Structure and Mechanism of the Pepsin-Like Family of Aspartic Peptidases

Ben M. Dunn*

Department of Biochemistry & Molecular Biology, University of Florida College of Medicine, Gainesville, Florida 32610-0245

Received May 15, 2002

Contents

I. Introduction

The last general review of the aspartic peptidase family was by David R. Davies in 1990.[1] Since then, several reviews on selected aspects of the structure and function of aspartic peptidases have appeared, and three books summarizing conferences[2−4] and a monograph[5] have been published. This review will focus on developments since 1990, which include the isolation, cloning, and expression of a large number of new members of this family, crystal structures, insights into mechanism, and studies of inhibitors of this class. Note: this review will use the terms "peptidase" and "endopeptidase" to describe enzymes that have been called either "protease" or "proteinase" in the past, as recommended by the editors of the *Handbook of Proteolytic Enzymes*.[6]

* To whom correspondence should be addressed. Phone: (352) 392-3362. Fax: (352) 846−0412. E-mail: bdunn@college.med.ufl.edu.

Ben M. Dunn received the B.S. degree in Chemistry from Delaware in 1967. He earned the Ph.D. in BioOrganic Chemistry from UC Santa Barbara in 1971, working under Prof. Tom Bruice. He was a postdoctoral fellow of the NIH from 1971−73 with Christian Anfinsen, and served one year as a Staff Fellow in the Laboratory of Chemical Biology. He began his independent career at the University of Florida as an Assistant Professor of Biochemistry and Molecular Biology and rose to his current rank of Distinguished Professor in the same institution. Prof. Dunn has served on many NIH review panels and edits two journals, *Protein and Peptide Letters* for rapid reports, and *Current Protein and Peptide Science* for reviews. He is a co-editor of *Current Protocols in Protein Science* and is a member of the Council of the American Peptide Society and of the International Proteolysis Society.

The MEROPS database [www.Merops.co.uk], now considered to be the authoritative compilation of information on peptidases in general, lists a number of families of the aspartic peptidases. However, if one considers the presence of the sequence, -Asp−Thr-(Ser)−Gly-, to be essential to be an aspartic peptidase, then we need consider only families A1 (pepsin-like) with 282 members, A2 (retroviral peptidases) with 93 members, A3a (cauliflower mosaic virus endopeptidase) with 16 members, A4b (bacilliform virus putative endopeptidase) with three members, A4 (aspergillopepsin) with four members, A9 (human spumaretrovirus) with one member, A11a (copia transposon) with 51 members, A12 (retrotransposons bs1 endopeptidase) with a single member, and A16 (Tas retrotransposons endopeptidase) with four members. As little is known about the latter categories, this review will focus on family A1, the pepsin-like enzymes. Table 1 presents a listing of enzymes whose three-dimensional structures have been determined and gives the PDB code for the structure of highest resolution in each group (discussed further in section III). The A2 family that includes HIV-1 peptidase has been the subject of extensive reviews[7] and will not be considered in detail here.

10.1021/cr010167q CCC: $39.75
Published on Web 11/09/2002

Table 1. Number of Structures of Members of the Aspartic Proteinase Family Found in the Publicly Accessible Databases[a]

class	name	source	#	PDB code
mammalian	pepsin	human	3	1QRP
		pig	7	4PEP
		Atlantic cod	1	1AM5
	gastricsin	human	2	1HTR
	renin	human	7	1HRN
		mouse	1	1SMB
	cathepsin D	human	3	1LYA
	BACE	human	1	1FKN
	uropepsin	human	1	1FLH
	chymosin	calf	4	3CMS
fungal	endothiapepsin	*E. parasitica*	27	1EPM
	penicillopepsin	*P. janthinellum*	14	1BXO
	rhizopuspepsin	*R. chinensis*	5	3APR
	SAPs	*C. albicans*	2	1EAG
		C. tropicalis	1	1J71
	proteinase A	*S. cerevisiae*	9	1DPJ
	mucorpepsin	*M. pussillus*	2	1MMP
		R. miehei	2	2ASI
	aspergillopepsin	*A. phoenicis*	1	1IBQ
protozoan	plasmepsins	*P. falciparum*	2	1PFZ
		P. vivax	1	1QS8
		C. chabaudi	1	model
plants	phytepsin	*H. vulgare*	1	1QDM
	cardosin	*C. cardunculus*	1	1B5F
viral	HIV-1		211	1DAZ
	HIV-2		19	1IDA
	SIV		6	1AZ5
	FIV		7	4FIV
	EIAV		2	2FMB

[a] The principal source is the Protein Data Bank (PDB) and the PDB codes of the highest-resolution structure in each category are shown. The reference for HIV-1, HIV-2, and SIV proteases is HIVdb (http://srdata.nist.gov/HIVdb), which, in addition to the structures available in the PDB, provides a number of additional ones as well. Abbreviations: BACE = β-amyloid converting enzyme; SAPs = secreted aspartic peptidases of *Candida*; SIV = simian immunodeficiency virus peptidase; FIV = feline immunodeficiency virus peptidase; EIAV = equine infectious anemia virus peptidase.

The listing above does not include the presenilins, as they do not have the signature sequence and clearly do not have the characteristic three-dimensional structure of the aspartic peptidase family. However, they and other families listed in the MEROPS database may be new types of proteolytic enzymes of unknown mechanism. Deciphering their properties will be an exciting new chapter in the world of proteolysis, and the presenilins are discussed briefly in this review.

This review will consider first recent studies that have led to the discovery of new members of the family in a wide variety of organisms. Next, the structure of the typical member of the family will be described and recent structural determinations will be highlighted. Following this, the interesting properties of the N-terminal prosegment of the family members will be discussed. Recent studies of the role of specific residues in and around the active site cleft as probed by mutagenesis will be described and various inhibitors will be presented. Finally, new information on the mechanism of these enzymes will be presented and discussed.

II. Cloning and Expression of New Aspartic Endopeptidases

A. Parasite Peptidases

It is now recognized that peptidases play an important role in the life cycle of parasites of man and other animals. Jean et al. have provided an important summary of the gene structure of aspartic peptidases from yeast, fungi, nematodes, vertebrates, plants, and protozoans.[8] Coombs et al. discussed the complex situation in *Plasmodium falciparum* where a total of 10 genes have been discovered[9] in the genome sequencing project. These are designated plasmepsins 1−10 and have overall homology to each other and to other aspartic peptidase family members. However, specific sequences characterize each one. Four of the gene products have been found[10] in the parasite food vacuole where hemoglobin is digested during the inter-erythrocytic phase of the life cycle. Plasmepsins I and II have been described in detail.[11−16]

Banerjee et al. used immunoaffinity chromatography to separate the histo-aspartic-peptidase or HAP from preparations of the food vacuole.[10] This enzyme of unusual sequence,[17] with His replacing Asp at position 32 (pepsin numbering) and Ala replacing Gly at position 34, was shown by Banerjee et al.[10] to be active, but with a pH optimum that is shifted to about pH 6, at least one pH unit higher than the optimum for other plasmepsins that have been studied. Curiously, this enzyme could not be expressed in an active form by recombinant methods.[17]

In addition to studies of native HAP, Banerjee et al. also expressed plasmepsin IV from a pET15b vector in *Escherichia coli* and determined[10] the kinetic parameters for cleavage of a synthetic substrate representing the α33−34 cleavage site in hemoglobin. Plasmepsin IV was more efficient than HAP at cleaving this peptide, while HAP was more efficient at cleaving globin chains. Banerjee et al. also searched for transcripts of plasmepsin genes with RT-PCR on total RNA from asynchronous parasites grown in erythrocytes. In this study, plasmepsins VI, VII, and VIII could not be detected. However, plasmepsin VI has been cloned from an mRNA preparation from *P. falciparum* in our laboratory (B. Beyer and B. Dunn, unpublished results).

Wyatt and Berry have reported on the properties of plasmepsin IV expressed from the pET3b vector in *E. coli*.[18] They demonstrated that this enzyme was able to cut hemoglobin at the α33−34 cleavage site by sequence analysis of an early product in digestion of the protein substrate. Plasmepsin IV was also reasonably efficient in cleavage of several synthetic substrates that had been used earlier to study plasmepsins I and II,[19] and was inhibited by similar low molecular weight compounds. However, the K_i values differed from those obtained against plasmepsin II.

LeBonniec et al. found that plasmepsin II was active at neutral pH by finding evidence for digestion of the erythrocyte membrane protein, spectrin, and thus raising the possibility that plasmepsin II has

other roles beside hemoglobin degradation in the parasite food vacuole.[20] Wyatt and Berry demonstrated that plasmepsin IV also has activity against spectrin.[18]

Humphreys et al. have reported on the catalytic properties of a recombinant *Plasmodium berghei* plasmepsin.[21] This species of the malarial parasite selectively infects rodents. As such, this may provide a good small animal model for studies of antimalarial treatments in vivo. The *P. berghei* plasmepsin had a sequence with great similarity to other plasmepsins known at that time and demonstrated catalytic properties consistent with those of the other plasmepsins. It remains to be determined if the single enzyme of this type found in *P. berghei* is the only one present. The fact that *P. falciparum* contains 10 genes indicates that additional *P. berghei* enzymes may exist.

A series of reports from Brindley and colleagues[22] has established that other parasites that feed on red blood cells contain aspartic peptidase activity. The cloning, expression, and characterization of a cathepsin D-like peptidase from *Schistosoma japonicum* (Sjpasp) were reported by Becker et al. Comparison of the sequence of the cloned gene to aspartic peptidase known at that time revealed the closest identity (55%) to the cathepsin D group of enzymes; however, the *S. japonicum* enzyme lacks the internal processing sites that produces cathepsin D in the two-chain form. Brinkworth et al. have speculated on the origin of host range specificities in blood-feeding parasites.[23] Their hypothesis is based on molecular modeling that suggests that an aspartic peptidase from canine-specific hookworm would have better binding to the canine hemoglobin than for human hemoglobin. While this modeling exercise was carefully done and the high sequence identity to human cathepsin D favors homology modeling, the conclusions must be considered tentative until sufficient experiments are performed to define the active site specificity. In that regard, Brindley and colleagues have reported the localization of *S. japonicum* and *S. mansoni* activity to the gastrodermis and superficial digestive vacuoles, and the cleavage specificity on human hemoglobin for the *S. japonicum* enzyme.[24] Harrop et al. have also cloned a cathepsin D-like aspartic endopeptidase from the canine hookworm *Ancylostoma caninum* and identified elements in its amino acid sequence that predict differences in substrate specificity from other enzymes of this class.[25] The identity to the lysosomal aspartic peptidase of *Aedes aegypti* was 47%, with Sjpasp was 46%, and with mammalian cathepsin D was 47−50%.

Jolodar and Miller identified a cDNA from *Onchocerca volvulus*, the parasite that causes "river blindness", that had highest identity in the mature enzyme region to mammalian cathepsins D.[26] This sequence also had homology to sequences that had been identified in *Caenorhabditis elegans* and a sequence from an incomplete cDNA from the dog parasite *Ancyclostoma caninum*. However, some features of the *O. volvulus* sequence were more closely related to the cathepsin E subfamily. In particular, one of the N-glycosylation sites common to cathepsin D from several species that is involved in lysosomal targeting was missing. In addition, the nematode sequences have a large prosegment of more than 80 residues. This is intermediate between the prosegments of pepsinogen-like zymogens and the prosegment of malarial enzymes (ca. 124 amino acids).

Geier et al. utilized affinity chromatography on pepstatin-Agarose followed by two-dimensional gel electrophoresis to purify six proteins ranging in size from 26 to 52 kDa from homogenates of *C. elegans*.[27] Following N-terminal sequence analysis, five were identified to have significant identity (25−60%) to aspartic endopeptidases of other organisms. An additional seven genes of *C. elegans* were identified as members of this family by searches in the genomic DNA of the nematode. Two of the sequences are most closely related to the cathepsin D subfamily. A review by Hashmi et al. discussed[28] the current status of the *C. elegans* aspartic endopeptidase family of six members in the context of the genome-wide approach of functional genomics.

Tcherepanova et al. cloned one of the *C. elegans* aspartic peptidase genes (termed *asp-1* by these authors) and expressed the gene in insect cells.[29] This protein had a signal sequence of 16 amino acids, a prosegment of 41 amino acids, and a mature enzyme M_r of 35 556. A single N-linked glycosylation site apparently leads to addition of carbohydrate, which would account for the observed mature enzyme size of ca. 40 kDa on SDS−PAGE. The protein is exclusively transcribed in the intestinal cells of the nematode in the late embryonic stage and is not seen in adult worms or mature larvae.

Gallego et al. extended the search for aspartic endopeptidases to the human roundworm parasite *Strongyloides stercoralis*.[30] They identified a cDNA encoding a pepsinogen-like enzyme (36% identity to both human pepsinogen and gastricsin). This gene was expressed in the infective third stage larvae and the free-living larvae, suggesting that an inhibitor against this enzyme might be an effective antiparasitic agent.

B. Cathepsin D

Cathepsin D was shown to be an essential enzyme by the work of Saftig et al.[31] They created a knockout mouse with the gene for cathepsin D deleted and found homozygous mice develop normally for two weeks. The mice stop thriving in the third week and develop atrophy of the ileal mucosa followed by other degradation of intestinal organs. The authors speculate that limited proteolysis of growth factors is required for normal maturation of the mouse. In these knockout mice, lysosomal proteolysis was normal, indicating that proteolysis in that organelle is carried out by other enzymes, such as the cysteine peptidases.

Enzymes related to cathepsin D have been isolated from a number of organisms. De Stasio et al. isolated an enzyme, which was active at acidic pH values, of 40 kDa from lizard ovaries.[32] The cDNA was derived and the sequence revealed major identity with cathepsin D from chicken, trout, mouse, and human. This protein may play a role in processing of yolk

proteins in the ovaries. Northern blot analysis demonstrated that mRNA for the lizard cathepsin D was present at high levels during the reproductive period and especially in previtellogenic oocytes. A similar protein had been identified earlier in the rainbow trout, *Oncorhynchus mykiss*, by Brooks et al.[33] The trout cathepsin D was 81% identical to human cathepsin D.

Two proteins showing greater than 50% identity to aspartic peptidases were purified and cloned from the Antarctic icefish.[34,35] In the first report, one sequence was obtained from each of two related fish, *Chionodraco hamatus* and *Trematomus bernacchii*. These sequences had greater than 50% identity to cathepsin D, cathepsins E, pepsin, and gastricsin from various species, but the precise classification could not be made. The protein from *C. hamatus* identified in the second report from this group was more readily identified as a member of the cathepsin D subfamily of aspartic peptidases and was unusual in its high thermal stability and activity. While cold-adapted enzymes usually have reductions in the numbers of Pro, Arg, and aromatic amino acids, the cathepsin D from the icefish was not found to be different from other cathepsin Ds in the number of those amino acids.

The purification and characterization of cathepsin D from herring muscle was reported by Nielsen and Nielsen.[36] This enzyme had an amino terminal sequence that was very similar to the enzyme from the Antarctic icefish. Studies on the digestion of the B-chain of oxidized insulin revealed a specificity for hydrophobic amino acids in P1 and P1′.

Dragonetti et al. discussed[37] the sorting and secreting of cathepsin D in the rat in response to stimulation by molecules that trigger immune responses. This work suggests that targeting to the lysosome and the secretory pathway are interrelated and that cathepsin D can be secreted by a calcium-regulated pathway in a mature and active form.

C. Enzymes Involved in Cleaving the β-Amyloid Precursor Protein

The identification of a membrane-bound enzyme [β-secretase = beta amyloid cleaving enzyme (BACE) = Asp2 = memapsin 2] with the ability to cleave the β-amyloid precursor protein (APP) at the so-called β-processing site and with a sequential relationship to the pepsin family by five groups[38−43] provided a major impetus to the study of the causative factors in Alzheimer's disease. Starting from a genetics perspective, Acquati et al. reported that a gene that maps to the Down critical region of chromosome 21 is identical to BACE2, a second new human peptidase.[44] These findings also provided a boost to the aspartic peptidase field similar to that provided by the discovery of the HIV-1 peptidase more than a decade earlier. The subsequent crystal structure by Hong et al.[45] provided a target for drug design.

