

1

Evolution and Classification of the Serpin Superfamily

*James A. Irving, Lisa D. Cabrita, Dion Kaiserman,
Margaret M. Worrall and James C. Whisstock*

1. Introduction

In a pioneering study, the sequence similarity between three protein sequences (human antitrypsin, antithrombin, and chicken ovalbumin) established these proteins — two of them protease inhibitors — as a discrete family.¹ The importance of this family was highlighted by the work over the preceding 17 years, examining the association of the two human proteins with hereditary disease.^{2,3} On the basis of their sequence similarity, and the fact that these, and related, proteins were predominantly characterized as serine protease inhibitors, the label of serpin was applied to this family of high-molecular weight protease inhibitors.⁴

With the discovery of the plant serpin protein, barley Z,⁵ serpin genes extended into the termini of two early diverging lineages of the *eukarya*: the higher plants and the *amniota*. As sequence data were obtained for many new serpin variants in an ever-increasing spectrum of species, the databases were populated with sequences of serpins in flowering plants (*magnoliophyta*), amniotes (*amniota*) and viruses,^{6,7} but failed to fill the genetic gulf between these terminal branches. Key gaps in the species tree of the serpins included simple plants such as the green algae (phylum *chlorophyta*), early diverging animal lineages such as hydra (phylum *cnidaria*), fungi, and bacteria and archaea.⁸

More recently, the accumulation of large volumes of genome sequence data has led to the discovery of several serpins from unicellular organisms, such as thermopin from the moderate thermophilic bacterium *Thermobifida fusca*,^{9,10} and a serpin from the green alga *Chlamydomonas reinhardtii*.¹¹ A targeted approach seeking to address the paucity of data directly has also resulted in the discovery of the jellyfish serpin, jellypin.¹² Species coverage is inexorably expanding to enable a broader account of the family within an evolutionary context.^{9,11,13,14} However, the nature and timing of the evolutionary events that occurred at the inception of the serpin fold remain unclear. For example, the sparse distribution of prokaryotic serpins neither favors the possibility of an ancient, ancestral prokaryotic serpin, nor the possibility that serpins appeared after the eukaryote/prokaryote split and were exchanged between prokaryotes and eukaryotes by a lateral gene transfer event.^{9,11}

2. The Challenges of Serpin Classification

At the time of writing, a PSI-BLAST search of the non-redundant NCBI database identifies around 1600 full and partial serpin sequences and minor variants (Fig. 1). When the redundancy is minimized by removing sequences with greater than 95% identity, the large number remaining — 820 serpin genes in 189 species — is testament to the broad utility of the serpin fold in many biological contexts.

With such a large collection of sequences, it is useful to use some form of classification to partition the dataset into smaller sized relational groups. By far, the most common strategy applied to protein families relies explicitly or implicitly, due to a Trinitarian relationship between sequence, structure and function on evolutionary relationships. Sequence domain databases such as Pfam¹⁵ and the “superfamily” level of classification in the structural database SCOP¹⁶ reflect evolutionary considerations. Evolutionary relationships are in general a preferable means of classifying proteins over functional similarity for several reasons: (i) Within the context of a chosen scoring system, sequence homology or structural similarity is an intrinsic, objective property of two proteins. (ii) Furthermore, an analysis of protein similarity is based on data whose state of completeness can be determined: it is usually known whether a sequence is complete and whether

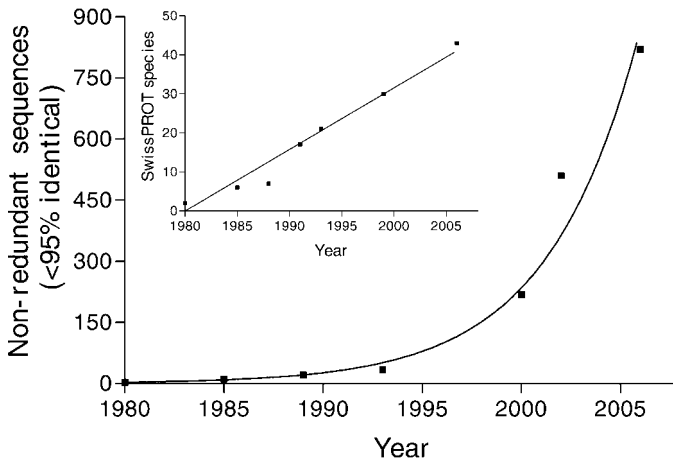


Fig. 1. Growth in the number of serpin sequences, taken from the literature^{1,4,6–8} and from PSI-BLAST searches of the non-redundant database (<http://www.ncbi.nlm.nih.gov/BLAST>). Inset graph shows the trend in the number of species in which serpin sequences are found, from the literature and PSI-BLAST searches of SwissPROT releases 11, 19, 24, 38 and 49.¹¹⁸ At 189, the number of species is considerably higher in the non-curated NCBI non-redundant database.

all potential paralogues can be accounted for (i.e. whether it comes from a fully sequenced genome). (iii) In contrast, the functional data collected of a protein will be much more subjective, and reflect factors as varied as how comprehensively it has been studied in the laboratory, the experimental protocols used, and how biological activity has been measured. Across a large dataset, knowledge of function will most likely be subject to marked inconsistency. (iv) Furthermore, the state of knowledge of a protein's function is likely to change and thus the resulting classifications may quickly become redundant.

Nevertheless, one common goal of protein classification is to enable the prediction of function for a group of proteins based on their association with one another.¹⁷ The basic unit describing a cluster of evolutionarily related sequences in a phylogenetic tree is known as a *clade*. There are many examples of clades coinciding with protein function. For example, the phylogeny of solute transporter proteins reflects subgroups of proteins related by mechanism, energy coupling and polarity, and has permitted the development of a detailed classification system.¹⁸ Similarly, the biological

role of the members of the kinesin superfamily of microtubule-associated motor proteins was found to be reflected by their deduced evolutionary relationships.¹⁹ However, function is a multi-layered concept that includes characteristics such as the mechanism used, reactions catalyzed, binding partners, and localization within an organism.¹⁷

Most serpins are protease inhibitors.²⁰ As a consequence, the function of interest usually concerns the specific protease inhibited, ligand bound, or the biological effect achieved. This detailed concept of “function” is poorly defined by the evolutionary relationships between serpins,⁸ with one exception: in general, the phylogenetic trees can identify direct orthologues as far back as the fish–bird/mammal split (in the order of roughly 450 MYA) where no significant gene expansion has occurred.^{8,14} One reason that evolutionary groupings fail to predict serpin function is the ease with which a functional transition is achieved. It is well known that the reactive center loop (RCL), the primary factor influencing a serpin’s protease target specificity, is subjected to the greatest adaptive evolutionary pressure in a serpin gene.²¹ The substitution of a single amino acid in the RCL can completely alter the protease target *in vivo*. For example, the variant antitrypsin Pittsburgh, a mutant in which the key specificity-determining methionine residue of this serpin has been substituted with arginine, results in a hemorrhagic disorder^{22,23} due to the inappropriate inhibition of thrombin. A similarly striking change of activity is seen upon substitution of threonine-345 in the so-called “hinge region” of the RCL by arginine, i.e. the serpin can no longer act as an inhibitor.²⁴ A change in the class of protease inhibited by a serpin is also achieved with only a few amino acid substitutions.^{25,26}

Differences in structural morphology, such as the presence or absence of a secondary structural element, binding loop or active site, can assist in distinguishing between protein subgroups. For example, the variations in mechanism and function that distinguish the cysteine endopeptidase cathepsin L from the exopeptidase cathepsin B, such as pH stability and exopeptidase activity, are reflected in structural differences and readily identifiable sequence motifs.^{27,28} Considerably greater structural divergence is seen with the calcium-activated protease, calpain.²⁹ These differences coincide with an evolutionary partition between the classes of protease early in the evolution of the eukaryotes.^{30,31}

In the case of serpins, there are few examples of structural characteristics that have become fixed *within* sub-families and permit differentiation *between* sub-families. Figure 2 shows a superposition of serpin structures spanning prokaryotes and eukaryotes with a diverse range of functions, which reveals a relatively immutable core fold that includes most secondary structural elements, surprisingly little variation in length, and a lack of structural element insertion within loop regions. In spite of sequence identities ranging from 25 to almost 90%, the three core sheets and 8-9 alpha

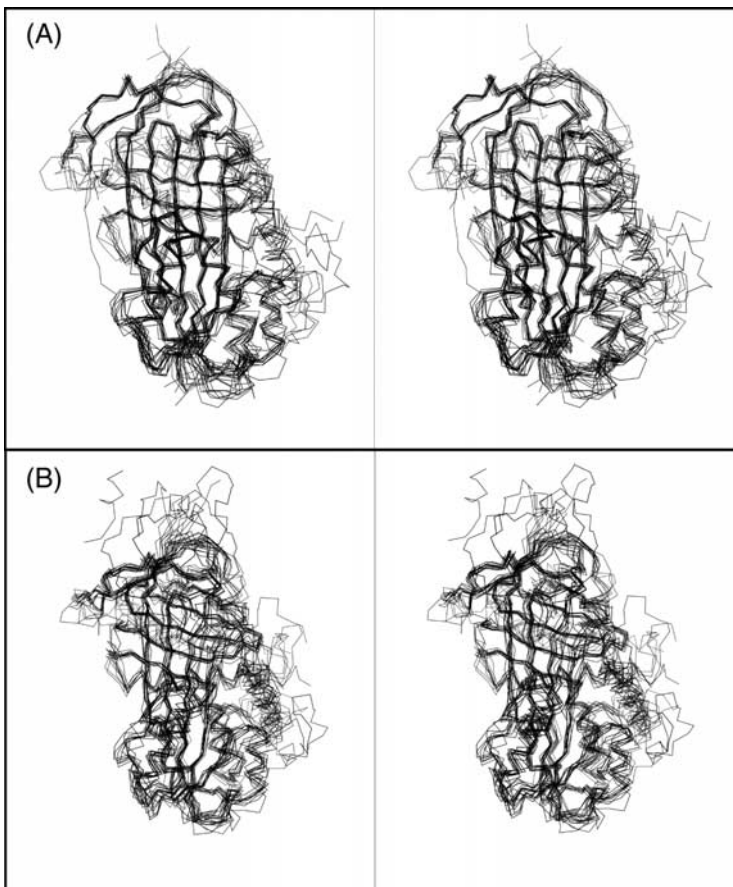


Fig. 2. Stereo figures showing the superposed structures of (A) cleaved serpins and (B) uncleaved serpins, as listed in Table 2. Figure prepared using Molscript.¹¹⁹

helices are present and they superimpose reasonably well. The notable exceptions are the reactive centre loop which can vary in length, conformation and flexibility; the region between the C- and D-helices usually known as the “C–D loop”; N-terminal extensions for some serpins, such as heparin cofactor II;³² and C-terminal extensions, as seen with the thermophilic serpin thermopin.¹⁰ Perhaps, the most pronounced correlation between these structural features and evolutionary grouping is found with the ov-serpin subfamily, whose members generally possess a C–D loop, lack N- and C-terminal extensions to the core domain and lack signal sequences.³³ Of the structures solved to date, a relatively short C–D loop is only visible in the case of ovalbumin.³⁴

