REVIEW ARTICLE

Evolutionary families of peptidase inhibitors

Neil D. RAWLINGS[1], Dominic P. TOLLE and Alan J. BARRETT

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, U.K.

The proteins that inhibit peptidases are of great importance in medicine and biotechnology, but there has never been a comprehensive system of classification for them. Some of the terminology currently in use is potentially confusing. In the hope of facilitating the exchange, storage and retrieval of information about this important group of proteins, we now describe a system wherein the inhibitor units of the peptidase inhibitors are assigned to 48 families on the basis of similarities detectable at the level of amino acid sequence. Then, on the basis of three-dimensional structures, 31 of the families are assigned to 26 clans. A simple system of nomenclature is introduced for reference to each clan, family and inhibitor. We briefly discuss the specificities and mechanisms of the interactions of the inhibitors in the various families with their target enzymes. The system of families and clans of inhibitors described has been implemented in the MEROPS peptidase database (http://merops.sanger.ac.uk/), and this will provide a mechanism for updating it as new information becomes available.

Key words: clan, compound inhibitor, domain repeat, MEROPS database, peptidase inhibitor, protease inhibitor.

INTRODUCTION

Proteolytic enzymes (best termed peptidases) are essential for the survival of all kinds of organisms, and are encoded for by approx. 2 % of all genes [4]. Despite their life-giving functions, enzymes that break down proteins are potentially very damaging in living systems, so their activities need to be kept strictly under control. Several distinct mechanisms exist for the control of excessive peptidase activity, important amongst which are the interactions of the enzymes with proteins that inhibit them. These proteins are the subject of the present review.

We have adopted a broad definition of 'peptidase inhibitor' in deciding what to include here. We think it likely that all of the proteins considered have the potential to attenuate the activities of peptidases both *in vitro* and *in vivo* by the formation of complexes with the enzymes, but we have not been able to apply any quantitative criterion to this assessment. Valuable proposals have been made as to how one can assess the physiological relevance of an inhibitor [5], but the appropriate data have often not been provided when an inhibitor was described.

The scientific study of the peptidase inhibitors is nearly as old as that of the peptidases themselves. Hundreds of protein inhibitors of peptidases are now known and they are the subjects of thousands of research communications. The research is driven by the many potential applications of knowledge about the inhibitors in medicine, agriculture and biotechnology. At the most fundamental level, an understanding of the mode of interaction of protein inhibitors with enzymes may suggest novel approaches to the design of synthetic inhibitors for use as drugs. Many naturally occurring inhibitors, such as the anticoagulant hirudin, are being used as the basis of engineered proteins for injection in their own right [6]. There are a number of inherited diseases that are attributable to abnormalities in peptidase inhibitors. These include forms of emphysema, epilepsy, hereditary angioneurotic oedema and Netherton syndrome [7–10]. Some such diseases may be susceptible to treatment with the inhibitors administered as drugs, with synthetic inhibitors that take over their function, or with the natural inhibitors made available by gene therapy. Excessive proteolytic activities may well contribute to a number of disease conditions and, again, gene therapy to introduce inhibitors is under consideration [11,12]. In agriculture, genetically modified crop plants expressing inhibitors of the digestive enzymes of their insect pests are already under study [13,14].

This active field of research generates a rapid flow of information, but the storage and retrieval of all the new information that is being obtained about the peptidase inhibitors are handicapped by difficulties of nomenclature. In what was perhaps the most significant review that has been written on the peptidase inhibitors, Laskowski and Kato [15] deplored the confusion of nomenclature that existed in the field in 1980. They pointed out that inhibitors are commonly discovered by their activity against readily available enzymes, most commonly trypsin, chymotrypsin or subtilisin, and then are named after the source organism or tissue, as '*Streptomyces* subtilisin inhibitor' or 'pancreatic trypsin inhibitor'. Such names give no clue to the relationships of the inhibitors, and make it difficult to know whether information that is available about the mechanism of action of one inhibitor can correctly be applied to another. It was evident to Laskowski and Kato [15] that peptidase inhibitors could best be classified in their homologous families, but the sequence information then available allowed only about a dozen families to be recognized.

The names used for peptidase inhibitors have not improved since 1980, but there is now a wealth of sequence data for these proteins, and the time seems right to make a new attempt at a systematic classification of them. In the text below, we first describe the methods we have used to classify the inhibitors in families and clans, and then describe and discuss the results we obtained.

Databases referred to in the text are: MEROPS (http://merops.sanger.ac.uk/) [1]; Entrez (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi); Pfam [2] (http://www.sanger.ac.uk/Software/Pfam/); and SCOP [3] (http://scop.mrc-lmb.cam.ac.uk/scop/).

[1] To whom correspondence should be addressed (e-mail ndr@sanger.ac.uk).

METHODS

By way of preamble, we must mention that a major obstacle to the classification of peptidase inhibitors by their amino acid sequences is the fact that many of the proteins contain multiple homologous inhibitor domains in a single polypeptide chain. These domains are not identical to each other, but are functional inhibitors in their own right. The varying number of these repeats makes it impossible to assemble meaningful alignments of the entire amino acid sequences of the inhibitors with which to obtain measures of sequence similarity, as can be done for most other proteins. Because of this, the system of classification that we now propose is, strictly speaking, one for the inhibitory domains or 'inhibitor units' as we term them here, and not for the whole protein. It is only with the inhibitor units that one can quantify the structural similarities in the way that is necessary for the construction of dendrograms. That said, the great majority of inhibitors contain units from only a single family, and for these it is legitimate, for some purposes, to think of the whole protein as belonging to the given family.

Assembling peptidase inhibitor families

From the biochemical literature we assembled a list of known peptidase inhibitor proteins. The amino acid sequences of those containing 14 or more amino acid residues were obtained from the SWISS-PROT database or its supplement TrEMBL [16], downloaded in FastA format [17] and copied into a library file. The program FASTA [17] (with the default BLOSUM50 matrix) was used in an all-against-all search of the library of inhibitors for sets of homologues. Homologous proteins were taken to be those for which the aligned amino acid sequences were related with an E value of 0.001 or less. With sequences as short as 14 residues it is potentially difficult to obtain matches at this level of significance, and the sunflower cyclic inhibitor was included in family I12 despite a less significant match on the basis of the study by McBride et al. [18]. The sets of homologous proteins formed provisional, or draft, families, and a well-studied inhibitor in each draft family was chosen as the 'type-example' for the family. The type-example was the representative member of the family to which all other members must be shown directly or indirectly (see below) to be homologous. If the provisional type-example was found to be a protein that contained more than one inhibitor unit (see below), one of these units was selected as the type-example. An 'inhibitor unit' was defined as the segment of the amino acid sequence containing a single reactive site (or bait region, for a trapping inhibitor) after removal of any parts that are known not to be directly involved in the inhibitory activity. A protein that contained only a single inhibitor unit was termed a simple inhibitor, and one that contained multiple inhibitor units was termed a compound inhibitor.

Each type-example-inhibitor-unit sequence was used to search the Entrez non-redundant protein sequence database by use of the BLASTP program [19] with the default BLOSUM62 matrix. Each sequence retrieved was considered to be a homologue, and a member of the same family as the type-example, if the E value of the alignment was 0.001 or less and the alignment spanned the inhibitor unit. A family was permitted to contain only a single member if no homologues were found. This process led to the identification of some proteins that were homologous to members of two draft families. Each such protein provided a link between the draft families and required them to be merged.

A number of well-known inhibitors did not show significant relationship directly to any type-example. For each of these, the sequence was run in a FASTA search against the entire collection of inhibitor sequences. When the result showed significant sequence relationship to one or more members of an inhibitor family other than the type-example, the query inhibitor was included in the family. A relationship such as this was described as a 'transitive relationship'.

Dendrograms

A dendrogram for each family was prepared as follows. An amino acid sequence alignment for the family was prepared by use of CLUSTAL W [20], and a distance matrix of percentage identities was calculated. The distance matrix was converted into accepted point mutations according to the method of Schwartz and Dayhoff [21]. The tree was constructed by use of the Fitch-Margoliash algorithm with contemporary tips as implemented in the KITCH program from the PHYLIP package [22].

Assembling of clans

We used the term 'clan' to designate a single evolutionary line of inhibitors defined by a single type of protein fold, very much as we did for peptidases [23]. A clan contains one or more complete families, since we assume that all members of a family have similar protein folds. We used definitions from the SCOP structural classification of proteins [24] to assemble clans of peptidase inhibitors. For these proteins, the superfamily level in the hierarchy of the SCOP database corresponds to our clan.

Terminology

Each clan of inhibitors was assigned an identifier (e.g. 'IA') formed from the letter 'I' or 'J' and an additional capital serial letter 'A' to 'Z'. Once the series 'IA' to 'IZ' had been completed, the next clan was named 'JA', and the names 'JB' to 'JZ' are available for future use. The identifier assigned to each family of inhibitor units was formed from the letter 'I' and a serial number, e.g. 'I25'. Many of the inhibitor units were assigned identifiers like 'I01.001', in which the first three characters represent the family identifier (padded with a zero when necessary) and the final three characters form a serial number. Inhibitors (or inhibitor units) that were found to have been characterized in some detail were assigned to the specific identifiers, and others were left as unassigned members of the given family. A compound inhibitor protein that contained several inhibitor units was termed a homotypic compound inhibitor, if the units were all from a single family, or a heterotypic compound inhibitor if they were from more than one family. The identifier for each homotypic compound inhibitor consisted of the letter 'L' followed by the family name of the inhibitor units (padded as necessary), a hyphen, and a three-digit serial number, e.g. 'LI01-001'. For a heterotypic compound inhibitor, no single family name could be used in the construction of the identifier, so it was constructed as 'LI9', followed by a one-digit serial number, a hyphen, and a three-digit serial number, e.g. 'LI90-001'.

RESULTS

Families of peptidase inhibitors

Our searches of the amino acid sequence databases led to the retrieval of 2500 sequences homologous to those of known peptidase inhibitors. From these, 48 families were built and were assigned identifiers (as described above). The families and the type-example for each are shown in Table 1, together with an indication of the families of peptidases that have been reported to be inhibited by proteins in the family.

