factor. Very small amounts of the initial factors are sufficient to

trigger the cascade because of the catalytic nature of the process. These numerous steps yield a large amplification, thus ensuring a *rapid* and *amplified* response to trauma [28]. A similar mechanism is involved in the dissolution of blood clots where activation of plasminogen activators leads to conversion of plasminogen to plasmin which is responsible for lysis of the fibrin clot. A third important example of the coordinated action of serine proteases is that of the intestinal digestive enzymes; coordinated control is achieved by the action of trypsin as the common activator of all pancreatic zymogens; trypsinogen, chymotrypsinogen, proelastase, and procarboxypeptidase. The apoptosis pathway is another important example of the coordinated action of proteases. More recently, a cascade mechanism is hypothesized for kallikrein involvement in cancer and inflammation [29,30]. The possible involvement of members of each cluster in such pathways warrants further investigations.

Clustering of two or more serine proteases might have a functional relevance. For instance, factors VII and X are both involved in the extrinsic coagulation pathway and factor X is known to have a positive feedback effect on factor VII; both genes are located next to each other on chromosome 13q34. Three other serine proteases expressed exclusively in hematopoietic cells [31] (cathepsin G, granzyme G and H) are also clustered on chromosome 14 (Fig. 3C). In addition, at least 7 adjacent members of the human kallikrein gene cluster (KLK5, 6, 7, 8, 10, 11, and 14) were found to be overexpressed in ovarian cancer [32–36] and the same four centromeric kallikreins (KLK15, 3, 2, and 4) were all shown to be highly expressed in the prostate and be up-regulated by androgens [20,37–39].

Upon completion of the Human Genome Project, research will be directed towards identification of all serine proteases and protease-like molecules, their functional characterization, as well as their regulation by specific and generalized inhibitors. The understanding of the physiology and pathobiology of serine proteases may lead to the utilization of these enzymes as biomarkers as well as therapeutic targets for various human diseases.

### References

- [1] N.D. Rawlings, A.J. Barrett, Evolutionary families of peptidases, *Biochem. J.* 290 (1993) 205–218.
- [2] R.M. Schultz, M.N. Liebman, Structure–function relationship in protein families, in: T.M. Devlin (Ed.), *Textbook of Biochemistry with Clinical Correlations*, fourth ed., Wiley-Liss, New York, 1997, pp. 1–2,116.
- [3] W.H. Horl, Proteinases: potential role in health and disease, in: M. Sandler, H.J. Smith (Eds.), *Design of Enzyme Inhibitors as Drugs*, Oxford University Press, Oxford, 1989, pp. 573–581.