The implication of this degree of structural conservation is that amino acid sequence-based approaches provide the greatest amount of discriminatory information in determining relationships between genes in the serpin superfamily. However, in some cases, the gene structure, defined by the position and phasing of intron/exon boundaries, can discern relationships that are not resolved at an appropriate level of statistical significance by protein sequence alone.^{14,33,35,36} This has proven to be the case for the mammalian serpins: six serpin clades whose associations were unreliable using a variety of sequence-based phylogenetic approaches^{8,37} could be reduced to three larger clades based on common gene structures.³⁶ The appearance of serpin genes at the same chromosomal locus has also been seen as evidence of a recent common ancestor.^{13,14,38} However, as a substantial number of serpin sequences lack information on gene structure and chromosomal position, these characteristics cannot form the basis of a universal serpin classification system in themselves. Hence, the classification convention that was adopted for the serpins is based on evolutionary relationships deduced from their amino acid sequences.³⁹

3. The Phylogeny of the Serpin Superfamily

In approaching a phylogenetic reconstruction of a protein family, a choice must be made between a number of approaches that: (i) use amino acid or nucleotide sequence; (ii) treat each sequence as a whole (sequence-based) or each amino acid/nucleotide individually (character-based); (iii) make different assumptions about how readily one amino acid type is substituted

with another, and whether this differs depending on the position in the sequence; (iv) measure the difference between sequences or amino acids/nucleotides differently; (v) produce trees according to different criteria, e.g. deriving a tree that achieves a minimum overall evolutionary score; and (vi) involve very different degrees of computation time.⁴⁰ Diverse approaches have been applied to the serpin superfamily. A character-based method that attempts to derive the tree with the least number of “evolutionary steps” between branches and nodes, the “maximum parsimony” approach,⁴¹ has been applied to nucleotide data of the whole family⁷ and to protein data within subfamilies;⁸ the character-based “maximum likelihood” approach,⁴² which attempts to find the branching arrangement that produces the most likely tree, has also been used.³⁷ The application of the “neighbor-joining” method, based on computed evolutionary differences between sequences, has been assessed as being in best agreement with other data such as serpin gene structure³⁷ and co-localization at chromosomal loci.⁸ Nevertheless, the statistically significant phylogenetic groups determined by the various methods are compatible with one another, and differ primarily in the proportion of sequences that can be assigned to those groups.

There are 35 known functional human serpin genes. A summary evolutionary tree depicting the established phylogenetic relationships between subfamilies of serpin genes containing at least one human member are shown in Fig. 3; the human sequences are listed in Table 1. The tree is based on a “neighbor-joining” approach,⁴³ with the non-parametric bootstrap technique⁴⁴ used to eliminate relationships with poor support; consequently, several branches radiate from the base of the tree because their relationships are uncertain.⁸ Where evidence of a common gene structure³⁶ or chromosomal position^{13,14} suggests the order of sequence divergence, these are also shown in Fig. 4. As the base of the tree indicates the lowest point at which groupings can be reliably determined, it does not reflect a fixed point in time. However, clades containing human serpin genes notably are restricted to sequences from vertebrate organisms; furthermore, non-vertebrate clades (which are not shown in this tree) lack vertebrate sequences. Hence, the deepest point depicted by the base of the tree post-dates the appearance of the vertebrate lineage, estimated to be approximately 800 MYA.^{45,46}

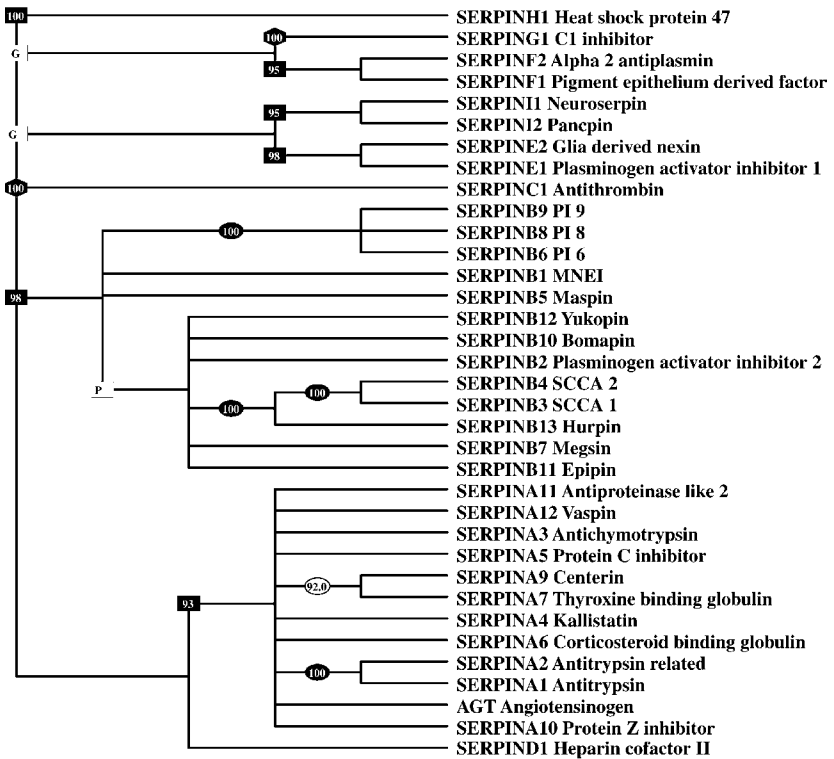


Fig. 3. A phylogenetic tree illustrating the relationship between the 35 full-length human serpins. Only relationships strongly supported by the data are shown. Black hexagons indicate clades with 100% support in a consensus neighbor-joining tree and ovals indicate subgroups within clades with >90% support;⁸ black squares indicate >90% support using the partition cluster consensus method;^{8,120} “G” indicates groups based on gene structure information.³⁷ and “P” indicates branches based on relative positions within chromosomal loci.^{13,14} Branch lengths are not to scale.

Clades C (antithrombin), D (heparin cofactor II), G (C1-inhibitor), H (heat shock protein 47), and I (neuroserpin) comprise few sequences which are mostly orthologues; hence, the evolutionary relationships between their members are consistent with common functions. The other clades mostly have members with markedly different properties. A dichotomy is apparent between clade B, which comprised of predominantly intracellular serpins, and the other vertebrate clades, which are predominantly populated by serpins with extracellular roles, although there

Table 1 Division of the serpin superfamily into clades.^a

Clade ^b	Locus	Member	Common name	Gene ^c
A Antitrypsin-like	14q32.1	SERPIN A13	Kallistatin-like	I
		SERPIN A3	Antichymotrypsin	I
		SERPIN A5	Protein C inhibitor	I
		SERPIN A4	Kallistatin	I
		SERPIN A12	Vaspin	I
		SERPIN A9	Centerin	I
		SERPIN A11	Antiproteinase-like 2	I
		SERPIN A1	Antitrypsin	I
		SERPIN A2	Antitrypsin-related	I
	Xq22.2	SERPIN A6	Corticosteroid-binding globulin	I
		SERPIN A10	Protein Z inhibitor	I
	1q42-q43	SERPIN A7	Thyroxine-binding globulin	I
		AGT	Angiotensinogen (SERPIN A8) ¹	I
B Ovalbumin-like	6p25	SERPIN BP1	<i>Pseudogene</i>	
		SERPIN B1	Monocyte/neutrophil elastase inhibitor	II
		SERPIN B9	PI-9	II
	18q21.3	SERPIN B6	PI-6	II
		SERPIN B8	PI-8	II
		SERPIN B10	Bomapin	III
		SERPIN B2	Plasminogen activator inhibitor-2	III
		SERPIN B7	Megsin	III
		SERPIN B11	Epipin	III
		SERPIN B3	Squamous cell carcinoma antigen-1	III
		SERPIN B4	Squamous cell carcinoma antigen-2	III
		SERPIN B13	Hurpin	III
		SERPIN B12	Yukopin	III
SERPIN B5	Maspin	II		
C Antithrombin	1q23-q25.1	SERPIN C1	Antithrombin	IV
D Heparin cofactor II	22q11.2	SERPIN D1	Heparin cofactor II	I

(Continued)

Table 1 (Continued)

Clade ^b	Locus	Member	Common name	Gene ^c
E PAI-1/GDN	7q21.3-q22	SERPIN E1	Plasminogen activator inhibitor-1	V
	2q33-q35	SERPIN E2	Glia-derived nexin	V
F PEDF	17p13.1	SERPIN F1	Pigment epithelium-derived factor	VI
	17pter-p12	SERPIN F2	Alpha-2-antiplasmin	VI
G C1-inhibitor	11q12-q13.1	SERPIN G1	C1-inhibitor	VI
H Heat shock protein 47	11q13.5	SERPIN H1	Heat-shock protein 47	VII
I Neuroserpin	3q26	SERPIN I1	Neuroserpin	V
		SERPIN I2	Pancpin	V

^aThe human serpins are shown, including the chromosomal location, nomenclature and gene structure. For multigene loci, genes are listed in the telomere-to-centromere direction.