Table 1 Families of peptidase inhibitors

The accession numbers, and the residue numbers for the inhibitor units, are from the SWISS-PROT or TrEMBL databases. Where available, we also include the family assignment of each type-example-inhibitor unit in the Pfam database. Under 'families of peptidases inhibited' we list the MEROPS identifiers of the peptidase families of which one or more members have been reported to be inhibited by a protein from the inhibitor family, together with illustrative references.

Family or subfamily	Common name	Type-example-inhibitor-unit name (source)	SWISS-PROT accession (inhibitor-unit range); Pfam accession	Families of peptidases inhibited
I1	Kazal	ovomucoid unit 3 (*Meleagris gallopavo*)	P01004 (135–185); PF00050	S1 [15]
I2	Kunitz (animal)	aprotinin (*Bos taurus*)	P00974 (36–93); PF00014	S1 [15]
I3A	Kunitz (plant)	soybean trypsin inhibitor (*Glycine max*)	P01070 (25–205); PF00197	Mainly S1 [15], but also C1 [25–27] and A1 [28]
I3B		proteinase inhibitor B (*Sagittaria sagittifolia*)	P07479 (25–181); PF00197	S1 [15]
I4	serpin	α_1-proteinase inhibitor (*Homo sapiens*)	P01009 (25–418); PF00079	Mainly S1 [29], but also S8 [30], C1 [31,32] and C14 [33]
I5	ascidian	ascidian trypsin inhibitor (*Halocynthia roretzi*)	P16589 (1–55)	S1 [34]
I6	cereal	ragi seed trypsin/α-amylase inhibitor (*Eleusine coracana*)	P01087 (1–122); PF00234	S1 [35]
I7	squash	trypsin inhibitor MCTI-1 (*Momordica charantia*)	P10294 (1–30); PF00299	S1 [36]
I8	Ascaris	nematode anticoagulant inhibitor (*Ascaris suum*)	P07851 (1–63); PF01826	S1 [37], but also M4 [38]
I9	YIB	protease B inhibitor (*Saccharomyces cerevisiae*)	P01095 (1–74)	S8 [39]
I10	marinostatin	marinostatin (*Alteromonas* sp.)	P29399 (1–14)	S1 [40]
I11	ecotin	ecotin (*Escherichia coli*)	P23827 (21–162); PF03974	S1 [41]
I12	Bowman-Birk	Bowman–Birk plant trypsin inhibitor (*Glycine max*) unit 1	P01055 (42–71); PF00228	Mainly S1 [42], but also C1 [43]
I13	pot 1	eglin C (*Hirudo medicinalis*)	P01051 (1–70); PF00280	Mainly S1 [44], but also S8 [45]
I14	hirudin	hirudin (*Hirudo medicinalis*)	P01050 (1–65); PF00713	S1 [46]
I15	antistasin	antistasin unit 1 (*Haementeria officinalis*)	P15358 (18–72); PF02822	S1 [47]
I16	SSI	subtilisin inhibitor (*Streptomyces albogriseolus*)	P01006 (32–144); PF00720	Mainly S8 [48], but also S1 [49] and M4 [50]
I17	elafin	mucus proteinase inhibitor unit 2 (*Homo sapiens*)	P03973 (26–83); PF00095	S1 [51]
I18	mustard	mustard trypsin inhibitor (*Sinapis alba*)	P26780 (31–93); PF05828	S1 [52]
I19	pacifastin	proteinase inhibitor LCMI I (*Locusta migratoria*)	P80060 (20–54)	S1 [53]
I20	pot 2	proteinase inhibitor II (*Solanum tuberosum*)	P01080 (28–86); PF02428	S1 [54]
I21	7B2	secretogranin V (*Homo sapiens*)	P05408 (27–212); PF05281	S8 [55]
I24	pinA	pinA endopeptidase La inhibitor (bacteriophage T4)	P07068 (1–161)	S16 [56]
I25A	cystatin 1	cystatin A (*Homo sapiens*)	P01040 (1–98); PF00031	C1 [57]
I25B	cystatin 2	ovocystatin (*Gallus gallus*)	P01038 (24–139); PF00031	Mainly C1 [58], but also C13 [59]
I25C	cystatin 3	metalloprotease inhibitor (*Bothrops jararaca*)	Q9DGI0 (28–141); PF00031	Not C1, but S8 [60], M12 [61]
I27	calpastatin	calpastatin unit 1 (*Homo sapiens*)	P20810 (170–222); PF00748	C2 [62]
I29	CTLA	cytotoxic T-lymphocyte antigen	P12399 (27–136)	C1 [63]
I31	thyropin	equistatin (*Actinia equina*)	P81439 (50–98); PF00086	C1 [64]
I32	IAP	BIRC-5 protein (*Homo sapiens*)	O15392 (1–142); PF00653	C14 [65]
I33	ascaris PI3	ascaris pepsin inhibitor PI-3 (*Ascaris suum*)	P19400 (21–169)	A1 [66]
I34	IA3	saccharopepsin inhibitor (*Saccharomyces cerevisiae*)	P01094 (1–68)	A1 [67]
I35	timp	timp-1 (*Homo sapiens*)	P01033 (24–207); PF00965	Mainly M10 [68], but also M12 [69]
I36	SMI	*Streptomyces* metalloproteinase inhibitor (*Streptomyces nigrescens*)	P01077 (30–131)	M4 [70]
I37	PCI	potato carboxypeptidase inhibitor (*Solanum tuberosum*)	P01075 (1–39)	M14 [46]
I38	aprin	metalloproteinase inhibitor (*Erwinia chrysanthemi*)	P18958 (20–120); PF02977	M10 [71]
I39	α_2M	α_2-macroglobulin (*Homo sapiens*)	P01023 (24–1474); PF00207	Numerous families including aspartic, cysteine, metallo and serine catalytic types [72]
I40	bombyx	*Bombyx* subtilisin inhibitor (*Bombyx mori*)	Q10731 (1–77)	S8 [73]
I42	chagasin	chagasin (*Leishmania major*)	Q9GY64 (1–179)	C1 [74]
I43	oprin	oprin (*Didelphis marsupialis*)	P82957 (83–291); PF00047	M12 [75]
I44	–	carboxypeptidase A inhibitor (*Ascaris suum*)	P19399 (1–65)	M14 [76]
I46	LCI	leech carboxypeptidase inhibitor (*Hirudo medicinalis*)	P81511 (16–81)	M14 [77]
I47	latexin	latexin (*Homo sapiens*)	Q9BS40 (1–222)	M14 [78]
I48	clitocypin	clitocypin (*Lepista nebularis*)	Q9P4A2 (1–150)	C1 [79]
I49	proSAAS	proSAAS (*Homo sapiens*)	Q9UHG2 (34–260)	S8 [80]
I50	p35	baculovirus p35 caspase inhibitor (Spodoptera litura nucleopolyhedrovirus)	O41275 (1–296); PF02331	C14 [81], but also C25 [82]
I51	IC	carboxypeptidase Y inhibitor (*Saccharomyces cerevisiae*)	P14306 (1–219); PF01161	S10 [83]
I52	TAP	tick anticoagulant peptide (*Ornithodorus moubata*)	P17726 (1–60)	S1 [84]
I57	–	staphostatin B (*Staphylococcus aureus*)	Q9EYW6 (1–109)	C47 [85]
I58	–	staphostatin A (*Staphylococcus aureus*)	Q99SX7 (1–108)	C47 [85]
I59	triabin	triabin (*Triatoma pallidipennis*)	Q27049 (19–160)	S1 [86]

The family for which the largest number of sequences was found is the serpin family, I4, with over 500 sequences. The serpins have been authoritatively reviewed by Silverman et al. [87]. Families I1, I2 and I25 each contain over 200 sequences, and the other families that have more than 100 members are I3, I12, I39 and I43. At the small end of the scale, families I5, I24, I34, I36, I40, I44, I46 and I58 are each represented by a single sequence at the time of writing. Some of the inhibitor units were very small: 14 residues each for marinostatin (MEROPS identifier I10.001) and sunflower cyclic inhibitor (I12.002), and units of less than 50 residues were found in families I7, I19, I31, I37 and I45.

It should be noted that many putative inhibitor units in some families may well not be functional inhibitors. They are homologous to known inhibitors, but the structural basis of inhibition is

not so well understood that we can predict with confidence that they are active inhibitors. For example, domains homologous to the family I1 (animal-type Kunitz) inhibitors occur in many multidomain proteins that are not known to be peptidase inhibitors. Again, the only known inhibitor in the large family I43 of immunoglobulin-like proteins is oprin, a snake venom metalloproteinase inhibitor from the Virginia opossum (*Didelphis marsupialis*) [75]. For these reasons, many uncharacterized homologues of inhibitors can be described only by the term 'unassigned inhibitor homologues'.

Subfamilies in two families

The dendrograms for families I3 and I25 (results not shown) show deep divergences that justify the recognition of subfamilies. Each subfamily had been a separate draft family until transitive relationships were revealed by the BLAST searches. In family I3, the arrowhead proteinase inhibitor B (SWISS-PROT P07479) in subfamily I3B was shown to be homologous to members of subfamily I3A other than the type-example (e.g. SWISS-PROT Q39488). Family I25, containing the cystatins, has three subfamilies. The cysteine peptidase inhibitors are in subfamilies I25A (the type-1 cystatins) and I25B (the type-2 cystatins). The sequence of soya phytocystatin (SWISS-PROT Q39842) provided a transitive link between subfamilies I25A and I25B (e.g. to Q852N1). Amongst the significant relationships linking subfamilies I25B to I25C were that of chicken ovocystatin (P01038) to the first cystatin-like unit of bovine H-kininogen (P01044, residues 19–135). The domains in subfamily I25C are mostly not inhibitors of peptidases, but include inhibitors of a subtilisin homologue in peptidase family S8 [60] and of a snake-venom metalloendopeptidase in family M12 [61].

Compound inhibitors

Inhibitor units in eleven families were found to occur in proteins that contain two or more inhibitor units and are therefore defined as compound inhibitors (Table 2). For example, turkey ovomucoid contains three inhibitor units from the Kazal family, I1 (MEROPS; I01.001, I01.002, I01.003). All but a handful of the compound inhibitors are homotypic, i.e. contain units only from a single family. The number of inhibitor domains in the compound inhibitors ranges from 2 to 15. The few heterotypic compound inhibitors we encountered contained units from families I2 and I17; these were the Red Sea turtle chelonianin, human eppin [88] and the human WFIKKN and WFIKKNRP proteins [89,90]. As can be seen in Figure 1, the inhibitor units from the two families are arranged differently in chelonianin as compared with the mammalian proteins.

