Proteases Universally Recognize Beta Strands In Their Active Sites

Joel D. A. Tyndall,*,† Tessa Nall, and David P. Fairlie

Centre for Drug Design and Development, Institute for Molecular Bioscience, University of Queensland, Brisbane, Qld 4072, Australia

Received July 30, 2004

Contents

1. Introduction

Among proteases or proteinases, the endopeptidases account for a significant proportion (∼2%) of the human genome with over 550 defined members and a further 100 or so predicted human proteases.[1,2] Proteases also represent 1−5% of the genomes of infectious organisms such as bacteria, parasites, and viruses. They are categorized by the nature of their active-site catalytic residue as metallo (34%), serine (30%), cysteine (26%), aspartic (4%), and the less characterized threonine (5%). Proteases have been assembled into clans based on structural and catalytic homology, further categorized into families according to sequence homology, and localized to specific chromosomes at least for the human genome. Literature and database searches, including MEROPS,[3] Ensembl,[4] InterPro,[5] NCBI,(http://www.ncbi.nlm.nih.gov/), and Celera,[6] are useful sources of this information although it seems that no single source represents a comprehensive listing.

Protease-mediated cleavages of specific peptide main-chain amide bonds are essential for the synthesis and turnover of all proteins. Proteases therefore determine protein folding and function, thereby regulating most physiological processes. Consequently, control over protease expression and function can potentially be an effective strategy for therapeutic intervention. Mammalian proteases, of which about 20% are membrane bound, are aberrantly expressed or under-regulated during many disease conditions, suggesting that inhibitors of specific proteases might be able to control irregular mammalian physiology. Alternatively, since there are now over 50 human proteases where one or more genetic mutations have been found to lead to hereditary diseases,[2] gene therapy may be a useful approach when it eventually becomes technologically feasible.

Infectious organisms such as viruses, bacteria, and other parasites use similar kinds of proteases which, once installed in infected mammalian hosts, compete for host nutrients and operate with cellular machinery to prolong infectivity. One approach to combating infectious disease is to selectively inhibit these foreign proteases within host cells, thereby retarding replication rates of infectious organisms and assisting normal immunological defense mechanisms involved in their eradication. Thus, the goal of protease inhibitor design is to develop *potent* inhibitors of the offending foreign or mammalian protease but with accompanying *selectivity* to avoid inhibiting very similar host proteases required for normal host physiology.

By virtue of controlling protein synthesis, turnover, and function, proteases are able to regulate most physiological functions including digestion and nutrient intake, ovulation and fertilization, embryo development, growth and differentiation, cell signaling and migration, tissue homeostasis, bone formation and resorption, neuronal outgrowth, antigen presentation, cell-cycle regulation, immunity, phagocytosis, inflammation, cytokine release, wound healing, angiogenesis, apoptosis, and aging. Mammalian proteases associated with such control are potential

* To whom correspondence should be addressed. Fax +64-3-4797034. E-mail: joel.tyndall@otago.ac.nz.
† Current mailing address: School of Pharmacy, University of Otago, P.O. Box 913, Dunedin, 9008, New Zealand.

10.1021/cr040669e CCC: $53.50
Published on Web 02/16/2005

Joel Tyndall obtained his B.Sc. (Hons) degree in Inorganic Chemistry from Monash University in 1994 and his Ph.D. degree in 2000 from the Centre for Drug Design and Development, The University of Queensland, which focused on molecular recognition in disease. He spent a year at the University of Edinburgh with Professor Malcolm Walkinshaw. Since 2001 he has worked with Professor Fairlie in the field of drug design. He took up a faculty position in 2004 at the School of Pharmacy at Otago University, New Zealand.

Tessa Nall was born in Brisbane, Australia. She graduated with her Bachelor of Biotechnology degree with first-class honors in the field of Drug Design and Development in 2003 from the Institute of Molecular Bioscience, The University of Queensland. Her honors thesis incorporated the analysis of protease–inhibitor structures as well as West Nile Virus NS3 protease enzymology. In 2004 she worked as Network Administrator for the protease networks www.protease.net and www.protease.net.au.

Professor David P. Fairlie received his B.Sc. (Honors) degree from the University of Adelaide and Ph.D. degree from the University of New South Wales and did postdoctoral research at Stanford University and the University of Toronto. He has held research/teaching appointments in six Australian universities and led the Chemistry Group in the Centre for Drug Design and Development, University of Queensland, since 1991. His research interests are in chemical synthesis (organic, medicinal, inorganic); molecular recognition (DNA-, RNA-, protein-, metal-binding compounds); peptide and protein mimetics; non-peptidic enzyme inhibitors and receptor antagonists for viral and parasitic infections, inflammatory disorders, cancers, and neurodegenerative diseases; and mechanisms of chemical reactions, biological processes, disease development, and drug action.

targets for therapeutic intervention as are non-mammalian proteases that play pivotal roles in disease-causing infections mediated by viruses, bacteria, and parasites in general.

The approach of selectively inhibiting specific proteases associated with aberrant physiology and disease has only matured during the past decade and now emerges as a promising therapeutic strategy for improving the human/mammalian condition.[7] Protease inhibitors are in trials for inflammatory, immunological, and respiratory conditions[8−11] and cardiovascular[12] and degenerative disorders including Alzheimer's disease.[13,14] HIV has been successfully targeted by protease inhibitors with nine in clinical use: Amprenavir, Atazanavir, Fosamprenavir, Indinavir, Lopinavir, Nelfinavir, Ritonavir, Saquinavir, and Tipranavir.[7] Human proteases, for which some physiological control with exogenous inhibitors has been sought, include serine proteases (e.g., thrombin, coagulation factors VIIa and Xa, elastases, complement convertases, tryptases, kallikreins), aspartic proteases (e.g., renin, cathepsin D, memapsin/BACE), metalloproteases (e.g., angiotensin-converting enzyme (ACE), neprilysin, matrix metalloproteases, carboxypeptidases, tumor necrosis factor α-converting enzyme (TACE), aminopeptidases), and cysteine proteases (e.g., cathepsins B, H, K, L, S, caspases, calpains). Viral proteases[15,16] are essential enzymes in the replication of viruses, such as human immunodeficiency virus (HIV), feline immunodeficiency virus (FIV), simian immunodeficiency virus (SIV), human T-cell leukemia virus-1 (HTLV-1),[17] Rous sarcoma (RSV),[18] hepatitis A, B, and C, West Nile virus, Dengue virus, cytomegalovirus (CMV), poliovirus, rhinovirus 3C, influenza, herpes, coronaviruses (e.g., SARS), and thus are potential drug targets.[19] Proteases are also being investigated as key targets for bacterial and fungal infections such as *C. albicans*[20] and *C. tropicalis*, and for parasitic infections such as malaria (plasmepsins), schistosomiasis (cathepsins), and hookworm infections (cathepsins).[21]

This review summarizes an analysis of over 1500 three dimensional crystal (X-ray) and solution (NMR) structures from the pdb of substrates, products and inhibitors bound in the active sites of aspartic, serine, metallo, cysteine, and threonine endopeptidases. The active sites of all five protease classes were found to recognize peptidic and non-peptidic ligands in an extended beta strand conformation, with few exceptions. Comparisons of protease-bound ligand conformations are illustrated by structural superpositions for a subset of structures, including 21 aspartic (145 inhibitor complexes), 44 serine (220 inhibitor complexes), 20 metallo (123 inhibitor complexes), 23 cysteine (89 inhibitor complexes), and 2 threonine (4 inhibitor complexes) proteases, among the protease−ligand structures that we have analyzed. The extended substrate-binding mode is also illustrated for 3 aspartic proteases (13 substrates), 1 serine protease, 1 cysteine protease and 1 metalloprotease. All

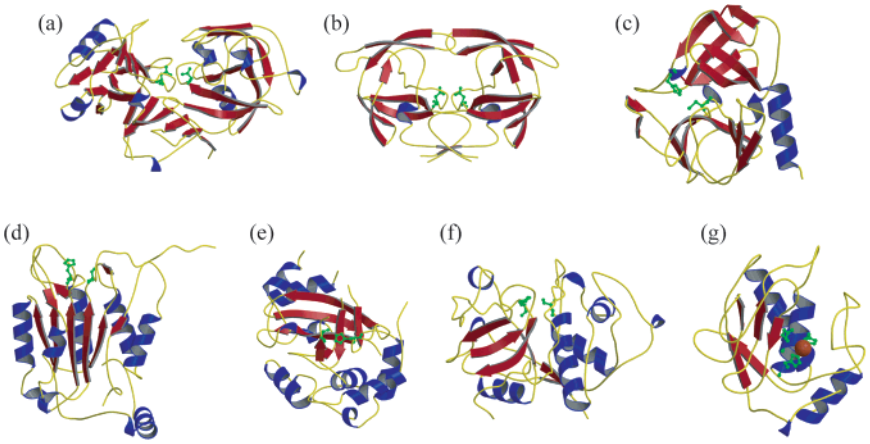


Figure 1. Seven common structural folds among proteases. From top left are secondary structure representations of the (a) pepsin-like aspartic protease β-barrel (1pso), (b) retropepsin aspartic protease β-barrel (7hvp), (c) trypsin-like fold β-barrel (1cqq), (d) subtilisin-like and caspase-like αβα sandwich or Rossman fold (1ice), (e) herpes virus serine protease αβ barrel (1 cmv), (f) papain-like cysteine protease αβ complex fold (1pap), (g) thermolysin-like catalytic domain (1mmq). Catalytic residues shown in green; the zinc atom is shown as an orange ball (thermolysin). Figures were generated using Molscript 2.0[176] and Raster3D v2.3.[177]

endoprotease complexes deposited in the Protein Data Bank (PDB http://www.rcsb.org/pdb mirrored at http://oca.wehi.edu.au:8383/oca/[22]) through April 2004 were included in this study, updated with only a few key structures beyond that date. The main databases used in this work other than the PDB were the Merops peptidase database (http://www.merops.co.uk),[3] Relibase (http://relibase.ccdc.cam.ac.uk/),[23] PDBsum (http://www.biochem.ucl.ac.uk/bsm/pdbsum/),[24−26] the HIV protease database (http://mcl1.ncifcrf.gov/hivdb/),[27] CATH (http://www.biochem.ucl.ac.uk/bsm/cath/),[28,29] and SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/mirrored at http://scop.wehi.edu.au/scop/).[30] The nomenclature of Schechter and Berger is used for the terminology of peptide substrate and inhibitor binding to proteases (e.g., P3, P2, P1, P1′, P2′, P3′ for inhibitor side chains, S3−S3′ for corresponding enzyme sites).[31]

2. Endopeptidase Architecture and Common Structural Folds

Proteases are many and varied in both their function and their specificity. In addition to assembling proteases into five classes according to their principal catalytic residue, proteases can also be categorized according to the folded structure of their polypeptide backbones.

Aspartic proteases tend to adopt one of two well-known folded structures, the cellular pepsin-like (Figure 1a) or viral retropepsin-like (Figure 1b) fold. The extensively studied pepsin-like fold is a bilobal structure with two similar β-barrel domains, each contributing one aspartic acid residue to the catalytic site.[32−34] Proteases that adopt this fold include pepsin, cathepsin D, endothiapepsin, chymosin, acid protease, and memapsin. A superimposition of the protein structure of a dozen pepsin-like proteases reveals a high degree of structural homology with only a few minor deviations (results not shown). Some differences can be attributed to separate movement of the two domains, as seen in endothiapepsin[35] and plasmepsin II.[36] The retroviral aspartic proteases or retropepsins are distinguished from the cellular pepsin-like proteases by their viral origin and structural differences. They consist of two identical β-barrel domains which form an active homodimer (Figure 1b).[18,37] Retropepsins of HIV, FIV, SIV, RSV, and equine infectious anaemia (EIA) also have much shorter sequences (~100 residues/monomer) than most pepsin-like proteases (>300 amino acids).

Serine proteases adopt two principal structural folds, trypsin-like or subtilisin-like. Most have the trypsin-like (also termed chymotrypsin-like) structure (Figure 1c) which consists of two β barrels with the catalytic triad coming together at the interface of the two domains, not dissimilar to the aspartic proteases.[38] Members of this structural superfamily include prokaryotic serine proteases (e.g., glutamic acid specific, α-lytic, lysyl endopeptidase, streptogrisin A, streptogrisin B), eukaryotic serine proteases (e.g., cathepsin G, crab collagenase, chymase, chymotrypsin, elastase, enteropeptidase, factor D, factor VIIa, factor IXa, factor Xa, granzyme B, kallikrein, matriptase, plasmin, plasminogen, proteinase C, thrombin, trypsin), and viral serine proteases (e.g., hepatitis C and Dengue NS3). Interestingly, there are also viral cysteine proteases that adopt this trypsin fold (see below). The second major fold for serine proteases is the subtilisin-like structure (Figure 1d), as exemplified by kumamolisin, proteinase K, mesentericopeptidase, subtilisin Carlsberg, subtilisin BPN, subtilisin DY, and thermitase. This fold consists of a three-layer αβα sandwich fold or Rossman fold with the catalytic serine and histidine positioned on the ends of adjacent helices (helix 6 and 3, respectively, pdb1scn). Herpes virus serine proteinases or assemblins are another smaller family of serine proteases that consists of a seven-stranded, mostly antiparallel, β-barrel surrounded by seven helices (Figure 1e). The three-dimensional structure of human cytomegalovirus (hCMV) protease was the first structure in this category and has little structural homology with other clans of serine proteases.[39]

Cysteine proteases fall into three main structural categories: viral, papain-like, or caspase-like. The viral cysteine proteases such as hepatitis A, rhinovi-

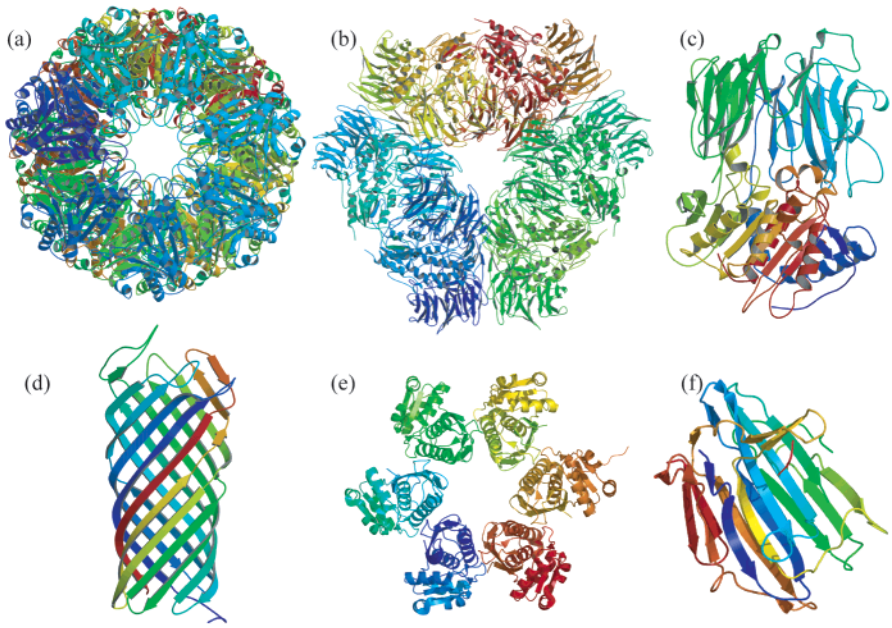


Figure 2. Unusual structural folds of proteases: (a) proteasome (1pma), (b) tricorn protease (1n6e), (c) prolyl oligopeptidase (1qfs), (d) outer membrane protease OmpT, (e) lon protease (1rr9), (f) scytalidoglutamic peptidase (1s2b).

rus, and tobacco etch virus (TEV) adopt a trypsin-like fold with two β-barrel domains as described above (Figure 1c). The papain-like cysteine proteases (e.g., papain, caricain, cathepsins B, H, K, L, S, V, X, Cruzipain, calpain, glycyl endopeptidase, staphopain A) adopt a combined α and β fold (Figure 1f) with an α-helical domain and a β-barrel-type domain. The V-shaped active site lies between the two domains, each providing one of the catalytic residues. The caspase-like cysteine proteases (e.g., caspases 1, 3, 7, 8, 9, gingipain) adopt a Rossman fold similar to the subtilisin-like protease (three-layer $\alpha\beta\alpha$ sandwich, Figure 1d) but differ in the positioning of the catalytic residues with the cysteine and histidine residues situated on adjacent β-strands within the central β-sheet structure.

Metalloproteases (zincins) are unique, as all zinc endoproteases contain a catalytic $\alpha\beta$ domain, known as the thermolysin catalytic domain (Figure 1g, Clan MA, where the zinc ion is coordinated to two histidines from the motif HEXXH).[40,41] Overall, these enzymes are composed of several domains including the thermolysin catalytic domain. Thermolysin itself contains the catalytic $\alpha\beta$ domain along with the α-helical C-terminal domain characteristic of the thermolysin-like family (SCOP). Serralysin is an unusual example that contains both the thermolysin catalytic domain as well as a large β-roll structure.

These 7-fold types share similarities in the positioning of the catalytic residues. In the pepsin-like, retropepsin-like, trypsin-like, and papain-like structural folds, one catalytic residue is presented by each of the two domains (Figure 1a, b, c, and f, respectively). In contrast, the Rossman fold of the subtilisin-like and caspase-like proteases presents the catalytic residues on adjacent structural motifs (helices and strands) from within a single domain (Figure 1d). The remaining two structural folds, herpes virus assemblins and metalloprotease catalytic domains (Figure 1e and g), are unique and present their catalytic machinery from within the β-barrel motif and upon a single helix, respectively.

Threonine proteases are a more recently identified category of proteases that contain a threonine nucleophile at the N terminus of the mature enzyme as their primary catalytic residue.[42,43] The first of this new class is the 20S proteasome (Figure 2a),[44] a large multidomain, multisubunit protease associated with protein degradation in many biological processes. The proteasome structure itself consists of 28 subunits arranged in four rings of seven, termed $\alpha\beta\beta\alpha$ with the interior of the β subunits responsible for proteolytic activity.[45]

Tricorn protease[46−48] and prolyl oligopeptidase[49,50] (Figure 2b and c) are two atypical serine proteases with a similar multidomain structure and degradative function to the proteasome. Following the action of the proteasome, tricorn protease degrades cytosolic proteins and particularly 7−9 residue peptides to produce di- and tripeptides,[46,47] which are then degraded to amino acids by tricorn interacting factors (e.g., tricorn-interacting aminopeptidase F1, an exoprotease with an α/β hydrolase fold, not covered within the scope of this review). The tricorn protease is a 720 KDa hexameric enzyme that can further assemble into an icosahedral capsid. Each subunit of the hexamer is divided into a further five subdomains: a six-bladed β propeller followed by a seven-bladed β propeller with a PDZ-like domain interspersed between two mixed α/β domains. The β-propeller unit covers the active site of the protease. The prolyl oligopeptidase preferentially cleaves peptides less than 30 amino acids in length with specificity for proline at P1. It is involved in the maturation and degradation of neuropeptides and peptide hormones. The structure contains a peptidase domain adopting an α/β hydrolase fold[51] with the catalytic triad being covered by a seven-bladed β-propeller structure which features a central tunnel (Figure 2c). These three proteases, the proteasome, tricorn protease, and

prolyl oligopeptidase, share similarities in their multidomain structures. They form molecular cages which act to exclude tertiary folded substrates from their centrally located active sites.[47] Another unusual protease fold is that of the outer membrane protease OmpT. Its architecture is represented by the large β-barrel structure in Figure 2d.[52] This is the only known protease with a structure of this type and the only current structure of a membrane-spanning protease.

The Lon protease (Figure 2e) is another heterooligomeric serine protease possessing a catalytic Ser-Lys diad.[53] This ATP-dependent protease has been identified in almost all living organisms from Archaea to eubacteria to humans. It is a multidomain enzyme with a single peptide chain of 784 residues. The three functional domains consist of a variable N-terminal domain, an ATPase domain, and the C-terminal protease (P) domain. The P domain (Figure 2e) has a unique fold and forms a hexameric ring and is understood to resemble oligomerization of the holoenzyme.

Glutamic proteases have only very recently been identified as a sixth family of proteases, with the structural determination of scytalidoglutamic peptidase (Scytalidocarboxyl peptidase-B, SCP-B).[54] Formerly described as pepstatin-insensitive carboxyl proteinases,[55] the catalytic diad of this protease consists of glutamic acid and glutamine residues. SCP-B is the first member of the family of eqolisins (derived from the active-site residues E and Q), and its fold is unusual for a protease. The structure, illustrated in Figure 2f, is that of a β-sandwich formed by two seven-stranded antiparallel β-sheets.

Despite the evolution of these quite different folded backbone structures, it will now be shown that all proteases share remarkably similar conformational requirements for ligand recognition in their active sites. Proteases commonly recognize the extended or β-strand backbone conformation in substrates, inhibitors, and products.