^bClade designation — as shown in the phylogenetic tree (Fig. 3).

^cRepresentative intron/exon patterns are shown in Fig. 4.

are exceptions to this observation. Non-inhibitory serpins do not cluster together and several clades contain serpins that are known to be activated by ligands.

3.1. The nomenclature of the serpin superfamily

The International Committee on the Nomenclature of the Serpins has adopted a system of naming that refers to the clade from which a given serpin arises⁸ and includes a systematic identifier which is applied as the family grows.³⁹ The gene symbol used is “SERPIN”, followed by a letter designating the subfamily (or clade) from which the particular protein arises, the letter “P” if the gene in question is a pseudogene, and a number indicating that particular paralogue. To indicate the species of origin, the three-letter prefix designation of that species by the Committee on Standardization in Human Cytogenetics is used. The 35 human sequences act as “reference sequences” by which orthologues from other organisms are named. Hence, as human antitrypsin is referred to as (HSA)SERPINA1, mouse antitrypsin is referred to as (MMU)SERPINA1. Where

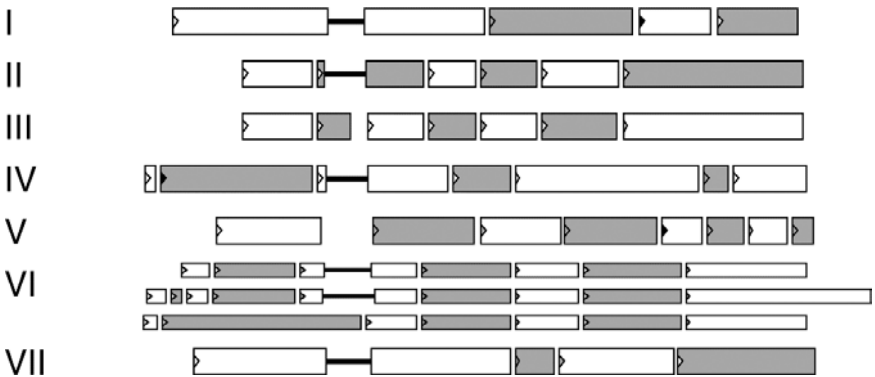


Fig. 4. The gene structure of the human serpins are shown grouped into representative “families” numbered as in Table 1: boxed regions indicate the average relative size and alignment of exons, with white and grey shading used for contrast; a solid horizontal bar indicates a large (>20 amino acid) region of gaps in the alignment that occurs within an exon, while gap regions outside exons are blank. Triangles are used to indicate the phase of an exon: white indicates the exon is in frame with the ATG initiation codon; grey indicates a +1 frameshift; and black indicates a +2 frameshift. As members of family VI have different numbers of exons, individual gene structures are shown.

there is no human orthologue — for example, a mouse sequence that is a member of clade A but has no strong evolutionary relationship with a particular human serpin, that gene would have a unique numerical designation greater than the highest numbered human serpin in the clade.

Clade A

This is the largest clade and contains the antitrypsin-like serpins. These serpins are involved in a diverse range of processes that are either commonly associated with the inhibition of serine proteinases (kallistatin, RASP-1, antitrypsin, and antichymotrypsin) or function in non-inhibitory roles. For instance, antitrypsin has a role in preventing lung damage by inhibiting elastase,⁴⁷ while protein C inhibitor has an active role in the coagulation pathway.^{48,49} Non-inhibitory serpins such as thyroxine binding globulin are involved in hormone transport,⁵⁰ while angiotensinogen has been shown to regulate blood pressure⁵¹ and UTMP has a proposed sequestering role in pregnancy.⁵² Antichymotrypsin has been implicated as having a role in Alzheimer’s disease due to its presence in the plaques of the brains

of Alzheimer's patients.⁵³ In addition, it has also been shown to interact with Alzheimer's A β peptide,^{1–42} which is then released during RCL cleavage and able to readily fibrillize.⁵⁴ The protein currently designated SERPINA13 appears to lack a functional RCL, due to the presence of a premature stop codon, and hence would not be expected to exhibit inhibitory activity, even though transcripts have been detected in the liver.⁵⁵ A recent study has indicated that this stop codon is not present in homologues from other primate species, and has been acquired since the human-chimpanzee divergence.⁵⁶ In the mouse, in spite of the overall conservation of gene order at the clade A locus (12F1) with respect to the syntenic region of the human locus (14q32), SERPINA1 and SERPINA3 have undergone significant expansion to yield 5 and 14 homologues, respectively.^{57,58} There is evidence of functional diversification among these genes.^{59,60}

Clade B

The serpins that form part of this clade are often referred to as the ov-serpins, due to the sequence similarity they share with ovalbumin.³³ These serpins are inhibitors of serine or cysteine proteases, with the exception of maspin and ovalbumin, which have no defined inhibitory functions. Clade B serpins partake in a myriad of regulatory roles involving inflammation, angiogenesis, apoptosis and fibrinolysis, and collectively, they provide a protective role within the cell.⁶¹ Ov-serpins lack an N-terminal signal sequence and mostly exert their biological effect within the cell; however, they have also been detected in the extracellular space. Most also characteristically possess a loop between the C- and D-helices (the "C–D loop"); in PAI-2, this loop affects protein polymerization and cell survival,^{62,63} while in the avian protein MENT, this loop plays a role in chromatin remodeling.⁶⁴ Comparison with non-human genomes suggests that the repertoire and position of 13 human clade B serpins have developed by gene duplication and chromosomal breakage from four progenitor genes,¹⁴ present at a single locus in the last shared ancestor with amphibians (SERPINs B1, B5, B6, and B12); that this expanded to six at the divergence point of mammals from birds (SERPINs B1, B2, B5, B6, B10, and B12),^{13,14} and that the mouse orthologues of three genes at the human 6p25 locus (SERPINs B1, B6, and B9) have undergone further propagation to yield more than 15 apparently functional serpins on mouse chromosome 13.^{65,66}

Clade C

Antithrombin is the major anticoagulant serpin and inhibits a number of serine proteinases of the coagulation pathway, even though the principal targets are generally regarded to be thrombin and factor Xa.⁶⁷ Antithrombin is activated against its target proteases by the binding of heparin to helix D.⁶⁸ Antithrombin has been found to exhibit antiviral properties,⁶⁹ as well as being involved in inflammation,⁷⁰ and antiangiogenesis.⁷¹ An analysis of sequence conservation patterns strongly suggests that the mechanism of activation of this protein is conserved in a wide variety of vertebrate species.⁷²

Clade D

Heparin cofactor II is a potent inhibitor of thrombin; however, unlike antithrombin, it does not inhibit other proteases of the coagulation system. It possesses an uncharacteristic N-terminal extension which interacts with the anion-binding exosite of thrombin. As with antithrombin (clade C), heparin cofactor II also relies on activation by glycosaminoglycans.³²

Clade E

Plasminogen activator inhibitor-1 has a defined role in the plasminogen/plasmin system, and thus contributes to a multitude of physiological processes; for example, it is associated with common syndromes such as atherosclerosis, diabetes, and hypertension. It is able to inhibit urokinase-type plasminogen activator and is thus said to have a role in cell migration and tissue remodeling.⁷³ Glia-derived nexin-1 (or protease nexin-1) is a potent inhibitor of thrombin, even though its physiological role has not been properly defined. It is a specific regulator of several proteases and has been implicated in roles within the central nervous system, the vasculature and the extracellular matrix.^{74,75}

Clade F

Pigment epithelium derived factor functions both as an inhibitor of angiogenic processes and is a neurotrophic factor.⁷⁶ In contrast, its closest relative, alpha-2-antiplasmin, regulates fibrinolysis during clot formation.⁷⁷

Clade G

C1-inhibitor regulates the activation of proteases in both the complement and the contact pathways, and its deficiency leads to angioedema. It has been shown that C1-inhibitor is involved at sites of inflammation and also in the migration of leukocytes across the endothelium.⁷⁸

Clade H

Heat shock protein 47 is not a protease inhibitor but a procollagen/collagen-specific binding protein that functions as a molecular chaperone during collagen biosynthesis.⁷⁹ It has also been implicated in a number of diseases including arteriosclerosis, myocardial infarction and a number of fibroses.⁸⁰

Clade I

Neuroserpin inhibits tissue-type plasminogen activator and has a prominent role in the nervous system, including neurite outgrowth and synaptogenesis. It also has a role in the vascular system and plays a therapeutic role in protecting the brain during ischaemia.⁸¹ Familial variants of neuroserpin have been shown to polymerise and aggregate as Collins bodies within neurons, leading to the neurodegenerative disease FENIB.⁸²

4. The Structural Phylogeny of the Serpin Domain

By 1993, sufficient serpin structures had been solved to clearly demonstrate the high degree of structural similarity between members. Cleaved ovalbumin, antichymotrypsin, and antitrypsin structures showed an overall root mean square deviation (between C α atoms) of 0.67–1.71 Å.⁷ In each case, the overall native fold remained the same, i.e. reactive centre loop protruding from the top of the molecule, three β -sheets, and nine α -helices.

In the last few years, there has been an increasingly extensive coverage of the serpin superfamily. Nine of the sixteen evolutionary branches of the serpin family now have one or more representative crystal structures (Table 2). The structural biology of this family is somewhat complicated by the ability of the fold to adopt alternative topologies, including the

Table 2 Highest-resolution structures of native and cleaved members from the serpin superfamily.