While the structure of the β-secretase is very similar to pepsin, there are small differences in the positions of several surface loops that may impact substrate and inhibitor specificity. In addition, the C-terminal extension, not revealed in the crystallographic analysis, provides a membrane anchoring location. This transmembrane domain has been shown to lead to localization in late Golgi compartments where the protein will have access to the APP.[46] In addition, the disulfide pairings of the protein are atypical for pepsin family members. Disulfides are found at positions Cys 216−Cys 420, Cys 278−Cys 443, and Cys 330−Cys 380, numbering from the beginning of the N-terminal segment.[47] Disulfide bridges in the pepsin-like enzymes tend to connect residues that are near neighbors in the sequence; in this respect, the beta-secretase is unusual in that it connects amino acids separated by from 50 to 204 amino acids. Pepsin, for example, has disulfides between Cys residues at 45−50, 206−210, and 248−282.

In addition to the action of the β-secretase, a second enzyme is necessary to cleave the intermediate formed to produce the amyloidogenic Aβ peptide. This second activity was termed γ-secretase, and early reports suggested that cathepsin D might be involved in this reaction.[48] This planted the idea that aspartic peptidases might be involved in Aβ production. While this has proven to be true for the β-secretase, opinion is still divided about the γ-secretase. A number of studies using peptidomimetic probes have been published,[49−51] and this area has been the subject of several comprehensive reviews.[52,53] Early studies using peptidomimetics based on cleavage site sequences found inhibition, but only in the micromolar region;[51,54] however, Shearman et al. found a compound by high-throughput screening, L-685,458, that had a 17 nM IC_{50} value.[50] This compound showed some inhibition of cathepsin D (1 μM) and papain (2 μM), although it was inactive against a panel of other enzymes. Easler et al. also demonstrated[51] that a compound with a bromoacetyl group attached at the amino terminus was able to react with a protein of 21 kDa that was suggested to be presenilin, a known transmembrane protein. Mutations within presenilin are known to lead to decreases in the production of the Aβ and increases in the precursors to that peptide. Presenilin has been proposed to exist within the membrane with eight transmembrane segments.

While the evidence is compelling, several aspects of these studies are troubling. First, a model has been proposed[53] whereby the transmembrane segment of the APP would bind in a central cavity in the presenilin structure. Two aspartic acids are suggested to be contributed by transmembrane domains 6 and 7 and to form an aspartic peptidase-like active site. Second, it is well-known that peptides that are folded up into a helical conformation are not accessible to proteolytic enzymes.[55] Therefore, it is difficult to see how cleavage can occur within a transmembrane complex. Second, while it is clear that two aspartic acids are important in interactions with presenilins, the structure of the transmembrane protein bears no resemblance to the structure of the pepsin family of enzymes discussed in this review. It may be that the presenilins are a new type of proteolytic enzyme, but it would be incorrect to

describe them as "aspartic peptidases", as has been done in the literature.

The use of peptidomimetic compounds to block the activity of enzymes in complex mixtures is also difficult to interpret. It is good to remember that a peptidomimetic compound by definition will resemble the transition state of cleavage of a peptide bond. The transition state for all types of proteolytic enzymes will, to some degree, resemble a tetrahedral structure following attack by either an enzyme nucleophile or by a water molecule as directed by the enzyme. Thus, the observation that a peptidomimetic compound inhibits a proteolytic process does not, per se, imply a particular type of enzyme is involved. Absolute proof will require cocrystallization of the compound with the protein target for structural determination.

D. Napsins (Novel Aspartic Peptidases)

The human genome sequencing project provided access to a large number of ESTs (extended sequence tags). By examining these, Tatnell et al.[56] were able to identify and subsequently clone two related sequences of genes for new enzymes from human tissue. Napsin A and B have identity to other human enzymes, 48.3% with human cathepsin D, 48.4% with cathepsin E, 43.9% with pepsin, 43% with renin, and 41.9% with gastricsin. Napsin A has an insert that includes an RGD sequence near the C-terminal end, which suggests that this protein may play a unique role. Schauer-Vukasinovic et al. expressed the napsin A gene and used the protein to produce a sequence-specific antibody against a region predicted to be surface exposed by a molecular model built on homology to cathepsin D.[57] The antibody detected napsin A protein in kidney and lung tissue. These authors also found that expression of napsin A in HEK293 cells resulted in the majority of the protein associated with membrane fractions. The napsin B gene is unusual in that it does not seem to contain a stop codon. Studies of expression levels in tissues revealed that lung had the highest level of mRNA for napsin A/B, followed by kidney, spleen, and leukocytes. Further analysis demonstrated that napsin B is expressed exclusively in spleen tissue, which may be consistent with a role in the immune system.

Chuman et al. reported that napsin A is highly expressed in normal lung and kidney tissue and is highly expressed in primary lung adenocarcinomas.[58] They suggest a functional role in processing of lung surfactant proteins. Napsin has also been identified in the mouse[59] and the rat.[60]

E. Cathepsin E

Among the human aspartic endopeptidases, pepsin and gastricsin are secreted into the lumen of the stomach, while renin is secreted into the blood. The intracellular enzymes cathepsin D and cathepsin E are distinct from the first three in that the former is targeted to the lysosomes of cells while the latter is neither secretory nor lysosomal but can be found in the endoplasmic reticulum, the trans-Golgi network, and the endosomal compartments of selected cells. Various groups have identified cDNAs for cathepsin E from humans,[61] rabbit,[62] guinea pig,[63] rat,[64] and mouse.[65,66] The structure of cathepsin E is also unique in that it has a Cys residue near the N-terminus that leads to the formation of an intermolecularly disulfide-linked dimer.[67] Cook et al. studied the regulation of gene expression of cathepsin E in various human tissues and came to the conclusion that expression is controlled by tissue-specific transcription factors.[68]

Cathepsin E involvement in antigen processing has been studied in antigen-presenting B cell lymphoblasts[69] and in primary cultured murine microglia cells.[70] The protein has been located in early endosomes of cells as well as associated with the plasma membrane, endoplasmic reticulum, and the Golgi apparatus.

Fowler et al.[71] created a Cys-to-Ala mutation in human cathepsin E in the region between the prosegment and the mature enzyme portion and obtained an active enzyme preparation. This unique Cys residue is present in all known cathepsin E species and is believed to form an intermolecular disulfide bond with a second molecule of cathepsin E. The dimer formation was demonstrated to lead to better pH and temperature stability, consistent with the biological niche of this protein in the endosomal vesicles.

F. Plant Aspartic Peptidases

In recent years, aspartic peptidases have been found in plants in increasing numbers.[72,73] Runeberg-Roos et al.[74] identified an activity in barley grain that was shown to be an aspartic peptidase present in the vacuolar compartment.[75] Verissimo et al.[76] purified two aspartic peptidases from flowers of the cardoon plant, *Cynara cardunculus*, and Ramalho-Santo et al.[77] reported on the specificity determined by using α-casein as substrate. These enzymes were termed cardosin A and B by this group and exhibited broad specificity for the residues acceptable in the P1 and P1′ positions. White et al.[78] described the properties of a different recombinant enzyme, cyprosin, from the same plant, with respect to the binding of a series of peptide inhibitors. When the gene was expressed in the yeast *Pichia pastoris*, the resulting protein purified from the culture medium was shown to consist of a two-chain enzyme where the unique plant specific insert had been processed out of the sequence. When the plant specific insert was deleted at the DNA level, the resulting expression gave the precursor, procyprosin. As processing to the mature, active form had not occurred, this result suggested that the plant specific insert may play a role in folding so that the final product is active or in targeting the nascent protein through the correct secretory pathway to ensure activation can take place. This conclusion agrees with that of Faro et al.[79] who studied cardosin A by molecular methods in plant tissue. After cloning the gene for cardosin A, they studied expression levels. An mRNA was found in young flower buds but not in mature or senescent pistils, which may indicate that expression is regulated at a developmental level. The precursor, procardosin, was found associated with the microsomal membrane of flower buds, while the mature two-chain form was found to be soluble.

D'Hondt et al.[80] cloned the cDNA for aspartic peptidase genes from the plants *Arabidopsis thaliana* and *Brassica napus* and reported one gene in *Arabidopsis* and four in *B. napus*. Subsequently, Mutlu et al.[81] demonstrated that the peptidase in *Arabidopsis* is present in seed tissue, specifically in the protein bodies. Domingos et al.[82] have studied a related enzyme from *Centaurea calcitrapa*, a thistle related to cardoon. The cDNA was cloned and the enzyme was expressed at highest levels in fresh flowers. However, it was also detected in seeds and leaves, unlike the cardoon enzymes. Again, extracts of fresh flowers yielded the proenzyme form, while extracts from dried flowers yielded the mature two-chain form. Leaves of the common bean, *Phaseolus vulgaris* L., and cowpea, *Vigna unguiculata* L. Walp, were subjected to water stress by withholding irrigation.[83] This caused the induction of expression of an aspartic peptidase in the cowpea, which was cloned and sequenced. The proteolytic activity of the leaves were tested using a series of class-specific inhibitors, and the results showed that aspartic peptidase activity accounts for a major portion of the total proteolysis when the plants are under stress. An aspartic peptidase was identified in rice by Asakura et al.,[84] purified,[85] and subsequently expressed in *E. coli*.[86] Payie et al.[87] purified a new aspartic proteinase from leaves of *Medicago sativa L.* (alfalfa) 3320-fold. They also demonstrated inhibition by pepstatin A and cleavage of synthetic substrates containing -Phe*Phe- bonds.

1. Plant Specific Inserts

In addition to a preprosequence and the normal 320−330 residue mature region, plant enzymes have an additional segment of approximately 100 residues known as the plant-specific insert (PSI). This segment is inserted into the C-terminal domain of the mature enzyme and seems to play a role in targeting the enzyme to the plant vacuole. The PSI is processed out of the enzyme following arrival in the lysosomal compartment. Ramalho-Santos et al. studied the processing events using antibodies against the PSI as well as against the prosegment.[88] They found that extracts of immature pistils of the plant *Cynara cardunculus* contained a protein of 64 kDa that reacted with both antibodies. Extracts of mature flowers contained a protein of 35 kDa, larger than that expected for the mature heavy chain of the two-chain native enzyme. This evidence, buttressed by sequencing of protein fragments, revealed a processing scheme where the 64 kDa procardosin A is cleaved first between Asn309 and Gly310, between the end of the 31 kDa heavy chain and the PSI. The next step is the removal of the PSI by a second cleavage between Ser 414 and Thr 415. In a third step, the prosegment is removed to yield the fully mature, two-chain enzyme. Asakura et al.[86] compared the activity of the aspartic peptidase from rice, oryzasin, in two forms: a GST-fusion protein with the full length promature enzyme, and a similar GST-fusion with the promature enzyme lacking the PSI. The catalytic activity in hemoglobin digestion was 1710 U mg^{-1} for the protein containing the PSI and 1130 U mg^{-1} for the version without the PSI. Both of these numbers were slightly lower than naturally occurring enzyme, 2936 U mg^{-1}, but seem to indicate that the PSI is not involved in the catalytic activity. Egas et al.[89] found that the PSI is able to interact with phospholipid vesicles and to induce leakage in a pH-dependent manner.

III. Structural Aspects of Aspartic Peptidases

The occurrence of aspartic peptidases in a variety of organisms is reflected in the number of structures that have been determined. Enzymes from retroviruses, fungi, mammals, protozoa, plants, and fish have been studied by crystallography and structures obtained. A total of 370 structures are available in the Protein Data Bank (Table 1). Of these, 65% are structures of the retroviral family of enzymes and nearly 52% are of HIV-1 peptidase. Approximately 19% are structures of enzymes from fungal sources, while just fewer than 12% are from mammalian sources. Protozoa have contributed three proteins (nine structures), plants have yielded two proteins (four structures), and one report of the structure of a protein from the Atlantic cod has appeared.

In the listings in Table 1, many are of inhibitor complexes of the various enzymes. In fact, comparisons of inhibitor binding to the aspartic peptidase family has been very important in developing the field of structure-based drug design, and this continues to be a fertile area. This will be discussed further in section VI.

The structure of an aspartic peptidase can be described in several ways. On one hand, these enzymes can be thought of as two-domain proteins where a gene duplication event has repeated the sequence. In addition, comparison of the structures of several different forms of one protein with, for example, different inhibitors bound, has revealed that the two halves of the enzyme can act as independent folding units that move relative to each other.[90]

The structure of each domain can be described as a series of secondary structure elements that are repeated within the domain to build up the structure. This is illustrated in Figure 1A,B and Table 2.

The crystal structure of porcine pepsinogen was reported by Sielecki et al.[91] and at higher resolution by Hartsuck et al.[92] The proenzyme is numbered 1p to 44p for the prosegment and 1 to 326 for the mature enzyme portion of the sequence (Figure 2A). A recent review by Kageyama[93] has discussed the sequential relationships between the zymogen form of aspartic peptidases as well as the activation mechanism.

The pepsin portion of pepsinogen strongly resembles the structure of porcine pepsin reported by Abad-Zapatero et al.,[94] Sielecki et al.,[95] and Cooper et al.[96] In particular, the catalytic apparatus appears intact, with Asp32 and Asp 215 positioned within 3 Å in the center of the active site cleft. Thus, the reason for the inactivity of the precursor form is not due to any problem with the catalytic site. Most importantly, the prosegment, which is comprised of one β-strand and three helices, interacts with the enzyme to result in complete blockage of access to the active site. The β-strand of the prosegment (residues 2p to 9p) occupies the position of residues

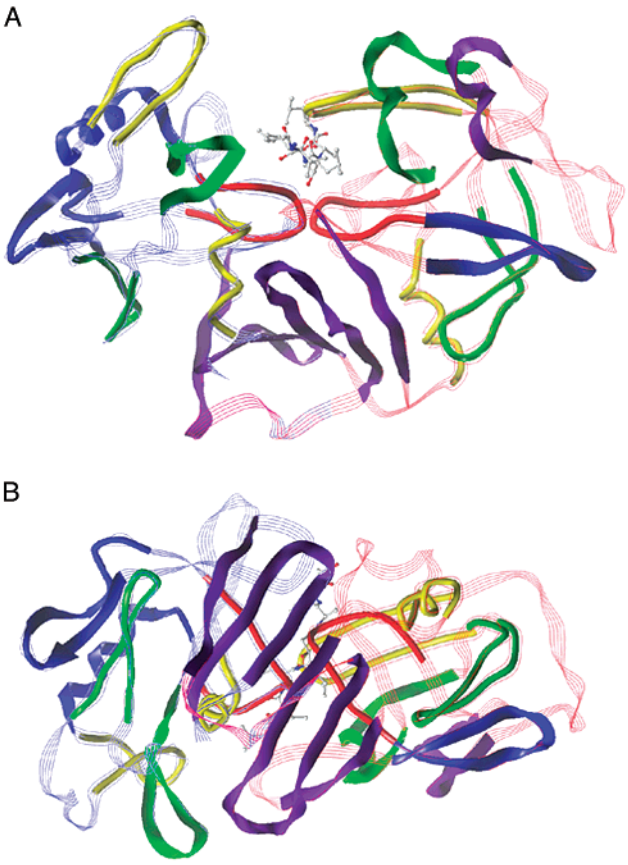

Figure 1. (A/B) Representation of the structural elements of all aspartic peptidases. File 1PSO[117] with human pepsin and pepstatin was used to construct the figure. A red ribbon (five lines) is drawn through the N-terminal domain and a blue ribbon is drawn through the C-terminal domain. Characteristic structure elements are noted by color and shape (specific identifications listed in Table 2). Each element in the N-terminal domain is mirrored by a similar element in the C-terminal domain. At the bottom of the molecule, a six-stranded β-sheet comprises a third domain in addition to the N- and C-terminal domains. Some prefer to divide the sheet into two separate three-stranded segments and include those with the N- and C-terminal domains. (A) View of the structure looking "end-on" to the active site with pepstatin bound. (B) View of the structure rotated 90° front-to-back to illustrate the six-stranded β-sheet. This is the view from the "bottom" of the molecule.