- [4] B.R. Henderson, W.P. Tansey, S.M. Phillips, I.A. Ramshaw, R.F. Kefford, Transcriptional and posttranscriptional activation of urokinase plasminogen activator gene expression in metastatic tumor cells, *Cancer Res.* 52 (1992) 2489–2496.
- [5] C.J. Froelich, X. Zhang, J. Turbov, D. Hudig, U. Winkler, W.L. Hanna, Human granzyme B degrades aggrecan proteoglycan in matrix synthesized by chondrocytes, *J. Immunol.* 151 (1993) 7161–7171.
- [6] G.M. Yousef, E.P. Diamandis, Expanded human tissue kallikrein family—a novel panel of cancer biomarkers, *Tumour Biol.* 23 (2002) 185–192.
- [7] G.M. Yousef, E.P. Diamandis, The new human tissue kallikrein gene family: structure, function, and association to disease, *Endocr. Rev.* 22 (2001) 184–204.
- [8] E.P. Diamandis, G.M. Yousef, Human tissue kallikreins: a family of new cancer biomarkers, *Clin. Chem.* 48 (2002) 1198–1205.
- [9] L. Stryer, *Biochemistry*, fourth ed., W.H. Freeman, New York, 1995.
- [10] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [11] C. Lopez-Otin, C.M. Overall, Protease degradomics: a new challenge for proteomics, *Nat. Rev. Mol. Cell Biol.* 3 (2002) 509–519.
- [12] C. Southan, A genomic perspective on human proteases, *FEBS Lett.* 498 (2001) 214–218.
- [13] G.M. Yousef, E.P. Diamandis, Human kallikreins: common structural features, sequence analysis and evolution, *Curr. Genom.* 4 (2003) 147–165.
- [14] G.M. Yousef, A. Chang, A. Scorilas, E.P. Diamandis, Genomic organization of the human kallikrein gene family on chromosome 19q13.3–q13.4, *Biochem. Biophys. Res. Commun.* 276 (2000) 125–133.
- [15] G.M. Yousef, M. Ordon, G. Foussias, E.P. Diamandis, Genomic organization of the Siglec gene locus on chromosome 19q13.4 and cloning of two new siglec pseudogenes, *Gene* 286 (2002) 259–270.
- [16] G.M. Yousef, M. Diamandis, K. Jung, E.P. Diamandis, Molecular cloning of a novel human acid phosphatase gene (ACPT) that is highly expressed in the testis, *Genomics* 74 (2001) 385–395.
- [17] M.M. Krem, E. Di Cera, Molecular markers of serine protease evolution, *EMBO J.* 20 (2001) 3036–3045.
- [18] M.M. Krem, T. Rose, E. Di Cera, The C-terminal sequence encodes function in serine proteases, *J. Biol. Chem.* 274 (1999) 28063–28066.
- [19] G.M. Yousef, E.P. Diamandis, The expanded human kallikrein gene family: locus characterization and molecular cloning of a new member, KLK-L3 (KLK9), *Genomics* 65 (2000) 184–194.
- [20] P.S. Nelson, L. Gan, C. Ferguson, P. Moss, R. Gelinas, L. Hood, K. Wang, Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression, *Proc. Natl. Acad. Sci. USA* 96 (1999) 3114–3119.
- [21] A.Y. Olsson, A. Lundwall, Organization and evolution of the glandular kallikrein locus in *Mus musculus*, *Biochem. Biophys. Res. Commun.* 299 (2002) 305–311.
- [22] A.J. Barrett, N.D. Rawlings, J.F. Woessner, *Handbook of Proteolytic Enzymes*, Academic Press, London, 1998.
- [23] J.D. Hooper, J.A. Clements, J.P. Quigley, T.M. Antalis, Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes, *J. Biol. Chem.* 276 (2001) 857–860.
- [24] D.R. Kim, S. Sharmin, M. Inoue, H. Kido, Cloning and expression of novel mosaic serine proteases with and without a transmembrane domain from human lung, *Biochim. Biophys. Acta* 1518 (2001) 204–209.