Clade ^a	Serpin	Species	Form	PDB ^b	Å ^c	Ref
A	Antitrypsin	<i>Homo sapiens</i>	Native	1qlp	2.0	105
			Cleaved, complex	1ezx AB	2.6	87
	Antichymotrypsin	<i>Mus musculus</i>	Native	1yxa	2.1	90
		<i>Homo sapiens</i>	Cleaved	1as4	2.1	106
	Protein C inhibitor	<i>Homo sapiens</i>	Cleaved	1lq8 AB	2.4	107
B	Plasminogen activator inhibitor 2	<i>Homo sapiens</i>	Native, mutant	1by7	2.0	85
	Leukocyte elastase inhibitor	<i>Equus caballus</i>	Cleaved	1hle	1.95	108
	Ovalbumin	<i>Gallus gallus</i>	Native	1ova	1.95	34
			Cleaved, mutant	1jti	2.3	95
Maspin	<i>Homo sapiens</i>	Native	1wz9	2.1	109	
C	Antithrombin	<i>Homo sapiens</i>	Native	1t1f	2.75	—
			Latent ^d	1e05 L	2.62	110
D	Heparin cofactor II	<i>Homo sapiens</i>	Native	1jmj	2.35	32
E	Plasminogen activator inhibitor 1	<i>Homo sapiens</i>	Native	1dvm	2.4	111
			Cleaved	9pai	2.7	112
F	Pigment epithelium derived factor	<i>Homo sapiens</i>	Native	1imv	2.85	113
I	Neuroserpin	<i>Mus musculus</i>	Cleaved	1jjo	3.06	114
K	Alaserpin (K variant)	<i>Manduca sexta</i>	Native	1sek	2.1	115
N	crmA	Cowpox virus	Cleaved	1f0c	2.26	104
n.d.	Thermopin	<i>Thermobifida fusca</i>	Native	1sng	1.76	92
			Cleaved	1mtp	1.5	116

^an.d. — no clade designation.

^bThe PDB accession of the crystal structure;¹¹⁷ if it has multiple components, the relevant serpin chains are listed.

^cResolution of the structure; *italics* denote structures higher than 2.5 Å resolution.

^dLatent antithrombin is used in place of the 3.2 Å cleaved antithrombin structure.

metastable native state, the stable cleaved and latent forms, serpin–protease complexes, and loop-sheet polymers.⁸³ In particular, this highlights the flexibility of two important features of the serpin mechanism: the RCL and the A β -sheet.

In spite of this conformational variation, superpositions of cleaved and uncleaved structures highlight the well-conserved nature of the serpin fold (Fig. 2). This suggests that the fold is well-suited to its biological function and is most likely intolerant to substantial structural changes within the molecule. For example, extending the length of the reactive center loop promotes the formation of an inactive latent conformation⁸⁴ and mutation of underlying residues in the shutter region can lead to folding defects.⁸⁵ Conservation of structure may also reflect the fact that serpins exist as “two-in-one” proteins, in that they adopt two different conformations with different packing arrangements.⁸⁶

Broadly speaking, there are two major conformations of serpin: RCL fully inserted (due to cleavage, transition to latency, or polymerization) and RCL mostly expelled (ranging from partially inserted in the case of non-activated antithrombin, to fully expelled as seen with maspin). Most serpins are required to achieve a delicate balance between these two states; it is the transition from RCL-expelled to RCL-inserted that permits serpins to act as covalent protease inhibitors.⁸⁷ Hence, this requires that during the course of evolution, a serpin sequence is under pressure to both favor folding to the active state and maintaining the ability to adopt an energetically favorable RCL-inserted form, without leading to inappropriate conformations such as polymers and the latent form. Indeed, the path of the reactive center loop as it inserts into the central A- β sheet of the molecule is mirrored by the pattern of highly conserved residues among members of the serpin superfamily;⁸ and regions that play a role in the transition between loop-expelled and loop-incorporated, termed the hinge, the breach and the shutter, appear to be *co-evolving* at a slower than average rate.⁸⁸ These residues are shown in Fig. 5 and it can be seen that the association between them is best understood in the context of the cleaved structure of antitrypsin, and not the active, inhibitory, uncleaved form, illustrating the influence of different structural contexts on evolutionary behavior.

Figure 6 shows the influence of evolutionary divergence on structural movements within the serpin fold. It is clear from the uncleaved serpin

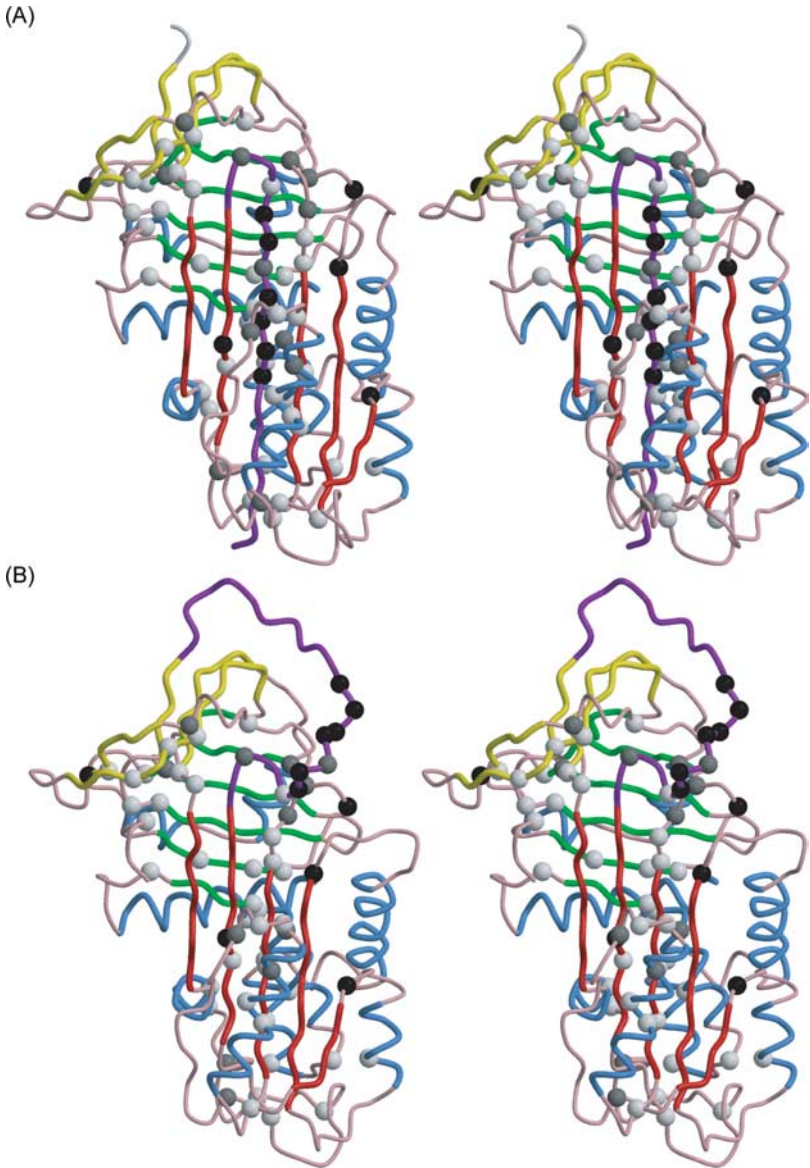


Fig. 5. The position of residues identified as conserved in more than 75% of serpin sequences⁸ (white spheres) and those representing co-evolving sites⁸⁸ (black spheres) are shown mapped onto the crystal structures of (A) cleaved and (B) uncleaved human anti-trypsin (see Table 2 for accessions). Gray spheres indicate residues common to the two datasets. The stereo figure was prepared using Molscript¹¹⁹ and Raster3D.¹²¹

(A)