3. Protease-Bound Substrate and Product Structures

3.1. Protease-Bound Substrate Structures

Catalytically active proteases do not normally cocrystallize with substrates, without proteolytic action cleaving the substrate. However, when a catalytic residue is mutated to an isosteric residue, such as Asp for Asn in the case of HIV proteases or Cys for Ala in the tobacco etch virus protease (Table 1), it is possible to crystallize protease−substrate complexes with intact substrate trapped in the active site of the enzyme. Some reported examples are listed in Table 1.

HIV-1 protease, which is important for replication of human immunodeficiency virus and a primary target for anti-AIDS drugs, processes the gag and gag-pol polyproteins in 10 known non-homologous sites. Twelve substrates complexed to both HIV-1 and HIV-2 protease enzymes (proteases not displayed) are shown in Figure 3a with all substrates occurring in an extended conformation. Six of these are for HIV-1 protease complexed to peptides corresponding to the cleavage sites CA-p2, MA-CA, p2-NC, p1-p6, RT-RH, and RH-IN, where abbreviations are capsid (CA), matrix (MA), nucleocapsid (NC), reverse transcriptase (RT), RNAse H (RH), and integrase (IN). Three others are for peptides corresponding to the substrate cleavage sites MA-CA, CA-p2, and p1-p6 complexed with inactive/drug-resistant mutant (D25N/V82A) HIV-1 protease.[56] Two other peptide−enzyme complexes contain apparently intact peptide substrates bound in the active site of the catalytically active protein (pdb2hpe, pdb2hpf), presumably the result of autodigestion, although the authors stated that the identity of the side chains was highly ambiguous, suggesting several different peptide fragments with chains oriented in opposite directions in relation to the protease homodimer (Mulichak, A. Private communication). Another peptidic substrate has been cocrystallized with a mutant FIV protease[57] and the structure illustrates that the peptide binds in an almost identical extended backbone (not shown).

Table 1. Known Structures for Protease−Substrate Complexes[a]

protease	PDB	substrate
HIV-1	1a94	RVL-FEANle
	1f7a	KARVL-AEAM
	1kj4	VSQNY-PIVA
	1kj7	ATIM-MQRG
	1kjf	APGNF-LQSRP
	1kjg	AETF-YVDGA
	1kjh	IRKAL-FLDGI
	1mt7	VSQNY-PIV
	1mt8	ARVL-AEAM
	1mt9	RPGNF-LQSRP
HIV-2	2hpe	AAAAA-AAAG
	2hpf	AAAGG-AAG
FIV	3fiv	Ac-(Nal)VL-AE(Nal)-NH2
TEV	1lvb	TENLYFQ-SGT
prolyl oligopeptidase	1e8n	(Abz)GFGP-FG
anthrax lethal factor[b]	1pww	MLARRKKVYP-YPMEPTIAEG-NH2

[a] Nle = norleucine, Ac = acetyl, Nal = naphthylalanine, Abz = 6-aminobenzoic acid. [b] Complex formed in the absence of catalytic zinc.

Recently two tobacco etch virus 3C protease complexes were released.[58] One structure is that of an inactive cysteine protease (C151A) with the substrate TTENLYFQ-SGT (Figure 3b, gray = carbon atoms), the other being that of a catalytically active mutant (S219D) in complex with the product ENLYFQ. In both cases the ligand is present in the active site of the protease in an extended conformation. This study illustrates the importance of residues at P6, P3, P1, and P1′ for specificity. These results also give insight into substrate binding of other 3C-type proteases such as that of poliovirus.

The crystal structure of the anthrax lethal factor in complex with a peptide substrate, based on a consensus sequence of the mitogen-activated protein kinase kinase (MAPKK), has been recently updated (Figure 3c).[59] The 20-residue peptide (LF20) is found to bind in an extended conformation along the 40-Å long substrate binding groove and displays the enzymes' preference for hydrophobic residue s at P1′ (tyrosine) and P2 as well as unusual selectivity for basic residues at several positions toward the N terminus of the substrate.

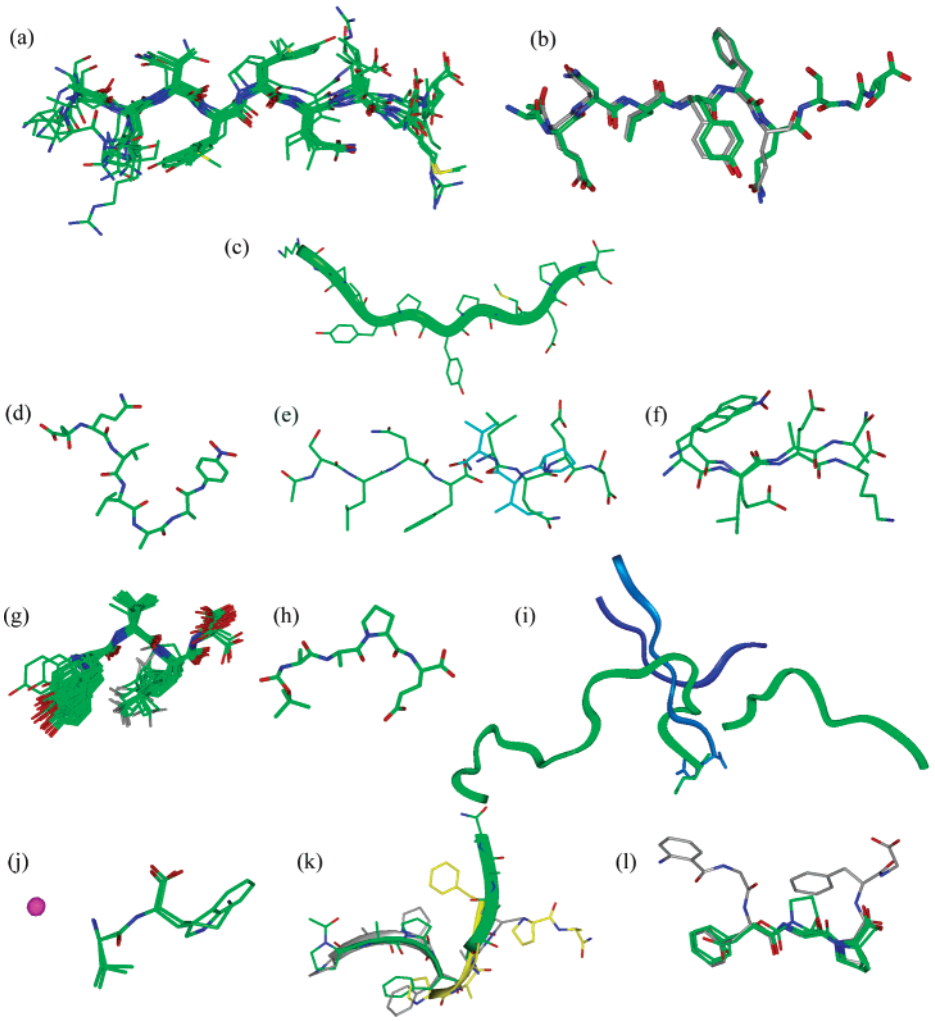


Figure 3. Protease-bound substrates and hydrolysis products (proteases not shown; color coding green = carbon, blue = nitrogen, red = oxygen, yellow = sulfur). (a) Twelve substrates bound to HIV protease; HIV-1 protease (D25N, D125N) 1a94, 1f7a, 1kj4, 1kj7, 1kjf, 1kjg, 1kjh, 1mt7, 1mt8, 1mt9; HIV-2 protease (D25, D125) 2hpe 2hpf. (b) Tobacco etch virus substrate and product (gray carbon atoms). (c) Anthrax lethal factor substrate (ribbon). (d) Papain-bound succinyl-QVVAA-pNA. (e) HIV-1 protease cleavage products (1hte, 1ytg light blue, 1yth). (f) SIV hydrolysis products. (g) Human cytomegalovirus product ensemble P4−P1 (Lys ϵN shown in gray). (h) Glutamyl endopeptidase II product. (i) Botulinum neurotoxin type B cleavage products of synaptobrevin-II (green ribbon, 1f83) and the product-bound state of BoNT serotype A (blue ribbon, 1e1h). (j) Thermolysin cleavage products with adjacent catalytic zinc ion. (k) Proteinase K products. (l) Prolyl oligopeptidase substrate and product complexes. All substrates, products, and inhibitors are presented with their N terminus to the left progressing to the C terminus on the right, which corresponds to the description of P3 to P3′ from left to right. All images created within the InsightII modeling environment unless otherwise stated.[178]

One unusual "substrate" complex is that between the wild-type cysteine protease papain and the substrate-like succinyl-QVVAA-pNA (Figure 3d).[60] The pNA substrate remains unprocessed because it is lodged in the active site in a non-extended conformation. This is the result of a mismatch between the designed peptidic substrate and the specificity of this cysteine protease (preference for a residue bearing a large hydrophobic side chain at the P2 position). The substrate was based on the well-conserved central loop (see below) of the cystatin superfamily (QVVAG), which acts as an endogenous inhibitor of cysteine proteases. The pNA substrate binds in the vicinity of active site but not in the expected position, which was predicted to correspond to the binding of the central cystatin loop. However, it occupies the non-prime site normally occupied by the N terminus of cystatins. When the pNA molecule is superimposed on the central loop region of cystatin A that it is intended to mimic (1nb3), it displays a very similar conformation (turn/loop, result not shown). This unusual binding can be explained by the tenet that sequence determines structure, and therefore, this substrate is not recognized by the protease for hydrolysis.

3.2. Protease-Bound Product Structures

Most attempts to crystallize catalytically active proteases with substrates have failed due to rapid hydrolysis of the substrate followed by product dissociation from the active site. There are a few cases (Table 2) where one or more *products* have been trapped in the active site long enough to permit cocrystallization with protease.

The structures of HIV protease, SIV protease, and TEV have all been solved with a product peptide in

Table 2. Known Structures for Protease−Product Complexes

protease	PDB	products in protease active site
HIV-1	1hte, 1ytg, 1yth,	P3−P1 and P1−P4, AcP4−P1,
SIV	1yti, 1ytj	P1−P4, P1−P4
TEV	1lvm	P6−P1
CMV	1bfz (NMR)	P4−P1 (hydrolysis product)
glutamyl endoprotease II	1hpg	P4−P1 (pNA product),
botulinum neurotoxin	1f83, 1e1h	Sb2[a] residues 33−88, self
thermolysin	3tmn, 8tln	P1−P2
proteinase K	1pfg, 1pj8, 1pek	P4−P4
prolyl oligopeptidase	1e8m, 1h2z, 1o6f, 1o6g	P3−P2

[a] Sb2 = Synaptobrevin-II.

the active site (Figure 3e, 3f, and 3b (gray carbon atoms), respectively). Each was bound within the active site in an extended β-strand conformation. Two products are bound to HIV-1 protease in an unexpected manner. The C-terminal product PIV-NH_2 (Figure 3e, light blue) binds in the reverse of what is expected (C-terminal amine interacting with the catalytic aspartates, Valine at P1′) and acts as an inhibitor following substrate cleavage.[61] Further evidence for the recognition of an extended conformation is the NMR study of the N-terminal cleavage products formed through autocatalysis by the aspartic protease of human cytomegalovirus (hCMV). The results show a well-defined ensemble of structures in an extended conformation from P1 to P4 when complexed to hCMV (Figure 3g).[62]

Another example of a product analogue complexed with a serine protease is that of Boc-AAPE-COOH and glutamyl endoprotease II.[63] The product of the cleaved peptide-pNA is bound from P4 to P1 of the active site in an extended format shown in Figure 3h.

One three-dimensional structure of a metalloprotease, Botulinum neurotoxin type (Bontoxolysin), has been determined as a complex of cleaved synaptobrevin (Figure 3i, green).[64] This was achieved by almost immediate freezing of the crystals after soaking in a solution containing the synaptobrevin substrate. The resulting structure is a "postcleavage" complex and shows atypical binding of the hydrolysis products. These two peptide products bind in the active site of the protease in a retro manner, where the C terminus binds to the non-prime side of the active site (P3−P1), opposite to what is normally seen. This could be due to a rearrangement of the products after hydrolysis or simply an exception to what is believed to be the binding mode. Bontoxolysin is unusual in that it requires binding of a substrate to two recognition sites, one of which necessitates a substrate of minimum length in addition to a nine-residue region termed the SNARE secondary region (SSR) sequence. A more recent structure of this neurotoxin has confirmed unusual binding and specificity for this metalloprotease. The structure shows a homodimer of the metalloprotease with each monomer representing a cleaved product bound deep within the active site (Figure 3i, blue). The cleaved loop is indeed oriented in the retro manner or opposite of the canonical direction for other metalloproteases and shows atypical ligand binding within the active site.[65]

Two almost accidental product complexes of the bacterial metalloprotease thermolysin have been published.[66,67] The first is a complex containing the cleavage product of the dipeptide Val-Trp from the intended inhibitor mercaptoacetyl-valyl-tryptophan, the latter appearing to be the autocatalytic cleavage product of the C-terminal residues of thermolysin itself, Val-Lys. Both complexes show binding of the dipeptides at S1′−S2′, which are too small to infer any recognition motif (Figure 3j, further discussed below).

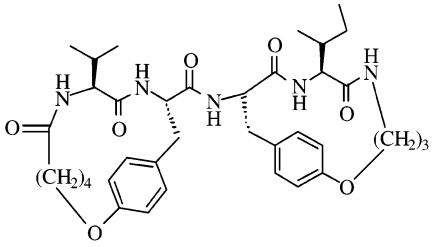
Three unusual product complexes are shown for proteinase K. Two of these complexes are of hydrolyzed peptidic inhibitors, AcPAPF-D-A-A-NH_2 and AcPAPF-D-A-AAA-NH_2 (Figure 3k; gray and green ribbons, respectively).[68,69] The complexes are noncovalent but stabilized by an extensive hydrogen-bonding network, with their inhibition most likely imparted by the presence of the D-alanine at position P1′. These transition-state complexes do show each product in an isolated extended conformation.[69] The third complex is also of a hydrolyzed substrate analogue but is complicated by the inhibitory presence of mercury (Figure 3j, yellow ribbon, Hg not visible).[70] This product complex is quite different from the previous two examples, the N-terminal portion being anchored in the S1 subsite via the first proline residue of the sequence (PAPF-PA).

Prolyl oligopeptidase has the unusual specificity requirement of a proline residue at P1. One substrate complex (S554A mutant) and four product complexes are shown in Figure 3l. The product complexes consist mainly of short hydrolyzed pNA substrates that span from P3 to P1. The substrate (gray carbon atoms) shows the six residues Abz-GFGPFG-OH with recognition and binding limited to the P3−P2′ region. A hydrogen bond is formed between the amide of P1′ and the carbonyl of P2, forming a reverse γ turn centered around the proline at P1.

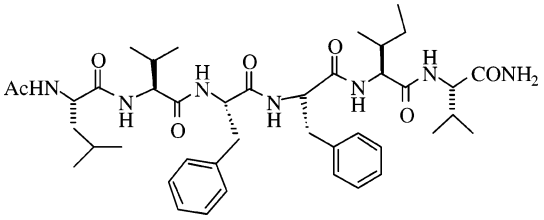
3.3. Cyclic Substrates and Products

A novel category of synthetic protease substrates is macrocyclic substrates, which preserve a fixed β-strand peptide conformation in the substrate. We previously tested the proposition that cyclization might force inhibitors or substrates into a protease-binding conformation, namely, the β strand. For example, we created the novel bis-macrocyclic molecule **1** as a mimic of the acyclic peptide **2**, Ac-LVF↓FIV−NH_2.[71] Each macrocycle in **1** was designed to serve as a structural mimic of a tripeptide in a β strand,[72−75] constrained by the presence of only 15 or 16 ring atoms, 2 trans amide bonds, and a para-

substituted aromatic ring. The two chiral centers in each cycle were derived simply from L-amino acids. This minimalist approach to locking substrates into a protease-binding shape was quite successful and also enabled the cleavage products to be locked into β strands.[71]



1 (Ki 0.8 μM)



2 (Ki 60 μM)

We were particularly interested in using **1** to compare its relative affinity for HIV-1 protease versus its more flexible acyclic hexapeptide analogue **2**. Although substrates are more traditionally examined for K_m and k_{cat}, the ratio k_{cat}/K_m being indicative of substrate efficiency, K_m (= $(k_{-1} + k_2)/k_1$) is a measure of protease affinity of the substrate but also includes the rate of decomposition of the enzyme−substrate complex, and the kinetics were difficult to measure in water because such small molecules are not very water soluble and are only processed very slowly.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

We therefore instead examined substrates **1** and **2** as competitive inhibitors of the processing of another substrate, the fluorogenic 2-Abz-Thr-Ile-Nle-Phe(p-NO_2)-Gln-Arg-NH_2 (AbzNF*-6; K_m = 26 μM), by HIV-1 protease (pH 6.5, I = 0.1 M, 37 °C). By determining K_i (= k_1/k_{-1}) for **1** (0.8 μM) and **2** (60 μM) in these experiments[71] we could obtain protease affinities for the substrates. K_i is the equilibrium constant defining the affinity of inhibitor (I) for enzyme (E). Bicyclic compound **1** was found to have ~80-fold higher affinity for HIV-1 protease than acyclic **2**, corresponding ($\Delta G = -RT \ln K$) to an ~11 kJ/mol difference in binding energy.[71]

$$E + I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} EI$$

Using HPLC to monitor peptide cleavage products, **1** was however found to be a poorer substrate (K_m = 110 μM) than **2** (K_m = 20 μM) as defined by the Michaelis−Menten equilibrium constant, yet **1** is more constrained to an extended conformation and has the higher affinity for the protease as indicated above. We attributed this difference to slower decomposition of the enzyme−substrate complex (ES) for **1** than **2**, the better substrate binding less tightly to the enzyme to enable faster turnover and slow dissociation of the two cyclic products which have higher affinity for the enzyme than acyclic peptides that dissociate readily from the enzyme to enable catalytic turnover.[71]

4. Aspartic Protease Inhibitors

Aspartic proteases[18,76,77] tend to bind 6−10 peptide residues of polypeptide substrates in their active site and cleave them using two catalytic aspartic acid residues from the protease. Most inhibitors tend to bind in the active site of the protease by flanking both sides of the catalytic residues. The majority of known aspartic proteases also have one or more flaps that close down on top of the inhibitor to complete the active site.

A total of 21 aspartic proteases were analyzed in detail here using more than 300 three-dimensional protease−inhibitor structures (Table 3). Figure 4 displays the active-site binding conformation of inhibitors from superimposed protease−inhibitor complexes (protease omitted) for a subset of 20 representative aspartic proteases. More than 300 protease−inhibitor structures were visually inspected for 21 aspartic proteases, 145 of which are displayed in Figure 4. The vast majority of these inhibitors were located in the substrate-binding groove of the protease spanning both sides of the central catalytic aspartate residues (e.g., S4−S3′), the large number of interactions with protease imparting selectivity. Clearly Figure 4 indicates a high degree of conformational homogeneity in the protease-bound inhibitors, with an extended β-strand conformation observed in all but two cases (see below), despite the many differences between these peptidic, peptidomimetic, and non-peptidic inhibitors.

4.1. Viral Aspartic Proteases (Retropepsins)

Human immunodeficiency virus retropepsins (HIV-1&2) have been the most extensively studied aspartic proteases in recent years with over 200 PDB entries for HIV-1 and HIV-2 retropepsin−inhibitor complexes with nine inhibitors used clinically to treat HIV-AIDS. Previously[71] we reported superimposed structures for HIV-1 protease bound to 20 acyclic and 2 cyclic inhibitors, which displayed rmsd values of 0.57 ± 0.15Å. We have now analyzed ~200 such structures but show only a subset of 12 acyclic inhibitors in Figure 4a (left). All structures exhibited an extended β-strand conformation for the inhibitor backbones, as observed in Figure 4a.