2−9 of the mature enzyme, becoming part of a six-stranded β-sheet at the "bottom" of the molecule. Residues 1−13 of pepsin are displaced by the binding of 2p to 9p and extend from the "top" of the enzyme to make a connection to residue 44p of the prosegment. Following exposure to low pH, cleavage events occur that lead to a conformational rearrangement of residues 1−11 of mature pepsin. This segment must undergo a major movement to end up with residues 2−9 occupying the β-strand position at the "bottom" of the molecule. This appears to involve a hinging motion around residues 13−14 of the mature enzyme. Between residues 9p and 44p the prosegment can be divided into three helical segments: 11p to 19p, 22p-29p, and 34p-37p. The latter of these provides a Lys36p−Tyr37p sequence that plays a major role in the interactions between the prosegment and the active site of the enzyme portion of the structure. The side chain of Lys36p points down into

Table 2. Amino Acids of Human Pepsin that Comprise the Structural Elements Presented in Figure 1A,B[a]

N-terminal domain			C-terminal domain	
sequence	segment	display	sequence	segment
2−9	β-strand	purple/flat	178−183	β-strand
17−28	A1	blue/flat	196−209	A1′
29−38	B1	red/tube	212−221	B1′
46−55	C1	blue/flat	223−235	C1′
72−81	D1	yellow/tube	237−247	D1′
85−100	A2	green/tube	257−269	A2′
106−118	B2	green/flat	275−286	B2′
135−144	C2	yellow/tube	301−309	C2′
157−168	D2	purple/flat	311−324	D2′

[a] The elements repeat within each domain and between domains, suggesting that these enzymes arose through two separate gene duplication events. The same coloring and display type is used for similar elements in each domain. The β-strands at the beginning of each domain complete the six-stranded β-sheet at the bottom of the enzyme, seen in Figure 1B.

the center of the catalytic site so that the ϵ-amino group hydrogen bonds to both catalytic aspartic acids. The hydroxyl group of Tyr37p hydrogen bonds to the outer oxygen of Asp 215, and the aromatic ring occupies subsite S1′. The hydroxyl group of Tyr 9 of the mature portion forms a similar hydrogen bond to the outer oxygen of Asp 32, and the aromatic ring occupies subsite S1. These two interactions follow the 2-fold symmetry seen in the mature enzyme. In both cases, the interactions with the binding pockets are similar to what is seen for small molecule inhibitor binding, but differs from substrate binding in that the side chains are coming from helical segments of the protein. The interactions seen in the pepsinogen structure provide a rationale for the dual function of pro-parts in assisting folding as well as blocking premature activity of the enzyme.[97]

Khazanovich-Bernstein et al.[98] have described the structure of truncated proplasmepsins 2, the precursor to plasmepsin 2. In this protein, the large natural prosegment of 124 amino acids was truncated to a size that was similar to that of the gastric proenzymes. This facilitated expression[15] so that sufficient amounts could be prepared for crystallization. The structure of this proenzyme was dramatically different from that of the gastric zymogens (Figure 3). The prosegment, rather than passing through the active site cleft as was seen with pepsinogen and progastricsin, winds instead around the C-terminal domain. The net effect of this unusual arrangement is that the two domains are pulled apart by about 4 Å at the catalytic site. This destroys the arrangement that leads to hydrogen bonding to a water molecule and prevents catalysis. Additionally, the flap is disordered due to the presence of the N-terminus of the mature enzyme in a different site than found in the mature enzyme. Therefore, even though the active site is open and could interact with substrates, the flap is unable to close down on a substrate to assist in achieving the correct geometry for the attack of a water molecule on the peptide bond to be cleaved.

It remains an open question as to the structure and role of the long amino terminal segment of proplasmepsins. A membrane spanning sequence is contained within the 124-residue segment, so it is

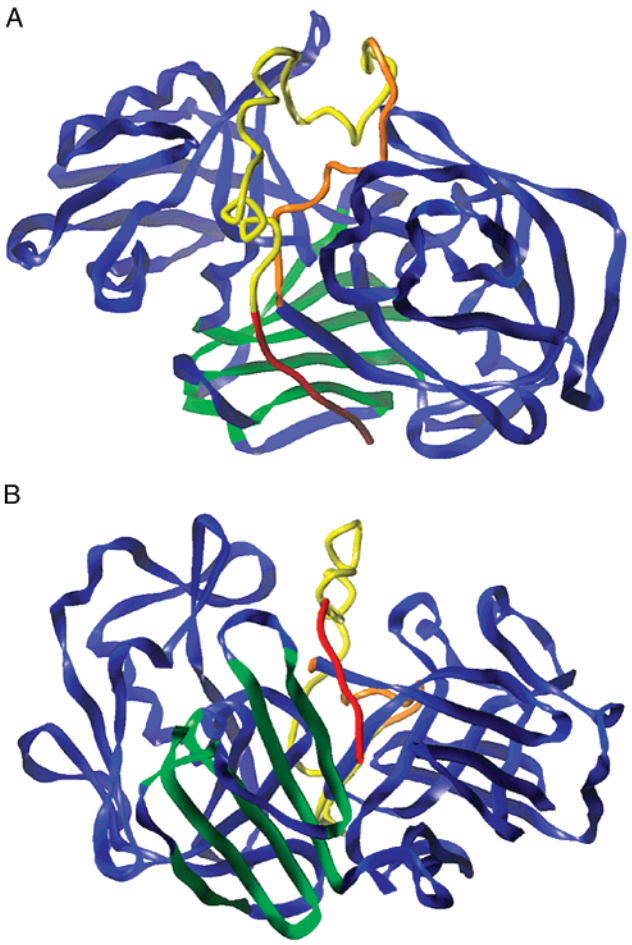

Figure 2. (A/B) Representation of porcine pepsinogen structure. PDB file 3PSG[92] was used to create this figure. The prosegment, residues 1p−44p, is represented as a tube. The first 11 residues of the prosegment are colored red. This segment provides the first strand of the six-stranded β-sheet at the bottom of the molecule. The other five strands are rendered as a green ribbon. The rest of the prosegment is represented as a yellow tube. The first 13 residues of the mature portion of pepsin are represented in orange. The connection between the end of the prosegment and the beginning of the mature enzyme is at the top in part A of this figure. Residue 1 of the mature enzyme must move more than 30 Å upon removal of the prosegment to take up its position at the beginning of the first strand of the mature enzyme (Figure 1). The main body of the mature enzyme is rendered as a blue ribbon. (B) View from the "bottom" of the molecule to emphasize the position of the first 11 residues of the prosegment as the first strand of the β-sheet.

reasonable to suggest[98] that this is a type 1 transmembrane protein. It is known that the proenzyme becomes membrane anchored and then transported to the food vacuole before activation. However, the relation between the membrane anchor and the rest of the proenzyme is not clear, as no structural work has been done on any extended proenzyme.

Another variation on the structure of the prosegments of aspartic peptidases may be provided by the prodomain of β-secretase. Shi et al.[99] have reported that the prosegment of this enzyme does not interfere with either catalytic activity or binding of an inhibitor, thus suggesting that the prosegment does not make a strong interaction with the mature enzyme

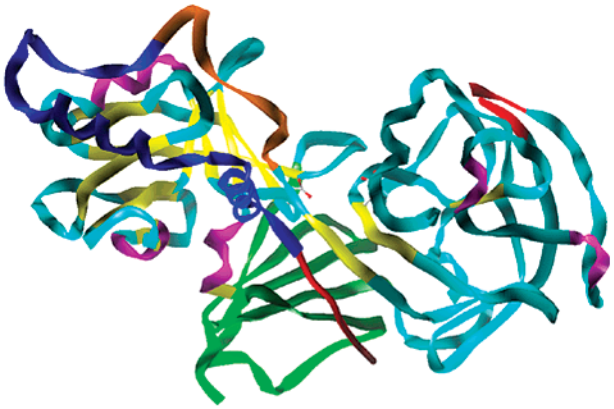

Figure 3. Representation of the structure of the truncated proenzyme of *P. falciparum* plasmepsin 2. PDB file 1PFZ[98] was used to create this figure. In this view, the first 11 residues of the prosegment are colored orange and occupy the position of the first strand of the β-sheet at the bottom of the molecule. The rest of the prosegment is rendered as a blue ribbon, while the first 13 residues of the mature portion of the enzyme are rendered as an orange ribbon. The rest of the enzyme is rendered in colors to represent secondary structure, with β-strands in yellow. The β-sheet at the bottom of the molecule is in green ribbon.

portion. These studies were done at pH 4.5, where one would expect strong electrostatic interactions between a prosegment and mature enzyme segment. Inspection of the sequence of the prosegment, however, reveals that promemapsin 2 has 10 fewer positive charges than pepsinogen in the proenzyme segment.

Moore et al.[100] have reported the structure of human progastricsin at 1.62 Å resolution. Gastricsin is a second gastric enzyme of the pepsin family, and is present at about one-fifth the level of pepsin in man. The structure of progastricsin is very similar to that of pepsinogen, consistent with their 45% sequence identity, with only changes in the "flap" region and several surface loops. One important difference is that the end of the prosegment of progastricsin and the beginning of the mature enzyme sequence are in different conformations than the corresponding region of porcine pepsinogen. This results in the flap adopting a different conformation, and, consequently, loop 127−134 must move.

Khan et al.[101] provided a stimulus to studies of proenzyme activation by solving the structure of an intermediate along the pathway of activation of human progastricsin. In their structure, obtained by switching a solution of progastricsin at the low pH (2.0) needed for activation to higher pH (above 5.0) at a time where the first few steps have occurred but before the final events happen, a peptide spanning residues 27p−43p has been removed, presumably by a combination of intramolecular and, possibly, intermolecular cleavage events (Figure 4). The remainder of the prosegment, residues Ala1p−Phe26p, remain bound to the mature enzyme and occupies the same binding points as in the progastricsin structure. Specifically, the first 10 residues remain as the first β-strand of the six-stranded sheet at the bottom of the protein. Apparently, the *middle* section of the prosegment is able to dissociate away from the position it occupies in the proenzyme structure so

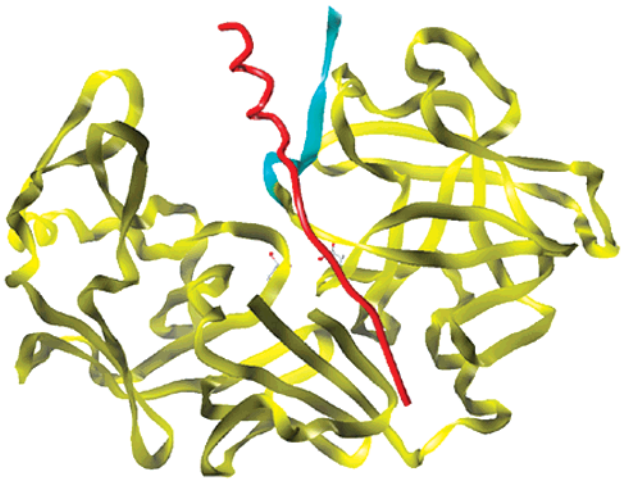

Figure 4. Representation of "intermediate 2" during the activation of progastricsin. PDB file 1AVF[101] was used to create this figure. The body of the mature enzyme is shown in yellow ribbon. The first 26 residues of the prosegment are shown as a red tube and the first 13 residues of the mature enzyme are shown as a blue ribbon. In this intermediate, the sequence between residue 26p and the first residue of the mature enzyme has been cleaved out. The first 11 residues of the 1p−26p segment remain anchored in the β-sheet as the first strand.

that it may be bound in a productive manner in the active site, resulting in cleavage of the Phe26p−Leu27p bond. Cleavage of the Leu43p−Ser1 bond between the prosegment and the mature enzyme releases the 27p−43p peptide that dissociates away from "intermediate 2" of the activation pathway.

As a consequence of the position of the prosegment 1p−26p, and especially the first 10 residues, the amino terminal end of the mature protein remains in a position different from its position in the final state. Movement of Ser1−Met10 to form the first strand of the β-sheet at the bottom of the enzyme requires the prosegment to dissociate. This can happen upon exposure to lower pH (below pH 5.0) as ionic interaction between Asp 11 of the mature portion and Arg14p of the prosegment becomes disrupted with protonation. A general discussion of the steps involved in proenzyme activation was presented.[102]

Relevant to this discussion of the structural position of the amino terminal end of the mature enzyme is the report by Lee et al.[103] Cathepsin D was shown to form a distinct conformation at pH 7.5, as compared to structures done at lower pH where the enzyme is fully active (Figure 5). The most significant difference is that the N-terminal segment of the mature enzyme, residues 3−7 in the structure by Lee et al., has moved from the normal position as the first strand of the six-stranded sheet at the bottom of the protein to a new position in the active site cleft. This is achieved by rotation around a hinge at residue 13, leading to a maximum displacement of 30 Å. The side chain of Lys 8 places the ϵ-amino group between the two catalytic Asp residues forming a salt bridge. The side chain of Tyr 10 is found in the S1 specificity pocket and Pro 4 and Glu 5 force the β-hairpin flap to move upward by up to 8 Å. As a consequence of these new positions, the active site is blocked for

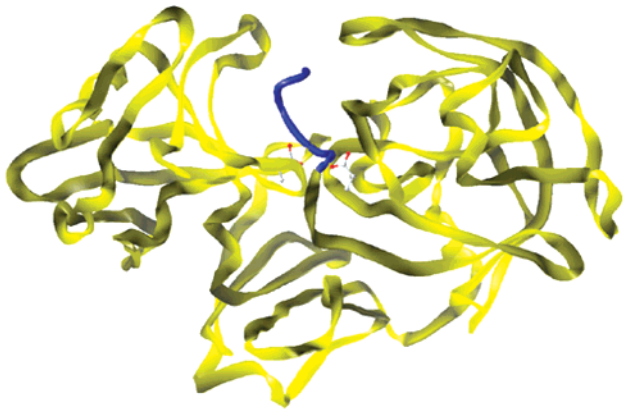

Figure 5. Representation of the "high pH" form of human cathepsin D. PDB file 1LYW[103] was used to create this figure. The body of the enzyme is shown as a yellow ribbon, while the first 11 residues are rendered as a blue tube. As the pH is raised from the pH optimum for catalysis, 3.5, to a pH of 7.5, the first 11 residues leave their normal position as the first strand of the β-sheet and move into the active site cleft, blocking entry of substrates and preventing catalysis.

entry by substrates, and this form of the enzyme is inactive. This process is reversible, as switching the enzyme back to pH 5.0 or below leads to catalytic activity. The high pH form of cathepsin D resembles the inhibited proenzyme structures of pepsinogen and progastricsin as, in all cases, a Lys residue forms ionic contacts with the active site aspartic acids and a Tyr residue is inserted into the S1 pocket.