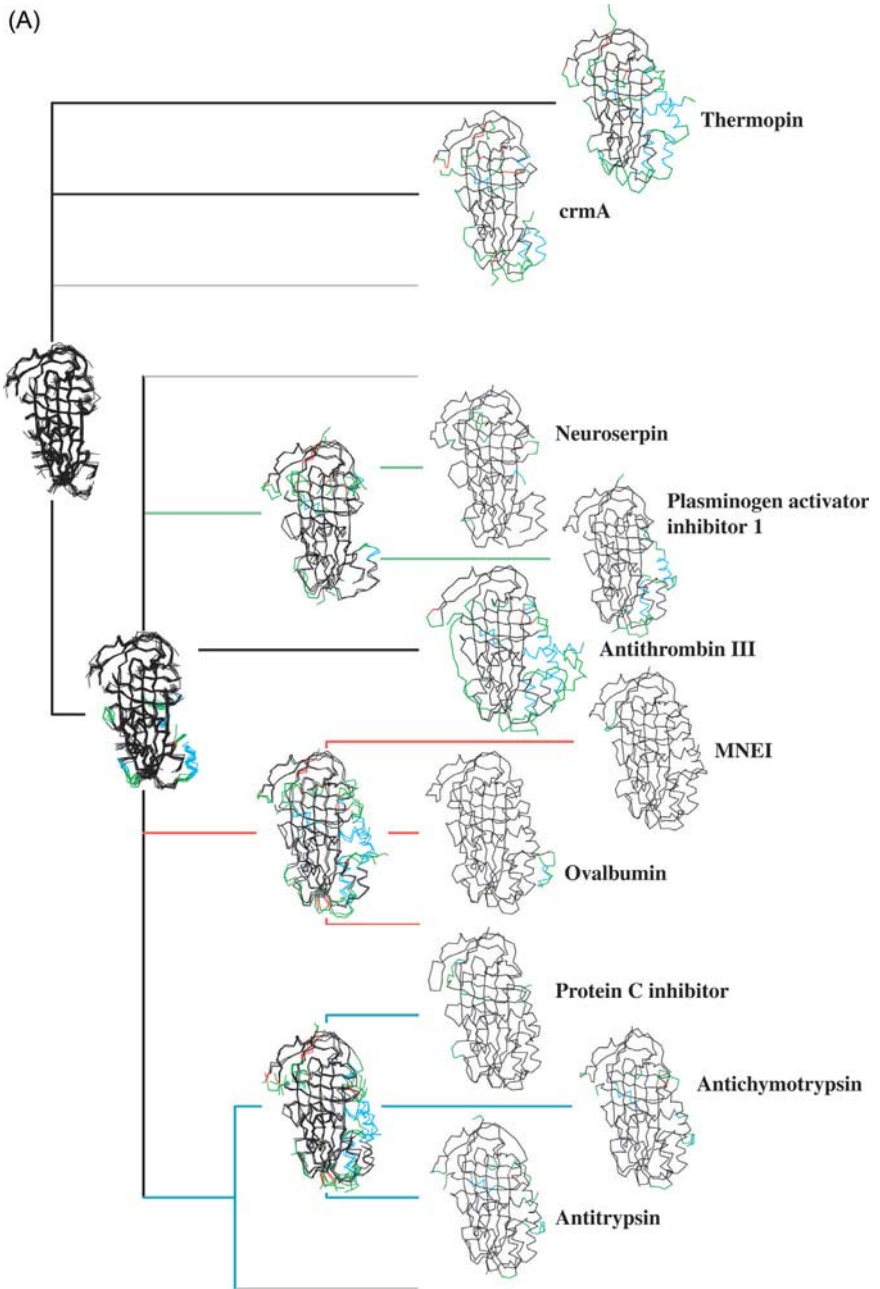


Fig. 6. (Continued)

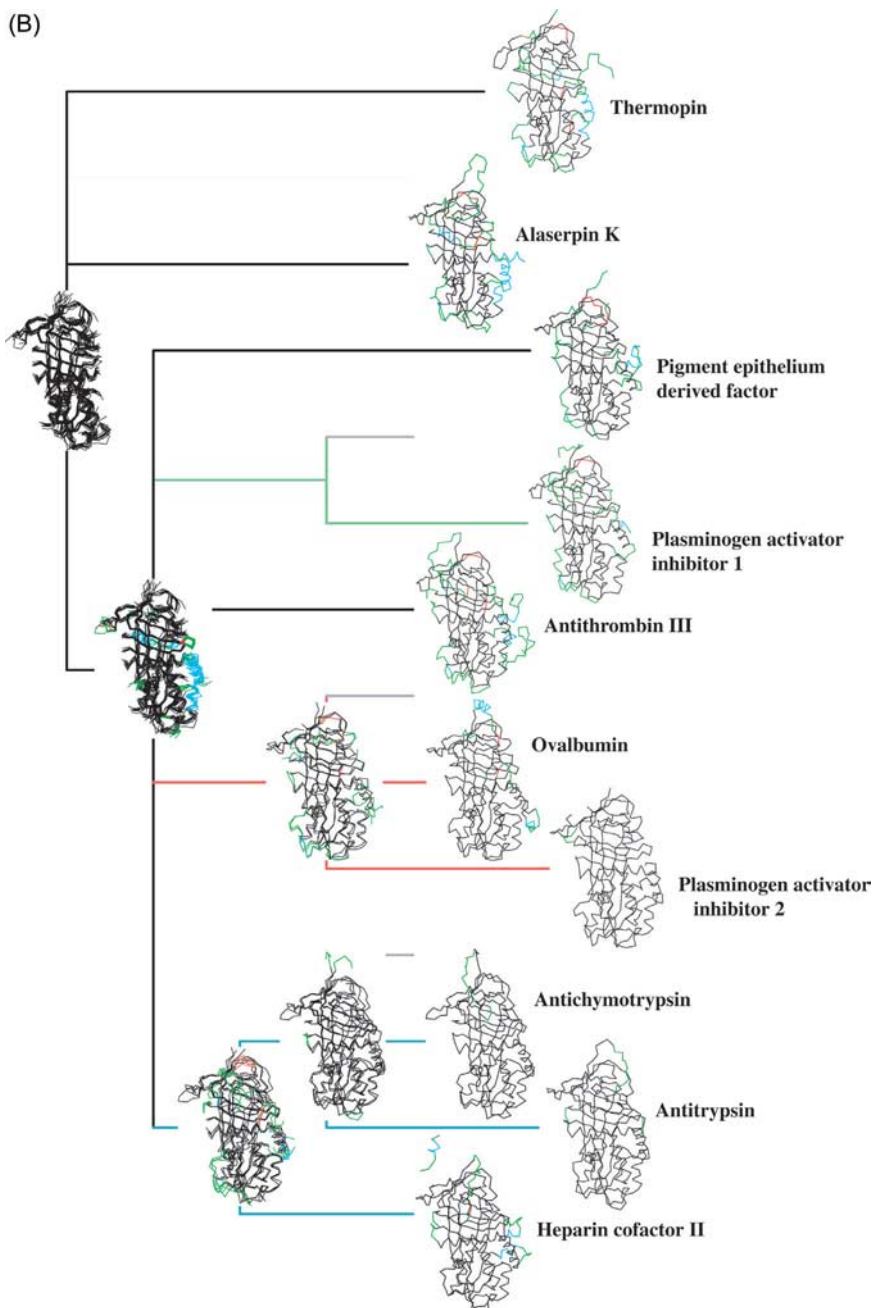


Fig. 6. (Continued)

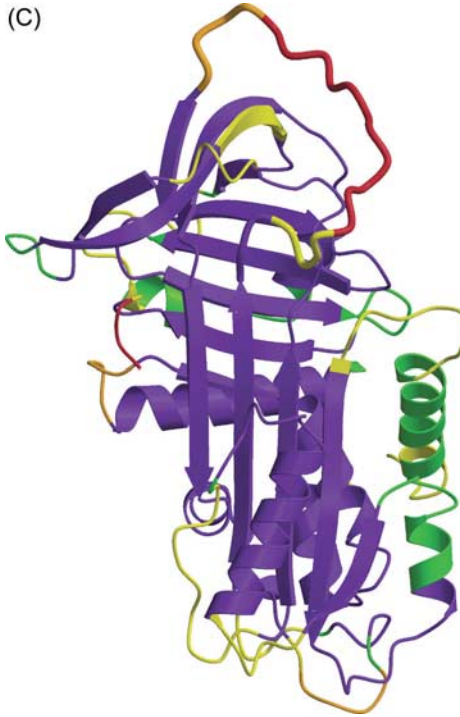


Fig. 6. A tree depicting the phylogenetic relationship between (A) cleaved and (B) native serpins of known structure, as listed in Table 2. Multiple pairwise superpositions were performed using the DaliLite program.¹²² At each ancestral node on the tree, the residues constituting the “common core”¹²³ of all structures that descend from that node were calculated from these pairwise superpositions. A “common core” excludes regions that are not consistently structurally aligned. Each cartoon shown represents the “common core” where two or more branches meet; colored lines indicate those regions that are absent from the common core at the next ancestral node (cyan, alpha-helix; red, beta-sheet; green, and loop region). For clarity, the multi-gene clades A/D, B, and E/I are colored blue, red and green, respectively. (C) The structure of native antitrypsin, indicating the order in which regions are lost from the common core when moving from the terminal branch to the most ancestral node; red indicates the first region to be lost, followed by orange, yellow, and green; blue-violet residues are present in the common core at the base of the tree. Structures were drawn using Molscript.¹¹⁹

structures (Fig. 6A and B) that the reactive centre loop is the most mutable element, along with the CD-loop of the clade B serpin, ovalbumin, and the C- and N-termini of thermopin and the N-terminus of heparin cofactor II. This is followed closely by the loop joining helix D to strand 1A, the loop

joining the F-helix to strand 3A, and the N-terminal portion of the G-helix. Deeper into the tree, more loops joining structural elements are lost, such as the residues between strands 1B and 2B. Finally, the D-helix, E-helix, and G-helix are found to superpose poorly near the base of the tree (Fig. 6C).

4.1. Sequence plasticity in the reactive center loop

The reactive center loop (RCL) of serpins has long been recognized as being hypervariable,²¹ i.e. with a non-synonymous mutation rate above the background rate, and this can be easily seen when examining the multi-gene clusters that have appeared due to extensive gene duplication in the mouse.^{58,59} The RCL is uniquely varied in its structure and when comparing one serpin with another, it is evident that it is able to adopt a number of different conformations; for example, the so-called canonical conformation of antitrypsin around the specificity-determining residues⁸⁹ and the RCL of the closely related murine antichymotrypsin, which is almost perpendicular to that of antitrypsin and is partially inserted into the A-sheet.⁹⁰ The serpin fold is largely well ordered. The central region of the RCL represents a frequent exception to this and typically makes few contacts with the body of the serpin; thus, it is often not visible in crystal structures due to a high degree of disorder such as seen with plasminogen activator inhibitor-1⁹¹ and thermopin.⁹² Presumably, this permits the acquisition of optimal target specificity on the basis of sequence alone, without requiring compensating mutations for the maintenance of structural integrity.