A further 15 inhibitors of HIV protease are shown in Figure 4a (right), 11 of which are macrocyclic inhibitors. One particular non-peptidic inhibitor worthy of mention, UCSF8,[78] binds in two separate orientations in two different crystal forms. One

Table 3. Crystal Structure Listing for 145 Inhibitors Complexed with 21 Aspartic Endoproteases

aspartic protease	EC code	PDB codes
aspergillopepsin I (aspartic proteinase)	3.4.23.18	1ize
candidapepsin (secreted aspartic protease)	3.4.23.24	1j71, 1eag, 1zap
cathepsin D	3.4.23.5	1lyb
chymosin (rennin)	3.4.23.4	1czi
endothiapepsin	3.4.23.22	1e5o, 1e80, 1e81, 1e82, 1eed, 1ent, 1epl, 1epm, 1epn, 1epo, 1epp, 1epq, 1epr, 1er8, 1gkt, 1vt, 1gvu, 1gvv, 1gvw, 1gvx, 1od1, 1oex, 2er0, 2er6, 2er7, 2er9, 3er3, 3er5, 4er1, 4er2, 4er4, 5er1, 5er2
equine infectious anaemia	3.4.23.16	1fmb, 2fmb
FIV protease	3.4.23.-	1b11, 1fiv, 2fiv, 3fiv, 4fiv, 5fiv, 6fiv
HIV-1 protease[a]	3.4.23.16	1dif, 1g2k, 1htg, 1htf, 1hvc, 1hvj, 1a3o, 1g35, 1hvh, 1hpx, 1pro, 1hwr, 1cpi, 1mtr, 1d4k, 1d4l, 1f7a, 3aid, 2aid, 1aid, 1b6p, 1b6m, 1b6l, 1b6n, 1b6o, 1b6k, 1b6j.
HIV-2 protease	3.4.23.47	1ida, 1idb, 1hii, 1hsh, 1ivp, 1ivq, 1jld, 2mip, 2hpe, 2hpf, 3upj, 4upj, 5upj, 6upj
memapsin(BACE)	3.4.23.46	1fkn, 1m4h
mucorpepsin	3.4.23.23	2rmp
penicillopepsin	3.4.23.20	1apt, 1apu, 1apv, 1apw, 1bxo, 1bxq, 1ppk, 1ppl, 1ppm, 2wea, 2web, 2wec, 2wed
pepsin A	3.4.23.1	1f34, 1psa, 1pso, 1qrp
plasmepsin-I	3.4.23.38	1qs8
plasmepsin-II	3.4.23.39	1lf2, 1lf3, 1lee, 1m43, 1me6, 1sme
plasmepsin-IV	3.4.23.-	1ls5
renin	3.4.23.15	1bil, 1bim, 1hrn, 1rne, 1smr, 1pr7*, 1pr8*
rhizopuspepsin	3.4.23.21	3apr, 4apr, 5apr, 6apr
rous sarcoma virus	3.4.23.-	1bai
saccharopepsin	3.4.23.25	1dp5, 1dpj, 1fq4, 1fq5, 1fq6, 1fq7, 1fq8, 1g0v, 2jxr
SIV protease	3.4.23.-	1siv, 1tcw, 1yti, 1ytj, 1yth, 1ytg, 2sam

[a] The PDB codes listed denote only the structures presented in this review. As of January 2003 214 structures of HIV protease had been deposited on the PDB. All structures have been analyzed for this study and only a representative number shown.

conformation is consistent with the extended conformation seen for other HIV protease inhibitors (Figure 4a, right; purple, ball-and-stick, pdb:2aid); the other conformation lies essentially above and perpendicular to the expected binding conformation, preventing closure of the flaps of HIV protease.

Numerous other related retroviral proteases have been studied, many being homologues of the autoimmune disease HIV proteases (e.g., FIV, SIV, EIA, HTLV). Feline immunodeficiency virus (FIV) retropepsin is exemplified in Figure 4b by six inhibitor-protease structures and one substrate complex bound to the mutant FIV protease D30N (3fiv). Simian immunodeficiency virus (SIV) retropepsin is illustrated in Figure 4c and shows five inhibitor−protease and two product−protease complexes. Equine infectious anaemia (EIA) retropepsin binding is characterized in Figure 4d by two inhibitors, and Rous sarcoma virus (RSV) retropepsin, identified as a target in tumor-producing viruses, is represented by one inhibitor complex (Figure 4e). Ligands bound to all these retropepsins are in a single conformation, the extended β-strand. This is necessitated by the nature of the active site which, when closed, is in the form of a thin cylinder that can only accommodate peptides and related molecules in a strand conformation, helical or turn-like conformations being too large to occupy the active site.

4.2. Human and Mammalian Aspartic Proteases

Cathepsin D is a lysosomal protease with a broad range of house-keeping functions including the degradation of proteins, antigen processing, and formation of peptide hormones.[79,80] Figure 4f displays the extended conformation of pepstatin, a general aspartic protease inhibitor, bound in the active site of cathepsin D. Chymosin (rennin) is responsible for the physiological hydrolysis of the milk protein κ-casein, and it has historically been used extensively in the manufacture of cheeses. The single inhibitor of chymosin is shown in the extended β-strand conformation bound across the S4−S1 pockets of the active site (Figure 4g).[81] Memapsin-2 (BACE-2 or β-secretase) is one of the key enzymes associated with the processing of amyloid precursor protein (APP) to generate β-amyloid peptide, the causative agent of amyloid plaques in Alzheimer's disease.[82] Figure 4h illustrates the active-site binding conformation of two inhibitors of memapsin. Pepsin A is found in the stomach of vertebrates and is a key enzyme in digestion of proteins.[83] Pepsin A inhibitors bind to the active site of pepsin in the same extended conformation. One unusual proteinaceous inhibitor (PI-3, Figure 4i, ribbon) from the intestinal parasite *Ascaris* binds to pepsin in an unexpected mode,[84] occupying active-site positions P1′−P3′. A large β sheet is also formed between the flap region of pepsin and PI3, stabilizing this flexible β-hairpin common among aspartic proteases. The protease renin is involved in the production of angiotensin II within the renin−angiotensin system, a key mediator in hypertension.[80] Five active-site-bound inhibitors of renin are shown in Figure 4j in an extended manner.

4.3. Parasitic, Plant, and Fungal Aspartic Proteases

Malaria is a disease of endemic proportions throughout the world. Three parasitic proteases from *Plasmodium vivax* and *Plasmodium falciparum*, plasme-

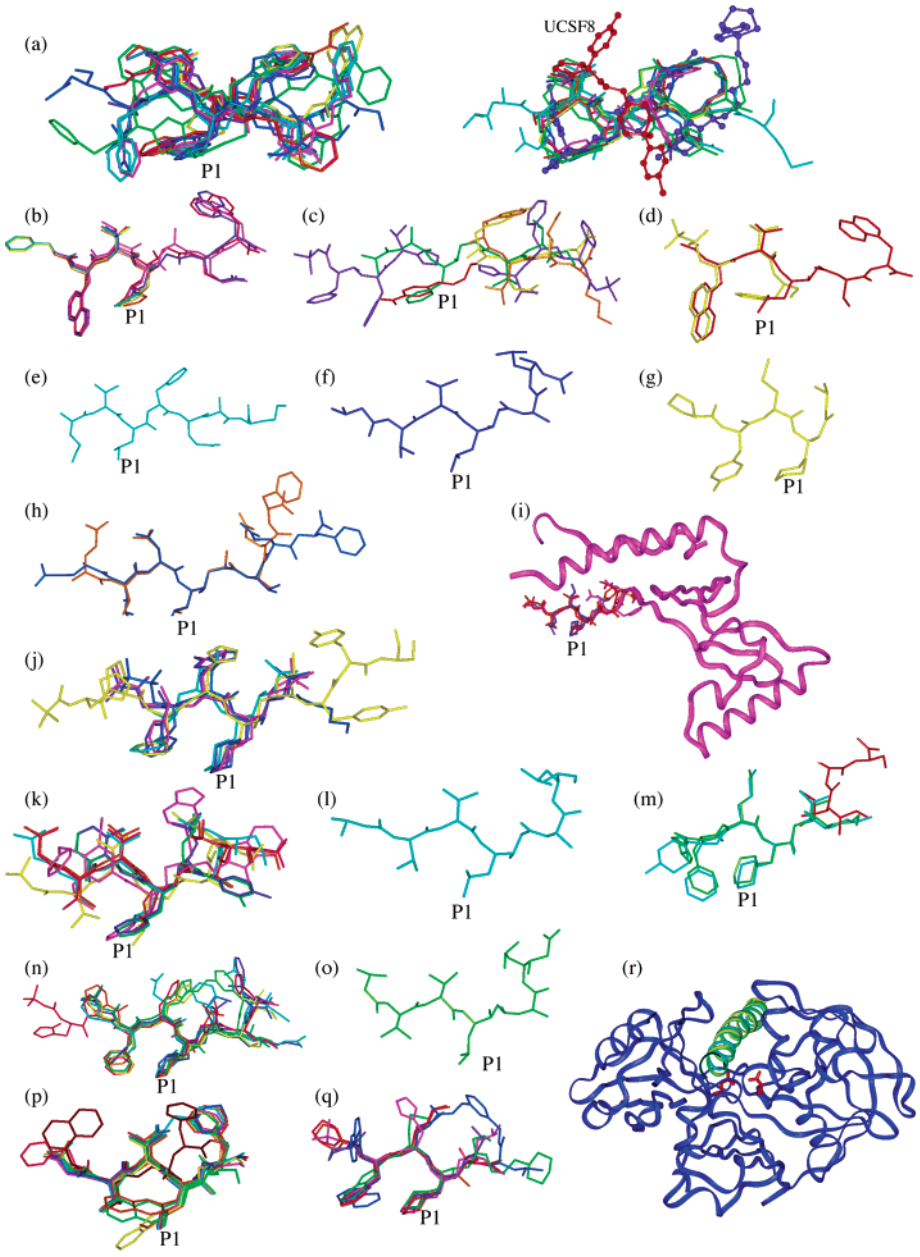


Figure 4. Superimposed aspartic protease−inhibitor structures (proteases omitted; RMSDs for superimposition of protein component of complex; PDB code in parentheses; ligands are represented with main chains running horizontally N to C terminal from left to right) for viral aspartic proteases: (a) [left] human immunodeficiency virus (HIV) retropepsin, RMSD 0.63 Å (1dif, 1g2k, 1htg, 1htf, 1hvc, 1hvj, 1a3o, 1g35, 1hvh, 1hpx, 1pro, 1hwr); [right] human immunodeficiency virus (HIV) retropepsin; RMSD 0.53 Å (1cpi, 1mtr, 1d4k, 1d4l, 1f7a, 3aid, 2aid, 1aid, 1b6p, 1b6m, 1b6l, 1b6n, 1b6o, 1b6k, 1b6j). Red and purple inhibitors (ball-and-stick) display the different binding orientations of UCSF8. (b) Feline immunodeficiency virus (FIV) retropepsin, RMSD 0.20 Å (1b11, 1fiv, 2fiv, 3fiv, 4fiv, 5fiv, 6fiv). (c) Simian immunodeficiency virus (SIV) retropepsin; RMSD 0.68 Å (1siv, 2sam, 1yti, 1ytj, 1yth, 1ytg, 1tcw). (d) Equine infectious anaemia (EIA); RMSD 0.16 Å (1fmb, 2fmb). (e) Rous sarcoma virus (RSV), (1bai). Human and mammalian aspartic proteases: (f) cathepsin D, (1lyb); (g) chymosin (1czi); (h) memapsin (BACE), RMSD 0.22 Å (1m4h, 1fkn); (i) pepsin A RMSD 0.65 Å (1f34 (Ribbon), 1psa, 1pso, 1qrp). (j) Renin, RMSD 0.63 Å (1bil, 1bim, 1hrn, 1rne, 1smr). Parasitic aspartic proteases: (k) plasmepsins I, II, and IV RMSD 1.00 Å (1lee, 1lf2, 1lf3, 1ls5, 1m43, 1qs8, 1sme). Fungal aspartic proteases: (l) aspergillopepsin I (1ize); (m) candidapepsin (secreted aspartic protease) RMSD 0.63 Å (1eag, 1j71, 1zap); (n) endothiapepsin RMSD 0.44 Å (1e50, 1gkt, 1gvt, 1gvu, 1gvv, 1gvw, 1gvx, 5er2), (o) mucorpepsin (2rmp); (p) penicillopepsin; RMSD 0.27 Å (1apu, 1apv, 1apw, 1apt, 1ppk, 1ppm, 1ppl, 1bxo, 1bxq, 2wea, 2web, 2wec, 2wed); (q) saccharopepsin RMSD 0.30 Å (1fq4, 1fq5, 1fq6, 1fq7, 1fq8, 2jxr); (r) saccharopepsin (1dp5, enzyme shown as blue ribbon with catalytic aspartates shown in red and helical inhibitor in yellow, 1dpj, 1g0v). RMSDs correspond to protease component.

psins I, II, and IV, have been crystallographically characterized with multiple inhibitors. Figure 4k shows the superimposition of all seven inhibitor−protease structures of these enzymes, illustrating extended inhibitor backbones in the active site of the enzyme.

Two fungal proteases, aspergillopepsin I and candidapepsin (secreted aspartic protease), are represented in Figure 4l and 4m by one- and three-ligand structures, respectively. One ligand is a hydrolysis product of unknown origin, bound from P1′ to P4′ in the active site of candidapepsin (Figure 4m, red).

Endothiapepsin has been extensively studied, and a representative subset of eight enzyme complexes shows inhibitors in the classical extended conformation (Figure 4n). An inhibitor of mucoropepsin is shown bound to the enzyme in an extended conformation (Figure 4o). Penicillopepsin is another fungal enzyme of the pepsin-like family of proteases, and its homology to mammalian enzymes forms the basis for extensive structural studies. The protease-bound structures are shown for 13 inhibitors in Figure 4p.

The yeast protease saccharopepsin represents the final aspartic protease discussed in this section and reveals the exceptional binding mode of an endogenous 68-residue, proteinaceous inhibitor IA3. Figure 4q displays six inhibitors bound in the typical extended fashion, but inhibitor IA3 forms an almost perfect α helix and binds within the active-site cleft of the enzyme (Figure 4r).[85] The large size and helical shape of this inhibitor causes the flexible flap region to remain in an open conformation, some 8 Å away from a corresponding complex with a small peptidic inhibitor. This is an exception to the hypothesis that proteases bind substrates, inhibitors, and products in only an extended β-strand conformation.

5. Metalloprotease Inhibitors

All known metalloproteases[86−89] use a Zn^{2+} ion to catalyze the hydrolysis of a peptide bond. The metal is tetrahedrally coordinated to three donor groups from the enzyme and a water molecule that is also hydrogen bonded to the carboxylate side chain of a glutamic acid, which activates it for nucleophilic attack. Most metalloprotease inhibitors contain a zinc-binding ligand like hydroxamate, carboxylate, phosphinate, or a thiol and target only one side of the active site.

Figure 5 (Table 4) depicts 123 inhibitor structures in their protease-bound conformations as revealed by crystal structures for 20 metalloproteases. The inhibitors are mostly in an extended strand conformation, except for those of thermolysin and one inhibitor of neutrophil collagenase (MMP-8), both discussed below.

5.1. Bacterial Metalloproteases

One family of secreted bacterial metalloproteases is the serralysins. These belong to the metzincins, which are characterized by their catalytic zinc-binding domain motif HEXXHXXGXXH with an additional conserved methionine.[40] *Pseudomonas* and *Serratia* (endogenous endoproteases aeruginolysin and serralysin, respectively) are bacterial pathogens responsible for many infections including pneumonia and surgical wound infection as well as acting as virulence factors associated with tissue damage leading to anaphylactic products.[90,91] These gram-negative bacteria secrete a proteinaceous inhibitor which is believed to protect periplasmic proteins against the protease. Figure 5a and b show the three-dimensional binding conformations of the endogenous inhibitor in complex with each bacterial protease (ribbon). The main mode of inhibition is via the five N-terminal residues, which bind to the prime side of the active site in an extended conformation. Two small peptidic inhibitors are also shown binding to the non-prime side in an extended conformation.

Several toxic bacterial metalloproteases, anthrax lethal factor and bontoxilysin, are associated with the pathogenesis of anthrax and botulism, respectively, and their substrate and product complexes (Figure 5c and d) have been discussed above. Thermolysin is a bacterial thermostable metalloprotease which has been extensively characterized. The majority of thermolysin−inhibitor complexes contain short inhibitors that only span from S1′ to S2′ and include a zinc-binding moiety. Thermolysin has a specificity for hydrophobic residues on the prime side of the active site, not the non-prime side as is generally the case with other protease families. On close inspection these ligands appear to orient away from the active site of the enzyme, probably due to truncation and blockage of the active-site cleft by a calcium-binding loop not seen in other metalloproteases. Figure 5e shows seven inhibitors of thermolysin exhibiting this unusual binding mode.

5.2. Human and Mammalian Metalloproteases

Angiotensin-converting enzyme (ACE) is part of the renin−angiotensin system which produces angiotensin II. It is not strictly an endopeptidase but rather a carboxyl dipeptidase as it cleaves the two C-terminal residues of angiotensin I. Figure 5f shows the receptor-bound conformation of the current hypertensive drug Lisinopril bound to angiotensin-converting enzyme-I.[92] A subtype, ACE-II, is also now known (pdb1r42, pdb1r4l).[93]

TACE (tumor necrosis factor α-converting enzyme, also known as ADAM17) is an endopeptidase responsible for release of TNFα, a major immunomodulatory and proinflammatory cytokine. Its catalytic domain is similar to other metzincins (HEXXHXXGXXH) as well as the thermolysin-like fold. Figure 5g shows a hydroxamate inhibitor binding in the prime side of the active site via a distinct extended conformation.[94]

The matrix metalloproteases (MMPs, matrixins) are a large family of related proteases involved in extracellular matrix degradation.[87,89] This degradation is part of many pathological and physiological processes including the hydrolysis of collagen. Endogenous inhibitors maintain normal control of these processes. The major specificity determinant of MMP's appears to be amino acid residue recognition at the S1′ subsite. Figure 5h shows seven inhibitors of collagenase-1 (MMP-1) in their protease-bound conformations. Collagenase-2 (MMP-8) is represented in Figure 5i by 13 inhibitors bound across the active site, the majority in an extended manner. Several inhibitors of neutrophil collagenase bind from P1 to P3′ (Figure 5i, purple, yellow, and dark blue), and two bind from P3 to P1 (Figure 5i, red and green) with all inhibitors adopting the extended strand conformation. However, 1-hydroxylamine-2-isobutyl-malonyl-Ala-Gly-NH_2 (Figure 5i, light blue; pdb 1jaq) appears to be an exception,[95] binding in an unusual turn conformation (hydroxylamine oxygen to Cα of Gly distance is <7 Å). This is attributed to a simple

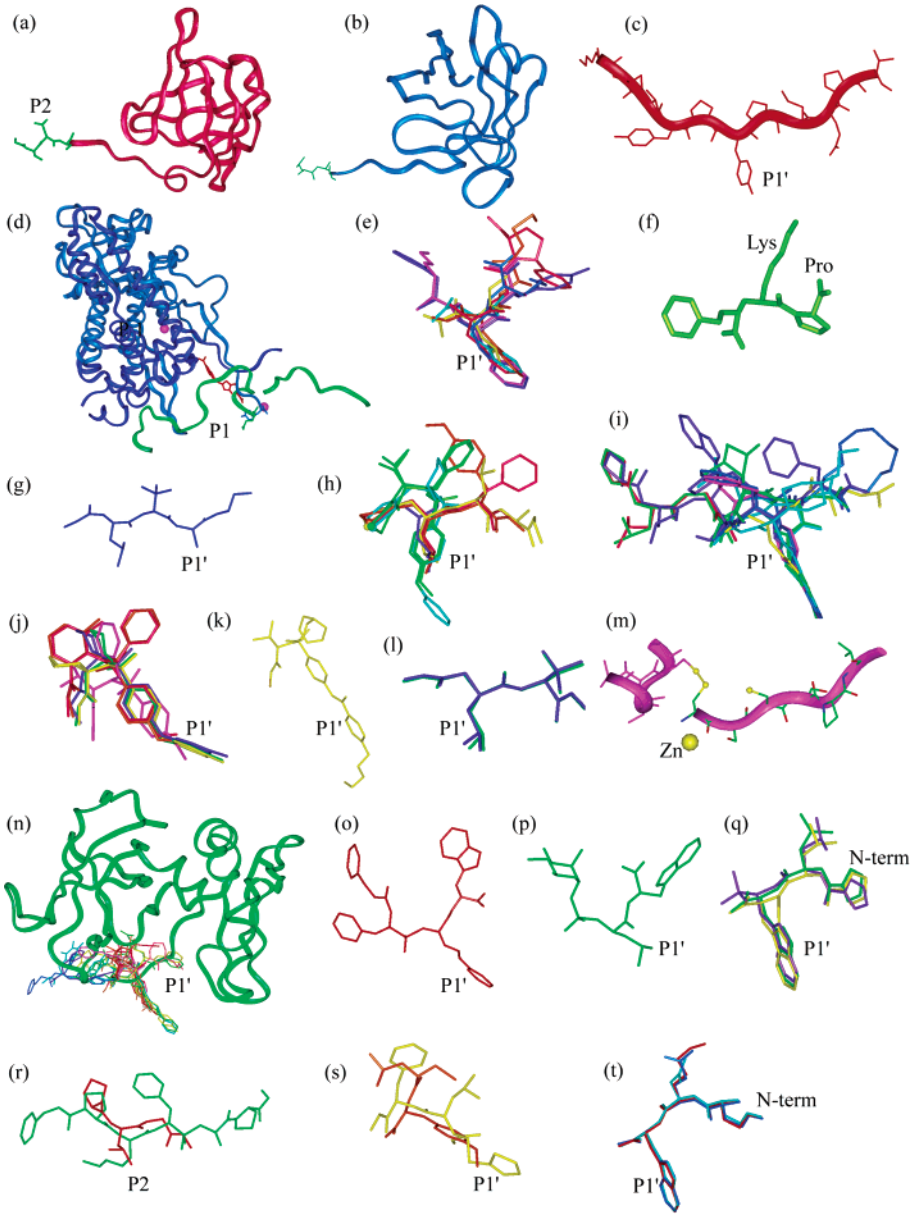


Figure 5. Superimposed metalloprotease−inhibitor structures (RMSDs quoted for superimposition of protein component of complex; PDB code in parentheses, protease omitted; ligands are represented with main chains running horizontally N to C terminal from left to right where appropriate) for bacterial metalloproteases: (a) aeruginolysin RMSD 0.52 Å (1jiw (ribbon), 1kap); (b) serralysin RMSD 0.46 Å (1smp (ribbon), 1af0); (c) anthrax lethal factor; (1pww); (d) bontoxilysin; RMSD 0.51 Å (1f83, 1fqh, substrate (blue) and inhibitor (red)); (e) thermolysin RMSD 0.189 Å (1hyt, 1lnd, 1qf0, 1qf1, 1qf2, 1tln, 7tln). Human and mammalian metalloproteases: (f) angiotensin-converting enzyme-1 inhibitor lisinopril (1o86); (g) ADAM17 endopeptidase (TNF α-converting enzyme), (1bkc); (h) collagenase-1/MMP-1 RMSD 0.77 Å (1cgl, 1fbl, 1hfc, 2tcl, 3ayk, 4ayk, 966c); (i) collagenase 2/MMP-8 RMSD 0.29 Å (1a85, 1a86, 1bzs, 1i73, 1i76, 1jan, 1jao, 1jap, 1jj9, 1kbc, 1mmb); (j) collagenase-3/MMP-13 (830c, 456c, 1cvx, 1eub, 1fls, 1fml); (k) gelatinase A/MMP-2 (1hov); (l) gelatinase B/MMP-9 (1gkc, 1gkd); (m) MMP-14 (1bqq, 1buv) inhibitor is broad-spectrum proteinaceous tissue inhibitor of matrix metalloproteinase (TIMP); only the active-site binding residues are shown (disulfide bridge yellow, zinc yellow ball); (n) stromelysin-1/MMP-3 RMSD 0.98 Å (1d8f, 2usn, 3usn, 1b8y, 1bqo, 1caq, 1ciz, 1d5j, 1sln, 1biw, 1bm6, 1g49, 1usn, 1d8m, 1d7x, 1hfs, 1uea (TIMP, ribbon)); (o) stromelysin 3/MMP-11 (1hv5); (p) Neprilysin (1dmt). Other metalloproteases: (q) adamalysin RMSD 0.26 Å (2aig, 3aig, 4aig); (r) astacin (1qji, 1qjj); (s) atrolysin C (1atl, 1dth); (t) atrolysin E (1kug, 1kui, 1kuk).