Assignment of families to clans

It is sometimes clear from similarities in tertiary structure that members of two families are distantly related to each other, and this justifies grouping the families in a clan. Similarities in protein fold are apparent between members of families I1, I5, I8 and I20, and these families are grouped in clan IA (Table 3). Similarly, families I2 and I52 are grouped in clan IB, and families I7 and I37 in clan IE. For 23 other families, three-dimensional structures show no relationship, and each family is assigned to a separate clan. Fifteen families cannot be assigned to any clan because no tertiary structure is available. The large proportion of single-family clans contrasts with the situation for peptidases, where most clans contain more than one family.

DISCUSSION

We set out on the task of establishing a system for the classification of peptidase inhibitors, expecting to use methods very similar to those we have developed for peptidases. But we became aware that there are at least three important differences between peptidases and their inhibitors, such that the inhibitors require special treatment. The crucial differences can be summarized as follows. (1) An effective inhibitor unit may contain as few as 14 amino acid residues, whereas the peptidase units commonly contain approx. 200 residues. (2) Reactive-site residues in inhibitor units are often not conserved in the way that active-site residues are in peptidases, so it is seldom possible to tell from the sequence whether or not an inhibitor homologue is likely to have inhibitory activity. (3) Inhibitor units in a dozen families have duplicated and reduplicated during evolution so that modern-day proteins exist that contain multiple divergent copies of them. In contrast, multiple peptidase units within a single polypeptide occur in only a very few peptidases (amongst which angiotensin-converting enzyme and metallocarboxypeptidase D are best known).

The style of identifiers for inhibitor families

The families of peptidases are divided into five main groups according to the chemistry of the catalytic sites of the enzymes [92]. In contrast with this, we have proposed here a single series for the families of inhibitors. The possibility was considered that the families could be usefully classified according to the catalytic types of the peptidases inhibited, in sets of serine peptidase inhibitors, cysteine peptidase inhibitors, and so on. Such classification has often been used in the past, and has been authoritatively advocated [93]. There are difficulties with this approach, however. As is shown in Table 1, a number of families contain inhibitors of peptidases of more than one catalytic type. For example, the proteins in family I3 (the plant Kunitz-type inhibitors) generally inhibit serine peptidases of family S1, but also include inhibitors of cysteine peptidases (C1) and the aspartic peptidase cathepsin D (A1) (Table 1) as well as, possibly, subtilisins [93a]. Similarly, family I4, the serpin family, has been thought of as a family of serine peptidase inhibitors, but it also contains CrmA and other important inhibitors of cysteine peptidases (Table 1). We concluded that it would not be feasible to classify and name the families of inhibitors according to the catalytic types of the peptidases that they inhibited, and simply numbering them in one series.

Classification at the level of inhibitor units

In the Results section, and specifically in Table 2, we drew attention to the fact that many of the proteins that inhibit peptidases contain multiple inhibitory domains. This clearly raised an issue of policy for the classification; were we to attempt a classification of the entire multidomain proteins, or only of the individual inhibitor units? It was clear that only the isolated inhibitor units could be handled by methods similar to those we have used successfully for peptidases, and that these should be the primary objects in the classification. This decision committed us to a kind of classification of the peptidase inhibitors that has not been previously used, as far as we are aware.

Notwithstanding the decision to make inhibitor units the objects that would populate the inhibitor level of the hierarchical classification of peptidase inhibitors, there remained a need for consistent identifiers for the compound inhibitor proteins. These must indicate the classification of the inhibitor units contained in

Table 2 Families of inhibitor units that are found in compound inhibitors

The inhibitor units that occur in compound inhibitors belong to the twelve families listed here. The examples include proteins that are simple inhibitors (number of units = 1) and homotypic compound inhibitors, which contain units from only one family. Square brackets contain the SWISS-PROT identifiers and MEROPS identifiers.

Family	Number of inhibitor units	Example (source) [identifier]
Family I1	1	elastase inhibitor (*Anemonia sulcata*) [P16895; I01.108]
	2	bikazin (*Canis familiaris*) [P01002; LI01-003]; rhodniin (*Rhodnius prolixus*) [Q06684; LI01-007]
	3	ovomucoid (*Meleagris gallopavo*) [P01004; LI01-001]
	4	dipetalogastin (*Dipetalogaster maximus*) [O96790; LI01-006]
	5	agCP6264 protein (*Anopheles gambiae*) [EAA11963; putative]
	7	ovoinhibitor (*Gallus gallus*) [P10184; LI01-002]
	9	agrin (*Rattus norvegicus*) [P25304; putative]
	15	serine protease inhibitor Kazal type 5 [Q9NQ38; LI01-004]
Family I2	1	aprotinin (*Bos taurus*) [P00974; I02.001]
	2	α-1-microglobulin [P02760]; bikunin [P02760; LI02-001]; hepatocyte growth factor activator inhibitor 1 (all *Homo sapiens*) [O43278; LI02-004]
	3	tissue-factor-pathway-inhibitor-1 (*Homo sapiens*) [P10646; LI02-002]; K10D3.4 protein (*Caenorhabditis elegans*) [CAA99886; putative]
	5	ZC84.6 protein (*Caenorhabditis elegans*) [CAA79570; putative]
	8	C34F6.1 protein (*Caenorhabditis elegans*) [CAB03944; putative]; W01F3.3 protein (*Caenorhabditis elegans*) [CAB07294; putative]
	9	Y43F8B.3 protein (*Caenorhabditis elegans*) [CAA21511; putative]
	12	Ac-KPI-1 I inhibitor (*Ancylostoma caninum*) [AAN10061; LI02-008]
Family I8	1	*Ascaris* chymotrypsin/elastase inhibitor [P07851; I08.001]
	2	C09F9.2 protein (*Caenorhabditis elegans*) [CAB03861; putative]
	3	HmEGFL-1 protein (*Herdmania momus*) [AAB67704; putative]
Family I12	1	sunflower cyclic inhibitor (*Helianthus annuus*) [I12.002]
	2	Bowman–Birk inhibitor (*Gycine max*) [P01055; LI12-001]
	4	Bowman–Birk trypsin inhibitor (*Hordeum vulgare*) [P12940; LI12-002]
	5	Bowman–Birk trypsin inhibitor P0037C04.14 (*Oryza sativa*) [BAB55527; putative]
	6	Bowman–Birk trypsin inhibitor P0037C04.11 (*Oryza sativa*) [Q9ARS4; putative]
Family I15	1	hirustasin (*Hirudo medicinalis*) [P80302; I15.001]
	2	antistasin (*Haementeria officinalis*) [P15358; LI15-002]; ghilanten (*Haementeria ghilianii*) [P16242; LI15-003]
	6	(*Hydra magnipapillata*) [P38977; putative]
Family I17	1	elafin (*Homo sapiens*) [P19957; I17.002]
	2	mucus proteinase inhibitor (*Homo sapiens*) [P03973; LI17-001]
Family I19	1	proteinase inhibitor II (*Schistocerca gregaria*) [O46163; I19.001]
	2	proteinase inhibitor LCMI I (*Locusta migratoria*) [P80060; I19.001]
	3	pacifastin-related peptide precursor (*Schistocerca gregaria*) [Q8MYK3; putative]
	4	agCP6937 protein (*Anopheles gambiae*) [EAA05233; putative]
	9	pacifastin (*Pacifastacus leniusculus*) [P91776; LI19-001]
Family I20	1	proteinase inhibitor PTI (*Solanum tuberosum*) [P01079; putative]
	2	proteinase inhibitor II (*Solanum tuberosum*) [P01080; LI20-001]
	3	proteinase inhibitor II (*Nicotiana tabacum*) [Q40561; LI20-002]
	4	proteinase inhibitor (*Nicotiana alata*) [Q9SQ77; putative]
	6	proteinase inhibitor (*Nicotiana alata*) [Q40378; putative]
Family I25	1	cystatin A (*Homo sapiens*) [P01040; I25.001]
	2	metalloproteinase inhibitor (*Bothrops jararaca*) [Q9DGI0; I25.026]
	3	kininogen (*Homo sapiens*) [P01042; LI25-002]; multicystatin (*Helianthus annuus*) [BAA95416; LI25-006]
	8	multicystatin (*Solanum tuberosum*) [P37842; LI25-001]
Family I27	2	calpastatin Calp1 (*Xenopus laevis*) [CAB48419; putative]
	4	calpastatin (*Homo sapiens*) [P20810; LI27-001]
Family I29	1	cytotoxic T-lymphocyte antigen (*Mus musculus*) [P12399]
	4	salarin [91]
Family I31	1	MHC II invariant chain p41 form [P04233; I31.002]
	2	saxiphilin (*Rana catesbeiana*) [P31226; putative]
	3	equistatin (*Actinia equina*) [P81439; LI31-001]

the compound inhibitor as far as possible, but not risk confusion with identifiers that truly fit into the hierarchical system. We adopted identifiers like 'LI01-001' in which the initial 'L' and the hyphen as the fifth character are common to all the identifiers for compound inhibitors. In this example, 'I01' indicates that this is a homotypic compound inhibitor containing units of family I1, and '001' is simply the first in a sequence of serial numbers. The few heterotypic compound inhibitors we found all contain units from families I2 and I17, so there is no single family name that can be used to construct an identifier in exactly this way. We arbitrarily used 'I90' in place of the family identifier in constructing identifiers for the WFIKKN and WFIKKNRP

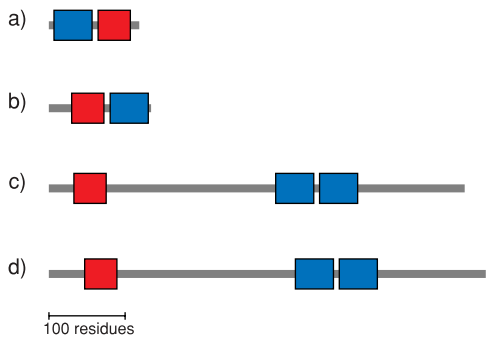

Figure 1 Diagrammatic structures of heterotypic compound inhibitors

Inhibitor units from family I2 are shown as blue blocks, and those from family I17 are red blocks. The proteins, represented approximately to scale, are (**a**) chelonianin (MEROPS identifier LI90-003, P00993, 110 residues, units I02.022 and I17.004), (**b**) eppin (LI90-004, O95925, 133 residues, I17 unassigned and I2 unassigned), (**c**) WFIKKN (LI90-001, Q96NZ8, 548 residues, I17 unassigned, I2 unassigned and I02.033) and (**d**) WFIKKNRP (LI90-002, Q8TEU8, 576 residues, I17 unassigned, I2 unassigned and I2 unassigned).

proteins as well as chelonianin and eppin, e.g. LI90-003 for chelonianin.