As discussed above, this is a two-edged sword; a point mutation in this region can detrimentally alter the biological properties of the serpin, and genetic disease ensues. The N-terminal portion of the RCL is somewhat more constrained in sequence as it becomes incorporated into the A-sheet during the inhibitory process, and is required to contain a string of small polar and aliphatic amino acids known as the “hinge region motif”.⁹³ This is exemplified by ovalbumin, a serpin that does not undergo conformational change.⁹⁴ With the requirement for a hinge region compatible with the A-sheet insertion discarded, residue 339 of the RCL has undergone a change from threonine to arginine with respect to related serpins. Introducing this mutation into other serpins results in their inability to effectively inhibit proteases, and mutating this residue back to a threonine in ovalbumin permits RCL insertion into the A-sheet.⁹⁵

Similar plasticity is seen in the inhibition of different protease classes by serpins. While the majority of serpins characterized to date have been shown to be inhibitors of trypsin-like serine proteinases,^{6,96} the family also has inhibitory members that target other classes of proteinase. The viral crmA protein inhibits caspase-1,⁹⁷ and squamous cell carcinoma antigen-1 inhibit papain-like cysteine proteinases,⁹⁸ but minimal changes in the specificity-determining region of the RCL are required to alter the target protease class.^{25,26,99}

5. Concluding Statements: The Prokaryotic Serpins and the Root of the Serpin Superfamily

At the root of the “tree of life” lies the dichotomy between prokaryotic and eukaryotic organisms, and the divergence of plants, animals, and fungi. The order in which these branches diverged at this period in natural history is a well-debated topic. An ancient gene represented in a significant number of prokaryotes and eukaryotes implies universal importance to cellular life; whereas there are a significant number of eukaryotic genes that appeared only after the divergence from prokaryotic ancestors.¹⁰⁰

Due to the high throughput sequencing of whole prokaryotic organisms from a diverse range of environments and genetic backgrounds, a host of serpin genes have been identified that suggest the root of the gene tree may lie prior to the divergence of bacteria from archaea.^{9,11} However, it is clear that serpin genes are not *indispensable* to cellular life, as there are a large number of fully sequenced prokaryotic organisms, as well as yeasts, lacking a recognizable member of this family.⁹ Whether this arises due to some other proteins providing an analogous function cannot be definitively known; the primary role of serpins is to inhibit proteases, but it is also clear that the presence of a protease does not predicate the presence of a serpin.⁹ The prokaryotic organisms known to contain serpin genes include extremophiles (such as *Pyrobaculum aerophilum*, *Thermoanaerobacter tengcongensis*, and *Thermobifida fusca*) and mesophiles. Based on current data, there is no evidence to suggest the capture of eukaryotic serpin genes by bacteria. However, there are interesting examples of bacteria that have an association with an eukaryotic host: *Ruminococcus albus*, which resides in the foregut of ruminant animals and produces

cellulolytic enzymes; and *Nostoc punctiforme*, which lives as an endosymbiont in the fungus *Geosiphon pyriformis*. It is possible that a shared lineage exists with these organisms but remains undetectable due to gene divergence.

Multiple serpin genes have been identified in plants,¹⁰¹ but as yet cognate target proteases have not been identified. It was clear from a study involving serpins from *Arabidopsis thaliana* and *Hordeum vulgare* that for plants, the diverse repertoire of serpins evolved from a single progenitor serpin, for each of the species, *prior* to species divergence.⁸ It seems likely then that a single serpin gene existed at least as far back as the plant/animal split, and its recruitment by complex, multicellular organisms has led to the large number of genes seen today.

Based on the connectivity of secondary structure elements in serpins, two separate studies have countenanced the possible integrity of a C-terminal domain composed solely of the B- and C-sheets and G- and H-helices, and lacking helices A–F and strands s1A–s3A.^{102,103} It has been suggested on this basis that the serpin fold may have developed from the joining of the two domains near the conserved Trp-194 residue, followed by an insertion event generating two strands of the A-sheet.¹⁰² Furthermore, there is structural evidence that the D-helix and the G-helix are dispensable secondary structural components.^{10,104} It may be that only with the knowledge of the earliest precursors of the serpin fold will the timing of the origin of the serpin superfamily become clear.

References

1. Hunt LT and Dayhoff MO (1980) A surprising new protein superfamily containing ovalbumin, antithrombin-III, and alpha 1-proteinase inhibitor. *Biochem Biophys Res Commun* **95**:864–871.
2. Laurell CB and Eriksson S (1963) The electrophoretic alpha 1-globulin pattern of serum in alpha 1-antitrypsin deficiency. *Scand J Clin Lab Invest Suppl* **15**:132–140.
3. Egeberg O (1965) Inherited antithrombin deficiency causing thrombophilia. *Thromb Diath Haemorrh* **13**:516–541.
4. Carrell RW and Travis J (1985) Alpha-1-antitrypsin and the serpins: Variation and countervariation. *Trends Biochem Sci* **8**:20–24.

5. Hejgaard J, Rasmussen SK, Brandt A and Svendsen I (1985) Sequence homology between barley endosperm protein Z and protease inhibitors of the alpha-1-antitrypsin family. *FEBS Lett* **180**:89–94.
6. Huber R and Carrell RW (1989) Implications of the three-dimensional structure of alpha-1-antitrypsin for structure and function of serpins. *Biochemistry* **28**:8951–8966.
7. Marshall CJ (1993) Evolutionary relationships among the serpins. *Philos Trans R Soc Lond B Biol Sci* **342**:101–119.
8. Irving JA, Pike RN, Lesk AM and Whisstock JC (2000) Phylogeny of the serpin superfamily implications of patterns of amino acid conservation for structure and function. *Genome Res* **10**:1845–1864.
9. Irving JA, Steenbakkens PJM, Lesk AM, Op den Camp HJM, Pike RN and Whisstock JCW (2002) Serpins in prokaryotes. *Mol Biol Evol* **19**:1881–1890.
10. Irving JA, Cabrita LD, Rossjohn J, Pike RN, Bottomley SP and Whisstock JC (2003) The 1.5 Å crystal structure of a prokaryote serpin: Controlling conformational change in a heated environment. *Structure* **11**:387–397.
11. Roberts TH, Hejgaard J, Saunders NF, Cavicchioli R and Curmi PM (2004) Serpins in unicellular Eukarya, Archaea, and Bacteria: Sequence analysis and evolution. *J Mol Evol* **59**:437–447.
12. Cole EB, Miller D, Rometo D, Greenberg RM, Bromme D, Cataltepe S, Pak SC, Mills DR, Silverman GA and Luke CJ (2004) Identification and activity of a lower eukaryotic serine proteinase inhibitor (serpin) from *Cyanea capillata*: Analysis of a jellyfish serpin, jellypin. *Biochemistry* **43**:11750–11759.
13. Benarafa C and Remold-O'Donnell E (2005) The ovalbumin serpins revisited: Perspective from the chicken genome of clade B serpin evolution in vertebrates. *Proc Natl Acad Sci USA* **102**:11367–11372.
14. Kaiserman D and Bird PI (2005) Analysis of vertebrate genomes suggests a new model for clade B serpin evolution. *BMC Genomics* **6**:167.
15. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C and Eddy SR (2004) The Pfam protein families database. *Nucleic Acids Res* **32**:D138–D141.
16. Murzin AG, Brenner SE, Hubbard T and Chothia C (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**:536–540.
17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM

- and Sherlock G (2000) Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**:25–29.
18. Saier MH Jr, Tran CV and Barabote RD (2006) TCDB: The transporter classification database for membrane transport protein analyses and information. *Nucleic Acids Res* **34**:D181–D186.
 19. Dagenbach EM and Endow SA (2004) A new kinesin tree. *J Cell Sci* **117**: 3–7.
 20. Gettins PG (2002) Serpin structure, mechanism, and function. *Chem Rev* **102**:4751–4804.
 21. Inglis JD and Hill RE (1991) The murine Spi-2 proteinase inhibitor locus: A multigene family with a hypervariable reactive site domain. *Embo J* **10**:255–261.
 22. Lewis JH (1978) Antithrombin Pittsburgh: An alpha1-antitrypsin variant causing hemorrhagic disease. *Blood* **51**:129–137.
 23. Owen MC, Brennan SO, Lewis JH and Carrell RW (1983) Mutation of antitrypsin to antithrombin alpha 1-antitrypsin Pittsburgh (358 Met leads to Arg), a fatal bleeding disorder. *N Engl J Med* **309**:694–698.
 24. Schulze AJ, Huber R, Degryse E, Speck D and Bischoff R (1991) Inhibitory activity and conformational transition of alpha 1-proteinase inhibitor variants. *Eur J Biochem* **202**:1147–1155.
 25. Luke C, Schick C, Tsu C, Whisstock JC, Irving JA, Bromme D, Juliano L, Shi AUGP, Chapman HA and Silverman GA (2000) Simple modifications of the serpin reactive site loop convert SCCA2 into a cysteine protease inhibitor: A critical role for the P3' proline in facilitating RSL cleavage. *Biochemistry* **39**:7081–7091.
 26. Irving JA, Pike RN, Dai W, Bromme D, Worrall DM, Silverman GA, Coetzer TH, Dennison C, Bottomley SP and Whisstock JC (2002) Evidence that serpin architecture intrinsically supports papain-like cysteine protease inhibition: Engineering alpha(1)-antitrypsin to inhibit cathepsin proteases. *Biochemistry* **41**:4998–5004.
 27. Karrer KM, Peiffer SL and DiTomas ME (1993) Two distinct gene subfamilies within the family of cysteine protease genes. *Proc Natl Acad Sci USA* **90**:3063–3067.
 28. Turk B, Turk V and Turk D (1997) Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol Chem* **378**: 141–150.
 29. Hosfield CM, Elce JS, Davies PL and Jia Z (1999) Crystal structure of calpain reveals the structural basis for Ca(2+)-dependent protease activity and a novel mode of enzyme activation. *Embo J* **18**:6880–6889.