rotation about the Cα−C of Ala, so that the main-chain Ala-Gly segment is oriented into S1′ while the Ala side chain projects along the extended axis. Thus, in this case the main chain fulfils the role of a side chain by occupying the extended format.

Collagenase-3 (MMP13) binds shorter inhibitors with a large hydrophobic group at P1′, and the structures of six inhibitors (Figure 5j) illustrate this well. Figure 5k shows one inhibitor of gelatinase A (MMP-2) which also prefers smaller inhibitors with large hydrophobic groups occupying S1′. Gelatinase B (MMP-9) binds a more classical peptidic inhibitor and can be seen in Figure 5l in an extended fashion.

The active-site binding domain of a specific endogenous inhibitor (tissue inhibitor of metalloproteases, TIMP-2) is shown in complex with membrane-type matrix metalloprotease 1 (MMP-14, Figure 5m). The N-terminal residues occupy the prime side of the active site in an extended fashion with coordination to the catalytic zinc mediated by the free amine

Table 4. Crystal Structure Listing for 123 Inhibitors Complexed with 20 Metalloendoproteases

metalloprotease	EC code	PDB codes
ADAM17 endopeptidase(TNF α-converting enzyme)	3.4.24.86	1bkc
adamalysin	3.4.24.46	2aig, 3aig, 4aig
aeruginolysin	3.4.24.40	1jiw, 1kap
angiotensin-converting enzyme-1 (ACE)	3.4.15.1	1o86
anthrax lethal factor	3.4.24.83	1pwv, 1pww (substrate)
astacin	3.4.24.21	1qji, 1qjj
atrolysin C	3.4.24.42	1atl, 1dth
atrolysin E	3.4.24.44	1kug, 1kui, 1kuk
bontoxilysin	3.4.24.69	1g9a, 1g9b, 1g9c, 1g9d, 1e1h, 1f83
collagenase-1 (interstitial, MMP-1)	3.4.24.7	1cgl, 1fbl, 1hfc, 2tcl, 3ayk, 4ayk, 996c
collagenase-2 (Neutrophil, MMP-8)	3.4.24.34	1a85, 1a86, 1i73, 1i76, 1jan, 1jao, 1jap, 1jaq, 1jh1, 1jj9, 1kbc, 1mmb, 1mnc,
collagenase-3 (MMP-13)	3.4.24.-	830c, 456c, 1cxv, 1eub, 1fls, 1fm1
gelatinase A (MMP-2)	3.4.24.24	1hov
gelatinase B (MMP-9)	3.4.24.35	1gkc, 1gkd
matrilysin (MMP-7)	3.4.24.23	1mmp, 1mmq, 1mmr
membrane-type matrix metalloproteinase 1 (MMP-14)	3.4.24.80	1bqq, 1buv
nepralysin	3.4.24.11	1dmt
serralysin	3.4.24.40	1af0, 1smp
stromelysin-1 (MMP-3)	3.4.24.17	1b3d, 1b8y, 1biw, 1bm6, 1bqo, 1c3i, 1c8t, 1caq, 1ciz, 1d5j, 1d7x, 1d8f, 1d8m, 1g05, 1g49, 1g4k, 1hfs, 1hy7, 1kn8, 1sln, 1uea, 1ums, 1umt, 1usn, 2srt, 2usn, 3usn
stromelysin-3 (MMP-11)	3.4.24.-	1hu5
thermolysin	3.4.24.27	1fjt, 1fjw, 1gxw, 1hyt, 1kei, 1kjo, 1kjp, 1kkk, 1kl6, 1kr6, 1kro, 1ks7, 1kto, 1lna, 1lnb, 1lnc, 1lnd, 1lne, 1lnf, 1ln0, 1os0, 1qf0, 1qf1, 1qf2, 1stl, 1tlp, 1tlx, 1tmn, 2tlx, 2tmn, 3tmn, 4tln, 4tmn, 5tln, 5tmn, 6tmn, 7tln, 8tln

nitrogen and the carbonyl oxygen of the N-terminal cysteine residue. The key to inhibition is the disulfide-bonded portion, which occupies the non-prime side of the active site. These protein inhibitors are able to satisfy the requirements of extended strand recognition without presenting an amide bond for hydrolysis.[96] The three-dimensional structure of stromelysin-1 (MMP-3) was the first to be solved in complex with a TIMP (TIMP-1)[97] and is displayed in Figure 5n (green ribbon) along with 16 conventional peptidic inhibitors binding across the active site in an extended conformation. The large P1′ substituent is also utilized in these inhibition protocols. One inhibitor is shown for stromelysin-3 (MMP-11), which shows the characteristically large hydrophobic group at P1′ (Figure 5o). This conformation does appear to look more like a turn or loop and is probably due to the cocrystallization of a CHAPS molecule (not shown), which makes hydrogen-bonding interactions with the phosphinic inhibitor as well as blocking further extension into S1 and S2 subsites of the protease.

Neprilysin is a non-matrix zinc endoprotease and is associated with the inactivation and degradation of signal peptides including the enkephalins, bradykinin and endothelin.[98] Figure 5p shows the binding conformation of the generic metalloprotease inhibitor phosphoramidon (*N*-((α)-rhamnopyranosyloxyhydroxyphosphinyl)-L-Leu-L-Trp) complexed to the catalytic zinc deep within a large cavity of neprylisin. The binding of the C-terminal portion to S1′−S2′ is reminiscent of thermolysin inhibitor complexes. Indeed, the conformation of phosphoramidon is almost identical to that of the inhibitor complexed to thermolysin (not shown). This metalloprotease is made up of two large domains that come together to form a large internal cavity reminiscent of the degradative serine proteases, tricorn protease and prolyl oligopeptidase as well as ACE.

5.3. Other Metalloproteases

Snake venom contains metalloproteases known as adamalysins or reprolysins. Their hydrolytic action can result in extensive hemorrhage in prey. These proteases form the metzincin family, which also includes the astacins as well as the serratia and matrixin (MMP) families are discussed above.[40,41] Adamalysin II is inhibited by three molecules adopting a retro mode of binding (reverse C−N terminus). These inhibitors are based on endogenous pyroglutamate containing tripeptides that act as competitive inhibitors.[99] Figure 5q shows three inhibitors binding across the prime side of the active site in an extended manner with the C-terminal tryptophan occupying the S1′ subsite. Astacin is a digestive enzyme found in crustaceans, and two inhibitors are shown in Figure 5r bound in a classical extended conformation with one particular phosphinic inhibitor spanning both the nonprime and prime subsites. Two hemorrhagic snake venom metalloproteases, atrolysin C and atrolysin E, complete this section on metalloproteases and their ligands. Batimistat, a potent inhibitor, is bound to atrolysin C (Figure 5s, yellow) and binds in a relatively extended conformation. Three pyroglutamate-containing peptides are bound within the prime side of atrolysin E in a retro manner, a fashion similar to the adamalysin inhibitors discussed above (Figure 5t).

To date, inhibitors of metalloproteases have incorporated a zinc-binding moiety to coordinate to the catalytic metal ion. This is a unique requirement within the family of proteases and creates a complication when determining the secondary structure based on the classical definition of the dihedral angles phi and psi centered around the α carbon of a peptide.

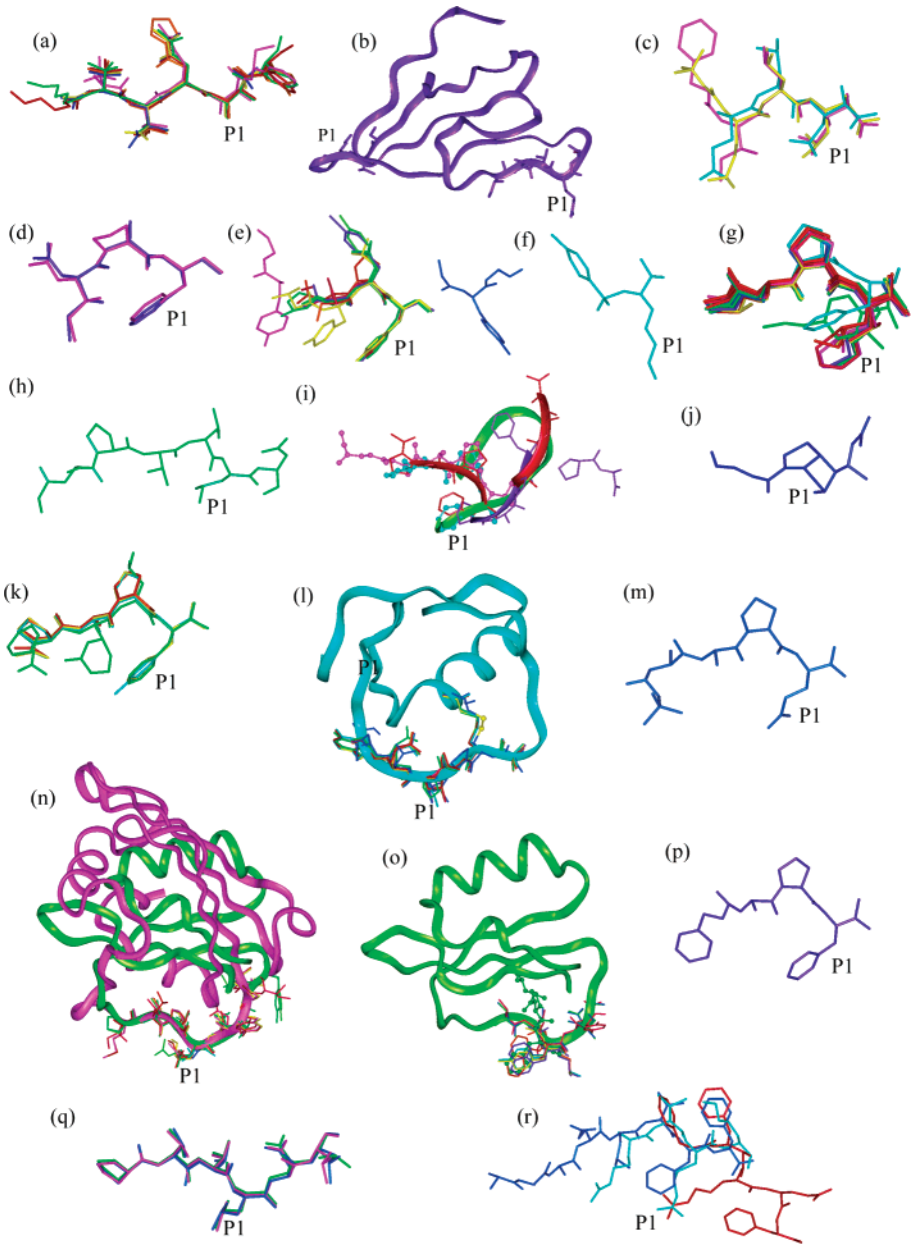


Figure 6. Superimposed serine protease−inhibitor structures (RMSDs quoted for superimposition of protein component of complex; PDB codes in parentheses, protease omitted; ligands are represented with main chains running horizontally N to C terminal from left to right where appropriate) for viral serine proteases. (a) Human cytomegalovirus assemblin RMSD 0.56 Å (1jq7, 1njt, 1nju, 1nkk, 1nkm, 2wpo). (b) Dengue NS3 protease (1df9, ribbon). (c) Hepacivirin RMSD 0.19 Å* (1dxw, 1dy8, 1dy9). Bacterial serine proteases: (d) kumamolisin RMSD 0.27 Å (1gtj, 1gtl); (e) sedolisin (pseudomonapepsin) 0.17 Å (1ga1, 1ga4, 1ga6, 1kdv, 1kdy, 1kdz, 1ke1, 1ke2); (f) lysyl endopeptidase (1arc); (g) α-lytic endopeptidase RMSD 0.16 Å (1gbb, 1gbc, 1gbf, 1gbh, 1gbi, 1gbk, 1gbl, 1gbm, 1p03, 1p05, 1p10, 2lpr, 3lpr, 3pro, 5lpr, 6lpr, 7lpr, 8lpr, 9lpr); (h) mesentericopeptidase (1mee); (i) proteinase K RMSD 0.38 Å (1bjr (green ribbon), 1pek, 1pfg (red ribbon), 1pj8 (purple ribbon), 3prk (ball-and-stick)); (j) signal peptidase I (1b12); (k) streptogrisin A RMSD 0.12 Å (1sgc, 3sga, 4sga, 5sga); (l) streptogrisin B RMSD 0.23 Å (1cso (ribbon), 1ct2, 1ct4, 1ds2, 1sgp, 1sgq, 1sgr, 2sgp, 3sgb, 4sgb); (m) steptogrisin E glutamyl endopeptidase II (1hpg); (n) subtilisin BPN 0.41 Å (1sbn (green ribbon), 1sib, 1sua, 1sue, 1sup, 2sic (ribbon), 2sni, 3sic, 5sic); (o) subtilisin Carlsberg RMSD 0.34 Å (1a10, 1av7, 1avt, 1cse (ribbon), 1scn, 1vsb, 2s, 3vsb); (p) subtilisin DY (1bh6); (q) thermitase RMSD 0.56 Å (1tec, 2tec, 3tec); (r) tricorn protease (1n6d, 1n6e, 1n6f). (Asterisk (*) indicates RMSD between 1dy8 and 1dy9 as 1dxw in NMR structure.)

6. Serine Protease Inhibitors

Serine proteases[100−105] are one of the largest classes of proteases studied with almost 800 structures deposited in the PDB with at least one-third of these are structures of thrombin and trypsin. Serine proteases are classically categorized by their substrate specificity, notably by the residue at P1 being trypsin-like (Lys/Arg preferred at P1), chymotrypsin-like (large hydrophobic residues such as Phe/Tyr/Leu at P1), or elastase-like (small hydrophobic residues such as Ala/Val at P1). Serine proteases that deviate from these categories are the subtilisins, which possess broad specificity, the herpes virus type serine proteases, and the unusual glutamic acid specific endoprotease. Figures 6−8 illustrate 220 structures containing ligands bound to the active site of 44 serine

Table 5. Crystal Structure Listing for 220 Inhibitors Complexed with 44 Serine Endoproteases

serine protease	EC code	PDB codes
brachyurin C	3.4.21.32	1azz
cathepsin G	3.4.21.20	1kyn, 1au8, 1cgh
chymase	3.4.21.39	1klt, 1pjp
α-chymotrypsin	3.4.21.1	1acb, 1cgi, 1cgj, 1cho, 1gl0, 1gl1, 1hja, 1mtn, 2cha, 6cha
γ-chymotrypsin	3.4.21.1	1ab9, 1afq, 1ca0, 1cbw, 1gcd, 1gct, 1gg6, 1ggd, 1gha, 1ghb, 1gmc, 1gmd, 1gmh, 1k2i, 1n8o, 1vgc, 2gct, 2gmt, 2vgc, 3gch, 3gct, 3vgc, 4gch, 4vgc, 6cha, 6gch, 7gch
δ-chymotrypsin		1dlk
complement factor D	3.4.21.46	1bio, 1dfp, 1dic
cytomegalovirus assemblin	3.4.21.97	1jq7, 2wpo, 1njt, 1nju, 1nkk, 1nkm
dengue NS3 protease	3.4.21.91	1df9
human neutrophil elastase	3.4.21.37	1b0f, 1h1b, 1hne, 1ppf, 1ppg
pancreatic elastase I	3.4.21.36	1b0e, 1bma, 1btu, 1e34, 1e35, 1e36, 1e37, 1e38, 1eai, 1eas, 1eat, 1eau, 1ela, 1elb, 1elc, 1eld, 1ele, 1elf, 1elg, 1esb, 1est, 1fle, 1gvk, 1h9l, 1hax, 1haz, 1inc, 1jim, 1mcv, 1mmj, 1nes, 1okx, 1qgf, 1qix, 1qr3, 2est, 4est, 5est, 7est, 9est
pancreatic elastase II	3.4.21.71	1bru
enteropeptidase	3.4.21.9	1ekb
coagulation factor VIIa	3.4.21.21	1cvw, 1dan, 1fak, 1qfk
coagulation factor IXa	3.4.21.22	1pfx, 1rfn
coagulation factor Xa	3.4.21.6	1ezq, 1f0r, 1f0s, 1fax, 1fjs, 1fxp, 1g2l, 1g2m, 1ioe, 1iqe, 1iqf, 1iqg, 1iqh, 1iqi, 1iqj, 1iqk, 1iql, 1iqm, 1iqn, 1j17, 1kig, 1ksn, 1kye, 1lqd, 1lpg, 1lpk, 1lpz, 1mq5, 1mq6, 1nfu, 1nfw, 1nfx, 1nfy, 1p0s, 1xka, 1xkb,
granzyme B (natural killer cell protease)	3.4.21.79	1fi8, 1iau
hepatitis C NS3 4a protease(hepacivirin)	3.4.21.98	1dxw, 1dy8, 1dy9
kallikrein	3.4.21.35	1hia, 1l2e, 1lo6 2kai, 2pka
kexin	3.4.21.61	1ot5, 1r64
kumamolisin	3.4.21.-	1gtj, 1gtl,
lysyl endopeptidase	3.4.21.50	1arc
α-lytic protease	3.4.21.12	1gbb, 1gbc, 1gbd, 1gbf, 1gbh, 1gbi, 1gbk, 1gbl, 1gbm, 1p01, 1p02, 1p03, 1p04, 1p05, 1p06, 1p10, 1p11, 1p12, 1tal, 1ull, 2lpr, 3lpr, 3pro, 5lpr, 6lpr, 7lpr, 8lpr, 9lpr
matripase	3.4.21.-	1eaw
mesentericopeptidase	3.4.21.62	1mee
plasmin	3.4.21.7	1bui
t-plasminogen activator	3.4.21.68	1bda, 1rtf, 1a5i, 1a5h
prolyl oligopeptide	3.4.21.26	1e5t, 1e8m, 1e8n, 1h2y, 1h2z, 1o6f, 1o6g, 1qfm, 1qfs
proteinase K	3.4.21.64	1bjr, 1pek, 1pfg, 1pj8, 1oyo, 3prk
protein C (activated)	3.4.21.69	1aut
sedolisin (pseudomona-pepsin)	3.4.21.100	1ga1, 1ga4, 1ga6, 1kdv, 1kdy, 1kdz, 1ke1, 1ke2
signal peptidase I	3.4.21.89	1b12
streptogrisin A (proteinase A)	3.4.21.80	1sgc, 3sga, 4sga, 5sga
streptogrisin B (proteinase B)	3.4.21.81	1cso, 1ct0, 1ct2, 1ct4, 1ds2, 1sgd, 1sge, 1sgn, 1sgp, 1sgq, 1sgr, 1sgy, 2sgd, 2sge, 2sgf, 2sgp, 2sgq, 3sgb, 3sgq, 4sgb
streptogrisin E (glutamyl endopeptidase II)	3.4.21.82	1hpg
subtilisin BPN	3.4.21.62	1a2q, 1gns, 1gnv, 1lw6, 1sbn, 1sib, 1sua, 1sue, 1sup, 1ubn, 2sic, 2sni, 3sic, 5sic
subtilisin Carlsberg	3.4.21.62	1a10, 1av7, 1avt, 1be6, 1be8, 1cse, 1oyv, 1r0r, 1scn, 1vsb, 2s, 3vsb
subtilisin DY	3.4.21.62	1bh6
thermitase	3.4.21.14	1tec, 2tec, 3tec
thrombin[a]	3.4.21.5	1a46, 1abj, 1ay6, 1bth, 1de7, 1fph, 1k22, 1nrn, 1nro, 1nrp, 1nrq, 1nrs, 1tbq
tricorn	3.4.21.-	1n6d, 1n6e, 1n6f
trypsin[a]	3.4.21.4	1aut, 1ezx, 1sbw, 1tyn, 3btg,
urokinase plasminogen	3.4.21.73	1c5w, 1c5x, 1c5y, 1ejn, 1f5k, 1f5l, 1f92, 1fv9, 1gi7, 1gi8, 1gi9, 1gj7, 1gj8, 1gj9, 1gja, 1gjb, 1gjc, 1gjd, 1lmw, 1owd, 1owe, 1owh, 1owi, 1owj, 1owk
venom plasminogen	3.4.21.-	1bqy

[a] Both thrombin and trypsin structures deposited on the PDB number in the hundreds, and a representative example is listed here.

proteases of viral, bacterial, human, mammalian, and other organism sources analyzed in this study (Table 5).