- [25] G.M. Yousef, A. Chang, E.P. Diamandis, Identification and characterization of KLK-L4, a new kallikrein-like gene that appears to be down-regulated in breast cancer tissues, *J. Biol. Chem.* 275 (2000) 11891–11898.
- [26] G.M. Yousef, A. Scorilas, K. Jung, L.K. Ashworth, E.P. Diamandis, Molecular cloning of the human kallikrein 15 gene (KLK15). Up-regulation in prostate cancer, *J. Biol. Chem.* 276 (2001) 53–61.
- [27] A. Magklara, A. Scorilas, D. Katsaros, M. Massobrio, G.M. Yousef, S. Fracchioli, S. Danese, E.P. Diamandis, The human KLK8 (neuropsin/ovasin) gene: identification of two novel splice variants and its prognostic value in ovarian cancer, *Clin. Cancer Res.* 7 (2001) 806–811.
- [28] D.U. Silverthorn, *Human Physiology. An Integrated Approach*, Prentice-Hall, New Jersey, 1995.
- [29] G.M. Yousef, E.P. Diamandis, Human tissue kallikreins: a new enzymatic cascade pathway?, *Biol. Chem.* 383 (2002) 1045–1057.
- [30] K. Bhoola, R. Ramsaroop, J. Plendl, B. Cassim, Z. Dlamini, S. Naicker, Kallikrein and kinin receptor expression in inflammation and cancer, *Biol. Chem.* 382 (2001) 77–89.
- [31] R.D. Hanson, P.A. Hohn, N.C. Popescu, T.J. Ley, A cluster of hematopoietic serine protease genes is found on the same chromosomal band as the human alpha/delta T-cell receptor locus, *Proc. Natl. Acad. Sci. USA* 87 (1990) 960–963.
- [32] H. Kim, A. Scorilas, D. Katsaros, G.M. Yousef, M. Massobrio, S. Fracchioli, R. Piccinno, G. Gordini, E.P. Diamandis, Human kallikrein gene 5 (KLK5) expression is an indicator of poor prognosis in ovarian cancer, *Br. J. Cancer* 84 (2001) 643–650.
- [33] E.P. Diamandis, A. Scorilas, S. Fracchioli, M. Van Gramberen, H. de Bruijn, A. Henrik, A. Soosaipillai, L. Grass, G.M. Yousef, U.-H. Stenman, M. Massobrio, A.G. Zee, I. Vergote, D. Katsaros, Human kallikrein 6 (hK6): a new potential serum biomarker for diagnosis and prognosis of ovarian carcinoma, *J. Clin. Oncol.* 21 (2003) 1035–1043.
- [34] G.M. Yousef, E.P. Diamandis, Kallikreins, steroid hormones and ovarian cancer: is there a link?, *Minerva Endocrinol.* 27 (2002) 157–166.
- [35] L. Luo, P. Bunting, A. Scorilas, E.P. Diamandis, Human kallikrein 10: a novel tumor marker for ovarian carcinoma?, *Clin. Chim. Acta* 306 (2001) 111–118.
- [36] E.P. Diamandis, A. Okui, S. Mitsui, L.Y. Luo, A. Soosaipillai, L. Grass, T. Nakamura, D.J. Howarth, N. Yamaguchi, Human kallikrein 11: a new biomarker of prostate and ovarian carcinoma, *Cancer Res.* 62 (2002) 295–300.
- [37] G.M. Yousef, A. Scorilas, A. Magklara, N. Memari, R. Ponzone, P. Sisoni, N. Biglia, M.A. Said, E.P. Diamandis, The androgen-regulated gene human kallikrein 15 (KLK15) is an independent and favorable prognostic marker for breast cancer, *Br. J. Cancer.* 87 (2002) 1294–1300.
- [38] H.G. Rittenhouse, J.A. Finlay, S.D. Mikolajczyk, A.W. Partin, Human kallikrein 2 (hK2) and prostate-specific antigen (PSA): two closely related, but distinct, kallikreins in the prostate, *Crit. Rev. Clin. Lab. Sci.* 35 (1998) 275–368.
- [39] G.M. Yousef, C.V. Obiezu, L.Y. Luo, M.H. Black, E.P. Diamandis, Prostase/KLK-L1 is a new member of the human kallikrein gene family, is expressed in prostate and breast tissues, and is hormonally regulated, *Cancer Res.* 59 (1999) 4252–4256.