30. Ward W, Alvarado L, Rawlings ND, Engel JC, Franklin C and McKerrow JH (1997) A primitive enzyme for a primitive cell: The protease required for excystation of *Giardia*. *Cell* **89**:437–444.
31. Huang X, Czerwinski E and Mellgren RL (2003) Purification and properties of the Dictyostelium calpain-like protein. *Cpl. Biochemistry* **42**:1789–1795.
32. Baglin TP, Carrell RW, Church FC, Esmon CT and Huntington JA (2002) Crystal structures of native and thrombin-complexed heparin cofactor II reveal a multistep allosteric mechanism. *Proc Natl Acad Sci USA* **99**:11079–11084.
33. Remold-O'Donnell E (1993) The ovalbumin family of serpin proteins. *FEBS Lett* **315**:105–108.
34. Stein PE, Leslie AG, Finch JT and Carrell RW (1991) Crystal structure of uncleaved ovalbumin at 1.95 Å resolution. *J Mol Biol* **221**:941–959.
35. Scott FL, Eyre HJ, Lioumi M, Ragoussis J, Irving JA, Sutherland GA and Bird PI (1999) Human ovalbumin serpin evolution: Phylogenetic analysis, gene organization, and identification of new PI8-related genes suggest that two interchromosomal and several intrachromosomal duplications generated the gene clusters at 18q21–q23 and 6p25. *Genomics* **62**:490–499.
36. Ragg H, Lokot T, Kamp PB, Atchley WR and Dress A (2001) Vertebrate serpins: Construction of a conflict-free phylogeny by combining exon-intron and diagnostic site analyses. *Mol Biol Evol* **18**:577–584.
37. Atchley WR, Lokot T, Wollenberg K, Dress A and Ragg H (2001) Phylogenetic analyses of amino acid variation in the serpin proteins. *Mol Biol Evol* **18**:1502–1511.
38. Rollini P and Fournier RE (1997) A 370-kb cosmid contig of the serpin gene cluster on human chromosome 14q32.1: Molecular linkage of the genes encoding alpha 1-antichymotrypsin, protein C inhibitor, kallistatin, alpha 1-antitrypsin, and corticosteroid-binding globulin. *Genomics* **46**:409–415.
39. Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin, PB, Gettins PG, Irving JA, Lomas DA, Luke CJ, Moyer RW, Pemberton PA, Remold-O'Donnell E, Salvesen GS, Travis J and Whisstock JC (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.
40. Nei M and Kumar S (2000) *Molecular Evolution and Phylogenetics, 1st Edition*. Oxford University Press: Oxford.
41. Eck RV and Dayhoff MO (1966) *Atlas of Protein Sequence and Structure, 1st Edition*. National Biomedical Research Foundation: Silver Spring, MD.
42. Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol* **17**:368–376.

43. Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**:406–425.
44. Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**:783–791.
45. Gu X (1998) Early metazoan divergence was about 830 million years ago. *J Mol Evol* **47**:369–371.
46. Feng DF, Cho G and Doolittle RF (1997) Determining divergence times with a protein clock: Update and reevaluation. *Proc Natl Acad Sci USA* **94**:13028–13033.
47. Beatty K, Bieth J and Travis J (1980) Kinetics of association of serine proteinases with native and oxidized alpha-1-proteinase inhibitor and alpha-1-antichymotrypsin. *J Biol Chem* **255**:3931–3934.
48. Rezaie AR, Cooper ST, Church FC and Esmon CT (1995) Protein C inhibitor is a potent inhibitor of the thrombin-thrombomodulin complex. *J Biol Chem* **270**:25336–25339.
49. Elisen MG, von dem Borne PA, Bouma BN and Meijers JC (1998) Protein C inhibitor acts as a procoagulant by inhibiting the thrombomodulin-induced activation of protein C in human plasma. *Blood* **91**:1542–1547.
50. Refetoff S, Murata Y, Mori Y, Janssen OE, Takeda K and Hayashi Y (1996) Thyroxine-binding globulin: Organization of the gene and variants. *Horm Res* **45**:128–138.
51. Gardes J, Bouhnik J, Clauser E, Corvol P and Menard J (1982) Role of angiotensinogen in blood pressure homeostasis. *Hypertension* **4**:185–189.
52. McFarlane JR, Foulds LM, O'Connor AE, Phillips DJ, Jenkin G, Hearn MT and de Kretser DM (1999) Uterine milk protein, a novel activin-binding protein, is present in ovine allantoic fluid. *Endocrinology* **140**:4745–4752.
53. Abraham CR, Selkoe DJ and Potter H (1988) Immunochemical identification of the serine protease inhibitor alpha 1-antichymotrypsin in the brain amyloid deposits of Alzheimer's disease. *Cell* **52**:487–501.
54. Janciauskiene S, Eriksson S and Wright HT (1996) A specific structural interaction of Alzheimer's peptide A beta 1-42 with alpha 1-antichymotrypsin. *Nat Struct Biol* **3**:668–671.
55. Marsden MD and Fournier RE (2005) Organization and expression of the human serpin gene cluster at 14q32.1. *Front Biosci* **10**:1768–1778.
56. Hahn Y and Lee B (2006) Human-specific nonsense mutations identified by genome sequence comparisons. *Hum Genet* **119**:169–178.
57. Barbour KW, Wei F, Brannan C, Flotte TR, Baumann H and Berger FG (2002) The murine alpha(1)-proteinase inhibitor gene family: Polymorphism, chromosomal location, and structure. *Genomics* **80**:515–522.

58. Forsyth S, Horvath A and Coughlin P (2003) A review and comparison of the murine alpha1-antitrypsin and alpha1-antichymotrypsin multigene clusters with the human clade A serpins. *Genomics* **81**:336–345.
59. Barbour KW, Goodwin RL, Guillonneau F, Wang Y, Baumann H, and Berger FG (2002) Functional diversification during evolution of the murine alpha(1)-proteinase inhibitor family: Role of the hypervariable reactive center loop. *Mol Biol Evol* **19**:718–727.
60. Horvath AJ, Forsyth SL and Coughlin PB (2004) Expression patterns of murine antichymotrypsin-like genes reflect evolutionary divergence at the Serpina3 locus. *J Mol Evol* **59**:488–497.
61. Silverman GA, Whisstock JC, Askew DJ, Pak SC, Luke CJ, Cataltepe S, Irving JA and Bird PI (2004) Human clade B serpins (ov-serpins) belong to a cohort of evolutionarily dispersed intracellular proteinase inhibitor clades that protect cells from promiscuous proteolysis. *Cell Mol Life Sci* **61**: 301–325.
62. Wilczynska M, Lobov S, Ohlsson PI and Ny T (2003) A redox-sensitive loop regulates plasminogen activator inhibitor type 2 (PAI-2) polymerization. *Embo J* **22**:1753–1761.
63. Dickinson JL, Bates EJ, Ferrante A and Antalis TM (1995) Plasminogen activator inhibitor type 2 inhibits tumor necrosis factor alpha-induced apoptosis. Evidence for an alternate biological function. *J Biol Chem* **270**:27894–27904.
64. Springhetti EM, Istomina NE, Whisstock JC, Nikitina T, Woodcock CL and Grigoryev SA (2003) Role of the M-loop and reactive center loop domains in the folding and bridging of nucleosome arrays by MENT. *J Biol Chem* **278**:43384–43393.
65. Kaiserman D, Knaggs S, Scarff KL, Gillard A, Mirza G, Cadman M, McKeone R, Denny P, Cooley J, Benarafa C, Remold-O'Donnell E, Ragoussis J and Bird PI (2002) Comparison of human chromosome 6p25 with mouse chromosome 13 reveals a greatly expanded ov-serpin gene repertoire in the mouse. *Genomics* **79**:349–362.
66. Mallon AM, Wilming L, Weekes J, Gilbert JG, Ashurst J, Peyrefitte S, Matthews L, Cadman M, McKeone R, Sellick CA, Arkell R, Botcherby MR, Strivens MA, Campbell RD, Gregory S, Denny P, Hancock JM, Rogers J and Brown SD (2004) Organization and evolution of a gene-rich region of the mouse genome: A 12.7-Mb region deleted in the Del(13)Svea36H mouse. *Genome Res* **14**:1888–1901.
67. Pratt CW and Church FC (1993) General features of the heparin-binding serpins antithrombin, heparin cofactor II and protein C inhibitor. *Blood Coagul Fibrinolysis* **4**:479–490.

68. Jin L, Abrahams JP, Skinner R, Petitou M, Pike RN and Carrell RW (1997) The anticoagulant activation of antithrombin by heparin. *Proc Natl Acad Sci USA* **94**:14683–14688.
69. Geiben-Lynn R, Brown N, Walker BD and Luster AD (2002) Purification of a modified form of bovine antithrombin III as an HIV-1 CD8+ T-cell antiviral factor. *J Biol Chem* **277**:42352–42357.
70. Oelschlager C, Romisch J, Staubitz A, Stauss H, Leithauser B, Tillmanns H and Holschermann H (2002) Antithrombin III inhibits nuclear factor kappaB activation in human monocytes and vascular endothelial cells. *Blood* **99**:4015–4020.
71. O'Reilly MS, Pirie-Shepherd S, Lane WS and Folkman J (1999) Antiangiogenic activity of the cleaved conformation of the serpin antithrombin. *Science* **285**:1926–1928.
72. Backovic M and Gettins PG (2002) Insight into residues critical for antithrombin function from analysis of an expanded database of sequences that includes frog, turtle, and ostrich antithrombins. *J Proteome Res* **1**: 367–373.
73. Gils A and Declerck PJ (2004) The structural basis for the pathophysiological relevance of PAI-I in cardiovascular diseases and the development of potential PAI-I inhibitors. *Thromb Haemost* **91**:425–437.
74. Baker JB, Low DA, Simmer RL and Cunningham DD (1980) Protease-nexin: A cellular component that links thrombin and plasminogen activator and mediates their binding to cells. *Cell* **21**:37–45.
75. Festoff BW, Rao JS and Hantai D (1991) Plasminogen activators and inhibitors in the neuromuscular system: III. The serpin protease nexin I is synthesized by muscle and localized at neuromuscular synapses. *J Cell Physiol* **147**:76–86.
76. Steele FR, Chader GJ, Johnson LV and Tombran-Tink J (1993) Pigment epithelium-derived factor: Neurotrophic activity and identification as a member of the serine protease inhibitor gene family. *Proc Natl Acad Sci USA* **90**:1526–1530.
77. Coughlin PB (2005) Antiplasmin: The forgotten serpin? *Febs J* **272**: 4852–4857.
78. Davis AE 3rd, Cai S and Liu D (2004) The biological role of the C1 inhibitor in regulation of vascular permeability and modulation of inflammation. *Adv Immunol* **82**:331–363.
79. Dafforn TR, Della M and Miller AD (2001) The molecular interactions of heat shock protein 47 (Hsp47) and their implications for collagen biosynthesis. *J Biol Chem* **276**:49310–49319.