6.1. Viral Serine Proteases

Human cytomegalovirus assemblin (hCMV) is the encoded protease vital for viral replication of this herpes virus. This pathogen is a danger to immunosuppressed individuals (AIDS patients and organ-transplant recipients) and congenitally infected infants. Figure 6a displays six inhibitors exhibiting the classic β-strand conformation when bound to the active site of this viral protease.[39,62] Viral proteases of the flaviviridae family include the Dengue NS3 protease and hepacivirin (hepatitis C NS3/4a protease). Like other viral proteases their function perpetuates replication of the virus and the inhibition of these proteases provides an ideal target for medical

intervention. Figure 6b shows the binding conformation of the Bowman−Birk inhibitor bound to the Dengue NS3 protease with two sets of active-site-binding residues displayed. This is one of numerous examples of proteinaceous serine protease inhibitors, including serpins (serine protease inhibitors) that act by mimicking the structure of the normal substrate, which have been reviewed elsewhere.[106−108] The Bowman−Birk inhibitor is bifunctional with two interaction sites, both of which bind to an individual Dengue NS3 protease molecule. The inhibitor forms part of a large β-sheet between itself and the protease with the interface at the active site. Hepacivirin has three representative inhibitors both within the active site (Figure 6c), and clearly both Dengue NS3 protease and hepacivirin recognize ligands in an extended β-strand conformation.

6.2. Bacterial Serine Proteases

Kumamolisin is a bacterial protease from the family of sedolisins (serine−carboxyl proteases).[109] They are characterized by their unique catalytic triad of Ser-Glu-Asp and were previously thought to be aspartic proteases as the first enzyme identified was named *Pseudomonas* pepstatin-insensitive carboxyl protease. Figure 6d shows two kumamolisin inhibitors in their active-site conformation. Sedolisin (pseudomonapepsin)[109,110] is a homologue of the human enzyme CLN2 that, when mutated, leads to the neurodegenerative disease, classical late-infantile neuronal ceroid lipofuscinosis.[111] A total of eight inhibitors are displayed in Figure 6e. Lysyl endopeptidase has been cocrystallized with a small peptide mimic that illustrates the trypsin-like specificity of the enzyme (Figure 6f). α-Lytic endopeptidase is well characterized with 19 cocrystallized inhibitors binding tightly in a conserved manner (Figure 6g). Mesentericopeptidase binds to eglin C, a 70-residue proteinaceous inhibitor from the medicinal leech. Eglin C (Figure 6n, green ribbon) is a general non-specific serine protease inhibitor, and Figure 6h shows only the residues that interact with the active site of mesentericopeptidase. The above ligands all display an extended conformation in the active site of the protease. However, proteinase K shows some very unusual peptide binding already mentioned above (Figure 3j). Two further complexes are shown here (Figure 6i, products shown for reference). One inhibitor binds in the expected manner for a peptidic chloromethyl ketone inhibitor (ball-and-stick). However, unusual binding is seen in the complex between the proteolytically generated lactoferrin fragment and proteinase K (green ribbon).[112] The hydrophobic C-terminal portion is inserted into the deep S1 subsite, and the N terminus is looped around to occupy the S2 subsite showing a turn-like conformation.

Bacterial signal peptidase I is a member of an important larger family of signal peptidases which cleave signal peptides from exported proteins within the cell.[113] The catalytic residues of these proteases consist of a Ser-Lys and form acyl enzyme complexes with inhibitors (Figure 6j).[114] Streptogrisin A (proteinase A), streptogrisin B (proteinase B), and streptogrisin E (glutamyl endopeptidase II) are represented in Figure 6k, l, and m, respectively. Of particular interest is streptogrisin B, which forms complexes with the proteinaceous inhibitors, ovomucoid inhibitor Omtky3 (Figure 6l, ribbon), and chymotrypsin inhibitor I from the potato (ribbon not shown). Both inhibitors bind to the enzyme via the active site and incorporate an intramolecular disulfide bridge for stability.

Subtilisins have been extensively studied over the years with PDB entries totaling 65. They evolved independently of the trypsin/chymotrypsin-like proteases illustrated by their different structural fold (section 2) and lack of cysteine residues. Three variants are discussed here: BPN′, Carlsberg, and DY (Figure 6n, o, and p, respectively). Subtilisin BPN′ inhibitors include the proteinaceous Streptomyces subtilisin inhibitor (SSI, magenta ribbon) and eglin C (green ribbon), which is also seen with subtilisin Carlsberg (ribbon). There is one unusual inhibitor conformation (Figure 6o, green, ball-and-stick) where the molecule protrudes outside the active site, stabilized by water molecules and not as part of the expected antiparallel β-sheet.[115] Thermitase is a subtilisin-like thermostable protease, and Figure 6q displays three complexes with eglin C (only active-site residues shown) displaying the classical extended mode of binding. The tricorn core protease, discussed above, is a C-terminal processing protease involved in protein degradation, and Figure 6r shows three inhibitors of the core protease, one of which shows a slightly bent conformation (red) compared with that of the other inhibitors.

6.3. Human and Mammalian Serine Proteases

Human and mammalian serine proteases are essential for life. Many serine proteases are associated with physiological processes including the coagulation cascade, complement cascade, as well as propeptide cleavage (convertases) in many different contexts. Cathepsin G is implicated in events of tissue remodeling and inflammation.[116] Unlike its lysosomal cysteine protease counterparts, cathepsin G is stored in azurophilic granules of neutrophils. Inhibitors of this enzyme are displayed in their enzyme-binding conformation in Figure 7a. Chymase is stored in mast cells and is reputed to be involved in numerous functions including peptide hormone processing. Figure 7b shows an inhibitor in an almost identical conformation to those of the previous enzyme cathepsin G.

Chymotrypsin is an active component in the digestive system of mammals, which incorporates gastric pepsin, pancreatic trypsin, chymotrypsin, elastase, and collagenase. Chymotrypsins are present in the intestine as the inactive precursors and four sequential cleavages after residues 13, 15, 146, and 148 produce the active π, δ, κ, and α forms, respectively.[117] Inhibitors of α-chymotrypsin are shown in Figure 7c, illustrating three proteinaceous inhibitors (eglin C, yellow; human pancreatic secretory trypsin inhibitor, green; bovine pancreatic trypsin inhibitor (aprotinin, see below, blue) as well as smaller peptidic inhibitors. Figure 7d shows inhibitors of γ-chymotrypsin with the non-specific proteinaceous inhibitor, ecotin shown

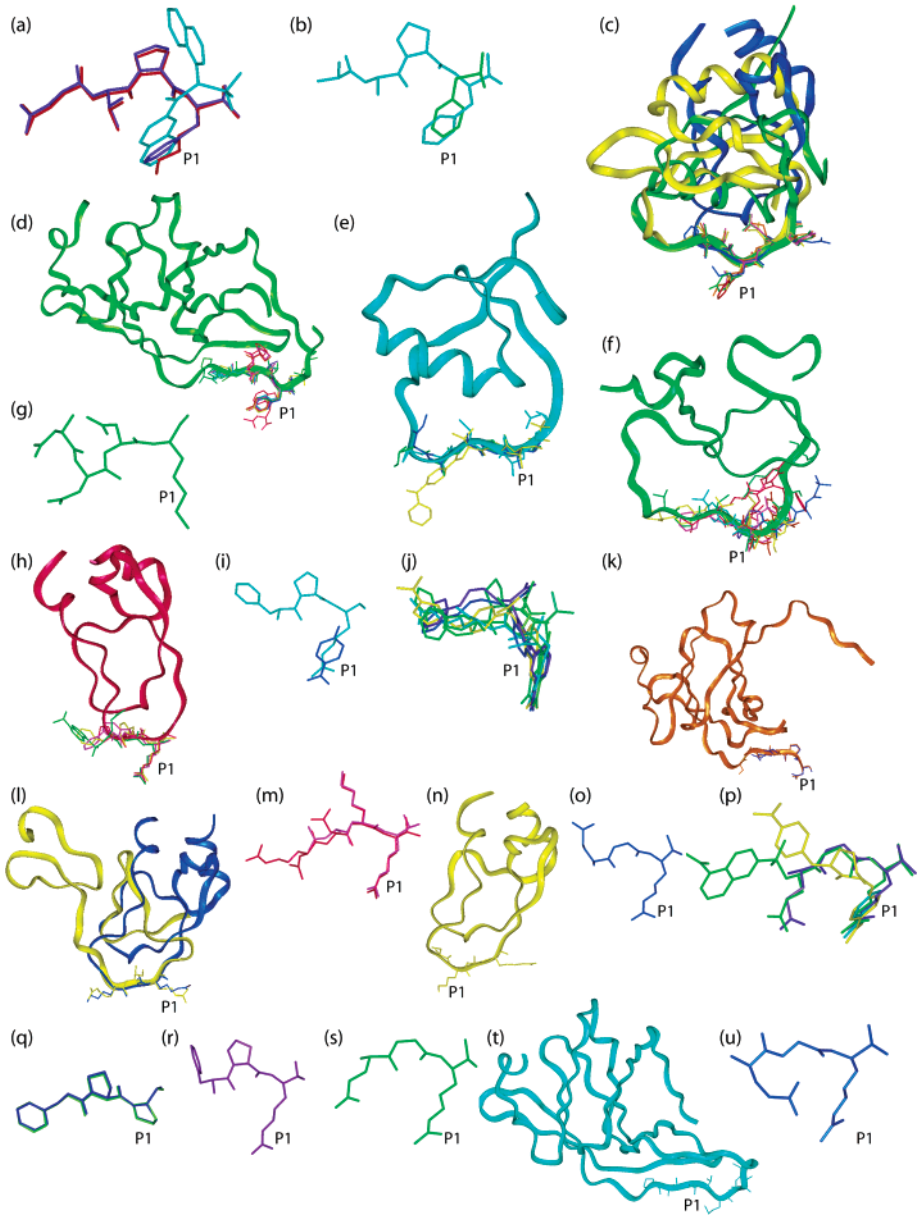


Figure 7. Human and mammalian serine proteases: (a) cathepsin G RMSD 0.40 Å (1kyn, 1au8, 1cgh); (b) chymase RMSD 0.54 Å, (1k*l*t, 1pjp); (c) α-chymotrypsin, 0.58 Å (1acb (yellow ribbon), 1cgi (green ribbon), 1cgj, 1hja, 1mtn (blue ribbon), 6cha); (d) γ-chymotrypsin (chymotrypsin A) RMSD 0.36 Å (1ab9, 1gct, 1gg6, 1ggd, 1gha, 1ghb, 1gmc, 1gmd, 1n8o (ribbon), 2gct, 3gct); (e) neutrophil/human leukocyte elastase RMSD 0.49 Å (1b0f, 1ppf (ribbon), 1ppg, 1hne); (f) pancreatic elastase RMSD 0.43 Å (1b0e, 1eai (ribbon), 1elb, 1elg, 1esb, 1hv7, 1qgf, 1qr3, 8est, 9est); (g) enteropeptidase (1ekb); (h) coagulation factor VIIa RMSD 0.58 Å (1cvn, 1dan, 1fak (ribbon), 1qfk); (i) coagulation factor IXa RMSD 0.61 Å (1pfx, 1rfn); (j) coagulation factor Xa RMSD 0.82 Å (1fax, 1fjs, 1fxy, 1kig, 1xkb, 1xka); (k) granzyme B RMSD 0.62 Å (1fi8 (ribbon), 1iau); (l) kallikrein I RMSD 0.62 Å (1hia (yellow), 2kai (blue)); (m) kexin (1ot5, 1r64); (n) matripase (1eaw); (o) plasmin (1bui); (p) t-plasminogen activator RMSD 0.65 Å (1bda, 1rtf, 1a5i, 1a5h); (q) prolyl oligopeptide RMSD 0.22 Å* (1h2y, 1qfs); (r) protein C [activated] (1aut); (s) u-plasminogen activator (1lmw). Other serine proteases: (t) brachyurin (1azz ribbon); (u) venom plasminogen (1bqy).

in green ribbon. The interaction site between ecotin and chymotrypsin is a clear extension of the existing β-sheets in each component. Pancreatic elastase (elastase 1) displays similar characteristics to the other digestive serine proteases in every aspect except their specificity, requiring small hydrophic residues at P1. Figure 7f presents 10 inhibitors of this enzyme including the ascaris chymotrypsin elastase isoinhibitor 1 (green ribbon) that binds to the active site via the flexible loop region characteristic of serpins.[107]

Neutrophil elastase (leucocyte elastase, Figure 7e; ribbon = ovomucoid inhibitor) is connected to tissue destruction, phagocytosis, and lung disease and shows all the characteristics of its digestive counterpart.[118,119] Enteropeptidase specifically activates pancreatic trypsinogen in the intestine, and only one inhibitor is presented in Figure 7g.

The coagulation cascade consists of numerous coagulation factors and culminates in thrombin releasing fibrin from fibrinogen. Inhibitors of coagulation factors VIIa, IXa, and Xa are illustrated in Figure 7h, i, and j, respectively. Aprotinin, a clinical antiplasmin agent, is seen binding to factor VIIa (red ribbon) via its exposed loop and mimics the binding of smaller peptidic molecules extremely well.

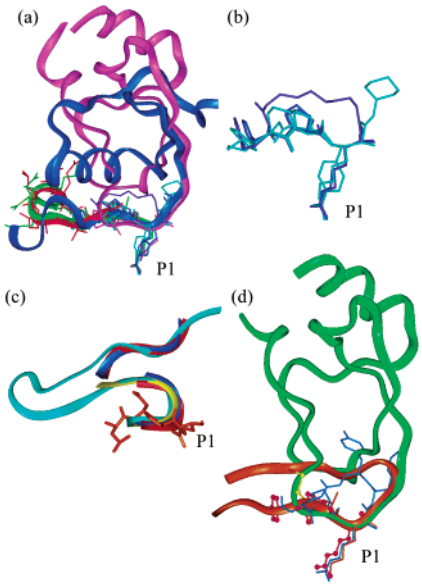


Figure 8. Thrombin and trypsin inhibitors: (a) thrombin (1a46, 1abj, 1ay6, 1bth (magenta ribbon), 1de7 (red ribbon), 1fph (green ribbon), 1k22, 1tbq (blue ribbon)); (b) binding conformation of PPACK (D-Phe-Pro-Arg-chloromethyl ketone, 1abj), cyclic inhibitor based on the cyclotheonamide (1ay6), a novel β-strand mimetic (pdb1a46), and the current pharmaceutical lead melagatran (1k22); (c) conformation of the activated N-terminal peptide of thrombin-activated receptor bound to thrombin (cyan (1nrn), dark blue (1nro), red (1nrp), yellow (1nrq), PPACK orange (1nrs)). (d) Trypsin inhibitors: BPTI (3btg (green ribbon)), PPACK (1aut, (ball-and-stick)), Mung bean inhibitor, (1sbw (orange ribbon)), and cyclotheonamide (1tyn (blue)).

Granzyme B (natural killer cell protease) is intrinsically involved in apoptosis and specifically associated with the activation of caspases (see below).[120] The proteinaceous inhibitor ecotin can be seen in Figure 7k, which interacts with the active site in the form of a β-sheet. Kallikreins are part of the kallikrein−kinin system, which is associated with the liberation of the proinflammatory mediators, bradykinin and kallidin. Kallikreins specifically hydrolyze kinins from their precursor forms.[121,122] Figure 7l shows two proteinaceous inhibitors, aprotinin (blue) and hirustasin (yellow), complexed with kallikrein I. Aprotinin is a nonspecific inhibitor, but hirustasin, derived from the leech, is a kallikrein-specific proteinaceous inhibitor.[123] Both inhibitors interact with the active site of the protease and form a β-sheet. Kexin (kex2) is a member of a large family of proteases responsible for the processing of a significant number of peptide products including hormones, neuropeptides, as well as coagulation and growth factors.[124,125] The dibasic inhibitors of kex2 are shown in Figure 7m in a classic extended strand conformation. Aprotinin appears again bound to matriptase (membrane-type serine protease 1), which can proteolytically activate hepatocyte growth factor, urokinase plasminogen activator, and protease-activated receptor-2 with the initial two being implicated in cancer (Figure 7n).[126,127]

Plasmin (activated plasminogen) is the enzyme responsible for the degradation of fibrin and other extracellular matrix proteins. There are two plasminogen activators, tissue type and urokinase type, both initiating enzymes in producing active plasmin. Figure 7o, p, and s displays the arginine-based inhibitors of these three related enzymes, plasmin, t-plasminogen activator, and u-plasminogen activator, respectively. Urokinase-type plasminogen activator has also been extensively studied using small molecule type inhibitors based on the arginine side-chain mimetic, benzamidine. This is very common in studies of trypsin-like serine proteases such as thrombin and trypsin (discussed below) as well as the coagulation factors kallikreins, matriptase, and tryptase.

The prolyl oligopeptidase enzyme is responsible for the cleavage of peptides having a proline residue at P1. This specificity correlates to the maturation and degradation of many neuropeptides and peptide hormones, and Figure 7q illustrates this with proline-containing inhibitors (product and substrate complexes presented earlier).

Protein C (activated) interferes in thrombin activity by hydrolyzing coagulation factors. The linear Phe-Pro-Arg motif inhibitor of this trypsin-like protease is shown in Figure 7r.

Two other mammalian serine proteases remain to be discussed and are probably the most widely studied of this class. Thrombin[128,129] (~150 PDB structures) and trypsin[101,102,130] (~220 PDB structures) are both trypsin-like proteases in their specificity toward basic residues at P1, and they are inhibited by many similar inhibitors. Thrombin is the key enzyme in the coagulation cascade where it activates factors XIII, V, VIII, and XI as well as itself and converts fibrinogen to fibrin.[131] Thrombin also has anticoagulant activity as it activates protein C.[128] Trypsin is involved in the digestion of proteins in the digestive system of mammals and is secreted in its precursor form in the pancreas. Its digestive action has also been heavily utilized in scientific research. A selection of inhibitors of thrombin and trypsin has been selected to illustrate their binding to these proteases. Figure 8a displays eight inhibitors in complex with thrombin. Aprotinin is shown as a magenta ribbon with the blue ribbon representing rhodniin, a highly specific thrombin inhibitor isolated from the blood-sucking assassin bug *Rhodnius prolixus*.[132,133] Two substrate analogues of thrombin, factor XIII activation peptide (residues 28−27) and fibrinogen (fibrinopeptide A residues 7−16), are shown as red and green ribbons, respectively (Figure 8a).