The low pH "active" form of cathepsin D was described in two reports in 1993. Metcalf and Fusek[104] described the structure of both bovine liver and human spleen cathepsin D. Baldwin et al.[105] described the structure of human liver cathepsin D and its complex with pepstatin. Metcalf and Fusek described the differences between the bovine and human enzymes and other aspartic peptidases focusing on a series of eight surface loops, including the flap and the polyproline loop that interact with ligands in the active site. Both Metcalf and Fusek and Baldwin et al. described the positions of N-linked carbohydrate sites on the surface of the enzymes. The placement of two regions near one another provides an explanation for the recognition by a phosphotransferase that leads to the targeting to the mannose-6-phosphate receptor resulting in lysosomal localization.[106] Baldwin et al. compared the binding of pepstatin to cathepsin D in their structure to the binding seen in the pepstatin−rhizopuspepsin complex reported by Suguna et al.[107] Cathepsin D exhibited a lower K_i value for pepstatin than did rhizopuspepsin (4 versus 150 nM) and this can be accounted for by smaller residues in cathepsin D; for example, Ser 80 is in the equivalent position of Asp 77 in rhizopuspepsin. While the Asp residue provides rhizopuspepsin with the ability to cleave substrates containing Lys in P1 or P2, it prevents the flap from closing down tightly on the bound pepstatin.

Canduri et al.[108] have crystallized a urinary version of human pepsin, termed uropepsin, and have reported the structure at 2.45 Å resolution. Human pepsinogen is synthesized in the chief cells of the

gastric mucosa and is stored in zymogen granules prior to release into the gastric lumen. However, some of the protein is secreted into the systemic circulation, where renal clearance leads to high levels of pepsinogen A but not pepsinogen C in the urine. Uropepsin isolated by Canduri et al. is one of four isozymes of pepsin A, containing the natural polymorphism Leu291Val. In addition to describing the structure, these authors also studied the catalytic activity of uropepsin relative to pepsin 3A, the major isoform from the stomach and found no difference in the catalytic parameters using a fluorescent substrate, Abz−Lys−Pro−Ile−Glu−Phe*Phe−Arg−Leu.

Frazao et al.[109] have reported the structure of cardosin A, a plant aspartic peptidase from flowers of *Cynara cardunculus*. Flowers of this plant have been used for centuries to provide an enzyme that is used in cheese making in Portugal. As in other plant aspartic peptidases, this enzyme has a plant-specific insert (PSI) between residues 238 and 240. The uncertainty here is because the PSI was removed from cardosin A prior to structural determination by the natural processing events, leading to two chains, one of 15 kDa and one of 30 kDa. The site of the insert is in the C-terminal domain and is surface exposed. The structure of cardosin A also contains two oligosaccharide units, one attached to Asn 67 in the 15 kDa subunit and one attached to Asn 257 in the 30 kDa subunit. Asn 67 in human cathepsin D is also glycosylated, as is the same position in yeast peptidase A. The second glycosylation site in cardosin A is similar to that in yeast peptidase A, Asn 256, but differs from human cathepsin D where Asn 183 is the site of carbohydrate attachment. Both positions are remote from the active site and unlikely to affect binding of substrates.

Kervinen et al.[110] solved the structure of prophytepsin, the precursor to phytepsin, from barley (Figure 6). This proenzyme structure is notable both for its prosegment, which interacts with the mature portion of the enzyme in a novel fashion, and the plant-specific insert in the C-terminal domain. Only residues 6p to 26p of the prosegment could be traced in the electron density for prophytepsin; however, the interactions these residues make are similar to those seen in pepsinogen and progastricsin. However, the characteristic Lys36p−Tyr37p sequence of animal proenzymes is absent in prophytepsin, as well as other plant aspartic peptidases (except the pumpkin proenzyme). Lys 11 of the mature enzyme sequence in phytepsin replaces the ionic interaction seen for Lys36p of pepsinogen with the catalytic aspartic acids at the bottom of the cleft. The amino terminal segment of the mature enzyme occupies a different position in the structure of the prophytepsin, as compared to pepsinogen, and is closer into the catalytic site. In addition, the side chain of Tyr 13 of the mature enzyme occupies the S1 binding subsite of the enzyme and Ile 7 side chain blocks the S1′ subsite. In pepsinogen, the function of Tyr 13 is played by Tyr 37p of the prosegment.

Kervinen et al.[110] also compared the structure of the mature segment of prophytepsin with the structure of cardosin A[109] described earlier and concluded that only small conformational rearrangements would be required to achieve full function following removal of the prosegment from prophytepsin.

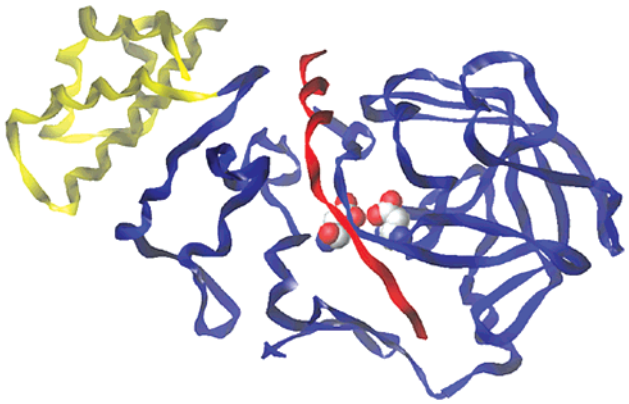

Figure 6. Representation of the structure of prophytepsin. PDB file 1QDM[110] was used to create this figure. The main body of the mature enzyme is rendered as a blue ribbon, while the "plant specific insert" is rendered as a yellow ribbon. Only 26 residues of the prosegment are seen in the crystal structure, but they occupy a normal position for a precursor, with the first 11 residues occupying the first position of the six-stranded β-sheet. The active site aspartic acids are rendered in space-filling mode to illustrate the close juxtaposition of these amino acids. This is the same position expected for the mature enzyme, thus indicating that the lack of activity of the proenzyme is not due to disorder of the catalytic machinery.

The structure of prophytepsin also gave the first glimpse of the plant specific insert, a 104-amino acid segment inserted into the C-terminal domain. This insert contains five helical segments, 4s−10s, 14s−22s, 27s−35s, 65s−85s, and 89s−103s and three interhelical disulfide bridges, Cys6s−Cys100s, Cys31s−Cys72s, and Cys37s−Cys69s, using the "s" to designate the residue number of the plant specific insert. Kervinen et al.[110] have proposed that the signal that leads to vacuolar targeting of prophytepsin involves components of the hydrophobic region of the plant specific insert. This includes residues Leu91s, Tyr95s, and Leu99s, surrounded by a ring of positively charged residues including Lys72, Lys87s, and Arg102s from the PSI and Lys 209 and Lys 242 of the mature enzyme.

Groves et al.[111] have described the structure of an inhibited complex of bovine chymosin. This has been an elusive target for many years, as initial attempts to crystallize a complex of chymosin with a bound inhibitor failed. The native structure of chymosin[112,113] has an unusual conformation for the flap hairpin[114] with the conserved Tyr 75 flipped by 180° leading to blocking of the S1 pocket. This interaction must be stable, so that inhibitor binding is blocked. Through a systematic search, Groves et al. found conditions where one, but only one, inhibitor, Pro-*p*-I-Phe-(S-Me)Cys-Cha-CHOH-COOiPr, where Cha is cyclohexylalanine, bound to chymosin in the crystalline state. This complex was compared to the native chymosin structure and to the structure of a renin−inhibitor complex. While the inhibitor-bound structure of chymosin was similar to other inhibitor complexes with aspartic peptidases, the differences between the native and inhibited forms are more

dramatic. Large adjustments along the active site cleft are necessary to accommodate the bound inhibitor; thus, this comparison is a good example of the malleability of enzyme residues within a binding cavity.

Silva et al.[115] described the three-dimensional structure of plasmepsin 2, an enzyme from the malarial parasite, *P. falciparum*. This enzyme was found to have a structure that was very similar to that of other aspartic peptidases, with two well-ordered domains, each contributing one Asp residue to the catalytic site. Two molecules were seen in the crystallographic asymmetric unit, and these had slightly different structures. One molecule exhibited a significant (5.2°) rotation of the C-terminal domain relative to the N-terminal domain. This type of movement has been observed in other enzymes of this family. These authors also reported on studies of several pepstatin-related compounds that were found to be low nanomolar inhibitors of plasmepsin 2. They utilized a similar cyclization strategy as Khan et al.[116] and connected the P1 and P1 positions, yielding a 0.2 nM inhibitor. Cyclization of the derivatives through the P2 and P3′ positions gave a compound that inhibited cathepsin D at a level of 1.4 nM, but exhibited a K_i value of 1500 nM for plasmepsin 2.

In an important paper, Fujinaga et al.[117] reported the structure of human pepsin alone and in complex with pepstatin. They also summarized the information available at the time concerning the subsite specificity of pepsin, gastricsin, renin, cathepsin D, and cathepsin E. For example, pepsin was found to prefer smaller residues in the P3 position of a substrate, while all other enzymes preferred a larger amino acid. Fuginaga et al.[117] also stated that the "...specificity in S3, for example, will depend on the residue at P1...." They also pointed out that subsites S2 and S1′ are contiguous so that occupancy in one will affect the preferred residue in the neighboring subsite. The human pepsin structure showed an essentially featureless binding cleft so that it was apparent that some side chains of a substrate or inhibitor could adopt different conformations. This has been observed in binding of inhibitors to endothiapepsin[118] and to rhizopuspepsin.[107] On the basis of the size, shape, and charges of the subsites along the cleft, Fujinaga et al.[117] rationalized the information available at that time on the preferences of the aspartic peptidases for substrate and inhibitor binding.

Hong et al.[45] have described the structure of memapsin 2 (or BACE-1, or Asp-2, depending on the nomenclature favored by different groups). This enzyme is found in a membrane-bound state and is responsible for cleavage at the β-site in the amyloid precursor protein. The peptidase domain was expressed without the membrane anchor and crystallized with an active site directed inhibitor. A total of four insertions into the sequence relative to human pepsin led to the structure of memapsin being larger in volume than pepsin. Hong et al. make the reasonable suggestion that these insertions may be involved in association with other proteins on the cell surface. Memapsin lacks six residues found between Pro 292 and Glu 297 of pepsin and, thus, the active site is more open for the former enzyme. This structure provided many details regarding binding of the inhibitor and permitted an analysis of why pepstatin does not provide a strong inhibition of the enzyme.

Candida albicans is a major source of fungal infections in humans. We are all exposed to this species and others on a daily basis, but normal immune systems are able to fight off infection. With any depression in immune function, such as those caused by chemotherapy for cancer, immune suppression to avoid transplant rejection, or infection by agents such as HIV, systemic opportunistic infection by *Candida* species becomes life-threatening. The multiplicity of peptidases present in various *Candida* species has been reviewed.[119,120] The secreted aspartic peptidases of *Candida* have been shown to play a role in oral[121] and vaginal[122] invasion. Hube et al.[123] showed by disruption of the genes of SAP 1, 2, or 3 that each was involved in virulence. In 1996, Abad-Zapatero and colleagues[124] reported the three-dimensional structure of the major secreted aspartic peptidase (or SAP) of *C. albicans* strain val-1. This enzyme is now known as SAP2 to differentiate it from other SAPs present in the organism. SAP2 is secreted in high amounts into culture fluid and so can be readily purified. Simultaneously, Cutfield et al.[125] published the structure of SAP2 from strain ATTCC 10261. While the overall features of the structure placed SAP2 firmly in the aspartic endopeptidase family, several insertions and deletions seem to have an influence on active site specificity. First, a long insertion, between Cys 47 and Cys 59, defines a broad flap that extends toward the active site cleft, marking a wide entrance to the binding site. This flap is rigidified by interactions of Gln 54 with the main-chain of residues 50 and 51. As Gln 54 is replaced in other SAPs, it is likely that this difference may play a role in defining substrate specificity. A seven-residue deletion alters the enzyme pocket that defines the S3 subsite, resulting in a very open cavity. In fact, this accounts for the observation that an inhibitor with both a benzyl group and a methylpiperazine group in the S3 pocket was accommodated. This structure provided insights that allowed both rationalization of the results of substrate specificity studies[126] and the design of new antifungal inhibitors.

Symersky et al.[127] studied the structure of the extracellular aspartic peptidase from *Candida tropicalis*, a yeast that infects patients with depressed immune systems. The structure resembles other aspartic peptidases in general terms, but was found to have large differences in its active site cleft. Several insertions and deletions in the sequence relative to porcine pepsin lead to a more open active site, which is in agreement with specificity studies done by Fusek et al.[126] Surprisingly, a tetrapeptide was found bound in the active site cleft of the enzyme. This appears to have arisen from an adventitious event and may represent a product complex. The tetrapeptide is bound in the S1′−S3′ region, suggesting that the release of the fragment to the right of the cleavage point was the final step in the mechanism (exclusive of protonations and uptake of water).

Koelsch et al.[128] studied the properties of four SAPs, SAP 1−3 and SAP 6, as these were known to be expressed in *C. albicans*. As SAPs 1, 3, and 6 were difficult to isolate directly from the fungus, these authors cloned the genes and expressed the proteins in *E. coli*. A novel method was used to examine substrate specificity. Mixtures of peptides were prepared where either P1 or P1′ contained 5−8 different amino acids. Following incubation with one of the SAPs, products were separated and identified by HPLC and mass spectroscopy. At the P1 position, SAP 1, 2, and 6 preferred Phe, with Leu, Tyr, and Ala following at lower levels. SAP3, in contrast, preferred Leu followed by Arg at P1. In the P1′ position, all the enzymes exhibited broad specificity, with most hydrophobic amino acids being acceptable. Curiously, Glu showed up in all cases as a marginally acceptable substitution in P1′. The positively charged amino acids, Pro, and Asn were not found at P1′ in the products of cleavage. This method provides a powerful way to compare specificity among related enzymes.

New members of the SAP family have been reported from *C. tropicalis*.[129] Only four definite genes were identified by cloning and termed SAPT1−4. Pichova et al.[130] have compared the properties of the secreted enzymes from *C. albicans*, *C. tropicalis*, *C. parapsilosis*, and *C. lusitaniae* by exploring a series of pepstatin derivatives with changes in the terminal groups and the side chain at P1. Some differences were noted among the enzymes tested, but the compound, Boc-Val−Val−PhenylStatine(3R4R)-Phenylstatine-Ome, inhibited all the enzymes at nanomolar or subnanomolar concentrations.

IV. The Amino Terminal Prosegment Region of Aspartic Peptidases

Khan et al.[131] compared the zymogen structures of aspartic peptidases with those of serine, cysteine, and metallopeptidases. Tanaka and Yada[132] have studied the inactivation of porcine pepsin and a series of mutants. As pepsin is unstable at alkaline pH while the precursor, pepsinogen is stable at pH values above 7.0, they reasoned that the difference might originate from the difference in surface charge or the position of the N-terminal segment. They also benefited from earlier work by Bohak et al.[133] that showed that chicken pepsin was more stable. Chicken pepsin was found to contain a better balance of positive and negative charges. By replacing several residues, in regions where negatively charged amino acids were concentrated, by neutral amino acids, Tanaka and Yada hoped to reduce electrostatic repulsion at pH 7.0 or above. By assaying the rate of denaturation at pH 7.0, a measurable stabilization effect was seen when several mutations were combined within both the N- and C-terminal domains. However, the effect was not large and the authors concluded that the charge distribution was not the major factor in the stability or instability of the mature enzyme.

As described earlier, the N-terminal segment of the mature enzyme (1−12) moves a considerable distance in the conversion from the precursor form to the mature enzyme. Tanaka and Yada[132] reasoned that the reverse of this movement might provide the initial step in destabilization of the N-terminal domain and the beginning of global denaturation. Lin and colleagues[42] had shown that the initial denaturation occurred in the N-terminal domain, consistent with the hypothesis of Tanaka and Yada. In addition, the structural work on cathepsin D demonstrating that the N-terminal segment moves in going from pH 5 to pH 7.5 supported this notion. Accordingly, Tanaka and Yada constructed mutants where several residues in the N-terminal fragment were mutated to amino acids that were found in chicken pepsin. The maximal effect was found when a total of five mutations were made, leading to a 5.8-fold slower inactivation compared to a protein of the wild-type sequence. Finally, they constructed a disulfide bond by mutating one residue in the N-terminal segment and one in the body of the enzyme. However, as disulfide bonds are less stable at low pH, and in the precursor form the two cysteines were not close together, the disulfide bond was not formed until the enzyme sample was brought to pH 7.0. At that point, denaturation was competing with disulfide bond formation, resulting in denaturation of most of the enzyme, but the formation of a small percentage of stable enzyme species.