Disulphide-bond patterns

Many peptidase inhibitors contain disulphide bonds, and in some families these may stabilize reactive-site loops, facilitating the resynthesis of the reactive-site bond after it has been cleaved by the target enzyme. Historically these disulphides have been used to confirm assignments to families when sequence similarities were low, and to reveal the repetition of inhibitor units in compound inhibitors [15]. The arrangements of the disulphide bonds in the various clans and families are shown schematically in Figure 2. The conservation of disulphide-bonding patterns within clans is not striking, and we suggest that the many crystallographically determined structures now available demonstrate distant evolutionary relationships much more clearly.

Inhibitors and their homologues

Clearly the crucial property of all the proteins we are seeking to classify here is their inhibition of peptidases, and it is therefore appropriate to consider to what extent their grouping in families and clans reflects their inhibitory activities and mechanisms. First it has to be noted that there are many homologues of the known inhibitors for which there are no published experimental data as to their possible inhibitory activity. Nor are we able to predict such activity. For nearly all families of peptidases, conserved active-site residues are known that appear to be essential for catalysis. This means that a homologue of a known peptidase that contains all of these residues can reasonably be termed a putative peptidase, and any that lacks one or more of these residues can confidently be termed a non-peptidase homologue. In the MEROPS database the non-peptidase homologues are listed separately from the active, or potentially active, peptidases. But the situation is very different for the inhibitors. The reactive-site residues of the inhibitors are

Table 3 Clans of peptidase inhibitors and inhibitor units

Information about protein folds is available for 31 families and this allows them to be assigned to clans, although many clans of inhibitors contain only a single family. Fold descriptions are based on those from the SCOP database.

Clan	Type structure for clan (source; Protein Data Bank accession code)	Families included	Fold description for family
IA	ovomucoid domain 3 (*Coturnix coturnix*; 1OVO)	I1	disulphide-rich small proteins with α and β
		I5	disulphide-rich small proteins with α and β
		I8	disulphide-rich small proteins nearly all β
		I20	disulphide-rich small proteins with α and β
IB	aprotinin (*Bos taurus*; 4PTI)	I2	disulphide-rich small proteins with α and β
		I52	disulphide-rich small proteins with α and β
IC	soybean trypsin inhibitor (*Glycine max*; 1AVU)	I3	β-trefoil
ID	α_1-proteinase inhibitor (*Homo sapiens*; 1ATU)	I4	helix cluster with a β-sandwich
IE	trypsin inhibitor MCTI-1 (*Momordica charantia*; 1F2S chain I)	I7	knottin
		I37	knottin
IF	Bowman–Birk plant trypsin inhibitor (*Vigna angularis*; 1PI2)	I12	knottin
IG	eglin C (*Hirudo medicinalis*; 1EGP)	I13	α- and β-sandwich
IH	ovocystatin (*Gallus gallus*; 1CEW)	I25	α- and β-core; helix packs against coiled antiparallel β-sheet
IJ	ragi seed trypsin/α-amylase inhibitor (*Eleusine coracana*; 1B1U)	I6	all-helical (folded leaf)
IK	metalloproteinase inhibitor *Erwinia* (*Erwinia chrysanthemi*; 1SMP chain I)	I38	closed β-barrel
IL	α_2-macroglobulin (*Homo sapiens*) [108]	I39	
IM	hirudin (*Hirudo medicinalis*; 4HTC chain I)	I14	knottin
IN	ecotin (*Escherichia coli*; 1ECZ)	I11	β-sandwich
IO	antistasin inhibitor unit 1 (*Haementeria officinalis*)	I15	knottin
IP	elafin (*Homo sapiens*; 1FLE chain I)	I17	knottin
IQ	baculovirus p35 caspase inhibitor (*Spodoptera litura* nucleopolyhedrovirus; 1I3S)	I50	Greek key β-sandwich
IR	ascaris pepsin inhibitor PI-3 (*Ascaris suum*; 1F32)	I33	α and β duplication; meander β-sheets form barrel-like structure
IS	leech carboxypeptidase inhibitor (*Hirudo medicinalis*; 1DTD chain I)	I46	disulphide-rich small proteins with α and β
IT	TIMP-1 (*Homo sapiens*; 1UEA chain B)	I35	OB-fold
IU	*Streptomyces* metalloproteinase inhibitor (*Streptomyces nigrescens*; 1BHU)	I36	Greek key β-sandwich
IV	BIRC-5 protein (*Homo sapiens*; 1E31)	I32	metal-bound small proteins with α and β
IW	proteinase inhibitor LCMI I (*Locusta migratoria*; 1GL1)	I19	disulphide-rich small proteins all β
IX	MHC II invariant-chain p41 form (*Homo sapiens*; 1ICF)	I31	intertwined trimer of identical 3-helical subunits
IY	subtilisin inhibitor (*Streptomyces albogriseolus*; 3SSI)	I16	α- and β-sandwich
IZ	triabin (*Triatoma pallidipennis*; 1AVG)	I59	open or closed β-barrel (lipocalin-like)
JA	protease-A-inhibitor-3 (*Saccharomyces cerevisiae*; 1DPJ)	I34	non-globular all-α

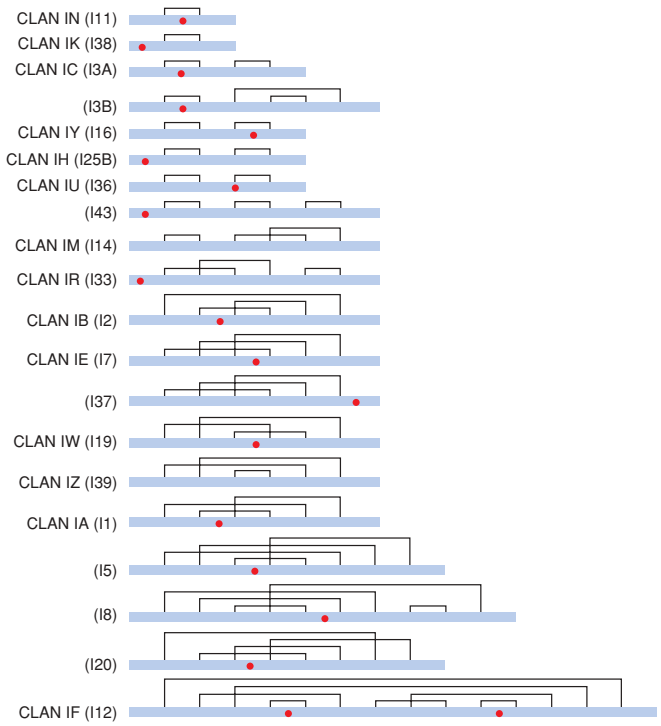

Figure 2 Patterns of disulphide bonds in type-examples from the inhibitor families

The distances between half-cystine residues in the polypeptide chains are not drawn to scale. Data are from the SWISS-PROT records cited in Table 1. The red circles indicate the positions of reported reactive-site residues, but do not exclude the possibility that there may be others elsewhere (because the disulphide-bonding pattern is not conserved between the two units in the example from family I12, both units are shown).

by no means strictly conserved, and indeed have been found to be amongst the most variable residues in several families [15,94]. Because of this, we are unable to distinguish non-inhibitory homologues from putative inhibitors simply by sequence, and can only describe a related protein that has not been demonstrated to have inhibitory activity as an 'unassigned inhibitor homologue'.

Mechanisms of inhibition

Having assigned the inhibitor units to families, one can ask to what extent the mechanisms of inhibition are constant within a family, and how it is that some families contain inhibitors of very different peptidases? Seven of the families and subfamilies in Table 1 (I3A, I4, I8, I12, I16, I25C, I39) can be seen to contain proteins that inhibit peptidases of more than one catalytic type, and four more (I13, I25B, I35, I50) contain inhibitors of peptidases from more than one family. This implies that either structurally dissimilar peptidases can be inhibited by a single mechanism, or structurally similar inhibitors inhibit by different mechanisms. The ways in which the inhibitors interact with their target enzymes vary enormously, but two general types of interaction can be recognized: irreversible 'trapping' reactions, and reversible tight-binding reactions.

1. Trapping reactions

The kind of interaction that depends most directly on the peptidase activity of the target enzyme is that which can be described as 'trapping'. This kind of reaction is specific for endopeptidases because it depends upon the cleavage of an internal peptide bond in the inhibitor that triggers a conformational change. Any catalytically inactive form of a peptidase, such as anhydrotrypsin, fails to enter into a trapping reaction, although it may well bind tightly to an inhibitor from one of the reversible classes. Trapping reactions are never truly reversible because unmodified inhibitor is not reformed, so the inhibitor can also be described as a suicide inhibitor. The three families that show trapping reactions are I4, I39 and I50. In these families there is some flexibility as to which peptide bond is cleaved to 'close the trap', and this broadens their inhibitory spectra. In families I4 and I50 the enzyme–inhibitor complex is normally covalent, and covalent complexes may also be formed by the macroglobulins in family I39.