These substrate analogues form an acyl enzyme complex and bind to thrombin in a clear extended conformation from P5 to P1. This is the most succinct evidence that thrombin and related molecules recognize their substrates in a β-strand conformation. Clearly proteinaceous inhibitors, both endogenous and exogenous, have evolved to interfere with this recognition by mimicking the smallest possible identification motif, in most cases P3−P1. Hirudin is a potent proteinaceous inhibitor derived from the medicinal leech currently used in the clinic.[134] Its mode of action is atypical of other thrombin inhibitors (not shown). The three N-terminal residues are inserted into the active site of thrombin in a retro fashion. In

addition, another major interaction occurs at the fibrinogen anion-binding exocite. The combination of these interactions blocks binding of fibrinogen prior to hydrolysis.[13]

Four small peptidic inhibitors are shown in Figure 8b binding in an extended manner, PPACK (D-Phe-Pro-Arg-chloromethyl ketone, a general trypsin-like serine protease inhibitor, pdb1abj), a cyclic inhibitor based on cyclotheonamide, a natural product isolated from a marine sponge (pdb1ay6), a novel β-strand mimetic (pdb1a46), and a current pharmaceutical lead melagatran (pdb1k22).

Thrombin is also responsible for cleavage and activation of the N terminus of the protease-activated receptor (thrombin receptor). A study of thrombin complexes with the self-activating amino terminal peptide from this receptor and its analogues has revealed an exception to the expected binding mode for proteases (Figure 8c).[136] The peptide itself was predicted to bind with the cleavage site (LDPR-SFLLRNPNDKYEPF) in the active site of the protease and the C-terminal region bound to the fibrinogen anion binding exo site. The observed binding mode shows an S-like peptide structure with the N-terminal portion in a turn-like conformation (Figure 8c, cyan ribbon, PPACK analogue (orange) shown P1−P3 for reference). The C-terminal portion does interact with the predicted anion-binding exo site; however, it is with another molecule of thrombin.

Almost identical in its binding of inhibitors is the digestive enzyme trypsin. A small representative subset of trypsin inhibitors is shown in Figure 8d including a proteinaceous inhibitor (bovine pancreatic trypsin inhibitor, BPTI, green ribbon), PPACK (ball-and-stick), and the cyclic peptide natural product, cyclotheonamide (blue). A fragment of a disulfide-bridged peptide, Mung bean inhibitor, can also be seen binding to the active site of trypsin (orange ribbon). This fragment embodies the inhibitor binding mode for trypsin inhibitors.

Serpins[107,108,137−139] are important endogenous serine protease inhibitors and bind to serine protease via a unique mechanism discussed here briefly. Antitrypsin, which binds to human neuterophil elastase, is discussed as the representative example. The mechanism of inhibition involves a major conformational change initiated by the catalytic serine with the reactive center of the serpin, which is subsequently cleaved. The acyl enzyme complex then moves some 70 Å to the opposite end of the serpin. This conformational change also facilitates loss of overall structure of the protease as well as perturbation of the active site, preventing release of the enzyme. The loss of structure within the protease also permits proteolytic destruction of the enzyme.[140,141] Figure 9 shows the structure of antitrypsin in complex with trypsin following the conformational change where the β strand in red represents the rearranged serpin.

6.4. Other Serine Proteases

Brachyurin C (crab collagenase) has the ability to cleave collagen, and the general pancreatic serine protease inhibitor ecotin is shown in complex with this enzyme (Figure 7t, enzyme not shown). One other non-mammalian protease, plasminogen activator from snake venom, has been cocrystallized with a chloromethyl ketone inhibitor (Figure 7u). This enzyme acts upon the host plasminogen and is resistant to inhibition by endogenous inhibitors such as plasminogen activator inhibitor.[142]

Figure 9. Complex between trypsin (green) and antitrypsin (yellow) where the red β strand is residues 343−358 and represents the major conformational change associated with inhibition by serpins (1ezx). Figure generated using Pymol.

7. Cysteine Protease Inhibitors

Cysteine or thiol proteases[143−145] hydrolyze peptide bonds in a manner similar to serine proteases and exist in three structurally distinct classes, papain-like (e.g., cathepsins),[146] ICE-like (caspases),[147] or picorna-viral (trypsin-like with cysteine replacing the catalytic serine). Like serine proteases, most cysteine proteases have fairly shallow, solvent-exposed active sites that can accommodate short substrate/inhibitor segments as β strands or protein loops (e.g., from endogenous inhibitors such as cystatins). Most peptidic inhibitors[104] tend to be 2−4 amino acids long, and nonpeptidic inhibitors are of equivalent length. They tend to interact with the nonprime subsites of cysteine proteases, terminating with various electrophilic isosteres. For 23 cysteine proteases (Figure 10, Table 6) the majority of the 89 inhibitor structures that are currently available are in an extended β-strand format with one notable exception.

7.1. Viral, Bacterial, and Parasitic Cysteine Proteases

Viral cysteine proteases[15,148] are commonly used as targets for the development of antiviral drugs. The cysteine proteases are found in viruses of positive-strand RNA including the poliomyelitis virus and rhinovirus (*Picornaviridae*), coronavirus (*Coronaviridae*), and hepacivirus (*Flaviviridae*). These proteases are responsible for the continuing viral replication cycle. Figure 10a and b shows one inhibitor in its protease-bound active-site conformation from hepatitis A virus-type picornain 3C and rhinovirus picornain 3C, respectively. Both inhibitors bind to their host enzymes in an extended conformation. Another viral cysteine protease, the tobacco etch virus NIa protease, was mentioned above and also shows excel-

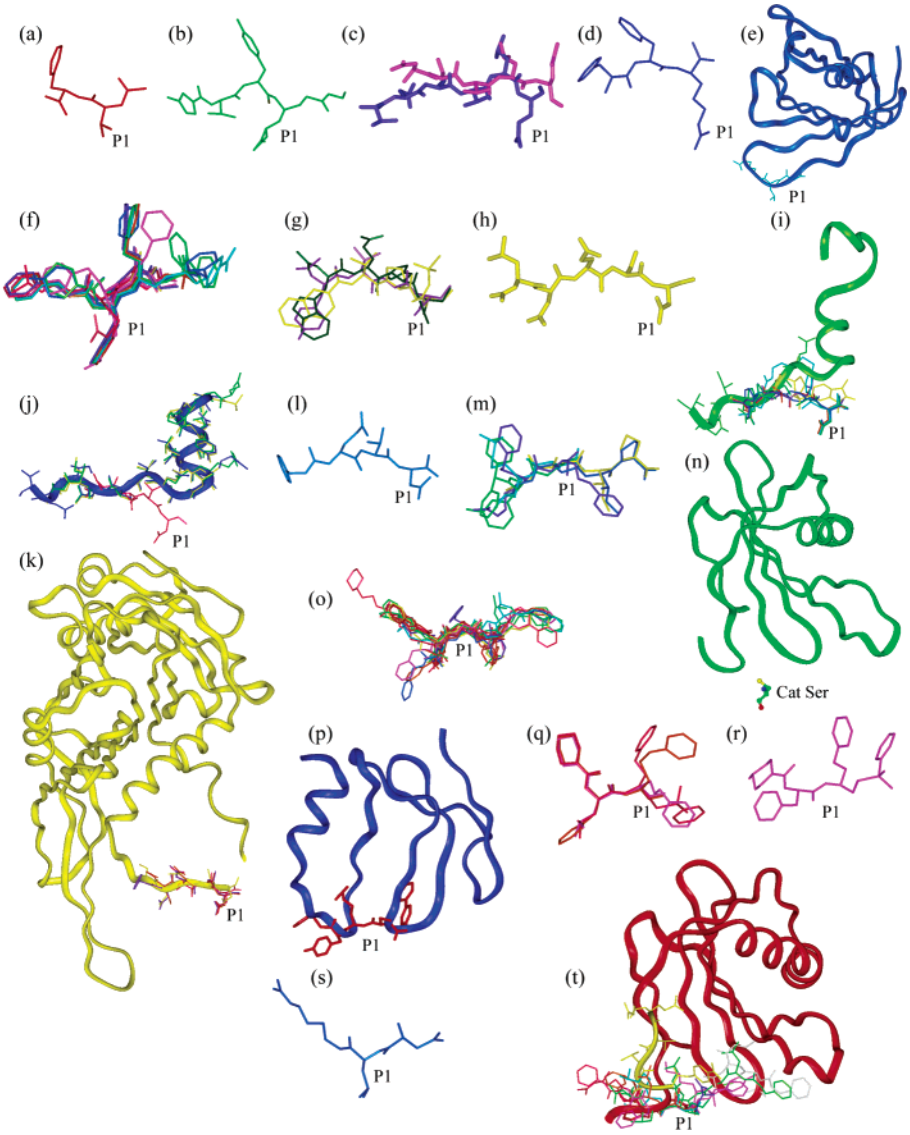


Figure 10. Superimposed cysteine protease−inhibitor structures (RMSDs quoted for superimposition of protein component of complex; PDB code in parentheses, protease omitted; ligands are represented with main chains running horizontally N to C terminal from left to right where appropriate) for viral cysteine proteases: (a) hepatitis A virus protease (1qa7); (b) rhinovirus picornain 3C protease (1cqq); (c) coronavirus 3cl-Pro (1p9u, 1uk4). Bacterial cysteine proteases: (d) gingipain (1cvr); (e) staphopain A (1cv8, 1pxv ribbon). Parasitic cysteine proteases: (f) cruzipain RMSD 0.29 Å (1aim, 2aim, 1ewl, 1ewm, 1ewo, 1ewp, 1f29, 1f2a, 1f2b, 1f2c, 1me3, 1me4). Human and mammalian cysteine proteases: (g) caspase-1 RMSD 0.41 Å (1ibc, 1ice, 1bmq); (h) caspase-2 (1puo); (i) caspase-3 (Apopain) RMSD 1.64 Å (1cp3, 1gfw, 1i3o (ribbon), 1nme, 1nmq, 1 nms, 1pau); (j) caspase-7 RMSD 0.40 Å (1i51, 1i4o, 1fij (red), 1kmc); (k) caspase-8 RMSD 0.52 Å (1i4e (ribbon), 1f9e, 1qdu, 1qtn); (l) caspase-9 (1jxq); (m) cathepsin B RMSD 0.56 Å (1csb, 1gmy, 1ito, 1qdq, 1ste); (n) cathepsin H (1nb3, 1nb5); (o) cathepsin K RMSD 0.36 Å* (1atk, 1au0, 1au2, 1au3, 1au4, 1ayu, 1ayv, 1ayw, 1bgo, 1mem, 1nl6, 1nlj); (p) cathepsin L (1icf (ribbon), 1mhw); (q) cathepsin S RMSD 0.27Å (1ms6, 1npz, 1nqc); (r) cathepsin V (1fh0). Plant cysteine proteases: (s) caricain (1meg); (t) papain RMSD 0.46 Å (1cvz, 1pad, 1pe6, 1pip (yellow ribbon), 1pop, 1ppd, 1ppp, 1stf (red ribbon), 2pad, 4pad, 5pad, 6pad, 1bp4, 1bq). Asterisk (*) indicates RMSD excludes pdb1nl6.

lent binding of both substrate and product in an extended conformation (Figure 3b).

Coronaviruses are responsible for respiratory tract infections, which can be fatal as seen in recent cases of severe acute respiratory syndrome (SARS). Two structures have been published reporting complexes between the SARS coronavirus main protease (SarsCoV $M^{pr°}$ or 3CL-$^{pr°}$) and chloromethyl ketone inhibitors, and both illustrate these inhibitors binding in an extended manner (Figure 10c).[19,149]

The two bacterial cysteine proteases, Gingipain R2, the major pathogen in periodontal disease, and Staphopain A isolated from *Staphylococcus aureus,* both recognize inhibitors in an extended conformation (Figure 10d and e, respectively). A recent structure of staphopain complexed with staphostatin, an endogenous inhibitor from *S. aureus,* shows the inhibitor binding in a substrate-like manner (extended conformation) within the active site (Figure 10e, ribbon).[150]

Cruzipain is the major protease associated with American trypanosomiasis or Chagas disease.[151] A large research effort has targeted this enzyme for inhibition with a total of 12 inhibitors shown in Figure 10f, all of which bind in a β-strand-like conformation.

Table 6. Crystal Structure Listing for 89 Inhibitors Complexed with 23 Cysteine Endoproteases

cysteine protease	EC code	PDB codes
actinidain	3.4.22.14	1aec
caricain		1meg
caspase-1 (ICE)	3.4.22.36	1bmq, 1ibc, 1ice, 1m72
caspase-2	3.4.22.-	1pyo
caspase-3 (apopain)	3.4.22.-	1cp3, 1gfw, 1i3o, 1nme, 1nmq, 1nms 1pau,
caspase-7	3.4.22.-	1i51, 1f1j, 1i4o, 1kmc
caspase-8	3.4.22.-	1f9e, 1i4e, 1qdu, 1qtn,
caspase-9	3.4.22.-	1jxq
cathepsin B	3.4.22.1	1csb, 1cte, 1gmy, 1ito, 1qdq, 1ste,
cathepsin H	3.4.22.16	1nb3, 1nb5
cathepsin K	3.4.22.38	1atk, 1au0, 1au2, 1au3, 1au4, 1ayu, 1ayv, 1ayw, 1bgo, 1bp4, 1bqi, 1mem, 1nl6, 1nlj
cathepsin L	3.4.22.15	1icf, 1mhw
cathepsin S	3.4.22.27	1ms6, 1npz, 1nqc
cathepsin V	3.4.22.43	1fh0
cruzipain	3.4.22.51	1aim, 1ewl, 1ewm, 1ewo, 1ewp. 1f29, 1f2a, 1f2b, 1f2c, 1me3, 1me4, 2aim
gingipain	3.4.22.37	1cvr
glycyl endopeptidase	3.4.22.25	1gec
hepatitis A 3C	3.4.22.-	1qa7
papain	3.4.22.2	1bp4, 1bqi, 1cvz, 1khp, 1khq, 1pad, 1pe6, 1pip, 1pop, 1ppd, 1ppp, 1stf, 2pad, 4pad, 5pad, 6pad
rhinovirus	3.4.22.-	1cqq
sars-coronavirus 3cl-Pro (Mpro)	3.4.22.-	1p9u, 1uk4
staphopainA	3.4.22.48	1cv8, 1pxv
tobacco etch virus NIa protease	3.4.22.-	1lvb, 1lvm

7.2. Human and Mammalian Cysteine Proteases

There are two large classes of human cysteine proteases, the caspases and the lysosomal proteases or cathepsins. Caspases were initially associated with two main functions, apoptosis and activation of proinflammatory mediators.[147,152−154] They all share a common fold as well as specificity for aspartate residues at P1. Caspase-1 (also known as interleukin 1-β-converting enzyme or ICE) proteolytically hydrolyzes prointerleukin 1β to produce a proinflammatory cytokine. Three covalently bound inhibitors of caspase-1 are shown in an extended conformation in Figure 10g. Caspase-2 was the first identified mammalian caspase associated with apoptosis. It is involved as an initiator in both the intrinsic and extrinsic pathways of apoptosis. The five-residue aldehyde inhibitor Ac-Leu-Asp-Glu-Ser-Asp-H is shown bound to caspase-2 in Figure 10h.[155] Caspase-3 is associated with the execution of apoptosis and is also a target for biological activation. Caspase activity is regulated in mammals by the protein family of inhibitor apoptosis proteins (IAPs). Six small inhibitors are shown in Figure 10i and act as classical cysteine protease inhibitors. Figure 10i also illustrates the binding of the BIR domain (conserved IAP module) of XIAP (green ribbon).[156] The inhibitor, XIAP, binds to caspase-3 via the interactions of three binding sites "hook" (helical region on right), "line", and "sinker". Three residues from V146-D148 bind across the active site in an extended fashion, but the direction of the polypeptide chain is the opposite of the standard binding mode, i.e., V146 corresponds to P2 and V147 corresponds to P3.

Four caspase-7 inhibitors are shown in Figure 10j, three of which are complexes with XIAP (ribbon), which bind across the active site in an elongated manner. Caspase-8 is associated with apoptosis initiation and covalently inhibited by baculovirus protein p53 (Figure 10k, yellow ribbon).[157] Inhibition is mediated by the formation of the acyl enzyme complex after cleavage of the p35 reactive-site loop at Asp87. Hydrolysis is further prevented by the positioning of the N terminus of p35 close to the active site, which eliminates solvent accessibility.[157] The C-terminal tail is bound within the nonprime side of the active site in an extended conformation almost identical to the three other inhibitors shown (Figure 10k). Caspase-9, also involved in apoptosis initiation, has been crystallized with a dehydroxymethylaspartic acid inhibitor, which binds in the classical extended fashion (Figure 10l).

Papain-like cysteine lysosomal proteases or cathepsins (from the Greek meaning to digest) are responsible for controlled protein digestion and degradation.[158] The known cathepsins act as endoproteases (cathepsins L, S, V, and F) and exoproteases (cathepsins B, C, H, and X) with the latter also possessing some endoproteolytic activity. Cathepsins of this type possess relatively short active-site clefts with three well-defined subsites (S2−S1′) but have no common substrate specificity. Figure 10m shows five inhibitors of cathepsin B bound covalently within the active site, two of which utilize the binding of the C-terminal carboxylate group of the inhibitor to a flexible loop region known as the occluding loop.[159] Cathepsin H is inhibited by stefins and cystatins, which are widespread nonspecific endogenous proteinaceous inhibitors involved in the control of enzymatic activity. Figure 10n displays the unusual binding mode of these molecules (ribbon, catalytic cysteine shown for reference). As well as causing a slight conformational change, stefin A binds to cathepsin H by wedging three loops across the active-site cleft. By doing so it isolates the catalytic cysteine from solvent, preventing hydrolysis of a peptide.[160] Cathepsin K has been found to be crucial in bone

Scheme 1. Nucleophilic Attack of Threonine Hydroxyl on Peptide Substrate

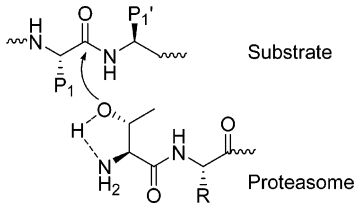


Table 7. Crystal Structure Listing for Four Inhibitors Complexed with Two Threonine Endoproteases

threonine protease	EC code	PDB codes
20S proteasome HsIUV	3.4.25.1	1g65, 1jd2, 1j2q 1kyi

remodeling, and Figure 10o shows 12 inhibitors bound within the active site of this endoprotease. These inhibitors are covalently linked to the enzyme and span both sides of the active site in an extended conformation. Cathepsin L is inhibited by a molecule similar to stefins, a fragment of the p41 form of MHC class II. Its mode of action is very similar to the stefins and illustrated in Figure 10p (blue ribbon).[161] The inhibitor shown in red exemplifies the contrasting binding modes of cystatins and linear compounds. The two remaining mammalian depictions in Figure 10 (parts q and r) represent small subsite-specific inhibitors of cathepsin S and cathepsin V, respectively.

7.3. Plant and Fungal Cysteine Proteases

Actinidain, caricain, and glycyl endopeptidase are cysteine proteases of plant origin, and all bind inhibitors in an extended conformation (Figure 10s). Figure 10t demonstrates the binding of 14 inhibitors to papain. Twelve of these inhibitors bind in the classical fashion with two other inhibitors binding in an unusual manner. Stefin B (red ribbon) binds in a fashion similar to that described for stefin A (Figure 10n) except that the N-terminal tail extends along the active site and does not form a "hook". The inhibitor shown with a yellow ribbon is a pNA substrate molecule acting as an inhibitor (discussed above).

8. Threonine Protease Inhibitors

Threonine endoproteases are one of the new classes of proteases and use an N-terminal threonine residue as its catalytic machinery. The side-chain hydroxyl group acts as the proteolytic nucleophile activated by the N-terminal amine moiety (Scheme 1).[162−164] The 20S proteasome (Table 7) has been extensively characterized as the archetypal threonine protease along with the 26S proteasome,[165,166] with HsIUV the prokaryotic equivalent of the 20S proteasome.[165]

The ubiquitin−proteasome pathway is a major avenue for degradation of proteins in the cytosol and nucleus of eukaryotic cells.[167] Proteins are marked for degradation via the attachment of a ubiquitin group whereby the proteasome will recognize, unfold, and digest these proteins. The complete 26S proteasome incorporates the ATP-dependent 19S caps,

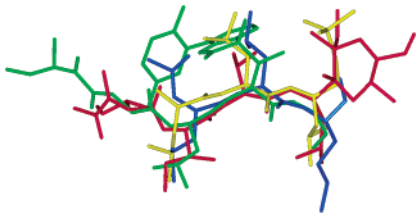
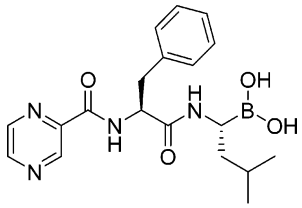


Figure 11. Superimposed threonine protease−inhibitor structures: proteasome inhibitors (1jd2 (green), 1j2q (blue), 1g65 (red)) with the HsIUV catalytic subunit inhibitor (1kyi (yellow)).

which are understood to recognize and unfold substrates, which are in turn fed to the 20S proteasome, where proteolysis occurs.[167] The 20S proteasome structure, as mentioned in section 2, is made up of 28 subunits arranged in four rings of seven, termed $\alpha\beta\beta\alpha$ with the interior of the β subunits responsible for proteolytic activity.[45]

Figure 11 shows four inhibitors of the proteasome and shows that, like its neighbors, the ligands are recognized in an extended conformation. Recently the proteasome has been identified as a target for cancer therapy[168] with the peptidic inhibitor Bortezomib[169] currently in clinical use for multiple myeloma.