An important study was provided by van den Hazel et al.[134] in which they found that the stability of the mature yeast peptidase A depended on the presence of the proenzyme segment. Expressing a construct with only the presequence fused to the mature enzyme resulted in degradation of the entire protein. However, some of the mature enzyme could be obtained in an active form by coexpressing a plasmid with the presequence followed by the prosequence. Increased amounts of expression of the prosegment construct lead to higher yields of active enzyme. This rescue in trans, where the presequence and prosegment were expressed separately from the mature enzyme sequence, was exploited by van den Hazel et al.[135] in a second report. They described the mutagenesis of the prosegment of yeast peptidase A. Random mutagenesis was used for the N-terminal and the C-terminal sections of the propeptide. Activity was examined by assaying the ability of these mutants to activate a second enzyme that could be analyzed with higher sensitivity. Two observations were made in this study: first, many different prosequences were acceptable, indicating that the sequence determinants were not stringent, and, second, in the mutants in the C-terminal half of the propeptide, Lys was observed in a high percentage of cases at position 53 of the sequence. It was concluded that this position was functionally equivalent to position 36p of pepsinogen, although there is no structural evidence to support this and the overall sequence identity between the prosegments of pepsinogen and yeast peptidase A is low (13.3%).

Richter et al.[136] have provided an excellent, thorough review of the data on the activation of the gastric aspartic peptidases (pepsinogen, progastricsin, and prochymosin). Key points include the presence of both direct and sequential pathways for

removal of the prosegment, the complex pH dependence of these processes, and the presence of both intramolecular and intermolecular steps in the maturation process.

Fukuda et al.[137] demonstrated that the prosequence of *Rhizopus niveus* aspartic peptidase is necessary for the secretion of active enzyme in a yeast expression system. They also found that the prosequence, expressed as a fusion with glutathione-*S*-transferase could support the secretion in trans and that a peptide of the prosegment sequence could assist the refolding of mature enzyme following denaturation in GuHCl.

Richter et al.[138] studied the effect of replacing Lys 36p with Arg, Met, or Glu. The Glu variant was unstable and could not be studied in detail. In the case of the Arg and Met substitutions, the rate of activation of the respective precursors was faster than that observed for the wild-type protein and the pH dependence of activation was broader. These experiments confirm the unique role of Lys 36p in the formation of the stable structure of the zymogen. In this work, the authors also reported that the activity of the resulting pepsins was decreased about half in the cases derived from the precursors with Arg or Met in position 36p. This observation is difficult to reconcile with the fact that the prosegment will be gone following activation. The authors suggest that the active site will be perturbed due to the folding with the mutant prosegments present. However, in these analyses, the binding of pepstatin was used to measure the amount of active enzyme present in the samples. The binding of pepstatin, or the substrates used in the assay procedures, will alter the conformation of the active site residues such that any effect of the mutant prosegment should be obliterated.

In striking contrast, Fortenberry and Chirgwin[139] have found that human cathepsin D can be expressed and secreted from mammalian cells without the prosegment present; however, lysosomal targeting was lost with deletion of the prosegment. The level of expression was reduced about two-thirds when compared to expression of the intact proenzyme form, a result that agreed with earlier studies on preprorenin and its deletion variants.[140−143] Fortenberry and Chirgwin[139] suggested that the results of van den Hazel et al.[134] indicating that the prosegment of yeast peptidase A was required might be due to the requirement for vacuolar targeting rather than for protein folding.

Human cathepsin D goes through a complex biosynthetic pathway. While procathepsin D can undergo a self-catalyzed cleavage to remove the first 26 residues of the prosegment, the final steps of maturation are catalyzed by other peptidases in the lysosome. Beyer and Dunn,[144] in an effort to create an expression system for human cathepsin D that would produce enzyme with properties identical to those of naturally occurring enzyme, engineered a cleavage site six residues upstream from the N-terminus of the mature enzyme. This construct self-activated to a species that was the predicted size and had properties that were kinetically identical to those of enzyme isolated from human liver. This engineered protein had six extra amino acids at the amino terminus and differed from the naturally occurring protein in two other ways. It lacked glycosylation as it was expressed in bacterial cells and it was a single chain enzyme as opposed to the two-chain enzyme found naturally. These differences are without effect upon the catalytic efficiency of the enzyme.

One of the least well-understood human enzyme is cathepsin E. It is nonlysosomal, believed to reside in the endosomal vesicles within the cytoplasm. The structure of this enzyme has not been solved yet. Azim and Zaidi[145] have described a model of the structure of the proenzyme based on the structure of porcine pepsinogen. One significant difference between cathepsin E and other aspartic peptidases is that the former can form a disulfide-linked dimer. In the model prepared by Azim and Zaidi, residue Cys 40p is solvent exposed and could readily achieve disulfide bond formation. The lack of Lys 36p and Tyr 37p in the procathepsin E sequence forced these authors to suggest a slightly different mode of interaction between the prosegment and the active site. Lys 36p in pepsinogen is replaced by Phe 36p in procathepsin E. In the model, this side chain is pointing between the two catalytic Asp residues. While it is possible to accommodate this side chain in steric terms, it is expected that the energy of the interaction will be significantly less favorable than the interaction of the Lys 36p in pepsinogen and progastricsin. Tyr 37p in pepsinogen is replaced by Thr 37p in procathepsin E, and the model suggests that this residue makes no contact with the active site as it is pointing in a different direction. Salt bridges have been considered to be a major part of the interaction between the prosegment of pepsinogen and the body of the mature enzyme. In procathepsin E, new ionic linkages replace many salt bridges that are present in pepsinogen due to sequence differences between the two enzymes. While the work of van den Hazel et al.[135] suggests that many different sequences are acceptable for the prosegment, some studies have indicated that there is specificity in the choice of sequence. For example, Fortenberry and Chirgwin[139] were unable to use the prosegments of pepsinogen or human prorenin to direct cathepsin D to the lysosomal compartment. An alternate explanation for the difference in sequence between pepsinogen and cathepsin E could be that the prosegment of procathepsin E wraps around the C-terminal domain of the enzyme as was seen for the proenzyme forms of the plasmepsins from *P. vivax* and *P. falciparum*.[98]

In contrast to the results of Richter et al.,[138] Francky et al.[146] showed that lamb prochymosin is capable of folding and activating with a Glu in position 36p. This amino acid is, in fact, the natural substitution. Calf prochymosin has Lys at position 36p as does many other gastric proenzymes; lamb prochymosin and similar enzymes from sheep, goat, and mouflon (subfamily *Caprinae*) have Glu at the same position. Francky et al. prepared the calf prochymosin Lys36pGlu and the lamb prochymosin Glu36pLys variants and compared the rates of acti-

vation of the native and mutant forms. All four proteins were activated at both pH 2.0 and 4.7, with both species with Glu in position 36p activating at higher rates than the two proteins with Lys at 36p. This agrees with the notion that with Glu at 36p the interaction with the catalytic center will be less strong and the prosegment will be more readily removed from its position, allowing easier access to the transition state for the initial cleavage.

V. Active Site Engineering

The availability of clones of a number of aspartic endopeptidases has made it possible to construct site-directed mutants for the testing of specific hypotheses about active site interactions. Probing binding and catalytic events within the active site can be viewed from two perspectives, that of the enzyme and that of the substrate.[147]

A. P1 Specificity

Fungal peptidases have been shown to have the ability to activate trypsinogen by cleaving a bond with Lys in the P1 position.[148] As most aspartic endopeptidases prefer to cleave between hydrophobic amino acids, this marks the fungal members of the family as unique. James et al.[149] rationalized this by demonstrating that, in penicillopepsin, Asp 77 on the flap was able to interact with a Lys side chain in the P1 position. Subsequently, Lowther et al.[150] used site-directed mutagenesis to establish that rhizopuspepsin follows other fungal peptidases in the ability to process peptides following Lys. They mutated Asp 77 to Ser and found a loss in the ability to cleave a specific peptide. Subsequently, Shintani et al.[151] used a panel of mutants to show that Asp 77 in aspergillopepsin from *Aspergillus saitoi* was essential for cleavage of Lys containing substrates. Furthermore, Shintani et al.[152] created mutants of porcine pepsin where Thr 77 was replaced by Asp, and Ser was inserted between Gly 78 and Ser 79. These mutant forms of pepsin were able to cleave peptide substrates containing Lys in the P1 position. Thus, a new function was inserted in to pepsin by a minor change in the amino acid sequence.

B. The Flap

Yada and colleagues studied the flap of the aspartic peptidases by making mutations in positions 77[153] and position 76[154] in porcine pepsin. Thr 77 was replaced by Ser, Val, and Gly. The Thr77Ser mutant was least affected by the substitution in assays of activation rates and catalytic efficiency. This led the authors to conclude that either Thr or Ser could occupy position 77 with approximately equal results. Substitution by Val or Gly, however, caused greater reductions in the rates of catalysis. The lack of a side chain hydroxyl group in the Val and Gly mutants eliminates several hydrogen bonds that are possible between amino acid 77 and substrate backbone atoms. These hydrogen bonds are believed to be responsible for anchoring the substrate in the active site in the proper geometry for attack by enzyme bound water. Hence, deletion of Thr 77 and replacement by either Val or Gly will disrupt one important feature of catalysis and the resulting loss in efficiency was not surprising. The same authors followed up the study of position 77 by making mutations in position 76. Replacing Gly 76 by Ala had the least effect on the activation rate and catalytic parameters, whereas substitution by either Val or Ser caused nearly the same decrease in cleavage rates. The authors argue that the major effect will be on the flexibility of the loop or flap. Andreeva and colleagues[155] had described the flap of aspartic peptidases as moving in a segmental fashion with the greatest motion at positions 76−78. Gly 76 and 78 are proposed to serve as hinges for the motion of the tip. Mutation of Gly 76 leads to reductions in the flexibility of the flap and both the loss of the ability of both Gly 76 and Thr 77 to form hydrogen bonds.

Tanaka et al.,[156] following the work of Beppu's group,[157−159] mutated the Tyr 75 residue of pepsin to Phe and Asn. The resulting derivatives of pepsin were active in four different assays: milk clotting, general proteolysis of hemoglobin, and the cleavage of two synthetic substrates. In all these assays, the resulting activities were, surprisingly, relatively unchanged from that of the native enzyme. In particular, the K_m values for the synthetic substrates were within experimental error of wild-type enzyme. The k_{cat} values were reduced in the order Tyr 75 (190 min^{-1}) > Phe 75 (70 min^{-1}) > Asn 75 (25 min^{-1}), and these results led Tanaka et al. to postulate that Tyr 75 is not involved in the binding of substrate, but might be involved in the correct positioning for catalysis. This result disagrees with the hypothesis put forth by Tang and Koelsch[160] that Tyr 75 plays a role in the "capture" of substrates by the enzyme.

In 1992, Andreeva and colleagues[114] published an analysis of the structure of chymosin determined in Gilliland's laboratory[161] focusing on the position of the Tyr on the flap (residue 77 in chymosin is identical to residue 75 in pepsin). His at position 76 on the flap β-hairpin is a unique residue in the sequence of chymosin. This His side chain forms hydrogen bonds with the carbonyl oxygen of Gly 80 and the side chain Oγ of Ser81. This shifts the conformation of the hairpin so that Tyr 77 cannot adopt a position similar to that seen for the flap Tyr residue in other aspartic peptidases. As a result, Tyr 77 in chymosin blocks access to the S1 and S3 binding subsites. Andreeva has described chymosin as a "self-inhibited" enzyme, and hypothesizes that this may account for the stricter substrate specificity exhibited by this enzyme. This analysis was a harbinger of the results described later for human cathepsin D at pH 7.5.

C. Active Site Mutations to Address Specificity Questions in Other Positions

Scarborough and Dunn[162] created mutants at position 287 of human cathepsin D, using clues from a molecular model published a year earlier.[163] Their conclusion that Met 287 was likely to be a point of interaction with the P2 residue of a substrate was

confirmed by the crystallographic analysis of Metcalf and Fusek[104] and Baldwin et al.[105] Replacement of Met 287 with Glu altered the substrate preferences slightly, with more hydrophilic P2 substitutions being acceptable for the mutant enzyme.

Beyer and Dunn[164] studied the specificity of the P2′ position of human cathepsin D and found that this site is very versatile, accepting both positively charged amino acids (Lys and Arg) and hydrophobic amino acids (Leu and Ile), but cleaving with lower efficiency peptides with Glu or Asp in P2′. By mutating a single residue, Ile 128 of the enzyme sequence, to Arg, the selectivity was altered dramatically. For substrates with positively charged amino acids in P2′ the efficiency of cleavage dropped, while for substrates with Leu or Ile the efficiency decreased slightly. The cleavage of a peptide with Asp in P2′ was improved so that the k_{cat}/K_m value was nearly as large as that for the Lys containing substrate.

Westling et al.[16] mutated three amino acids within the active site cleft of plasmepsin 2 from the malarial parasite *P. falciparum*. Changing Met 13, which is believed to be in the S3 subsite of the enzyme, to a Glu led to a decrease in cleavage of peptides containing hydrophobic residues in the P3 position. A substrate with Lys in P3 was cleaved 5.5-fold more efficiently by the Met13Glu mutant, with most of the effect coming from a significant decrease in the apparent K_m value (250 to 35 μM). Surprisingly, the mutant form also cleaved a substrate with Asp in P3 more efficiently. This may have been due to the formation of a Glu−Asp hydrogen bond at the pH of the assays (pH 4.4). Hydrophobic substitution in P3 resulted in less efficient cleavage by the Met13Glu mutant due to loss of hydrophobic contacts upon substitution. Position 287 in plasmepsin 2 was also probed by mutation to Glu. Alterations in the catalytic efficiency of the Ile287Glu mutant similar to those seen with mutation of Met 13 to Glu were observed, except that the effects were seen with substrates containing Lys or Asp residues in the P2 or P4 positions. Finally, the Ser at position 77 on the flap was replaced by Asp, resulting in a significant increase in the cleavage of a substrate with a Lys in the P1 position, similar to that observed with porcine pepsin by Shintani et al.[152]

Westling et al.[165] also reported on the comparison of three plasmepsins, one each from *P. vivax* and *P. malariae*, plus plasmepsin 2 from *P. falciparum*. While all three enzymes displayed the ability to cleave oligopeptide substrates in the same fashion as other aspartic peptidases, some distinctions were noted in the preferences for residues at several positions in the peptide chain. Most notably, plasmepsin 2 was less efficient a catalyst against peptides with basic residues in the P2 or P3 positions. Plasmepsin 2 is, however, able to cleave hemoglobin at a site that has an Arg residue two positions upstream of the cleavage point, i.e., at P3. This apparent discordance can be explained by the presence of other residues in the P2 and P1 positions. The sequences are compared below from P3−P1′ with positively charged residues in bold.

Synthetic peptides

Ile-**Lys**-Phe*Nph

Lys-Glu-Phe*Nph

Hemoglobin cleavage sites

Arg-Met-Phe*Leu

Leu-Val-Thr*Leu

Thr-Val-Leu*Thr

Arg-Leu-Leu*Val

It can be seen that the hemoglobin sites prefer smaller residues in P1*P1′ and none have a Glu or Lys in the P2 position. It is the combination of amino acids in the synthetic peptides used by Westling et al. that led to the observation of reduced efficiency of cleavage.

In addition, Westling et al.[165] compared the binding of a P2-to-P3′ side chain cyclized inhibitor to the binding of two noncyclized homologues with the three enzymes. While the plasmepsins from *P. vivax* and *P. malariae* bound well to all three inhibitors, the plasmepsin from *P. falciparum* interacted with the side chain cyclized inhibitor about 1000-fold less tightly. This effect was traced to the substitution of a Val at position 78 in the flap of the *P. falciparum* enzyme (originally noted by Silva et al.[115] in the description of the structure of plasmepsin 2), where all other plasmepsins have a Gly (Westling et al., unpublished).