In the large and widespread serpin (I4) family, the cleavage of an appropriate peptide bond in the reactive-site loop of the inhibitor triggers a dramatic conformational change, which is so rapid that catalysis proceeds only to the formation of an acyl enzyme and release of the C-terminal part of the reactive-site loop [29]. The N-terminal part of the loop inserts into a β-sheet, carrying the enzyme molecule still attached as the acyl enzyme to the opposite pole of the inhibitor molecule. This violent event disrupts the structure of the enzyme molecule and its catalytic site, so that hydrolysis of the acyl enzyme does not proceed and the covalent complex persists. The formation of an acyl enzyme is crucial to the serpin interaction, and the reaction is specific for the serine and cysteine peptidases that form acyl enzymes, whereas metallopeptidases may simply turn over the inhibitor without formation of any complex [95]. The reaction of serpins with cysteine endopeptidases of the papain family is apparently similar to that for serine peptidases [31]. Most serpins are inhibitors either of serine peptidases or cysteine peptidases, but not both. However, it has been shown that both antithrombin [32] and the mouse serpin SQN-5 are inhibitors of both serine and cysteine peptidases [31]. The use of non-identical but overlapping reactive sites in the serpin family is reviewed by Al-Khunaizi et al. [31], and the way in which the serpin architecture supports inhibition of cysteine peptidases has been systematically investigated by Irving et al. [96].

The second family of inhibitors that mediate trapping reactions is the macroglobulin family, I39. In these proteins, as illustrated by α_2-macroglobulin, cleavage of the inhibitor in the highly susceptible 'bait region' triggers the conformational change by which the target enzyme is trapped within the large re-folded inhibitor molecule [72]. The enzyme–inhibitor complex is stabilized largely by steric effects, although highly reactive thiol-ester groups can also form covalent links [97]. The target enzyme must be an endopeptidase to cleave the bait region of the inhibitor, and not too large a molecule to be enclosed by the macroglobulin, but any catalytic class of endopeptidase can be susceptible. Because there is no persistent interaction with the active site of the enzyme, the complexed enzyme molecule continues to hydrolyse accessible substrate molecules: these include artificial substrates and other small molecules, but not proteins. The exceptionally broad specificity of inhibition by α_2-macroglobulin, embracing endopeptidases of four catalytic types, is due in part to the fact that a great variety of bonds in the bait region of the molecule can be cleaved to trigger the trapping reaction [98].

Family I50 contains the baculovirus protein p35, which blocks the apoptosis of host cells by inhibition of the caspases. Cleavage of the caspase-sensitive bond in the reactive-site loop again leads to a conformational change that stabilizes the acyl-enzyme in an irreversible complex [81,99]. Protein p35 also inhibits gingipain K (family C25), and the cleavage site for inhibition of the gingipain is Lys-94, seven residues C-terminal to the caspase inhibitory site [82].

2. Reversible tight-binding interactions

Inhibitors in this group make high-affinity interactions with the active site of the target enzyme. The mechanism that has been studied in greatest detail is that termed the 'standard' mechanism by Laskowski and Qasim [100]. The inhibitory unit of a standard mechanism inhibitor has a single reactive-site peptide bond, and inhibition is caused by the binding of the inhibitor to the enzyme in a substrate-like fashion. The intact 'virgin' form of the inhibitor molecule exists in the complex in equilibrium with the 'modified' form of the inhibitor, in which the reactive-site peptide bond is cleaved, and the complex can dissociate to yield either virgin or modified inhibitor. The standard mechanism has been demonstrated conclusively only for inhibitors of serine peptidases. We recognize 19 families of such standard mechanism inhibitors. The reactive-site bonds occur in structurally similar reactive-site loops, and the conformation of the peptide containing the reactive site, which is the same in all of the inhibitors, is described as 'canonical' [46,100]. Crystallographic structures of enzyme–inhibitor complexes show that standard mechanism inhibitors are contained in families I1, I2, I3, I7, I8, I10, I11, I12, I13, I15, I16, I17, I18, I19, I20, I36 and I40. In addition, the structures of the uncomplexed ascidian trypsin inhibitor from *Halocynthia* (family I5; [101]) and the corn Hageman-factor inhibitor (family I6; [102]) suggest that they too probably inhibit by the standard mechanism. The 19 families of standard mechanism inhibitors fall into 13 different clans because of their different protein folds, and it therefore seems certain that the standard mechanism of peptidase inhibition has evolved on many separate occasions. Clearly, reactive-site loops that can adopt the canonical conformation, and thus inhibit serine endopeptidases by the standard mechanism, can be built on to a wide variety of supporting scaffold structures.

The families (or subfamilies) I3A, I8, I12 and I16 that generally inhibit serine peptidases by the standard mechanism each contain reversible inhibitors of peptidases of other catalytic types, but little is yet known about the structural basis for this. Intuitively, it seems likely that different reactive sites and mechanisms are responsible. However, the small Kunitz-type (I3) inhibitor of *Prosopis juliflora* has been reported to have overlapping inhibitory sites for trypsin and papain [27]. Also, there is evidence that in family I13 a single site is responsible for the inhibition of two very different types of serine peptidases, chymotrypsin (S1) and subtilisin (S8), by an inhibitor from wheat [103].

Inhibitors of thrombin have evolved in a wide variety of animals that feed on vertebrate blood and therefore need agents that can prevent it from clotting. The control of blood coagulation is also a medically important topic, so the natural inhibitors of thrombin have been the subjects of intense study. Hirudin and haemadin in family I14 are clearly not standard mechanism inhibitors. They enhance the specificity and affinity of their inhibition of thrombin by binding not only in the active site, but also to distant exosites [104,105]. The use of exosites is further extended by triabin (family I59) which interacts with thrombin exclusively via its fibrinogen-recognition exosite [86], and the complex that is inactive against fibrin retains activity on small-molecule substrates.

In the best-known family of cysteine peptidase inhibitors (I25), ovocystatin inhibits papain by binding to sites on either side of the active site, which becomes blocked. There is no interaction with the cysteine nucleophile but, instead, residues 8 and 9 interact with the S2 binding pocket in a substrate-like manner [5,58]. Remarkably, ovocystatin has a second independent inhibitory site for the non-papain-like cysteine peptidase, legumain (family C13, clan CD) [59]. The structure of the p41 Ii fragment (family I31) bound to cathepsin L [106] shows a three-loop arrangement of the p41 fragment reminiscent of the inhibitory edge of cystatins, but also suggests the reason why inhibitors based on the thyroglobulin type-1 fold inhibit much more selectively than the cystatins [5].

The inhibitors in family I29 are homologous to propeptides of several cathepsins including cathepsin L, and the crystal structures of the cathepsin proenzymes show how the inhibitors are likely to interact with the enzymes. The propeptides run through the active-site cleft in the reverse direction to substrates, blocking the active site and maintaining the catalytically inactive state of the proenzymes [5]. A further example of an inhibitor binding backwards in the active-site cleft of a cysteine peptidase is seen in XIAP (X-linked inhibitor of apoptosis protein) and its homologues in family I32. This tight-binding, reversible interaction has been reviewed by Stennicke et al. [99].

The crystal structure of staphostatin B (family I57), in complex with staphostatin, shows that the inhibitor runs through the active-site cleft in a substrate-like way, but the P1 Gly98 residue (conserved throughout the family) has a strained backbone conformation that would be sterically forbidden to any other residue [107].

Few protein inhibitors of aspartic peptidases are known, but *Ascaris* PI-3 (family I33) inhibits pepsin by blocking part of the substrate-binding site. The inhibitor molecule makes interactions with the pepsin 'flap' (subsites S1–S3) such that its three N-terminal residues occupy the S1′ to S3′ subsites, thus preventing substrates from binding [66]. Remarkably, the yeast IA3 inhibitor (I34) seems to lack secondary structure until it is bound to its natural target enzyme, saccharopepsin [67].

In family I35, the molecule of the tissue inhibitor of metallopeptidases, TIMP-1, is wedge-shaped, and the long edge occupies the active-site cleft of stromelysin-1 (family M10), binding either side of the catalytic zinc [68]. It seems probable that the mechanism of inhibition of ADAM (a disintegrin and metalloproteinase-like)-type (M12) metzincin metalloproteinases by TIMP-3 and other members of family I35 are similar [69].

The *Streptomyces* metalloproteinase inhibitor (family I36) is thought to inhibit thermolysin by the standard mechanism [70], but this awaits confirmation.

In family I38, the N-terminus of the *Pseudomonas* metallopeptidase inhibitor molecule binds to the peptidase along the active-site cleft in an extended conformation with the N-terminal serine acting as a zinc ligand, displacing the activated water molecule and thus inhibiting the peptidase [71].

Although it has a protein fold similar to that of the standard mechanism serine peptidase inhibitors from *Cucurbitaceae* (family I7), the potato carboxypeptidase inhibitor of family I37 (also in clan IE) inhibits by a very different mechanism. It binds the metallocarboxypeptidase A in a substrate-like fashion with its four C-terminal residues inserted into the active-site cleft. The C-terminal Gly-39 is slowly split off, but remains bound in the S1′ subsite where it is effectively buried by the remainder of the inhibitor. The complex has been described as 'an enzyme–product intermediate in the catalytic mechanism' [46]. Leech carboxypeptidase inhibitor (I46) interacts in a similar way, despite having a different protein fold [77].