Bortezomib

9. Exceptions to the Extended Conformation

Following a comprehensive analysis of the complete set of over 1500 protease structures deposited in the Protein Data Bank, only six structures show genuine deviation from the hypothesis that proteases recognize ligands in an extended conformation (Figure 12). The most profound example is the single inhibitor IA3 of saccharopepsin, which clearly binds in an almost perfect α helix (Figure 12a). However, there are also a number of other inhibitors of this protease that were found bound in an extended conformation within the active site. It is clear that the unusually large active site is able to accommodate a helix in this case, whereas most protease active sites are too small to permit a helix to bind.

Proteinase K also shows unusual binding of ligands in an almost turn-like conformation (Figure 12b). The metalloprotease *C. botulinum* neurotoxin shows unusual binding and specificity, in that it appears to bind and cleave its substrates in an orientation opposite to what is generally seen, as well as requiring large sequences for recognition (Figure 12c). The serine protease, thrombin, provides the fourth example of a deviation from the norm. This is seen in the complex between thrombin and the self-activating amino-terminal peptide of the thrombin receptor bound clearly within the active site in a turn conformation (Figure 12d). The two remaining examples

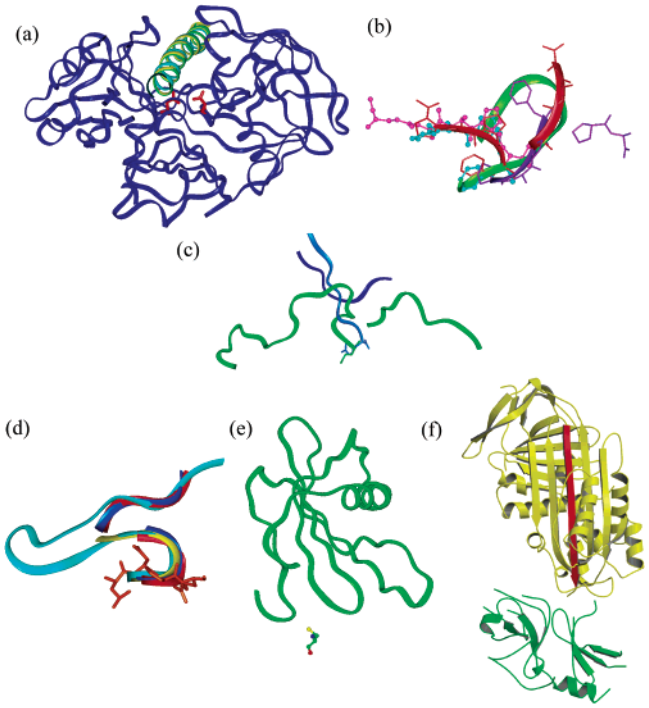


Figure 12. Exceptions to the extended conformation: (a) saccharopepsin (1dp5 (enzyme shown as blue ribbon with catalytic aspartates shown in red and helical inhibitor yellow), 1dpj (green), 1g0v); (b) proteinase K RMSD 0.38 Å (1bjr (green ribbon), 1pek, 1pfg (red ribbon), 1pj8 (purple ribbon), 3prk (ball-and-stick)); (c) botulinum neurotoxin type B cleavage products of synaptobrevin-II (1f83 (green ribbon)) and the product-bound state of BoNT serotype A (1e1h (blue ribbon)); (d) conformation of the activated N-terminal peptide of thrombin-activated receptor bound to thrombin (1nrn (cyan), 1nro (dark blue), 1nrp (red), 1nrq (yellow), 1nrs (PPACK orange)); (e) cathepsin H (1nb3, 1nb5); (f) complex between trypsin (green) and antitrypsin (yellow) where the red β strand is residues 343−358 and represents the major conformational change associated with inhibition by serpins (1ezx). Serpin complex generated using Molscript 2.0.

are the exceptions to the rule in that they are both endogenous inhibitors of their respective proteases. The class of stefins (cystatins) inhibits cysteine protease via a unique mechanism by binding several loops across the active site (Figure 12e). Finally, it is worth mentioning the unique and unusual binding mode of the serpins, such as antitrypsin, where a major conformational change follows initial complex formation which results in an irreversibly bound serpin−protease complex which subsequently causes proteolytic degradation of the complex (Figure 12f).

10. Conformational Selection and Proteolytic Processing

It is only in the past few years that a significant number of crystal structures, for each class of protease bound to small molecule ligands, have become available for the kind of comprehensive analysis made here. It is also worth noting that our analysis has involved peptidic, peptidomimetic, and nonpeptidic ligands bound to proteases. The above data clearly shows that the extended conformation, or nonpeptidic equivalent, is overwhelmingly preferred in the active site of over 100 protease enzymes. The almost universal recognition by proteases, documented herein, for substrates/inhibitors in a β-strand conformation raises some important questions. Why have proteases with different structural folds evolved to recognize the same ligand conformation in their active sites, and why is this the extended β-strand conformation?

First, one needs to consider that folded proteins normally consist of helices, turns/loops, and sheets, which together impart function. These structures need to be resistant to proteases—it would not be logical for cellular machinery to expend energy in folding polypeptides into such structural elements yet then be susceptible to cleavage and thus degradation by proteases. Hence, proteases do not normally recognize folded helices, turns/loops, or sheets, and we are unaware of examples of any such structures being cleaved by a protease. We would predict, on the basis of this review, that any cleavage site in a putative substrate would need to be in an extended region of the substrate, or the equilibrium between folded and unfolded substrate would need to be driven through the unfolded form in order to realize proteolytic cleavage.

Second, the stretching of the substrate amide bond toward the transition state is expected to be most efficient when these forces are directionally opposed, as optimally enforced by the extended strand due to the trans amide bond and the L-conformation of each flanking amino acid. This stretching is largely dictated by interactions between enzyme and 'pocket-filling' inhibitor/substrate substituents, that tend to be in close proximity on i, i + 2, i + 4, etc. positions on one face of the strand and i + 1, i + 3, i + 5, etc. positions on the opposite face.

Third, linear extension of a peptide maximally exposes all of the main-chain amide atoms to solvent/protease residues, otherwise shielded by side chains in intramolecularly hydrogen-bonded helix/turn/sheet structures, and promotes intermolecular hydrogen bonding with an enzyme. This may be very important since most protease active sites are shallow crevices or grooves that promote rapid association of substrates and facile dissociation of products, as required for multiple rapid turnovers. Product solvation may be particularly important in driving the equilibrium to cleaved products and minimizing intramolecular hydrogen bonding in polypeptide substrates/products.

Fourth, proteins that are known to be processed by proteases tend to be cleaved only *between* elements of secondary structure rather than within helical, turn, or sheet regions, and rates of processing tend to be greater after denaturation (breaking of intramolecular H bonds and unfolding) of protein substrates. Fifth, the diameter of the inhibitor-bound conformation of the active site of most proteases is too small to accommodate substrates in helical/turn/sheet conformations, which are consequently not recognized without substantial modification of the dimensions of the active-site groove.

These features all support conformational selection in which only the extended β strand is recognized from the conformational ensemble of ligand struc-

tures present in solution. It has been previously suggested[170,171] that entropic considerations do not support the idea that protein−ligand interactions, in general, proceed via conformational selection but rather are more likely to proceed via a two-/multistep 'zipper' process. This 'induced fit' model of protein−ligand interaction is more attractive than a rigid lock-and-key model for protease−ligand interaction, but the truth may well lie somewhere between this and the single conformation theory. Conformational restrictions that pre-organize the ligand structure close to its final protease-bound conformation can be expected to reduce the entropic barrier *en route* to a protease−ligand complex. It is important to realize that we have only presented data in this review consistent with proteases *ultimately* binding their ligands in an extended ligand conformation and cannot definitively conclude from the static crystal structure data alone that the *initial* interaction between protease and ligand requires the ligand to be in an extended conformation. However, it does not appear likely, from all the evidence to date, that other conformations are recognized by proteases and then rearrange to a different conformation that is cleaved. Biological properties of peptides and proteins are frequently interpreted on the basis of their observed solution or solid-state structures. However, the most abundant or lowest energy solution structure of a molecule is not necessarily the one recognized by an enzyme. The above results clearly demonstrate that inhibitors and substrate analogues at least end up bound in an extended conformation to the active site of a protease.

To date, inhibitor design has been specific for each protease, with general design approaches being hampered by the unpredictability of cooperative interactions between, and structural changes to, inhibitor and protease during complexation. Binding to proteases by their substrates, substrate analogues, and inhibitors in extended β-strand conformations has profound consequences for drug design. First, it suggests the under-utilized strategy of conformationally fixing protease *inhibitors* in an extended β strand in order to optimize their potency and selectivity. This approach has been highlighted here by conformationally constrained macrocycles, but there are many ways of stabilizing extended β strands.[172] Second, the design of protease inhibitors should be greatly simplified since new proteases, for which there is no or limited structural information, can still be effectively targeted based only upon knowledge of the class of protease and its substrate specificity. Compounds can then be designed in locked β-strand-mimicking structures. Third, proteolytic degradation of other bioactive peptides (agonists, antagonists, etc.) might be minimized by the reverse strategy of stabilizing their turn or helix conformations so that the extended strand conformation is less accessible. Conformational selection should now be widely examined for proteases, and attempts should be made to see if there are correlations between conformational equilibria and processing rates of polypeptide substrates and conformationally biased analogues.

11. Conclusions

This review is a comprehensive analysis in support of *ligand conformation* being important for recognition by all proteolytic enzymes and complements two earlier communications on the subject [71,173] On the basis of subsets of superimpositions of >1500 protease−ligand crystal structures, we conclude herein that aspartic, serine, cysteine, and metalloproteases commonly, almost universally, bind to the extended β-strand conformation of substrates, products, and inhibitors of both peptidic and nonpeptidic origin. Such *conformational selection* explains the resistance of folded/structured regions of proteins to proteolytic degradation, the susceptibility of denatured proteins to processing, and the higher affinity of conformationally constrained 'extended' inhibitors/substrates for proteases. Cyclic inhibitors[174] and a cyclic substrate[71] illustrate a minimalist approach to the strategy of constraining inhibitors/substrates in a protease-binding structure that has higher affinity for a protease than acyclic analogues. This preorganization of ligands for a protease can alternatively be achieved by replacing components of peptides with conformational constraints that fix the β-strand conformation[172,175] or by semirigid scaffolds that can project substituents into the indentations that line a substrate-binding active site of a protease.[172] On the basis of these results one would predict that the use of such approaches to fix an extended β-strand ligand conformation should produce high-affinity inhibitors even for proteases of unknown structure, provided that substrate specificity and protease type is known. This inference should be extremely valuable, particularly when the structure of a protease is not known and inhibitor design has to be based solely upon substrate preferences.

12. Acknowledgments

We thank the Australian funding agencies NHMRC and ARC for some financial support for this research.

13. References

(1) Yousef, G. M.; Kopolovic, A. D.; Elliott, M. B.; Diamandis, E. P. *Biochem. Biophys. Res. Commun*. **2003**, *305*, 28.
(2) Puente, X. S.; Sanchez, L. M.; Overall, C. M.; Lopez-Otin, C. *Nat. Rev. Genet.* **2003**, *4*, 544.
(3) Rawlings, N. D.; O'Brien, E.; Barrett, A. J. *Nucleic Acids Res.* **2002**, *30*, 343.
(4) Hubbard, T.; Barker, D.; Birney, E.; Cameron, G.; Chen, Y.; Clark, L.; Cox, T.; Cuff, J.; Curwen, V.; Down, T.; Durbin, R.; Eyras, E.; Gilbert, J.; Hammond, M.; Huminiecki, L.; Kasprzyk, A.; Lehvaslaiho, H.; Lijnzaad, P.; Melsopp, C.; Mongin, E.; Pettett, R.; Pocock, M.; Potter, S.; Rust, A.; Schmidt, E.; Searle, S.; Slater, G.; Smith, J.; Spooner, W.; Stabenau, A.; Stalker, J.; Stupka, E.; Ureta-Vidal, A.; Vastrik, I.; Clamp, M. *Nucleic Acids Res.* **2002**, *30*, 38.
(5) Mulder, N. J.; Apweiler, R.; Attwood, T. K.; Bairoch, A.; Barrell, D.; Bateman, A.; Binns, D.; Biswas, M.; Bradley, P.; Bork, P.; Bucher, P.; Copley, R. R.; Courcelle, E.; Das, U.; Durbin, R.; Falquet, L.; Fleischmann, W.; Griffiths-Jones, S.; Haft, D.; Harte, N.; Hulo, N.; Kahn, D.; Kanapin, A.; Krestyaninova, M.; Lopez, R.; Letunic, I.; Lonsdale, D.; Silventoinen, V.; Orchard, S. E.; Pagni, M.; Peyruc, D.; Ponting, C. P.; Selengut, J. D.; Servant, F.; Sigrist, C. J.; Vaughan, R.; Zdobnov, E. M. *Nucleic Acids Res.* **2003**, *31*, 315.
(6) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen,

L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C. *Science* **2001**, *291*, 1304.
(7) Abbenante, G.; Fairlie, D. P. *Med. Chem.* **2005**, *1*, 71−104.
(8) Bernstein, P. R.; Edwards, P. D.; Williams, J. C. *Prog. Med. Chem*. **1994**, *31*, 59.
(9) Tanaka, R. D.; Clark, J. M.; Warne, R. L.; Abraham, W. M.; Moore, W. R. *Int. Arch. Allergy Immunol.* **1995**, *107*, 408.
(10) Hugli, T. E. *Trends Biotechnol*. **1996**, *14*, 409.
(11) Fath, M. A.; Wu, X.; Hileman, R. E.; Linhardt, R. J.; Kashem, M. A.; Nelson, R. M.; Wright, C. D.; Abraham, W. M. *J. Biol. Chem.* **1998**, *273*, 13563.
(12) Stubbs, M. T.; Bode, W. *Thromb. Res.* **1993**, *69*, 1.
(13) Kim, T. W.; Pettingell, W. H.; Jung, Y. K.; Kovacs, D. M.; Tanzi, R. E. *Science* **1997**, *277*, 373.
(14) Selkoe, D. J. *Science* **1997**, *275*, 630.
(15) Maliar, T.; Balaz, S.; Tandlich, R.; Sturdik, E. *Acta Virol.* **2002**, *46*, 131.
(16) Tong, L. *Chem. Rev.* **2002**, *102*, 4609.
(17) Shuker, S. B.; Mariani, V. L.; Herger, B. E.; Dennison, K. J. *Chem. Biol.* **2003**, *10*, 373.
(18) Dunn, B. M.; Goodenow, M. M.; Gustchina, A.; Wlodawer, A. *Genome Biol.* **2002**, *3*, REVIEWS3006.
(19) Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. *Science* **2003**, *300*, 1763.
(20) Bein, M.; Schaller, M.; Korting, H. C. *Curr. Drug Targets* **2002**, *3*, 351.
(21) Leung, D.; Abbenante, G.; Fairlie, D. P. *J. Med. Chem.* **2000**, *43*, 305.
(22) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
(23) Hendlich, M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 1178.
(24) Laskowski, R. A.; Hutchinson, E. G.; Michie, A. D.; Wallace, A. C.; Jones, M. L.; Thornton, J. M. *Trends Biochem. Sci.* **1997**, *22*, 488.
(25) Luscombe, N. M.; Laskowski, R. A.; Westhead, D. R.; Milburn, D.; Jones, S.; Karmirantzou, M.; Thornton, J. M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 1132.
(26) Laskowski, R. A. *Nucleic Acids Res.* **2001**, *29*, 221.
(27) Vondrasek, J.; van Buskirk, C. P.; Wlodawer, A. *Nat. Struct. Biol.* **1997**, *4*, 8.
(28) Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. *Structure* **1997**, *5*, 1093.
(29) Pearl, F. M.; Lee, D.; Bray, J. E.; Sillitoe, I.; Todd, A. E.; Harrison, A. P.; Thornton, J. M.; Orengo, C. A. *Nucleic Acids Res.* **2000**, *28*, 277.
(30) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. *J. Mol. Biol.* **1995**, *247*, 536.
(31) Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157.
(32) Tang, J.; James, M. N.; Hsu, I. N.; Jenkins, J. A.; Blundell, T. L. *Nature* **1978**, *271*, 618.
(33) Davies, D. R. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 189.
(34) Blundell, T. L.; Cooper, J. B.; Sali, A.; Zhu, Z. Y. *Adv. Exp. Med. Biol.* **1991**, *306*, 443.
(35) Sali, A.; Veerapandian, B.; Cooper, J. B.; Moss, D. S.; Hofmann, T.; Blundell, T. L. *Proteins* **1992**, *12*, 158.
(36) Silva, A. M.; Lee, A. Y.; Gulnik, S. V.; Maier, P.; Collins, J.; Bhat, T. N.; Collins, P. J.; Cachau, R. E.; Luker, K. E.; Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Goldberg, D. E.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034.
(37) Wlodawer, A.; Gustchina, A. *Biochim. Biophys. Acta* **2000**, *1477*, 16.
(38) Perona, J. J.; Craik, C. S. *J. Biol. Chem.* **1997**, *272*, 29987.
(39) Tong, L.; Qian, C.; Massariol, M. J.; Bonneau, P. R.; Cordingley, M. G.; Lagace, L. *Nature* **1996**, *383*, 272.
(40) Bode, W.; Gomis-Ruth, F. X.; Stockler, W. *FEBS Lett.* **1993**, *331*, 134.
(41) Hooper, N. M. *FEBS Lett.* **1994**, *354*, 1.
(42) Seemuller, E.; Lupas, A.; Stock, D.; Lowe, J.; Huber, R.; Baumeister, W. *Science* **1995**, *268*, 579.
(43) Coux, O.; Tanaka, K.; Goldberg, A. L. *Annu. Rev. Biochem.* **1996**, *65*, 801.
(44) Groll, M.; Ditzel, L.; Lowe, J.; Stock, D.; Bochtler, M.; Bartunik, H. D.; Huber, R. *Nature* **1997**, *386*, 463.
(45) Groll, M.; Brandstetter, H.; Bartunik, H.; Bourenkow, G.; Huber, R. *J. Mol. Biol.* **2003**, *327*, 75.
(46) Brandstetter, H.; Kim, J. S.; Groll, M.; Huber, R. *Nature* **2001**, *414*, 466.
(47) Kim, J. S.; Groll, M.; Musiol, H. J.; Behrendt, R.; Kaiser, M.; Moroder, L.; Huber, R.; Brandstetter, H. *J. Mol. Biol.* **2002**, *324*, 1041.
(48) Brandstetter, H.; Kim, J. S.; Groll, M.; Gottig, P.; Huber, R. *Biol. Chem.* **2002**, *383*, 1157.
(49) Fulop, V.; Bocskei, Z.; Polgar, L. *Cell* **1998**, *94*, 161.
(50) Polgar, L. *Cell Mol. Life Sci.* **2002**, *59*, 349.
(51) Holmquist, M. *Curr. Protein Pept. Sci.* **2000**, *1*, 209.
(52) Vandeputte-Rutten, L.; Kramer, R. A.; Kroon, J.; Dekker, N.; Egmond, M. R.; Gros, P. *EMBO J.* **2001**, *20*, 5033.
(53) Botos, I.; Melnikov, E. E.; Cherry, S.; Tropea, J. E.; Khalatova, A. G.; Rasulova, F.; Dauter, Z.; Maurizi, M. R.; Rotanova, T. V.; Wlodawer, A.; Gustchina, A. *J. Biol. Chem.* **2004**, *279*, 8140.
(54) Fujinaga, M.; Cherney, M. M.; Oyama, H.; Oda, K.; James, M. N. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3364.
(55) Murao, S.; Oda, K. In *Aspartic Proteinases and their Inhibitors*; Kosta, V., Ed.; Walter de Gruyter: Berlin, 1985.
(56) Prabu-Jeyabalan, M.; Nalivaika, E. A.; King, N. M.; Schiffer, C. A. *J. Virol.* **2003**, *77*, 1306.
(57) Laco, G. S.; Schalk-Hihi, C.; Lubkowski, J.; Morris, G.; Zdanov, A.; Olson, A.; Elder, J. H.; Wlodawer, A.; Gustchina, A. *Biochemistry* **1997**, *36*, 10696.
(58) Phan, J.; Zdanov, A.; Evdokimov, A. G.; Tropea, J. E.; Peters, H. K., III; Kapust, R. B.; Li, M.; Wlodawer, A.; Waugh, D. S. *J. Biol. Chem.* **2002**, *277*, 50564.
(59) Turk, B. E.; Wong, T. Y.; Schwarzenbacher, R.; Jarrell, E. T.; Leppla, S. H.; Collier, R. J.; Liddington, R. C.; Cantley, L. C. *Nat. Struct. Mol. Biol.* **2004**, *11*, 60.
(60) Yamamoto, A.; Tomoo, K.; Doi, M.; Ohishi, H.; Inoue, M.; Ishida, T.; Yamamoto, D.; Tsuboi, S.; Okamoto, H.; Okada, Y. *Biochemistry* **1992**, *31*, 11305.
(61) Rose, R. B.; Craik, C. S.; Douglas, N. L.; Stroud, R. M. *Biochemistry* **1996**, *35*, 12933.
(62) LaPlante, S. R.; Aubry, N.; Bonneau, P. R.; Cameron, D. R.; Lagace, L.; Massariol, M. J.; Montpetit, H.; Plouffe, C.; Kawai, S. H.; Fulton, B. D.; Chen, Z.; Ni, F. *Biochemistry* **1998**, *37*, 9793.
(63) Nienaber, V. L.; Breddam, K.; Birktoft, J. J. *Biochemistry* **1993**, *32*, 11469.
(64) Hanson, M. A.; Stevens, R. C. *Nat. Struct. Biol.* **2000**, *7*, 687.
(65) Segelke, B.; Knapp, M.; Kadkhodayan, S.; Balhorn, R.; Rupp, B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6888.
(66) Holden, H. M.; Matthews, B. W. *J. Biol. Chem.* **1988**, *263*, 3256.
(67) Holland, D. R.; Tronrud, D. E.; Pley, H. W.; Flaherty, K. M.; Stark, W.; Jansonius, J. N.; McKay, D. B.; Matthews, B. W. *Biochemistry* **1992**, *31*, 11310.
(68) Betzel, C.; Singh, T. P.; Visanji, M.; Peters, K.; Fittkau, S.; Saenger, W.; Wilson, K. S. *J. Biol. Chem.* **1993**, *268*, 15854.
(69) Saxena, A. K.; Singh, T. P.; Peters, K.; Fittkau, S.; Betzel, C. *Protein Sci.* **1996**, *5*, 2453.
(70) Saxena, A. K.; Singh, T. P.; Peters, K.; Fittkau, S.; Visanji, M.; Wilson, K. S.; Betzel, C. *Proteins* **1996**, *25*, 195.
(71) Fairlie, D. P.; Tyndall, J. D. A.; Reid, R. C.; Wong, A. K.; Abbenante, G.; Scanlon, M. J.; March, D. R.; Bergman, D. A.; Chai, C. L.; Burkett, B. A. *J. Med. Chem.* **2000**, *43*, 1271.
(72) Abbenante, G.; March, D. R.; Bergman, D. A.; Hunt, P. A.; Garnham, B.; Dancer, R. J.; Martin, J. L.; Fairlie, D. P. *J. Am. Chem. Soc.* **1995**, *117*, 10220.
(73) March, D. R.; Abbenante, G.; Bergman, D. A.; Brinkworth, R. I.; Wickramasinghe, W.; Begun, J.; Martin, J. L.; Fairlie, D. P. *J. Am. Chem. Soc.* **1996**, *118*, 3375.
(74) Reid, R. C.; March, D. R.; Dooley, M. J.; Bergman, D. A.; Abbenante, G.; Fairlie, D. P. *J. Am. Chem. Soc.* **1996**, *118*, 8511.
(75) Reid, R. C.; Kelso, M. J.; Scanlon, M. J.; Fairlie, D. P. *J. Am. Chem. Soc.* **2002**, *124*, 5673.
(76) James, M. N. G.; Sielecki, A. R. In *Biological Macromolecules and Assemblies: Active sites of enzymes*; Jurnak, F. A., McPherson, A., Eds.; John Wiley & Sons: New York, 1987; Vol. 3.
(77) Dunn, B. M. *Chem. Rev.* **2002**, *102*, 4431.
(78) Rutenber, E.; Fauman, E. B.; Keenan, R. J.; Fong, S.; Furth, P. S.; Ortiz de Montellano, P. R.; Meng, E.; Kuntz, I. D.; DeCamp, D. L.; Salto, R.; et al. *J. Biol. Chem.* **1993**, *268*, 15343.
(79) Majer, P.; Collins, J. R.; Gulnik, S. V.; Erickson, J. W. *Protein Sci.* **1997**, *6*, 1458.
(80) Cooper, J. B. *Curr. Drug Targets* **2002**, *3*, 155.
(81) Groves, M. R.; Dhanaraj, V.; Badasso, M.; Nugent, P.; Pitts, J. E.; Hoover, D. J.; Blundell, T. L. *Protein Eng.* **1998**, *11*, 833.
(82) Hong, L.; Koelsch, G.; Lin, X.; Wu, S.; Terzyan, S.; Ghosh, A. K.; Zhang, X. C.; Tang, J. *Science* **2000**, *290*, 150.
(83) Fruton, J. S. *Q. Rev. Biol.* **2002**, *77*, 127.
(84) Ng, K. K.; Petersen, J. F.; Cherney, M. M.; Garen, C.; Zalatoris, J. J.; Rao-Naik, C.; Dunn, B. M.; Martzen, M. R.; Peanasky, R. J.; James, M. N. *Nat. Struct. Biol.* **2000**, *7*, 653.