Cao et al.[166] tackled an important question in the catalytic mechanism. Earlier work had shown that extending a peptide substrate so that an amino acid was present in the P3 position led to a large increase in k_{cat}/K_m [up to 50-fold for penicillopepsin,[167] 100 fold with endothiapepsin, 300-fold with rhizopuspepsin, and 1380-fold with pepsin[168]]. Examination of crystal structures of enzyme−inhibitor complexes revealed that a hydrogen bond was made in every case between the amino acid four residues downstream from the C-terminal catalytic Asp. At the time of the Cao et al.[166] publication, 111 sequences of aspartic peptidases were in the Merops database and only one of those, cockroach allergen, had a residue other than Thr, Ser, or Asn at the position in question, while today there are 288 sequences of aspartic peptidases and 20 have residues other than the three polar amino acids. Of the 20, only one, rat chymosin, is from an animal, while three sequences are from *C. elegans*, 13 are from *Arabidopsis thaliana* and one from *Nicotiana tabacum*. One additional cockroach sequence has been found. Most of the plant sequences in the Merops database have not been demonstrated yet to code for active enzymes. On the whole then, it is still clear that, predominantly, the residue four amino acids downstream of the C-terminal catalytic Asp is Thr, Ser, or Asn.

Cao et al.[166] tested the role of this residue directly by performing site directed mutagenesis on penicillopepsin-JT2, a second variant of penicillopepsin

cloned from the organism *Penicillium janthinellum*, placing Ser, Val, Gly, and Ala in position 219. All derivatives were catalytically active, with the Ser derivative exhibiting similar properties to that of the wild type enzyme. Specifically, the rate of cleavage of synthetic peptide substrates was significantly increased (500−1000-fold) by extension on either side of the cleavage site so that the S3 and/or S2′ subsites were filled by at least an Ala side chain. The other derivatives, lacking a hydroxyl group at position 219, did not show the same increase in activity with extension into P3. These data support the conclusion that hydrogen bonding of the side chain in position 219 with the P3 backbone NH is critical to the correct positioning of the substrate to permit maximal rates of cleavage. It remains to be seen whether Glu or Gln could support the same rate increase and whether rat chymosin, with Ala in position 219, exhibits the same rate enhancement in comparing substrates with and without amino acids in position P3. If rat chymosin lacks the rate enhancement, could placing a Thr residue there cause the rate increase to occur?

An alternative explanation for the large increase in catalysis upon occupying the S3 subsite was provided by Sali et al.[90] in their analysis of domain flexibility in aspartic peptidases. It had been known for some time that the C-terminal domain of these enzymes, specifically residues 190−302 in endothiapepsin, moved in relation to the rest of the protein molecule when inhibitors were bound into the active site. Furthermore, comparison of different aspartic peptidases reveals that the 190−302 segment is the most highly mobile. Calculation of the rms deviation between two aspartic peptidases is improved when considering the N- and C-terminal halves separately. Sali et al. showed that the C-terminal domain movement could be described by a screw axis rotation of between 2.9 and 17.6° for various pairs, along with a translation of up to 0.58 Å. By studying a series of inhibitors bound to endothiapepsin, Sali et al. concluded that occupancy of the S2 and S3 subsites were critical to the movement, and that this movement could adjust the binding of the substrate to increase turnover. Abad-Zapatero et al.[94] also described the rigid body movement in porcine pepsin. For this protein, the sequences 192−212 and 233−299 were identified as the parts that undergo movement. Sielecki et al.[91] described a similar movement when comparing the structure of porcine pepsinogen to the mature enzyme.

Williams et al.[169] constructed two mutants of chymosin B, Ala115Thr and Gly243Asp, to investigate the role of these positions in the unique reactions catalyzed by this bovine enzyme. Position 115 is in a helical region on the periphery of the active site and is believed to have a possible influence on binding in S1 or S3. Position 243 is on the other side of the active site cleft and has been postulated to be involved in the difference between chymosin A, with Asp at this position, and chymosin B, with Gly. In addition, these authors crystallized the derivatives and determined their three-dimensional structures with resolution of 2.3 Å for the Gly243Asp mutant and 2.8 Å for the Ala115Thr mutant. Kinetic analyses were done using two synthetic substrates, both based on the site within κ-casein that is cleaved specifically by chymosin during the process of milk clotting in the production of cheese. The sequences were

```
Leu-Ser-Nph*Nle-Ala-Leu-OMe

His-Pro-His-Pro-His-Leu-Ser-Phe*Nph-Ala-Ile-Pro-Pro-Lys-Lys
```

In these sequences, the asterisk (*) indicates the point of cleavage, Nph indicates *p*-NO_2Phe, and Nle indicates norLeucine. The peptides differ in the bond being cleaved, with Nph on the left side of the scissile bond in the first substrate and on the right side in the second substrate. The first substrate has been used previously by many authors to study the kinetics of chymosin, while the second substrate was designed to probe for effects occurring at positions remote from the cleavage site.

For the first substrate, the Ala115Thr mutation reduced the k_{cat}/K_m by a factor of 3.6 without changing the k_{cat} significantly. Inspection of the structure of the mutant enzyme reveals subtle, but important, changes in the positioning of Phe 112 that would have effects upon the interaction with substrate on the N-terminal side of the cleft. On the other hand, the Gly243Asp mutation reduced k_{cat} by 2.8-fold, while the k_{cat}/K_m was reduced by 3.4-fold. For the second substrate, a 2-fold higher K_m value was obtained and an increased k_{cat} value was observed with the Gly243Asp mutant compared to recombinant chymosin B (94 to 209 s^{-1}). However, the k_{cat}/K_m value was unchanged. The structural studies show convincingly that an earlier hypothesis[112] for interaction between His at P4 in the casein substrate and Asp at position 243 would be impossible; however, His at P8 would be within favorable distance to the loop containing residue 243. In addition, other negatively charged residues on the surface loops (Glu244−Asp250 and Asp278−Phe281) of chymosin could be involved in electrostatic interactions with an extended peptide or the casein substrate.

Bhatt and Dunn,[170] who created chimeric enzymes from two aspartic peptidases by connecting the N-terminal domain from one enzyme to the C-terminal enzyme from another, provided a novel approach to active site engineering. This was attempted first with porcine pepsin and rhizopuspepsin. The chimera with the N-terminal domain of pepsin and the C-terminal domain of rhizopuspepsin folded well and produced activity within a factor of 3−5 of that shown by the parent enzymes for the best substrates. The specificity for substrate sequences was observed to be consistent with the substitution at positions of the substrate that would point toward corresponding positions of the chimera. For example, substitution at P3 or P2′ would point toward the N-terminal domain and the rates reflected the selectivity of the subsites on that side of the cleft, while substitutions in P2 gave rates that reflected the selectivity of the C-terminal domain. Chimeras between plasmepsin 2 and pepsin were also constructed (D. Stewart, S. H. Hung, and B. M. Dunn, unpublished) and gave strongly active enzymes with the expected specificity.

Cotrell et al.[171] mutated the single Lys residue of porcine pepsin to either Met or Glu and found that the properties of the enzyme were altered. Although position 319 is outside of the active site cleft, a loss of catalytic activity and a decrease in stability of the mutant enzymes was observed. These authors rationalize these results on the basis of greater flexibility of the enzyme structure upon the substitution. This provides an example of an enzyme as a dynamic molecule, and sounds a warning for anyone attempting to alter the sequence of an enzyme at a "remote" position in order to create a point for tethering to a solid surface, for example.

VI. Substrate Design and Its Impact on Inhibitor Design

Gulnik et al.[172] improved the assay for cathepsin D by creating fluorescent derivatives of a good synthetic substrate using DABCYL (4-[4-dimethyl)-amino]naphthalene) and EDANS (5-[(2-aminoethyl)-amino]nappthalene-1-sulfonic acid) as the quencher and fluorophore, respectively. Cleavage of a Phe−Phe peptide bond between the two chromophores leads to an increase in fluorescence that can be accurately measured. This assay allows the use of less enzyme, which is a major advantage when studying the binding of inhibitors with K_i values below the nanomolar range. By substitution in the P2 position of the substrate sequence, Gulnik and colleagues were able to alter the specificity constant (k_{cat}/K_m) so that some peptides could be chosen to provide highly sensitive assays or for assays in which a long linear phase is needed, such as in multicell instruments with long cycle times.

Pimenta et al.[173] explored the specificity of cathepsin D using a large series of fluorescently labeled peptides with parent sequence based on the reactive site loop of kallistatin [Ala−Ile−Lys−Phe*Phe−Ser−Arg−Gln]. They found Tyr−Tyr to be the best substitutions at the scissile peptide bond (P1*Pl′), with k_{cat}/K_m of 29 μM^{-1} s^{-1}. With the ease of following kinetics of cleavage provided by the fluorescence method, Pimenta et al. were able to analyze an extensive series of peptides. They chose to use Phe−Phe as the peptide bond cleaved in this series, most likely due to synthetic considerations. In the series based on Abz−Ala−Ile−Lys−Phe*Phe−Ser−Arg−Gln(EDDnp), where Abz is *o*-aminobenzoic acid and EDDnp is *N*-[2,4-dinitrophenyl]ethylenediamine, 14 different amino acids were substituted into the P2 position, with the preferred residues in the order Ala > Glu > Leu > Gln,Ser. Lys in the P2 position of the parent peptide gave a k_{cat} of 4.9 s^{-1} and k_{cat}/K_m of 4.1 μM^{-1} s^{-1}, while the best substrate in the P2 substitution series (P2 Ala) gave a k_{cat} of 16 s^{-1} and k_{cat}/K_m of 60 μM^{-1} s^{-1}. The fact that a substrate with Lys in P2 was cleaved at all disagreed with the results of Scarborough et al.[162] who found that peptides with Lys were not cleaved by cathepsin D. However, the difference in sequence of the series of peptides considered by the two groups can account for the different observations. In fact, the substrate containing Lys in P2 in the Pimenta et al.[173] series was the slowest out of 14 derivatives considered. Pimenta et al. continued their analysis with a series of 12 substitutions in the P3 position with Leu > Val > Glu as the top substrates. In this series, replacement of the Ile in the parent sequence with the three mentioned above or with Ala, Thr, Asn, Gln, Glu, Lys, or Pro resulted in substrates with k_{cat}/K_m values greater than that of the parent peptide. In particular, their peptide SSSVIII with Lys in both P3 and P2 is surprising, as the S3 pocket in human cathepsin D is very hydrophobic. It is likely that in both S3 and S2 the long Lys side chain can place its methylene groups in the pocket and allow the terminal ϵ-amino group to extend into the aqueous exterior. Pimenta et al. also analyzed the P2′ (Val > Leu > Gln > Ala > Asn) and P3′ (Leu, Ile, Val > Ala) positions with series of peptides containing 13 different amino acids in each case. Considering all the results, human cathepsin D demonstrated a preference for hydrophobic amino acids in every position from P4 to P3′.

Gulnik et al.[174] extended their approach to the study of plasmepsin 2, an enzyme of the malarial parasite believed to be involved in degradation of hemoglobin in the food vacuole. Sensitive fluorogenic substrates were constructed that have improved the k_{cat}/K_m to a value of 36.5 μM^{-1} s^{-1}. In their series of peptide substrates, they found that Arg, when substituted in the P3 position, resulted in good rates of cleavage, which agrees with the observation that plasmepsin 2 cleaves hemoglobin at a site that has an Arg in the P3 position. This result is in contrast to the report of Westling et al.[16] who found that Arg or Asp in P3 gave substrates that were cleaved very slowly, if at all. The sequences of the two series of peptides that were compared by Gulnik et al. were, however, very different. The peptides compared had the following sequences from P4−P3′:

P4	P3	P2	P1	P1'	P2'	P3'	
-Pro-	Arg-	Val-	Leu-	Leu-	Arg-	Leu-	Gulnik et al.[158]
-Pro-	Arg-	Glu-	Phe-	Nph-	Arg-	Leu-	Westling et al.[16]

While P4−P3 and P2′−P3′ are identical in these two peptides, the central P2−P1−P1′ sequences are very different. As pointed out by Gulnik et al.,[174] the context of the peptide sequence is very important; thus, the presence of the larger residues in P2 (Glu versus Val) and P1 (Phe versus Leu) probably accounts for this difference in preference in the P3 position.

Turner et al.[175] studied the subsite preferences for both substrates and inhibitors of β-secretase [memapsin in the authors terminology; BACE,[39] Asp-2,[38,40] and β-site APP-cleaving enzyme,[41] for others]. Eight positions of a substrate were probed by adding mixtures of 19 amino acids (all common ones except Cys). The syntheses were done in sets where six or seven amino acids were combined in equimolar amounts for the synthesis. Thus, three syntheses were done for each of eight positions; this simplified identification of the products by mass spectroscopy following incubation with enzyme. The product analyses show that the following amino acids were pre-

ferred: P4 clearly prefers Glu, Gln, and Asp; P3 prefers Ile, Val, and Leu; P2 prefers Asp or Asn; P1 prefers Leu, Phe, and Met. On the other side of the cleavage point, the preferences were less strict, with Met, Glu, Gln, Ala, and Asp in P1′; Val, Ile, Ala, Leu, Glu, and Phe in P2′, and a wide range acceptable in P3′ and P4′. A synthetic substrate was designed based on the results of the analysis, Glu−Ile−Asp−Leu−Met−Val−Leu−Asp−Trp−His−Asp−Arg, where the underlined section is based on the product analysis. This peptide gave a k_{cat}/K_m value that was 14 times better than that measured with a peptide of the most efficiently cleaved natural substrate.

The results of the specificity screening generally agree with the structure obtained by Hong et al.[45] An inhibitor, Glu−Val−Asn−LeuΨ[CHOH−CH_2]-Ala−Ala−Glu−Phe was bound in the active site and makes contacts with several residues along the active site cleft. Subsites P4 and P2 appear to be open to the solvent and are able to accommodate hydrophilic residues. Subsites S3 and S1 are very hydrophobic, as is seen with most aspartic peptidases. Subsite S1 appears to be large, as the small Ala of the inhibitor does not make contact with any residue of the enzyme. Clearly, this position could accommodate larger side chains, and the observation that a variety of amino acids are acceptable in P1′ is in agreement with this. This analysis is in conflict with the report of Lin et al.[42] who suggested that a comparison of cleavage sites in peptides indicated that the specificity of memapsin 2 at the P1′ position was for small side chains such as Ala, Ser, and Asp. Subsite S2′ favors hydrophobic amino acids, but with acceptance of Glu and Thr as well. This is most likely due to residues such as Arg 126 and Tyr 198 in the vicinity. Residues P3′ and P4′ of the inhibitor make no contact with the enzyme, as the inhibitor makes an abrupt turn at P2′. The binding of the inhibitor, with mostly small side chains has defined the general features of the cleft, but clearly shows that there is sufficient room to design new compounds that will interact with more tighter binding and greater selectivity.

Taking advantage of the structural model, Ghosh et al.[176] modified the inhibitor by deleting the P3′ and P4′ residues as well as the P4 residue. Altering the residue in the P2 position resulted in an inhibitor (Boc-Val−Met−LeuΨ[CHOH−CH_2]Met−Val-benzamide with a 2.5 nM K_i.

Following up on their substrate specificity studies, Turner et al.[175] prepared a library of inhibitors, using Leuψ[CHOH−CH_2]Ala as the central isostere. By using a split synthesis solid-phase method, they were able to construct a library of ca. 130 000 compounds in which each bead contained a single sequence. After incubating memapsin with a sample of the beads large enough to represent all possible peptides, they used an immunological method to identify the beads where the enzyme bound. After picking the beads with most intense staining, they sequenced 10 beads. A consensus sequence was chosen, Glu−Leu−Asp−Leu*Ala−Val−Glu−Phe, where the * indicates the hydroxyethyl isostere. This gave a K_i value of 0.31 nM.