Distribution of families amongst organisms

We concluded from the large number of clans of peptidase inhibitors, that inhibitors have arisen on many different occasions during the evolution of living organisms. From the data presently available, it seems that most of these events have occurred in the course of the evolution of eukaryotes. Figure 3 shows how the families of inhibitors are distributed in the three superkingdoms of cellular organisms, and in viruses. We found reports of peptidase

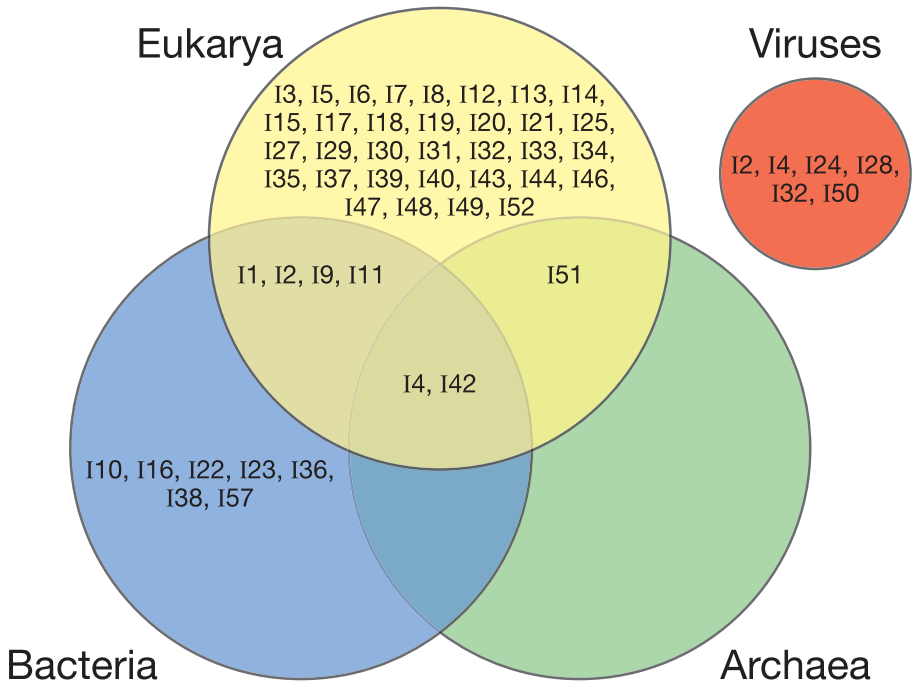

Figure 3 Distribution of inhibitor families throughout the superkingdoms of cellular organisms

The figure summarizes data for the distribution of the families that can be found in the MEROPS database. Although, for clarity, no intersection is shown between viruses and the other groups, it should be noted that family I4 occurs in all four groups.

inhibitors from all kinds of organisms, but the known families are, at present, numerous only in eukaryotes. Only three families are known so far from Archaea, two of which [I4 (serpins) and I42 (chagasin)] are present in all three superkingdoms; I4 is the most widespread of all, being found even in viruses. Distinct inhibitors, however, are most numerous in eukaryotes. None of the prokaryote genomes we examined contains more than six genes encoding members of the families of peptidase inhibitors so far recognized, whereas all of the eukaryotic genomes contain tens or hundreds of them (precise counts being shown in the MEROPS database).

Conclusions

In conclusion, we have assigned the inhibitory domains, termed inhibitor units, of the proteins that inhibit peptidases to families on the basis of their sequence relationships. We have also grouped together some families of distantly related inhibitors in clans. It was necessary to work at the level of the inhibitor units rather than the whole proteins, because many of the proteins contain multiple inhibitor units. A simple system of terminology for reference to the clans, families and inhibitor units within a hierarchical system is introduced. The complete compound inhibitors that contain multiple inhibitor units cannot be classified in any simple way, but a set of identifiers is used for these too. Not unexpectedly, we find that inhibitors in a single family tend to inhibit peptidases of a single catalytic type by a single kind of mechanism. But some families contain inhibitors that target peptidases in more than one family, or even catalytic type, and the mechanisms of these reactions may well be different, although few of them have yet been studied in detail. The new system of classification of the peptidase inhibitors has allowed them to be added to the MEROPS database, where additional details will be found.

We thank Dr Alex Bateman and our colleagues in the Pfam database team for helpful discussions, and the Medical Research Council and the Wellcome Trust for financial support.

REFERENCES

1 Rawlings, N. D., O'Brien, E. and Barrett, A. J. (2002) MEROPS: the protease database. Nucleic Acids Res. **30**, 343–346
2 Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. and Sonnhammer, E. L. (2002) The Pfam protein families database. Nucleic Acids Res. **30**, 276–280
3 Lo Conte, L., Brenner, S. E., Hubbard, T. J., Chothia, C. and Murzin, A. G. (2002) SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Res. **30**, 264–267
4 Barrett, A. J., Rawlings, N. D. and O'Brien, E. A. (2001) The MEROPS database as a protease information system. J. Struct. Biol. **134**, 95–102
5 Turk, B., Turk, D. and Salvesen, G. S. (2002) Regulating cysteine protease activity: essential role of protease inhibitors as guardians and regulators. Curr. Pharm. Des. **8**, 1623–1637
6 De Filippis, V., Colombo, G., Russo, I., Spadari, B. and Fontana, A. (2002) Probing the hirudin–thrombin interaction by incorporation of noncoded amino acids and molecular dynamics simulation. Biochemistry **41**, 13556–13569
7 Lomas, D. A., Lourbakos, A., Cumming, S. A. and Belorgey, D. (2002) Hypersensitive mousetraps, alpha1-antitrypsin deficiency and dementia. Biochem. Soc. Trans. **30**, 89–92
8 Ritchie, B. C. (2003) Protease inhibitors in the treatment of hereditary angioedema. Transfus. Apheresis. Sci. **29**, 259–267
9 Lehesjoki, A. E. (2003) Molecular background of progressive myoclonus epilepsy. EMBO J. **22**, 3473–3478
10 Bitoun, E., Chavanas, S., Irvine, A. D., Lonie, L., Bodemer, C., Paradisi, M., Hamel-Teillac, D., Ansai, S., Mitsuhashi, Y., Taieb, A. et al. (2002) Netherton syndrome: disease expression and spectrum of SPINK5 mutations in 21 families. J. Invest. Dermatol. **118**, 352–361
11 Krol, J., Kopitz, C., Kirschenhofer, A., Schmitt, M., Magdolen, U., Kruger, A. and Magdolen, V. (2003) Inhibition of intraperitoneal tumor growth of human ovarian cancer cells by bi- and trifunctional inhibitors of tumor-associated proteolytic systems. Biol. Chem. **384**, 1097–1102