## Appendix 1. Legend of Protein Symbols Used

<b>ACR:</b> Acrosin	<b>HAT:</b> Human airway trypsin-like enzyme	<b>PLAU:</b> u-Plasminogen activator
<b>APEH:</b> Acylaminoacyl-peptidase	<b>HGF:</b> Hepatocyte growth factor	<b>PLG:</b> Plasmin
<b>AZU1:</b> Azurocidin	<b>HGFAC:</b> Hepatocyte growth factor activator	<b>PPGB:</b> Lysosomal carboxypeptidase A
<b>BF:</b> Complement factor B	<b>HP:</b> Haptoglobin-1	<b>PRCP:</b> Lysosomal Pro-X carboxypeptidase
<b>C1R:</b> Complement component C1r (activated)	<b>HPN:</b> Hepsin	<b>PREP:</b> Prolyl oligopeptidase
<b>C1S:</b> Complement component C1s (activated)	<b>HPR:</b> Haptoglobin-related protein	<b>PROC:</b> Protein C (activated)
<b>C2:</b> Complement component 2	<b>HTRA3:</b> Serine protease HTRA3	<b>PROZ:</b> Protein Z
<b>CLN2:</b> Tripeptidyl peptidase I	<b>HTRA4:</b> Serine protease HTRA4	<b>PRSS1:</b> Cationic trypsin
<b>CLPP:</b> Endopeptidase Clp	<b>IF:</b> Complement factor I	<b>PRSS2:</b> Anionic trypsin
<b>CMA1:</b> Chymase	<b>IMMP2L:</b> Mitochondrial inner membrane protease 2	<b>PRSS3:</b> Mesotrypsin
<b>CPVL:</b> Vitellogenic carboxypeptidase-like protein	<b>KLK(x):</b> Human kallikrein (x)	<b>PRSS7:</b> Enteropeptidase
<b>CTRB1:</b> Chymotrypsin B	<b>KLKB1:</b> Plasma kallikrein	<b>PRSS8:</b> Prostatin
<b>CTRC:</b> Chymotrypsin C	<b>LACTB:</b> Animal D-Ala-D-Ala carboxypeptidase homologue	<b>PRSS11:</b> Osteoblast serine protease
<b>CTRL:</b> Chymopasin	<b>LOC51032:</b> Pancreatic elastase II form B	<b>PRSS12:</b> Neurotrypsin
<b>CTSG:</b> Cathepsin G	<b>LOC136242:</b> LOC136242 gene product	<b>PRSS15:</b> PIM1 endopeptidase
<b>DESC1:</b> DESC1 protease	<b>MASP(x):</b> Mannose-binding-protein-associated serine protease (x)	<b>PRSS16:</b> Thymus-specific serine peptidase
<b>DF:</b> Complement factor D	<b>MBTPS1:</b> Site-1 protease	<b>PRSS21:</b> Testisin
<b>DPP4:</b> Dipeptidyl peptidase 4	<b>MEST:</b> MEST gene product	<b>PRSS22:</b> Brain serine protease 2
<b>DPP6:</b> Dipeptidyl peptidase-like protein DPP6	<b>MPN:</b> Marapsin	<b>PRSS25:</b> OMI/HtrA2 endopeptidase
<b>DPP8:</b> Dipeptidyl peptidase 8	<b>MSP:</b> Membrane-type mosaic serine protease	<b>PRTN3:</b> Myeloblastin
<b>ELA1:</b> Pancreatic elastase	<b>MST1:</b> Macrophage-stimulating protein	<b>RHBDL:</b> Rhomboid-like protein 1
<b>ELA2:</b> Neutrophil elastase	<b>MSTP9:</b> Hepatocyte growth factor-like protein homologue	<b>SPUVE:</b> Umbilical vein protease
<b>ELA2A:</b> Pancreatic elastase-like protease	<b>PACE:</b> Furin	<b>ST14:</b> Matriptase
<b>ELA3A:</b> Pancreatic endopeptidase E	<b>PACE4:</b> PACE4 proprotein convertase	<b>TMPRSS(x):</b> Transmembrane serine protease (x)
<b>ELA3B:</b> Pancreatic endopeptidase E form B	<b>PCCSK(x):</b> Proprotein convertase (x)	<b>TPP2:</b> Tripeptidyl peptidase II
<b>EPHX1:</b> Epoxide hydrolase	<b>PLA:</b> Apolipoprotein	<b>TPS(x):</b> Trypsin alpha (x)
<b>F(x):</b> Blood coagulation factor (x)	<b>PLAT:</b> t-Plasminogen activator	<b>TPSB(x):</b> Trypsin beta (x)
<b>FAP:</b> Fibroblast activating protein alpha subunit		<b>TPSD1:</b> Trypsin delta 1
<b>GZM(x):</b> Granzyme (x)		<b>TPSG1:</b> Trypsin gamma 1
<b>HABP2:</b> Plasma hyaluronan-binding serine protease		<b>TRY(x):</b> Trypsin pseudogene (x)
		<b>TSP50:</b> testis-specific protein tsp50

## Appendix 2. Legend of functional domains identified in serine proteases

 Trypsin domain profile	 Calcium binding EGF-like domain signature	 Kringle
 Fibronectin	 Apple	 SRCR
 Frizzled	 LDLRA	 Sea Domain
 Cub Domain	 MAM domain signature:	 VWFA domain profile:
 PDZ domain profile	 CA-binding hand motif	 Transmembrane Domain
 Sushi domain	 Rhomboid domain	 DPP4 N-terminal sequence:
 GLA region	 Subtilase domain	 Proprotein convertase P domain
 B-propeller domain	 $\alpha/\beta$ hydrolase fold	 Signal peptidase
 Serine carboxypeptidase	 Phospholipase/carboxylesterase:	 IGFBP domain
 Kazal domain	 AAA	 Prolyl oligopeptidase
 Prolyl endopeptidase family serine active site		
 Atp-dependant serine protease, lon family, serine active site		
 Esterase/thioesterase/lipase active site serine		