Gustchina et al.,[177] following up on studies published by Visser et al.,[178] carried out an analysis of chymosin activity by studying the effects of lengthening the substrate peptide on the amino terminal end to give the peptide of sequence 98−109 of the κ-casein sequence that is cleaved by chymosin. Gustchina et al. also compared the effects of adding the fragment 90−102 on the kinetics of cleavage of 98−109 and 102−109, with the following results:

	Rate μMmin^{-1}
His-Pro-His-Pro-His-Leu-Ser-Phe*Met-Ala-Ile-Pro-NH_2	45
" + His-Pro-His-Pro-His	50
Leu-Ser-Phe*Met-Ala-Ile-Pro-NH_2	2.45
" + His-Pro-His-Pro-His	462

The longer peptide is cleaved about 18 times faster than the shorter version, in agreement with the results of Visser et al.,[178] and the authors suggest that the extension is able to interact with a negatively charged region on the surface of chymosin and alter the conformation of the enzyme, as suggested by Andreeva et al.[114] While the addition of the 98−102 segment had no effect on the cleavage rate of 98−109, adding the N-terminal His−Pro−His−Pro−His segment caused a 190-fold increase in the rate of cleavage of 102−109. It is surprising that the increase caused by the intermolecular effect is greater than the increase caused by covalently adding the His-rich sequence to the end of the shorter peptide. In the experiments of Gustchina et al.,[177] the added segment was used in a higher concentration (5 mM) that either the long substrate (0.25 mM) or the short substrate (0.5 M). Thus it is possible that additional effects might be present such as a general electrostatic effect.[179]

VII. Inhibitors of Aspartic Peptidases

A. Synthetic Inhibitors

This section will highlight several recent studies of inhibitors designed against aspartic peptidases. A review by Bursavich and Rich provides a historical overview of inhibitor design as well as specific comments on designing inhibitors for the active site of aspartic peptidases.[180] In the design of inhibitors against enzymes from pathogenic organisms, it is crucial to have information from the binding of potential compounds to related enzymes of the human body so as to avoid cross inhibition of essential functions. In a more general sense, comparisons with any enzyme of the same family would be valuable in deciphering the structure−function relationship of inhibitor binding. To this end, Cronin et al.[181] have compared the binding of five renin inhibitors to the active site of yeast peptidase A by crystallographic means. By comparing the interactions seen with the yeast enzyme to those seen in renin binding, the authors hoped to learn about the origin of the differences in binding affinity that have been observed. Yeast peptidase A has a three-dimensional structure[182] that is very similar to human cathepsin

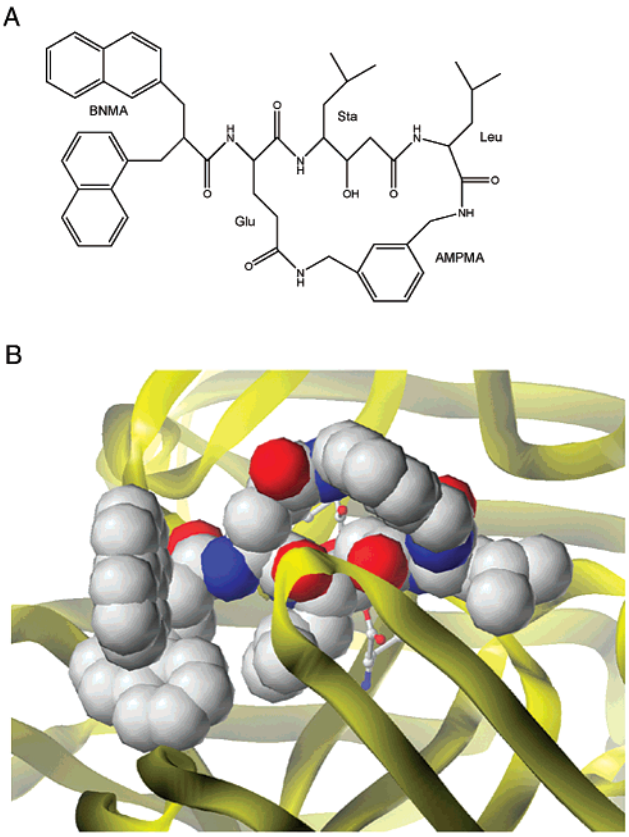

Figure 7. (A) The structure of PD-129,541. BNMA, bis-[(1-naphthyl)methyl)]acetic acid; Sta, statine (3-OH,4-NH_2,6-methylheptanoic acid); AMPMA, meta-di(aminomethyl)benzene. (B) Complex of PD-129,541 with yeast proteinase A. PDB file 1FQ5[181] was used to create this figure. The enzyme is represented as a yellow ribbon, with the catalytic aspartic acid residues rendered as ball-and-stick figures. The inhibitor is rendered in space-fill representation. In this orientation, one can see the cyclic structure connecting P2 to P3′ at the top. The bulk provided by this cyclic connection leads to differentiation in binding to different aspartic peptidases. In the lower part of the figure, one can see the Leu-like side chain that fills the S1 specificity pocket in the middle, and one of the naphthyl groups that occupies the S3 pocket.

D. In general, the five inhibitors bind through the active site cleft in a manner similar to that seen for binding of ligands to the family and, specifically, making backbone hydrogen bonds in the same pattern as noted by Bailey and Cooper.[183] These hydrogen bonds have been suggested to help position the scissile peptide bond of a substrate for attack by the bound catalytic water molecule.

Additionally, the inhibitors make a variety of other noncovalent interactions with the active site, leading to burial of large hydrophobic surfaces. Yeast peptidase A is similar to other aspartic peptidases with a relatively broad substrate specificity, such as the digestive enzyme pepsin, and the comparison of the binding surface with that of renin is particularly enlightening. For example, one inhibitor, PD-129,541 (Figure 7), had a naphthyl group in the P3 position. In yeast peptidase A, this group was accommodated in the S3 subsite by interaction with hydrophobic groups Phe 112 and Phe 117. The size of the yeast enzyme pocket was adequate for the naphthyl group, but in renin the S3 pocket is smaller due to substitution of several residues and a different position of a helical segment. This is one reason that this inhibitor binds less tightly to renin. The S2 subsite of the enzymes provides a major point for differentiation of binding of inhibitors. Residue 287 is a Thr in yeast peptidase A, but a His in renin, and a Glu in pepsin. Comparison of this region of the yeast enzyme with mouse submaxillary renin showed that the polyproline loop, residues 286−301 of renin, is closer to the binding cleft than the analogous loop of other enzymes, which in turn places the His at 287 closer to the P2 side chain of inhibitors. Endothiapepsin has an S2 subsite that is much more open that either yeast peptidase A or renin, due to a three-residue deletion (291−293).

The Parke-Davis compound PD-129,541 was also useful in defining the interaction at P3′. The inhibitor has a structure in which the P3′ side chain was cyclized to the P2 side chain by a *m*-di(aminomethyl)-benzene (AMPMA) group connected by amide bonds to the P2′ backbone carbonyl group and the Glu side chain at P2. This cyclization creates extra bulk on the side of the cleft that is defined by the C-terminal domain of the enzymes. Again, because renin has a distinctive polyproline loop at residues 286−301, and that this loop is closer to the active site cleft that in other enzymes, the space available for binding of inhibitor atoms is limited. At the same time, the space available to the yeast enzyme in the same region is larger, leading to better binding affinity than for binding to renin.

Recognizing the problem of conformational entropy in large inhibitors, Khan et al.[116] constructed a macrocyclic phosphonate inhibitor where the side chains at P2 and P1′ were covalently linked, based on earlier work on the bound conformation of an earlier version of the phosphonate compound.[184] This led to an improvement of 420-fold in the binding affinity to penicillopepsin for the cyclic compound compared to its acyclic analogue. Both compounds were cocrystallized with the enzyme and structures were determined to high resolution. Both compound adopted the same conformation in the bound state, thus establishing that the acyclic compound was capable of achieving that conformation. This study provides a measure of the benefit of restricting the conformation of an inhibitor.

Bailey and Cooper[183] compared the binding of 21 different inhibitors to endothiapepsin. Although this series covered a diverse group of chemical structures, all were bound in the extended β-strand conformation seen for other ligand complexes of aspartic peptidases. Bailey and Cooper have provided a very thoughtful analysis of the hydrogen bonding geometry and distances of each point in the inhibitor molecules, torsion angles along the inhibitor chains, hydrophobic and other nonbonding interactions, temperature factors, and changes in solvent accessibility for complex formation. Figure 3 of their paper gives donor−acceptor distances for 17 hydrogen bond pairs formed with nearly all of the inhibitors. Positions exhibiting a range of distances between 2.4 and 3.1 Å (therefore likely to be strong hydrogen bonds) include the side chain OH of Thr 219 and the

backbone NH of the residue at P3 of the inhibitor, the carbonyl group of residue P3 and a structural water molecule, the inner oxygen atom of Asp 32 and the OH of the isostere at P1, and the outer oxygen of Asp 215 and the same OH group. The NH of Gly 76 and the carbonyl oxygen of residue P1′, the carbonyl oxygen of Gly 34 and the backbone NH of residue P2′, and the inner oxygen of Asp 215 and the inhibitor OH group at P1 make important hydrogen bonds.

Majer et al.[185] studied the binding of a series of pepstatin derivatives to human cathepsin D. They were guided by the structure of Baldwin et al.,[105] which showed the interactions of pepstatin with the enzyme. Substituents were chosen in an attempt to increase steric bulk at several positions in hopes of increasing binding affinity. In almost all cases, binding was actually diminished by the substitutions. The best compound in the series replaced the P4 isovaleryl group with a *tert*-butylacetyl group, the P1 statine side chain (isobutyl) with a benzyl group, and the P2′ Ala with a Val side chain. However, this compound was not as good as pepstatin itself, causing the authors to conclude that pepstatin is a "highly optimized biological inhibitor of cathepsin D". They also noted that the results of substitution in pepstatin were not coincident with the results of substitution in oligopeptide substrates as reported by Scarborough et al.,[162] largely due to the differences in binding of the statine residue of the inhibitors.

Ripka et al.[186] used a modeling program, GrowMol, to design compounds that would fill up the binding site of the aspartic peptidase from *Rhizopus chinensis*. They successfully constructed a number of side chain cyclized structures that were cocrystallized with the fungal enzyme. The structures determined showed good agreement with the predicted structures although details of binding constants were not provided. The same group[187] applied the GrowMol technique to develop inhibitors against the active site of porcine pepsin, starting from a complex of Abbott compound A66702. After removing all of the inhibitor atoms except the critical hydroxyl group that interacts with the catalytic aspartic acids, the program added atoms to fill up the space with a limit of 30 atoms total. They pursued one class of structures predicted by this method, asymmetric peptidyl ureas with micromolar to subnanomolar potencies against pepsin and rhizopuspepsin. Two inhibitors were cocrystallized with rhizopuspepsin, one exhibiting a K_i of 1.5 μM and one yielding a K_i of 6 nM with the fungal enzyme. Both inhibitors were found to bind into the active site cleft in a manner consistent with predictions of the modeling method. In a third report, Bursavich et al.[188] extended the method further to non-peptide peptidomimetics. This process will be important in converting substrate-like compounds into stable inhibitors. A novel series of structures was generated by altering the conformation of the active site slightly: moving the flap β-hairpin up by 1 Å, rotating Tyr 75 by −120°, and moving Trp 39 slightly. These manipulations created space for modeling a piperidine derivative into the active site cavity. In this work, the authors were following the lead of earlier studies from the Roche group on the cocrystal structure of unusual structures, identified in a high-throughput screening experiment, in complex with renin.[189] The structure of the renin complex showed the structural rearrangements at the active site that were used by Bursavich et al.[188] in designing their inhibitor. The compound was made and found to have a micromolar K_i value; however, in this case no structural work was done to demonstrate that binding occurred in the fashion predicted.

Jiang et al.[190] utilized recombinant enzymes from *P. falciparum* and *P. vivax* to screen a library of chemical compounds in a database at Walter Reed Army Institute. A diphenylurea derivative was found that exhibited micromolar K_i values with both enzymes, while having much weaker binding to human cathepsin D. This compound killed two strains of *P. falciparum* in culture and provides a lead for further antimalarial drug development. Other studies have utilized inhibitors based on the allophenylnorstatine group,[191] or combinatorial libraries.[192−194]

B. Naturally-Occurring Protein and Peptide Inhibitors

New reports on the interaction of protein inhibitor binding have appeared. The first was provided by Li et al.[55] They studied the binding of the cytoplasmic 68-residue protein inhibitor[195] from yeast cells, IA_3, which has strong affinity to the vacuolar yeast peptidase A. Remarkably, this inhibitor is almost completely specific for the yeast enzyme over all other aspartic peptidases that have been examined.[196,197] Li et al.[55] determined the crystallographic structure of the complex between the IA_3 inhibitor and yeast peptidase A to 2.2 Å resolution for the inhibitor with Lys31−Lys32 replaced by Met31−Met32, and to 1.8 Å resolution for a synthetic peptide comprising residues 2−34 of the native inhibitor (Figure 8). Both derivatives of the naturally occurring IA_3 had K_i values within a factor of 3 of the 1.1 nM K_i found with the native sequence. Additionally, the structures observed were nearly identical, as the amino acids from 33 to the C-terminus of the inhibitor were disordered in the complex. IA_3 forms a nearly perfect amphipathic alpha helix from residues 2−32 and blocks the active site by binding along the cleft to prevent entry of substrates. The presence of the helical inhibitor forces the flap to move more than 8 Å from its position in the complex of yeast peptidase A with a small molecule inhibitor.[184] The inhibitor is bound is such a way that no peptide bond of IA3 is within 5 Å of the catalytic aspartic acid residues, thus ensuring that cleavage will not occur. Extensive studies were done to search for cleavage of the inhibitor protein with none detected.[55,197]

One remarkable feature of the complex of IA_3 and yeast peptidase A is that the catalytic water molecule, believed to be the nucleophile that attacks the scissile peptide bond in a productive substrate complex, remains bound in the inhibited enzyme structure. In all complexes with peptidomimetic inhibitors the catalytic water is displaced, by the -OH group of statine, for example. A Lys residue of IA_3 makes hydrogen-bonding contacts with Asp 32 (2.84 Å in the synthetic inhibitor complex). Asp 22 of the inhibitor

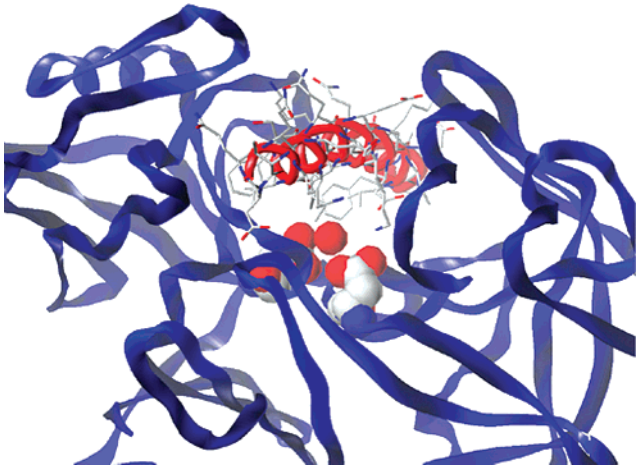

Figure 8. Representation of the structure of the yeast inhibitor, IA-3, bound to the active site cleft of yeast proteinase A. PDB file 1DPJ[55] was used to create this figure. The inhibitor used to form this complex is a synthetic peptide of residues 2−34 of the IA-3 sequence. Amino acids 2−32 only are seen in the diffraction pattern, as was also observed for the structure of the full-length IA-3 molecule. In this view, the enzyme is represented as a blue ribbon and the inhibitor as a red tube with side chains as capped sticks. The catalytic aspartic acids are shown as space-filling format and the catalytic water molecule is also given as a red sphere between the two Asp residues. This helical inhibitor does not come close enough to the catalytic residues to undergo hydrolysis, so this complex is stable with no cleavage of the peptide. A Lys side chain at position 18 of the inhibitor makes a hydrogen bonding contact to the outer oxygen of Asp 32 on the right-hand side. Asp 22 of the inhibitor also makes contact in the active site cleft. In general, however, most of the contacts are between the hydrophobic residues of the amphipathic helix and the hydrophobic active site cleft.