12 McKay, T. R., Bell, S., Tenev, T., Stoll, V., Lopes, R., Lemoine, N. R. and McNeish, I. A. (2003) Procaspase 3 expression in ovarian carcinoma cells increases survivin transcription which can be countered with a dominant-negative mutant, survivin T34A; a combination gene therapy strategy. Oncogene **22**, 3539–3547
13 Samac, D. A. and Smigocki, A. C. (2003) Expression of oryzacystatin I and II in alfalfa increases resistance to the root-lesion nematode. Phytopathology **93**, 799–804
14 Telang, M., Srinivasan, A., Patankar, A., Harsulkar, A., Joshi, V., Damle, A., Deshpande, V., Sainani, M., Ranjekar, P., Gupta, G. et al. (2003) Bitter gourd proteinase inhibitors: potential growth inhibitors of *Helicoverpa armigera* and *Spodoptera litura*. Phytochemistry **63**, 643–652
15 Laskowski, M. J. and Kato, I. (1980) Protein inhibitors of proteinases. Annu. Rev. Biochem. **49**, 593–626
16 Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. and Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. **31**, 365–370
17 Lipman, D. J. and Pearson, W. R. (1985) Rapid and sensitive protein similarity searches. Science (Washington, D.C.) **227**, 1435–1441
18 McBride, J. D., Watson, E. M., Brauer, A. B., Jaulent, A. M. and Leatherbarrow, R. J. (2002) Peptide mimics of the Bowman–Birk inhibitor reactive-site loop. Biopolymers **66**, 79–92
19 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. J. Mol. Biol. **215**, 403–410
20 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) 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**, 4673–4680
21 Schwartz, R. M. and Dayhoff, M. O. (1978) Matrices for detecting distant relationships. In Atlas of Protein Sequence and Structure 1978 (Dayhoff, M. O., ed.), pp. 353–358, National Biomedical Research Foundation, Washington D.C.
22 Felsenstein, J. (1989) PHYLIP – phylogeny inference package. Cladistics **5**, 164–166
23 Rawlings, N. D. and Barrett, A. J. (1993) Evolutionary families of peptidases. Biochem. J. **290**, 205–218
24 Lo Conte, L., Ailey, B., Hubbard, T. J., Brenner, S. E., Murzin, A. G. and Chothia, C. (2000) SCOP: a structural classification of proteins database. Nucleic Acids Res. **28**, 257–259
25 De Oliveira, C., Santana, L. A., Carmona, A. K., Cezari, M. H., Sampaio, M. U., Sampaio, C. A. M. and Oliva, M. L. V. (2001) Structure of cruzipain/cruzain inhibitors isolated from *Bauhinia bauhinioides* seeds. Biol. Chem. **382**, 847–852
26 Valueva, T. A., Revina, T. A. and Mosolov, V. V. (1997) Potato tuber protein proteinase inhibitors belonging to the Kunitz soybean inhibitor family. Biochemistry (Moscow) **62**, 1367–1374
27 Franco, O. L., Grossi, d. S., Sales, M. P., Mello, L. V., Oliveira, A. S. and Rigden, D. J. (2002) Overlapping binding sites for trypsin and papain on a Kunitz-type proteinase inhibitor from *Prosopis juliflora*. Proteins **49**, 335–341
28 Mares, M., Meloun, B., Pavlik, M., Kostka, V. and Baudys, M. (1989) Primary structure of cathepsin D inhibitor from potatoes and its structure relationship to soybean trypsin inhibitor family. FEBS Lett. **251**, 94–98
29 Huntington, J. A., Read, R. J. and Carrell, R. W. (2000) Structure of a serpin–protease complex shows inhibition by deformation. Nature (London) **407**, 923–926
30 Dufour, E. K., Denault, J. B., Hopkins, P. C. R. and Leduc, R. (1998) Serpin-like properties of alpha1-antitrypsin Portland towards furin convertase. FEBS Lett. **426**, 41–46
31 Al-Khunaizi, M., Luke, C. J., Askew, Y. S., Pak, S. C., Askew, D. J., Cataltepe, S., Miller, D., Mills, D. R., Tsu, C., Brömme, D. et al. (2002) The serpin SQN-5 is a dual mechanistic-class inhibitor of serine and cysteine proteinases. Biochemistry **41**, 3189–3199
32 Björk, I., Nordling, K., Raub-Segall, E., Hellman, U. and Olson, S. T. (1998) Inactivation of papain by antithrombin due to autolytic digestion: a model of serpin inactivation of cysteine proteinases. Biochem. J. **335**, 701–709
33 Komiyama, T., Ray, C. A., Pickup, D. J., Howard, A. D., Thornberry, N. A., Peterson, E. P. and Salvesen, G. (1994) Inhibition of interleukin-1β converting enzyme by the cowpox virus serpin CrmA. An example of cross-class inhibition. J. Biol. Chem. **269**, 19331–19337
34 Kumazaki, T., Ishii, S. and Yokosawa, H. (1994) Identification of the reactive site of ascidian trypsin inhibitor. J. Biochem. (Tokyo) **116**, 787–793
35 Hojima, Y., Pierce, J. V. and Pisano, J. J. (1980) Hageman factor fragment inhibitor in corn seeds: purification and characterization. Thromb. Res. **20**, 149–162
36 Wieczorek, M., Otlewski, J., Cook, J., Parks, K., Leluk, J., Wilimowska-Pelc, A., Polanowski, A., Wilusz, T. and Laskowski, Jr, M. (1985) The squash family of serine proteinase inhibitors. Amino acid sequences and association equilibrium constants of inhibitors from squash, summer squash, zucchini and cucumber seeds. Biochem. Biophys. Res. Commun. **126**, 646–652
37 Bernard, V. D. and Peanasky, R. J. (1993) The serine protease inhibitor family from *Ascaris suum*: chemical determination of the five disulphide bridges. Arch. Biochem. Biophys. **303**, 367–376
38 Griesch, J., Wedde, M. and Vilcinskas, A. (2000) Recognition and regulation of metalloproteinase activity in the haemolymph of *Galleria mellonella*: a new pathway mediating induction of humoral immune responses. Insect Biochem. Mol. Biol. **30**, 461–472
39 Maier, K., Muller, H. and Holzer, H. (1979) Purification and molecular characterization of two inhibitors of yeast proteinase B. J. Biol. Chem. **254**, 8491–8497
40 Takano, R., Imada, C., Kamei, K. and Hara, S. (1991) The reactive site of marinostatin, a proteinase inhibitor from marine *Alteromonas* sp. B-10–31. J. Biochem. (Tokyo) **110**, 856–858
41 Chung, C. H., Ives, H. E., Almeda, S. and Goldberg, A. L. (1983) Purification from *Escherichia coli* of a periplasmic protein that is a potent inhibitor of pancreatic proteases. J. Biol. Chem. **258**, 11032–11038
42 Odani, S. and Ikenaka, T. (1976) The amino acid sequences of two soybean double headed proteinase inhibitors and evolutionary consideration on the legume proteinase inhibitors. J. Biochem. (Tokyo) **80**, 641–643
43 Hatano, K., Kojima, M., Tanokura, M. and Takahashi, K. (1996) Solution structure of bromelain inhibitor VI from pineapple stem: structural similarity with Bowman–Birk trypsin/chymotrypsin inhibitor from soybean. Biochemistry **35**, 5379–5384
44 Heinz, D. W., Priestle, J. P., Rahuel, J., Wilson, K. S. and Grutter, M. G. (1991) Refined crystal structures of subtilisin novo in complex with wild-type and two mutant eglins. Comparison with other serine proteinase inhibitor complexes. J. Mol. Biol. **217**, 353–371
45 Bode, W., Papamokos, E., Musil, D., Seemüller, U. and Fritz, H. (1986) Refined 1.2 Å crystal structure of the complex formed between subtilisin Carlsberg and the inhibitor eglin c. Molecular structure of eglin and its detailed interaction with subtilisin. EMBO J. **5**, 813–818
46 Bode, W. and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. Eur. J. Biochem. **204**, 433–451
47 Rester, U., Bode, W., Moser, M., Parry, M. A., Huber, R. and Auerswald, E. (1999) Structure of the complex of the antistasin-type inhibitor bdellastasin with trypsin and modelling of the bdellastasin-microplasmin system. J. Mol. Biol. **293**, 93–106
48 Mitsui, Y., Satow, Y., Sakamaki, T. and Iitaka, Y. (1977) Crystal structure of a protein proteinase inhibitor, *Streptomyces* subtilisin inhibitor, at 2.3 angstrom resolution. J. Biochem. (Tokyo) **82**, 295–298
49 Taguchi, S., Yamada, S., Kojima, S. and Momose, H. (1998) An endogenous target protease, SAM-P26, of *Streptomyces* protease inhibitor (SSI): primary structure, enzymatic characterization and its interaction with SSI. J. Biochem. (Tokyo) **124**, 804–810
50 Kumazaki, T., Kajiwara, K., Kojima, S., Miura, K. and Ishii, S. (1993) Interaction of *Streptomyces* subtilisin inhibitor (SSI) with *Streptomyces griseus* metallo-endopeptidase-II (SGMPII). J. Biochem. (Tokyo) **114**, 570–575
51 Tsunemi, M., Matsuura, Y., Sakakibara, S. and Katsube, Y. (1993) Crystallization of a complex between an elastase-specific inhibitor elafin and porcine pancreatic elastase. J. Mol. Biol. **232**, 310–311
52 Menegatti, E., Tedeschi, G., Ronchi, S., Bortolotti, F., Ascenzi, P., Thomas, R. M., Bolognesi, M. and Palmieri, S. (1992) Purification, inhibitory properties and amino acid sequence of a new serine proteinase inhibitor from white mustard (*Sinapis alba* L.) seed. FEBS Lett. **301**, 10–14
53 Eguchi, M., Itoh, M., Nishino, K., Shibata, H., Tanaka, T., Kamei-Hayashi, K. and Hara, S. (1994) Amino acid sequence of an inhibitor from the silkworm (*Bombyx mori*) hemolymph against fungal protease. J. Biochem. (Tokyo) **115**, 881–884
54 Greenblatt, H. M., Ryan, C. A. and James, M. N. (1989) Structure of the complex of *Streptomyces griseus* proteinase B and polypeptide chymotrypsin inhibitor-1 from Russet Burbank potato tubers at 2.1 Å resolution. J. Mol. Biol. **205**, 201–228
55 Lindberg, I., Van den Hurk, W. H., Bui, C. and Batie, C. J. (1995) Enzymatic characterization of immunopurified prohormone convertase 2: potent inhibition by a 7B2 peptide fragment. Biochemistry **34**, 5486–5493
56 Hilliard, J. J., Simon, L. D., Van Melderen, L. and Maurizi, M. R. (1998) PinA inhibits ATP hydrolysis and energy-dependent protein degradation by Lon protease. J. Biol. Chem. **273**, 524–527
57 Green, G. D. J., Kembhavi, A. A., Davies, M. E. and Barrett, A. J. (1984) Cystatin-like cysteine proteinase inhibitors from human liver. Biochem. J. **218**, 939–946
58 Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. and Turk, V. (1988) The 2.0 Å X-ray crystal structure of chicken egg-white cystatin and its possible mode of interaction with cysteine proteinases. EMBO J. **7**, 2593–2599
59 Alvarez-Fernandez, M., Barrett, A. J., Gerhartz, B., Dando, P. M., Ni, J. A. and Abrahamson, M. (1999) Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. J. Biol. Chem. **274**, 19195–19203

60 Cornwall, G. A., Cameron, A., Lindberg, I., Hardy, D. M., Cormier, N. and Hsia, N. (2003) The cystatin-related epididymal spermatogenic protein inhibits the serine protease prohormone convertase 2. Endocrinology **144**, 901–908

61 Valente, R. H., Dragulev, B., Perales, J., Fox, J. W. and Domont, G. B. (2001) BJ46a, a snake venom metalloproteinase inhibitor – isolation, characterization, cloning and insights into its mechanism of action. Eur. J. Biochem. **268**, 3042–3052

62 Todd, B., Moore, D., Deivanayagam, C. C., Lin, G. D., Chattopadhyay, D., Maki, M., Wang, K. K. and Narayana, S. V. (2003) A structural model for the inhibition of calpain by calpastatin: crystal structures of the native domain VI of calpain and its complexes with calpastatin peptide and a small molecule inhibitor. J. Mol. Biol. **328**, 131–146

63 Guay, J., Falgueyret, J. P., Ducret, A., Percival, M. D. and Mancini, J. A. (2000) Potency and selectivity of inhibition of cathepsin K, L and S by their respective propeptides. Eur. J. Biochem. **267**, 6311–6318

64 Strukelj, B., Lenarcic, B., Gruden, K., Pungercar, J., Rogelj, B., Turk, V., Bosch, D. and Jongsma, M. A. (2000) Equistatin, a protease inhibitor from the sea anemone *Actinia equina*, is composed of three structural and functional domains. Biochem. Biophys. Res. Commun. **269**, 732–736

65 Riedl, S. J., Renatus, M., Schwarzenbacher, R., Zhou, Q., Sun, C. H., Fesik, S. W., Liddington, R. C. and Salvesen, G. S. (2001) Structural basis for the inhibition of caspase-3 by XIAP. Cell (Cambridge, Mass.) **104**, 791–800

66 Ng, K. K. S., Petersen, J. F. W., Cherney, M. M., Garen, C., Zalatoris, J. J., Rao-Naik, C., Dunn, B. M., Martzen, M. R., Peanasky, R. J. and James, M. N. G. (2000) Structural basis for the inhibition of porcine pepsin by *Ascaris* pepsin inhibitor-3. Nat. Struct. Biol. **7**, 653–657

67 Phylip, L. H., Lees, W. E., Brownsey, B. G., Bur, D., Dunn, B. M., Winther, J. R., Gustchina, A., Li, M., Copeland, T., Wlodawer, A. and Kay, J. (2001) The potency and specificity of the interaction between the IA_3 inhibitor and its target aspartic proteinase from *Saccharomyces cerevisiae*. J. Biol. Chem. **276**, 2023–2030

68 Gomis-Ruth, F. X., Maskos, K., Betz, M., Bergner, A., Huber, R., Suzuki, K., Yoshida, N., Nagase, H., Brew, K., Bourenkov, G. P., Bartunik, H. and Bode, W. (1997) Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. Nature (London) **389**, 77–81

69 Lee, M. H., Dodds, P., Verma, V., Maskos, K., Knauper, V. and Murphy, G. (2003) Tailoring tissue inhibitor of metalloproteinases-3 to overcome the weakening effects of the cysteine-rich domains of tumour necrosis factor-α converting enzyme. Biochem. J. **371**, 369–376

70 Hiraga, K., Seeram, S. S., Tate, S., Tanaka, N., Kainosho, M. and Oda, K. (1999) Mutational analysis of the reactive-site loop of *Streptomyces* metalloproteinase inhibitor, SMPI. J. Biochem. (Tokyo) **125**, 202–209

71 Feltzer, R. E., Trent, J. O. and Gray, R. D. (2003) Alkaline proteinase inhibitor of *Pseudomonas aeruginosa*: a mutational and molecular dynamics study of the role of N-terminal residues in the inhibition of *Pseudomonas* alkaline proteinase. J. Biol. Chem. **278**, 25952–25957