80. Razzaque MS and Taguchi T (1999) The possible role of colligin/HSP47, a collagen-binding protein, in the pathogenesis of human and experimental fibrotic diseases. *Histol Histopathol* **14**:1199–1212.
81. Yepes M and Lawrence DA (2004) Neuroserpin: A selective inhibitor of tissue-type plasminogen activator in the central nervous system. *Thromb Haemost* **91**:457–464.
82. Davis RL, Shrimpton AE, Holohan PD, Bradshaw C, Feiglin D, Collins GH, Sonderegger P, Kinter J, Becker LM, Lacbawan F, Krasnewich D, Muenke M, Lawrence DA, Yerby MS, Shaw CM, Gooptu B, Elliott PR, Finch JT, Carrell RW and Lomas DA (1999) Familial dementia caused by polymerization of mutant neuroserpin. *Nature* **401**:376–379.
83. Whisstock J, Skinner R and Lesk AM (1998) An atlas of serpin conformations. *Trends Biochem Sci* **23**:63–67.
84. Na YR and Im H (2005) The length of the reactive center loop modulates the latency transition of plasminogen activator inhibitor-1. *Protein Sci* **14**:55–63.
85. Gooptu B, Hazes B, Chang WS, Dafforn TR, Carrell RW, Read RJ and Lomas DA (2000) Inactive conformation of the serpin alpha(1)-antichymotrypsin indicates two-stage insertion of the reactive loop: Implications for inhibitory function and conformational disease. *Proc Natl Acad Sci USA* **97**:67–72.
86. Whisstock JC, Skinner R, Carrell RW and Lesk AM (2000) Conformational changes in serpins: I. The native and cleaved conformations of alpha(1)-antitrypsin. *J Mol Biol* **296**:685–699.
87. Huntington JA, Read RJ and Carrell RW (2000) Structure of a serpin-protease complex shows inhibition by deformation. *Nature* **407**:923–926.
88. Buck MJ and Atchley WR (2005) Networks of coevolving sites in structural and functional domains of serpin proteins. *Mol Biol Evol* **22**:1627–1634.
89. Elliott PR, Abrahams JP and Lomas DA (1998) Wild-type alpha 1-antitrypsin is in the canonical inhibitory conformation. *J Mol Biol* **275**:419–425.
90. Horvath AJ, Irving JA, Rossjohn J, Law RH, Bottomley SP, Quinsey NS, Pike RN, Coughlin PB and Whisstock JC (2005) The murine orthologue of human antichymotrypsin: A structural paradigm for clade A3 serpins. *J Biol Chem* **280**:43168–43178.
91. Harrop SJ, Jankova L, Coles M, Jardine D, Whittaker JS, Gould AR, Meister A, King GC, Mabbutt BC and Curmi PM (1999) The crystal structure of plasminogen activator inhibitor 2 at 2.0 Å resolution: Implications for serpin function. *Structure* **7**:43–54.
92. Fulton KF, Buckle AM, Cabrera LD, Irving JA, Butcher RE, Smith I, Reeve S, Lesk AM, Bottomley SP, Rossjohn J and Whisstock JC (2005) The high resolution crystal structure of a native thermostable serpin reveals the complex

- mechanism underpinning the stressed to relaxed transition. *J Biol Chem* **280**:8435–8442.
93. Hopkins PC, Carrell RW and Stone SR (1993) Effects of mutations in the hinge region of serpins. *Biochemistry* **32**:7650–7657.
 94. Stein PE, Tewkesbury DA and Carrell RW (1989) Ovalbumin and angiotensinogen lack serpin S-R conformational change. *Biochem J* **262**:103–107.
 95. Yamasaki M, Arii Y, Mikami B and Hirose M (2002) Loop-inserted and thermostabilized structure of P1-P1' cleaved ovalbumin mutant R339T. *J Mol Biol* **315**:113–120.
 96. Gettins PG (2000) Keeping the serpin machine running smoothly. *Genome Res* **10**:1833–1835.
 97. Ray CA, Black RA, Kronheim SR, Greenstreet TA, Sleath PR, Salvesen GS and Pickup DJ (1992) Viral inhibition of inflammation: Cowpox virus encodes an inhibitor of the interleukin-1 beta converting enzyme. *Cell* **69**:597–604.
 98. Schick C, Pemberton PA, Shi GP, Kamachi Y, Cataltepe S, Bartuski AJ, Gornstein ER, Bromme D, Chapman HA and A SG (1998) Cross-class inhibition of the cysteine proteases cathepsins K, L, and S by the serpin squamous cell carcinoma antigen 1: A kinetic analysis. *Biochemistry* **37**:5258–5266.
 99. Al-Khunaizi M, Luke CJ, Askew YS, Pak SC, Askew DJ, Cataltepe S, Miller D, Mills DR, Tsu C, Bromme D, Irving JA, Whisstock JC and Silverman GA (2002) The serpin SQN-5 is a dual mechanistic-class inhibitor of serine and cysteine proteinases. *Biochemistry* **41**:3189–3199.
 100. Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov, DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Rogozin IB, Smirnov S, Sorokin AV, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ and Natale DA (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol* **5**:R7.
 101. Dahl SW, Rasmussen SK and Hejgaard J (1996) Heterologous expression of three plant serpins with distinct inhibitory specificities. *J Biol Chem* **271**:25083–25088.
 102. Wright HT (1993) Introns and higher-order structure in the evolution of serpins. *J Molecular Evol* **36**:136–143.
 103. Whisstock JC, Irving JA, Bottomley SP, Pike RN and Lesk AM (1999) Serpins in the *Caenorhabditis elegans* genome. *Proteins* **36**:31–41.
 104. Renatus M, Zhou Q, Stennicke HR, Snipas SJ, Turk D, Bankston LA, Liddington RC and Salvesen GS (2000) Crystal structure of the apoptotic suppressor CrmA in its cleaved form. *Structure* **8**:789–797.

105. Elliott PR, Pei XY, Dafforn TR and Lomas DA (2000) Topography of a 2.0 Å structure of alpha1-antitrypsin reveals targets for rational drug design to prevent conformational disease. *Protein Sci* **9**:1274–1281.
106. Lukacs CM, Rubin H and Christianson DW (1998) Engineering an anion-binding cavity in antichymotrypsin modulates the “spring-loaded” serpin-protease interaction. *Biochemistry* **37**:3297–3304.
107. Huntington JA, Kjellberg M and Stenflo J (2003) Crystal structure of protein C inhibitor provides insights into hormone binding and heparin activation. *Structure* **11**:205–215.
108. Baumann U, Bode W, Huber R, Travis J and Potempa J (1992) Crystal structure of cleaved equine leucocyte elastase inhibitor determined at 1.95 Å resolution. *J Mol Biol* **226**:1207–1218.
109. Law RH, Irving JA, Buckle AM, Ruzyla K, Buzza M, Bashtannyk-Puhlovich TA, Beddoe TC, Nguyen K, Worrall DM, Bottomley SP, Bird PI, Rossjohn J and Whisstock JC (2005) The high resolution crystal structure of the human tumor suppressor maspin reveals a novel conformational switch in the G-helix. *J Biol Chem* **280**:22356–22364.
110. Skinner R, Abrahams JP, Whisstock JC, Lesk AM, Carrell, RW and Wardell MR (1997) The 2.6 Å structure of antithrombin indicates a conformational change at the heparin binding site. *J Mol Biol* **266**:601–609.
111. Stout TJ, Graham H, Buckley DI and Matthews DJ (2000) Structures of active and latent PAI-1: A possible stabilizing role for chloride ions. *Biochemistry* **39**:8460–8469.
112. Aertgeerts K, De Bondt HL, De Ranter C and Declerck PJ (1995) Crystallization and X-ray diffraction data of the cleaved form of plasminogen activator inhibitor-1. *Proteins* **23**:118–121.
113. Simonovic M, Gettins PG and Volz K (2001) Crystal structure of human PEDF, a potent anti-angiogenic and neurite growth-promoting factor. *Proc Natl Acad Sci USA* **98**:11131–11135.
114. Briand C, Kozlov SV, Sonderegger P and Grutter MG (2001) Crystal structure of neuroserpin: A neuronal serpin involved in a conformational disease. *FEBS Lett* **505**:18–22.
115. Li J, Wang Z, Canagarajah B, Jiang H, Kanost M and Goldsmith EJ (1999) The structure of active serpin 1K from *Manduca sexta*. *Structure Fold Des* **7**:103–109.
116. Irving JA, Cabrita LD, Rossjohn J, Pike RN, Bottomley SP and Whisstock JC (2003) The 1.5 Å crystal structure of a prokaryote serpin: Controlling conformational change in a heated environment. *Structure* **11**:387–397.

117. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE (2000) The Protein data bank. *Nucleic Acids Res* **28**:235–242.
118. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N and Yeh LS (2005) The universal Protein resource (UniProt). *Nucleic Acids Res* **33**:D154–D159.
119. Kraulis PJ (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of Protein structures. *J App Cryst* **24**:946–950.
120. Irving JA, Askew DJ and Whisstock JC (2004) Computational analysis of evolution and conservation in a protein superfamily. *Methods* **32**:73–92.
121. Merritt EA and Bacon DJ (1997) Raster3D: Photorealistic Molecular Graphics. *Meth Enzymol* **277**:505–524.
122. Holm L and Park J (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* **16**:566–567.
123. Lesk AM and Fordham WD (1996) Conservation and variability in the structures of serine proteinases of the chymotrypsin family. *J Mol Biol* **258**:501–537.