(85) Li, M.; Phylip, L. H.; Lees, W. E.; Winther, J. R.; Dunn, B. M.; Wlodawer, A.; Kay, J.; Gustchina, A. *Nat. Struct. Biol.* **2000**, *7*, 113.
(86) Pisabarro, M. T.; Ortiz, A. R.; Palomer, A.; Cabre, F.; Garcia, L.; Wade, R. C.; Gago, F.; Mauleon, D.; Carganico, G. *J. Med. Chem.* **1994**, *37*, 337.
(87) Bode, W.; Fernandez-Catalan, C.; Tschesche, H.; Grams, F.; Nagase, H.; Maskos, K. *Cell Mol. Life Sci.* **1999**, *55*, 639.
(88) Babine, R. E.; Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359.
(89) Massova, I.; Kotra, L. P.; Fridman, R.; Mobashery, S. *FASEB J.* **1998**, *12*, 1075.
(90) Hege, T.; Feltzer, R. E.; Gray, R. D.; Baumann, U. *J. Biol. Chem.* **2001**, *276*, 35087.
(91) Park, H. I.; Ming, L. J. *J. Biol. Inorg. Chem.* **2002**, *7*, 600.
(92) Natesh, R.; Schwager, S. L.; Sturrock, E. D.; Acharya, K. R. *Nature* **2003**, *421*, 551.
(93) Towler, P.; Staker, B.; Prasad, S. G.; Menon, S.; Tang, J.; Parsons, T.; Ryan, D.; Fisher, M.; Williams, D.; Dales, N. A.; Patane, M. A.; Pantoliano, M. W. *J. Biol. Chem.* **2004**, *279*, 17996.
(94) Maskos, K.; Fernandez-Catalan, C.; Huber, R.; Bourenkov, G. P.; Bartunik, H.; Ellestad, G. A.; Reddy, P.; Wolfson, M. F.; Rauch, C. T.; Castner, B. J.; Davis, R.; Clarke, H. R.; Petersen, M.; Fitzner, J. N.; Cerretti, D. P.; March, C. J.; Paxton, R. J.; Black, R. A.; Bode, W. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3408.
(95) Grams, F.; Reinemer, P.; Powers, J. C.; Kleine, T.; Pieper, M.; Tschesche, H.; Huber, R.; Bode, W. *Eur. J. Biochem.* **1995**, *228*, 830.
(96) Fernandez-Catalan, C.; Bode, W.; Huber, R.; Turk, D.; Calvete, J. J.; Lichte, A.; Tschesche, H.; Maskos, K. *EMBO J.* **1998**, *17*, 5238.
(97) 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.; Bode, W. *Nature* **1997**, *389*, 77.
(98) Oefner, C.; D'Arcy, A.; Hennig, M.; Winkler, F. K.; Dale, G. E. *J. Mol. Biol.* **2000**, *296*, 341.
(99) Cirilli, M.; Gallina, C.; Gavuzzo, E.; Giordano, C.; Gomis-Ruth, F. X.; Gorini, B.; Kress, L. F.; Mazza, F.; Paradisi, M. P.; Pochetti, G.; Politi, V. *FEBS Lett.* **1997**, *418*, 319.
(100) Powers, J. C.; Harper, J. W. In *Protease Inhibitors*; Barret, A. J. S., Salverson, G., Eds.; Elsevier: Amsterdam, 1986.
(101) Rawlings, N. D.; Barrett, A. J. *Methods Enzymol.* **1994**, *244*, 19.
(102) Barrett, A. J.; Rawlings, N. D. *Arch. Biochem. Biophys.* **1995**, *318*, 247.
(103) Krem, M. M.; Di Cera, E. *EMBO J.* **2001**, *20*, 3036.
(104) Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E. *Chem. Rev.* **2002**, *102*, 4639.
(105) Hedstrom, L. *Chem. Rev.* **2002**, *102*, 4501.
(106) Ye, S.; Goldsmith, E. J. *Curr. Opin. Struct. Biol.* **2001**, *11*, 740.
(107) Gettins, P. G. *Chem. Rev.* **2002**, *102*, 4751.
(108) Pike, R. N.; Bottomley, S. P.; Irving, J. A.; Bird, P. I.; Whisstock, J. C. *IUBMB Life* **2002**, *54*, 1.
(109) Wlodawer, A.; Li, M.; Gustchina, A.; Oyama, H.; Dunn, B. M.; Oda, K. *Acta Biochim. Pol.* **2003**, *50*, 81.
(110) Wlodawer, A.; Li, M.; Dauter, Z.; Gustchina, A.; Uchida, K.; Oyama, H.; Dunn, B. M.; Oda, K. *Nat. Struct. Biol.* **2001**, *8*, 442.
(111) Sleat, D. E.; Donnelly, R. J.; Lackland, H.; Liu, C. G.; Sohar, I.; Pullarkat, R. K.; Lobel, P. *Science* **1997**, *277*, 1802.
(112) Singh, T. P.; Sharma, S.; Karthikeyan, S.; Betzel, C.; Bhatia, K. L. *Proteins* **1998**, *33*, 30.
(113) Paetzel, M.; Karla, A.; Strynadka, N. C.; Dalbey, R. E. *Chem. Rev.* **2002**, *102*, 4549.
(114) Paetzel, M.; Dalbey, R. E.; Strynadka, N. C. *Nature* **1998**, *396*, 186.
(115) Steinmetz, A. C.; Demuth, H. U.; Ringe, D. *Biochemistry* **1994**, *33*, 10535.
(116) Greco, M. N.; Hawkins, M. J.; Powell, E. T.; Almond, H. R., Jr.; Corcoran, T. W.; de Garavilla, L.; Kauffman, J. A.; Recacha, R.; Chattopadhyay, D.; Andrade-Gordon, P.; Maryanoff, B. E. *J. Am. Chem. Soc.* **2002**, *124*, 3810.
(117) MacSweeney, A.; Birrane, G.; Walsh, M. A.; O'Connell, T.; Malthouse, J. P.; Higgins, T. M. *Acta Crystallogr, Sect. D: Biol. Crystallogr.* **2000**, *56* (Pt 3), 280.
(118) Ohbayashi, H. *IDrugs* **2002**, *5*, 910.
(119) Tremblay, G. M.; Janelle, M. F.; Bourbonnais, Y. *Curr. Opin. Investig. Drugs* **2003**, *4*, 556.
(120) Froelich, C. J. In *Nature Encyclopedia of Life Sciences.*; Nature Publishing Group: London, 1999.
(121) Regoli, D.; Calo, G.; Gobeil, F. In *Nature Encyclopedia of Life Sciences*; Nature Publishing Group: London, 1999.
(122) Yousef, G. M.; Diamandis, E. P. *Biol. Chem.* **2002**, *383*, 1045.
(123) Mittl, P. R.; Di Marco, S.; Fendrich, G.; Pohlig, G.; Heim, J.; Sommerhoff, C.; Fritz, H.; Priestle, J. P.; Grutter, M. G. *Structure* **1997**, *5*, 253.
(124) Zhou, A.; Webb, G.; Zhu, X.; Steiner, D. F. *J. Biol. Chem.* **1999**, *274*, 20745.
(125) Holyoak, T.; Wilson, M. A.; Fenn, T. D.; Kettner, C. A.; Petsko, G. A.; Fuller, R. S.; Ringe, D. *Biochemistry* **2003**, *42*, 6709.
(126) Oberst, M. D.; Williams, C. A.; Dickson, R. B.; Johnson, M. D.; Lin, C. Y. *J. Biol. Chem.* **2003**, *278*, 26773.
(127) Johnson, M. D.; Oberst, M. D.; Lin, C. Y.; Dickson, R. B. *Expert Rev. Mol. Diagn.* **2003**, *3*, 331.
(128) Narayanan, A. S.; Thiagarajan, P. In *Nature Encyclopedia of Life Sciences*; Nature Publishing Group: London, 2000.
(129) Van Aken, H.; Bode, C.; Darius, H.; Diehm, C.; Encke, A.; Gulba, D. C.; Haas, S.; Hacke, W.; Puhl, W.; Quante, M.; Riess, H.; Scharf, R.; Schellong, S.; Schror, T.; Schulte, K. L.; Tebbe, U. *Clin. Appl. Thromb. Hemost.* **2001**, *7*, 195.
(130) Powers, J. C.; Harper, J. W. In *Protease Inhibitors*; Barret, A. J. S., Salverson, G., Eds.; Elsevier: Amsterdam, 1986.
(131) Berkner, K. L. In *Nature Encyclopedia of Life Sciences*; Nature Publishing Group: London, 2000.
(132) Friedrich, T.; Kroger, B.; Bialojan, S.; Lemaire, H. G.; Hoffken, H. W.; Reuschenbach, P.; Otte, M.; Dodt, J. *J. Biol. Chem.* **1993**, *268*, 16216.
(133) van de Locht, A.; Lamba, D.; Bauer, M.; Huber, R.; Friedrich, T.; Kroger, B.; Hoffken, W.; Bode, W. *EMBO J.* **1995**, *14*, 5149.
(134) Weitz, J. I.; Crowther, M. *Thromb. Res.* **2002**, *106*, V275.
(135) Vitali, J.; Martin, P. D.; Malkowski, M. G.; Robertson, W. D.; Lazar, J. B.; Winant, R. C.; Johnson, P. H.; Edwards, B. F. *J. Biol. Chem.* **1992**, *267*, 17670.
(136) Mathews, II; Padmanabhan, K. P.; Ganesh, V.; Tulinsky, A.; Ishii, M.; Chen, J.; Turck, C. W.; Coughlin, S. R.; Fenton, J. W., II *Biochemistry* **1994**, *33*, 3266.
(137) Potempa, J.; Korzus, E.; Travis, J. *J. Biol. Chem.* **1994**, *269*, 15957.
(138) 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.; Whisstock, J. C. *J. Biol. Chem.* **2001**, *276*, 33293.
(139) Carrell, R. W.; Huntington, J. A. *Biochem. Soc. Symp.* **2003**, 163.
(140) Huber, R.; Bode, W. *Acc. Chem. Res.* **1978**, *11*, 114.
(141) Huntington, J. A.; Read, R. J.; Carrell, R. W. *Nature* **2000**, *407*, 923.
(142) Braud, S.; Le Bonniec, B. F.; Bon, C.; Wisner, A. *Biochemistry* **2002**, *41*, 8478.
(143) Rawlings, N. D.; Barrett, A. J. *Methods Enzymol.* **1994**, *244*, 461.
(144) Rasnick, D. In *Perspectives in Drug Discovery and Design*; McKerrow, J. H., James, M. N. G., Eds.; ESCOM: Leiden, 1996; Vol. 6.
(145) Otto, H. H.; Schirmeister, T. *Chem. Rev.* **1997**, *97*, 133.
(146) Lecaille, F.; Kaleta, J.; Bromme, D. *Chem. Rev.* **2002**, *102*, 4459.
(147) Denault, J. B.; Salvesen, G. S. *Chem. Rev.* **2002**, *102*, 4489.
(148) Ryan, M. D.; Flint, M. *J. Gen. Virol.* **1997**, *78* (Pt 4), 699.
(149) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13190.
(150) Filipek, R.; Rzychon, M.; Oleksy, A.; Gruca, M.; Dubin, A.; Potempa, J.; Bochtler, M. *J. Biol. Chem.* **2003**, *278*, 40959.
(151) Cazzulo, J. J. *Curr. Top. Med. Chem.* **2002**, *2*, 1261.
(152) Stennicke, H. R.; Salvesen, G. S. *Biochim. Biophys. Acta* **1998**, *1387*, 17.
(153) Thornberry, N. A.; Lazebnik, Y. *Science* **1998**, *281*, 1312.
(154) Wei, Y.; Fox, T.; Chambers, S. P.; Sintchak, J.; Coll, J. T.; Golec, J. M.; Swenson, L.; Wilson, K. P.; Charifson, P. S. *Chem. Biol.* **2000**, *7*, 423.
(155) Schweizer, A.; Briand, C.; Grutter, M. G. *J. Biol. Chem.* **2003**, *278*, 42441.
(156) Riedl, S. J.; Renatus, M.; Schwarzenbacher, R.; Zhou, Q.; Sun, C.; Fesik, S. W.; Liddington, R. C.; Salvesen, G. S. *Cell* **2001**, *104*, 791.
(157) Xu, G.; Cirilli, M.; Huang, Y.; Rich, R. L.; Myszka, D. G.; Wu, H. *Nature* **2001**, *410*, 494.
(158) Turk, D.; Guncar, G. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 203.
(159) Turk, D.; Podobnik, M.; Popovic, T.; Katunuma, N.; Bode, W.; Huber, R.; Turk, V. *Biochemistry* **1995**, *34*, 4791.
(160) Jenko, S.; Dolenc, I.; Guncar, G.; Dobersek, A.; Podobnik, M.; Turk, D. *J. Mol. Biol.* **2003**, *326*, 875.
(161) Guncar, G.; Pungercic, G.; Klemencic, I.; Turk, V.; Turk, D. *EMBO J.* **1999**, *18*, 793.
(162) Stock, D.; Ditzel, L.; Baumeister, W.; Huber, R.; Lowe, J. *Cold Spring Harbor Symp. Quant. Biol.* **1995**, *60*, 525.
(163) Groll, M.; Kim, K. B.; Kairies, N.; Huber, R.; Crews, C. M. *J. Am. Chem. Soc.* **2000**, *122*, 1237.
(164) Kisselev, A. F.; Songyang, Z.; Goldberg, A. L. *J. Biol. Chem.* **2000**, *275*, 14831.
(165) Sousa, M. C.; Kessler, B. M.; Overkleeft, H. S.; McKay, D. B. *J. Mol. Biol.* **2002**, *318*, 779.
(166) Pickart, C. M.; Cohen, R. E. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 177.
(167) Bochtler, M.; Ditzel, L.; Groll, M.; Hartmann, C.; Huber, R. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 295.
(168) Voorhees, P. M.; Dees, E. C.; O'Neil, B.; Orlowski, R. Z. *Clin. Cancer Res.* **2003**, *9*, 6316.

(169) Lara, P. N., Jr.; Davies, A. M.; Mack, P. C.; Mortenson, M. M.; Bold, R. J.; Gumerlock, P. H.; Gandara, D. R. *Semin. Oncol.* **2004**, *31*, 40.
(170) Burgen, A. S.; Roberts, G. C.; Feeney, J. *Nature* **1975**, *253*, 753.
(171) Jencks, W. P. In *Chemistry and Enzymology*; Dover Publications: New York, 1987.
(172) Loughlin, W. A.; Glenn, M. P.; Tyndall, J. D. A.; Fairlie, D. P. *Chem. Rev.* **2004**, *104*, 6085.
(173) Tyndall, J. D. A.; Fairlie, D. P. *J. Mol. Recognit.* **1999**, *12*, 363.
(174) Tyndall, J. D. A.; Fairlie, D. P. *Curr. Med. Chem.* **2001**, *8*, 893.
(175) Glenn, M. P.; Pattenden, L. K.; Reid, R. C.; Tyssen, D. P.; Tyndall, J. D. A.; Birch, C. J.; Fairlie, D. P. *J. Med. Chem.* **2002**, *45*, 371.
(176) Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946.
(177) Merritt, E. A.; Bacon, D. J. *Methods Enzymol.* **1997**, *277*, 505.
(178) Accelrys: San Diego, CA, 2004.

CR040669E