72 Barrett, A. J. (1981) α_2-Macroglobulin. Methods Enzymol. **80**, 737–754

73 Pham, T. N., Hayashi, K., Takano, R., Itoh, M., Eguchi, M., Shibata, H., Tanaka, T. and Hara, S. (1996) A new family of serine protease inhibitors (*Bombyx* family) as established from the unique topological relation between the positions of disulphide bridges and reactive site. J. Biochem. (Tokyo) **119**, 428–434

74 Monteiro, A. C. S., Abrahamson, M., Lima, A. P. C. A., Vannier-Santos, M. A. and Scharfstein, J. (2001) Identification, characterization and localization of chagasin, a tight-binding cysteine protease inhibitor in *Trypanosoma cruzi*. J. Cell Sci. **114**, 3933–3942

75 Neves-Ferreira, A. G. C., Perales, J., Fox, J. W., Shannon, J. D., Makino, D. L., Garratt, R. C. and Domont, G. B. (2002) Structural and functional analyses of DM43, a snake venom metalloproteinase inhibitor from *Didelphis marsupialis* serum. J. Biol. Chem. **277**, 13129–13137

76 Homandberg, G. A., Litwiller, R. D. and Peanasky, R. J. (1989) Carboxypeptidase inhibitors from *Ascaris suum*: the primary structure. Arch. Biochem Biophys. **270**, 153–161

77 Reverter, D., Fernandez-Catalan, C., Baumgartner, R., Pfander, R., Huber, R., Bode, W., Vendrell, J., Holak, T. A. and Aviles, F. X. (2000) Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. Nat. Struct. Biol. **7**, 322–328

78 Normant, E., Martres, M. P., Schwartz, J. C. and Gros, C. (1995) Purification, cDNA cloning, functional expression and characterization of a 26-kDa endogenous mammalian carboxypeptidase inhibitor. Proc. Natl. Acad. Sci. U.S.A. **92**, 12225–12229

79 Brzin, J., Rogelj, B., Popovic, T., Strukelj, B. and Ritonja, A. (2000) Clitocypin, a new type of cysteine proteinase inhibitor from fruit bodies of mushroom *Clitocybe nebularis*. J. Biol. Chem. **275**, 20104–20109

80 Basak, A., Koch, P., Dupelle, M., Fricker, L. D., Devi, L. A., Chrétien, M. and Seidah, N. G. (2001) Inhibitory specificity and potency of proSAAS-derived peptides toward proprotein convertase 1. J. Biol. Chem. **276**, 32720–32728

81 Xu, G., Rich, R. L., Steegborn, C., Min, T., Huang, Y., Myszka, D. G. and Wu, H. (2003) Mutational analyses of the p35-caspase interaction – a bowstring kinetic model of caspase inhibition by p35. J. Biol. Chem. **278**, 5455–5461

82 Snipas, S. J., Stennicke, H. R., Riedl, S., Potempa, J., Travis, J., Barrett, A. J. and Salvesen, G. S. (2001) Inhibition of distant caspase homologues by natural caspase inhibitors. Biochem. J. **357**, 575–580

83 Bruun, A. W., Svendsen, I., Sorensen, S. O., Kielland-Brandt, M. C. and Winther, J. R. (1998) A high-affinity inhibitor of yeast carboxypeptidase Y is encoded by *TFS1* and shows homology to a family of lipid-binding proteins. Biochemistry **37**, 3351–3357

84 St Charles, R., Padmanabhan, K., Arni, R. V., Padmanabhan, K. P. and Tulinsky, A. (2000) Structure of tick anticoagulant peptide at 1.6 Å resolution complexed with bovine pancreatic trypsin inhibitor. Protein Sci. **9**, 265–272

85 Rzychon, M., Sabat, A., Kosowska, K., Potempa, J. and Dubin, A. (2003) Staphostatins: an expanding new group of proteinase inhibitors with a unique specificity for the regulation of staphopains, *Staphylococcus* spp. cysteine proteinases. Mol. Microbiol. **49**, 1051–1066

86 Fuentes-Prior, P., Noeske-Jungblut, C., Donner, P., Schleuning, W. D., Huber, R. and Bode, W. (1997) Structure of the thrombin complex with triabin, a lipocalin-like exosite-binding inhibitor derived from a triatomine bug. Proc. Natl. Acad. Sci. U.S.A. **94**, 11845–11850

87 Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E., Salvesen, G. S., Travis, J. and Whisstock, J. C. (2001) The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions and a revised nomenclature. J. Biol. Chem. **276**, 33293–33296

88 Richardson, R. T., Sivashanmugam, P., Hall, S. H., Hamil, K. G., Moore, P. A., Ruben, S. M., French, F. S. and O'Rand, M. (2001) Cloning and sequencing of human *Eppin*: a novel family of protease inhibitors expressed in the epididymis and testis. Gene **270**, 93–102

89 Trexler, M., Bányai, L. and Patthy, L. (2001) A human protein containing multiple types of protease-inhibitory modules. Proc. Natl. Acad. Sci. U.S.A. **98**, 3705–3709

90 Trexler, M., Bányai, L. and Patthy, L. (2002) Distinct expression pattern of two related human proteins containing multiple types of protease-inhibitory modules. Biol. Chem. **383**, 223–228

91 Olonen, A., Kalkkinen, N. and Paulin, L. (2003) A new type of cysteine proteinase inhibitor – the salarin gene from Atlantic salmon (*Salmo salar* L.) and Arctic charr (*Salvelinus alpinus*). Biochimie **85**, 677–681

92 Barrett, A. J., Rawlings, N. D. and Woessner, J. F. (eds.) Handbook of Proteolytic Enzymes, edition 1, Academic Press, London, 1998

93 Laskowski, M., Jr., Qasim, M. A. and Lu, S. M. (2003) Interaction of standard mechanism, canonical protein inhibitors with serine proteinases. In Protein–Protein Recognition (Kleanthous, C., ed.), pp. 228–279, Oxford University Press, Oxford

93a Heibges, A., Salamini, F. and Gebhardt, C. (2003) Functional comparison of homologous membranes of three groups of Kunitz-type enzyme inhibitors from potato tubers (*Solanum tuberosum* L.). Mol. Genet. Genomics **269**, 535–541

94 Forsyth, S., Horvath, A. and Coughlin, P. (2003) A review and comparison of the murine α_1-antitrypsin and α_1-antichymotrypsin multigene clusters with the human clade A serpins. Genomics **81**, 336–345

95 Knauper, V., Reinke, H. and Tschesche, H. (1990) Inactivation of human plasma α_1-proteinase inhibitor by human PMN leucocyte collagenase. FEBS Lett. **263**, 355–357

96 Irving, J. A., Pike, R. N., Dai, W., Bromme, D., Worrall, D. M., Silverman, G. A., Coetzer, T. H., Dennison, C., Bottomley, S. P. and Whisstock, J. C. (2002) Evidence that serpin architecture intrinsically supports papain-like cysteine protease inhibition: engineering α_1-antitrypsin to inhibit cathepsin proteases. Biochemistry **41**, 4998–5004

97 Sottrup-Jensen, L. (1989) α-Macroglobulins: structure, shape and mechanism of proteinase complex formation. J. Biol. Chem. **264**, 11539–11542

98 Sottrup-Jensen, L., Sand, O., Kristensen, L. and Fey, G. H. (1989) The α-macroglobulin bait region. Sequence diversity and localization of cleavage sites for proteinases in five mammalian α-macroglobulins. J. Biol. Chem. **264**, 15781–15789

99 Stennicke, H. R., Ryan, C. A. and Salvesen, G. S. (2002) Reprieval from execution: the molecular basis of caspase inhibition. Trends Biochem. Sci. **27**, 94–101

100 Laskowski, M. and Qasim, M. A. (2000) What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? Biochim. Biophys. Acta **1477**, 324–337

101 Hemmi, H., Yoshida, T., Kumazaki, T., Nemoto, N., Hasegawa, J., Nishioka, F., Kyogoku, Y., Yokosawa, H. and Kobayashi, Y. (2002) Solution structure of ascidian trypsin inhibitor determined by nuclear magnetic resonance spectroscopy. Biochemistry **41**, 10657–10664

102 Behnke, C. A., Yee, V. C., Trong, I. L., Pedersen, L. C., Stenkamp, R. E., Kim, S. S., Reeck, G. R. and Teller, D. C. (1998) Structural determinants of the bifunctional corn Hageman factor inhibitor: X-ray crystal structure at 1.95 Å resolution. Biochemistry **37**, 15277–15288

103 Poerio, E., Di Gennaro, S., Di Maro, A., Farisei, F., Ferranti, P. and Parente, A. (2003) Primary structure and reactive site of a novel wheat proteinase inhibitor of subtilisin and chymotrypsin. Biol. Chem. **384**, 295–304

104 Rydel, T. J., Tulinsky, A., Bode, W. and Huber, R. (1991) Refined structure of the hirudin-thrombin complex. J. Mol. Biol. **221**, 583–601

105 Richardson, J. L., Kroger, B., Hoeffken, W., Sadler, J. E., Pereira, P., Huber, R., Bode, W. and Fuentes-Prior, P. (2000) Crystal structure of the human α-thrombin-haemadin complex: an exosite II-binding inhibitor. EMBO J. **19**, 5650–5660

106 Guncar, G., Pungercic, G., Klemencic, I., Turk, V. and Turk, D. (1999) Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. EMBO J. **18**, 793–803

107 Filipek, R., Rzychon, M., Oleksy, A., Gruca, M., Dubin, A., Potempa, J. and Bochtler, M. (2003) The staphostatin-staphopain complex: a forward binding inhibitor in complex with its target cysteine protease. J. Biol. Chem. **278**, 40959–40966

108 Kolodziej, S. J., Wagenknecht, T., Strickland, D. K. and Stoops, J. K. (2002) The three-dimensional structure of the human α_2-macroglobulin dimer reveals its structural organization in the tetrameric native and chymotrypsin α_2-macroglobulin complexes. J. Biol. Chem. **277**, 28031–28037

Received 28 November 2003/23 December 2003; accepted 5 January 2004
Published as BJ Immediate Publication 5 January 2004, DOI 10.1042/BJ20031825