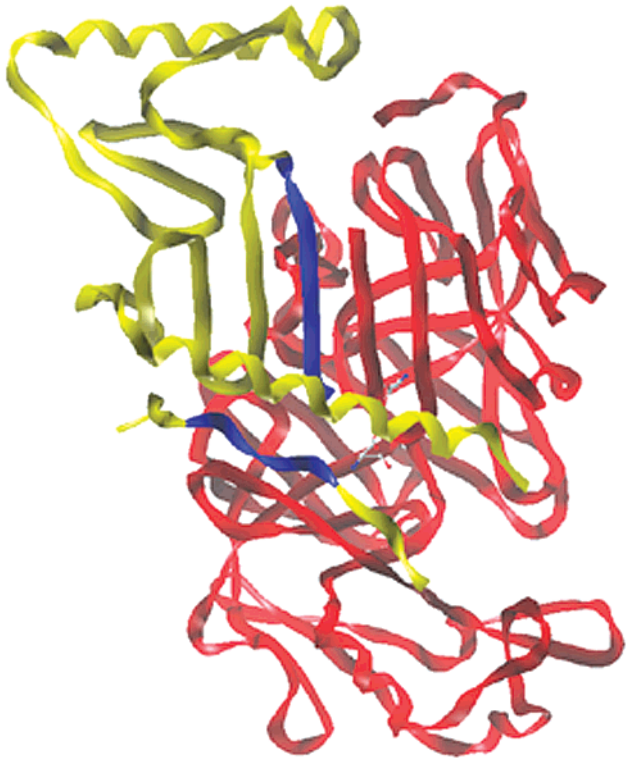

Figure 9. Representation of the structure of the complex between inhibitor PI-3 and porcine pepsin. PDB file 1F34[208] was used to create this figure. The enzyme is represented as a red ribbon. The inhibitor is represented as a yellow ribbon, except for two segments that make contact with the pepsin surface. These two segments residues 1−11 and residues 139−143 of the inhibitor are shown as blue ribbons. The catalytic aspartic acids are rendered as ball-and-stick figures and can be seen near the middle of the figure. However, no part of the inhibitor comes into contact with the catalytic apparatus.

makes hydrogen bonds to Lys 18 as well as to the side chain of Tyr 75 of the enzyme. The helix is held in the active site by hydrophobic contacts as well, especially between inhibitor residues Phe 12, Leu 19, Ala 23, Val 25, Val 26, Phe 30, and Met 31. Phe 12 and Leu 19 could make hydrophobic contacts with the S2 and S4 specificity pockets, respectively. These aspects were explored further by Phylip et al.[197] in a series of mutations of the IA_3 sequence. They found that two hydrophobic clusters, Val 8−X−X−hydrophobic−Phe 12 and Val 26-X−X−hydrophobic−Phe 30 could tolerate some substitutions of single residues by Ala, but replacement of two or more resulted in loss of inhibitory function.

Surprisingly, replacement of the Lys 18 by Met and of the Asp 22 by Leu resulted in an inhibitor that was more potent than the natural sequence. On the other hand, the conversion of the Leu residue at position 19 to an Ala led to a dramatic loss in inhibitor potency. These two observations support the idea that the amphipathic nature of the helix of IA_3 when bound to the enzyme is heavily dependent upon hydrophobic interactions with the generally hydrophobic cleft of yeast peptidase. The inability of the native inhibitor to bind to other members of the aspartic peptidase family, all of which have generally hydrophobic active site clefts, is most likely due to the absence of the precise geometric arrangement of pockets in enzymes other than yeast peptidase A that can interact with three elements of IA_3, namely, the two hydrophobic clusters described above and Leu 19. The inhibitor is unstructured when free in solution (T. Green, A. Edison, B. Dunn, unpublished); this feature may contribute to the high aqueous solubility of the protein, since if the molecule adopted the amphipathic helix in the absence of the enzyme, it most likely would self-associate.

An inhibitor of the digestive enzyme pepsin was described initially by Abu-Erreish and Peanasky,[198] sequenced by Martzen et al.,[199] and cloned by Kageyama.[200] PI-3, a 149-residue protein inhibitor, was shown to be specific for pepsin and gastricsin.[201,202] Ng et al.[203] described the structure of the pepsin-PI-3 complex (Figure 9). Novel features of this complex revealed by crystallography include the structure of the PI-3 inhibitor, consisting of two antiparallel β-sheets, each flanked by a single α-helix, and the unusual mechanism of inhibition. Unlike many peptidase inhibitors, such as bovine pancreatic trypsin inhibitor (BPTI, or aprotinin), the PI-3 inhibitor does not operate by placing an exposed loop directly into the active site cleft. PI-3 makes three important contacts with pepsin. The first region of contact is in the prime side of the active site cleft. Up to three residues at the amino terminus of PI-3 may be occupying subsites P1′ to P3′, although the electron density is not optimal at this point. The second point

of contact is provided by residues 4−8 of the inhibitor, which make an antiparallel β-strand next to the first half of the flap of the enzyme. This creates an eight-stranded β-sheet including three strands of PI-3 and five of the enzyme. The third point of contact is between residues 139−143 of the inhibitor, which are all Pro residues, and the 290−294 loop of pepsin, also known as the polyproline loop. The 290−294 sequence of the aspartic peptidase family is highly variable and this may account for the specificity of PI-3.

Equistatin is a protein of M_r 21 755 that was isolated from the sea anemone *Actinia equina*.[204] It has a structure composed of three repeated thyroglobulins type-1 domains, with the three domains having between 32 and 49% identity and was found to inhibit both cysteine and aspartic peptidases. Following an initial report using proteolytic digestion to separate the protein into two components: domain 1 and domains 2 and 3 together,[205] Strukelj et al.[206] expressed each domain separately and demonstrated that the first domain inhibited the activity of papain while the second domain inhibited cathepsin D. The K_i values determined were in the subnanomolar range and the activity against cathepsin D was specific as no other aspartic peptidases were inhibited. Structural analysis of the interaction between cathepsin D and this new inhibitor should reveal new modes of binding that might suggest strategies for drug design.

Christeller et al.[207] have isolated a mixture of proteinaceous inhibitors from squash phloem exudate and found that three proteins of similar sequence were present with slightly different amino terminal ends. Analysis of the cDNA of two clones revealed a 96% identity; thus, the different forms present may represent polymorphisms of the same inhibitor. Christeller et al. determined its molecular mass to be 10 552 Da by MALDI-TOF mass spectroscopy, although the protein appears to elute with an apparent molecular weight of double that on gel filtration. This might indicate an association of two molecules even though the inhibitor is lacking Cys residues. The K_i value of 3 nM, obtained for inhibition of pepsin, is impressive. A secreted aspartic peptidase from the yeast *Glomerella cingulata* was inhibited less strongly (34.5 nM) and the major secreted aspartic peptidase from *C. albicans* was inhibited to a lesser extent so that a K_i could not be determined.

A peptide inhibitor has been isolated from the extremophilic bacteria *Bacillus* sp.[208] It was shown to have a dual activity, inhibiting both xylanase and an aspartic peptidase from *Aspergillus satoi*.[209] The K_i value against the aspartic peptidase was modest at 3.25 μM; however, this is remarkable for a linear peptide of 11 residues with no unusual amino acids. The sequence of this inhibitor, Ala−Gly−Lys−Lys−Asp−Asp−Asp−Asp−Pro−Pro−Glu, is highly charged and could disrupt the hydrogen bonding network at the active site. Dash et al.[209] have presented a more thorough analysis of binding to pepsin as well as the *A. satoi* enzyme. Binding to pepsin is considerably tighter than to the fungal enzyme, with an overall K_i value of 55 pM. The binding mechanism was shown to be two-step, with an initial equilibrium dissociation constant of 17 nM for pepsin and 3.2 μM for the *Aspergillus* enzyme. The initial step was followed by an isomerization of the enzyme−inhibitor complex, with a faster forward and reverse rate for the *Aspergillus* enzyme than for pepsin.

VIII. Mechanistic Studies

Binding of the natural product pepstatin to a wide variety of aspartic peptidases has been taken to be a model for the tetrahedral intermediate in the catalytic reaction. The observation by Bott et al.[210] and James et al.[211] of the structures of fragments of pepstatin bound in the active sites of rhizopuspepsin and penicillopepsin, respectively, showed that the hydroxyl group of the first statine was bound between the Oδ2 of both Asp residues (32 and 215), effectively replacing the water molecule seen in the native enzyme. However, the statine moiety or the hydroxyethylene peptidomimetic that have been popular as transition state analogue inhibitors do not perfectly represent the transition state, as they lack one of the hydroxyls of the diol intermediate. The discovery by Rich et al.[212] that difluorostatones form the hydrates more readily than statones led to the discovery of potent inhibitors.[213−216] Veerapandian et al.[217] studied the binding of CP-81,282 [morpholino-4-carbonyl-Phe-Nle-cyclohexyldifluorostatone-*N*-methylamide], a tripeptide renin inhibitor containing a difluoroketone moiety. In this complex, the difluoroketone hydrates readily to yield the diol, and the structure shows the pro-R hydroxyl group is within hydrogen bonding distance to all four oxygen atoms of the two Asp residues, but is closest to the outer Oγ2 of Asp 215 (2.57 Å). The second hydroxyl (pro-(S)) is 2.48 Å from the Oγ2 of Asp 32. In this complex, the two inner Oγ1 atoms of Asp 32 and 215 are 3.34 Å apart, outside hydrogen bonding range. James et al.[218] studied isovaleryl−Val−Val−StoF_2NHCH_3 [StoF_2 = (*S*)-4-amino-2,2-difluoro-3-oxo-6-methylheptanoic acid], which is a derivative of a shortened version of pepstatin and have presented a mechanistic proposal that is similar to that of Veerapandian et al.[217] In the structure solved by James et al.,[218] the distance between the two inner oxygen atoms is 2.9 Å, and these authors also concluded that it was not necessary to place a hydrogen between these two oxygen atoms. In both proposals, Asp 215 acts as a general base to remove one proton from the water molecule while Asp 32 donates a proton to the carbonyl oxygen atoms of the scissile peptide bond. In the tetrahedral intermediate, Asp 215 is hydrogen bonded to the attacking oxygen atom, while the hydrogen remaining on that oxygen is hydrogen bonded to the inner oxygen of Asp 32. Transfer of the hydrogen from Asp 215 to the nitrogen of the scissile peptide bond is accomplished by inversion of configuration around the nitrogen atom. Following this step, the C−N bond breaks forming the two products. The carboxyl product (from the amino terminal side of the peptide bond) remains hydrogen bonded to Asp 32 and Asp 215 is in the negatively charged form and ready for the next round of catalysis.

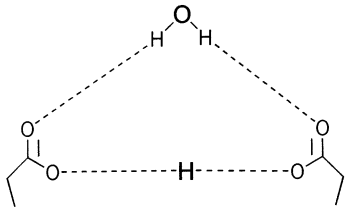

Figure 10. The symmetrical catalytic apparatus proposed by Northrop.[220] In this structure, the hydrogen between the two Asp residues is postulated to be in a low-barrier hydrogen bond. The two Asp residues have a net of one negative charge.

Coates et al.[219] studied the neutron diffraction of a hydroxyethylene based inhibitor bound to endothiapepsin. Neutron diffraction is performed on samples that have been soaked in deuterium oxide to exchange deuterium for hydronium (D for H) in proteins. The deuterium may be seen in the neutron diffraction pattern, thus indicating where the protons reside in these complexes. For the complex studied by Coates et al., the clear result was that Asp 32 was unprotonated, but formed hydrogen bonds to the hydroxyl group of the isostere bound in the active site. Asp 215 appeared to be protonated on the outer oxygen atoms, and these two observations are consistent with the proposals of both Veerapandian et al.[217] and James et al.[218] What is missing in the neutron diffraction study, of course, is the second hydroxyl group of the diol intermediate. Therefore, it seems that a structure should be done using the difluoroketone derivatives in place of the hydroxyethylene derivative.

Northrop[220] reviewed the kinetic mechanism of the aspartic peptidases and put forth a proposal that the key feature of the catalytic process is the formation of a low-barrier hydrogen bond between the two Asp residues. His proposal is based on the theoretical work of Piana and Carloni[221] on HIV PR who found the monoprotonated form to be most stable, with a distance of 2.5 Å between the Oδ1 atoms. In addition, it is well-known that a water molecule is hydrogen bonded to the Oδ2 atoms of the two Asp residues and that this water molecule is believed to be the nucleophile that attacks the carbonyl carbon of a peptide bond arranged in the active site.

Northrop has proposed a 10-membered cyclic structure involving the two carboxyl groups (COO^-), the hydrogen between Oδ1 of each Asp residue, and the water molecule (Figure 10). His mechanism involves features similar to those postulated by others, but includes a final isomerization step to return to the cyclic structure and re-form the low-barrier hydrogen bond.

A study by Hunkapillar and Richards[222] on porcine pepsin where the catalysis of hydrolysis of *N*-trifluoroacetyl-L-phenylalanine was studied in water and deuterium oxide was cited as evidence for a mechanism involving an isomerization step. While this work was done carefully, this is a very poor substrate for analysis of pepsin, with a k_{cat} value of 0.038 min^{-1} in water. The data for the pH profile in water were fit to a bell-shaped curve with pK_a values of 2.9 and 4.5. While the data in water appear adequate for the analysis, the data obtained in deuterium oxide are not sufficient to fit to a bell-shaped curve, with only five points near the top of the curve. The lack of data at both lower and higher pH values does not support the conclusion that the pK_a values in deuterium oxide are both shifted inward.

The suggestion of an isomechanism in which the enzyme finishes one catalytic cycle in the wrong ionic state and must lose a proton as well as gain a water molecule to return to the optimal configuration for catalysis does help to explain some kinetic and isotope data that have been collected over the years, as well as data on transpeptidation reactions. (See also the discussion of a recent report of pepsin inhibition below.) The most complete study of isotope effects was provided by Meek et al.[223] for the retroviral enzyme HIV-1 peptidase. Meek and colleagues presented solvent isotope effect data for two substrates of HIV-1 peptidase. On the basis of the deuterium isotope effect on k_{cat}/K_m of approximately 1.0 and the effect on k_{cat} of approximately 2.0, combined with solvent ^{18}O incorporation into substrate and an inverse nitrogen isotope effect, Meek concluded that the mechanism involves formation of a diol tetrahedral intermediate with general acid and base catalysis. Rate-limiting proton transfer to return the enzyme to the correct protonation state for the next round of catalysis was favored for the substrates they studied due to the lack of a deuterium solvent isotope effect on k_{cat}/K_m. However, Polgar et al.[224] have studied HIV-1 peptidase cleavage of two substrates (Lys−Ala−Arg−Val−Leu*Phe(NO_2)-P2′−Ala−Nle that differ only in the P2′ amino acid (Glu versus Gln). Solvent deuterium isotope effects differ for these substrates, leading to the conclusion that for the better substrate with Glu in P2′ the rate-limiting step is the chemical step of C−N bond scission, while for the substrate with Gln in the P2′ position, the rate-limiting step may be a conformational change. This study points out the difficulty of assigning mechanisms to a whole family of enzymes based on a single substrate. In addition, Porter et al.[225] have studied a substrate, 2-aminobenzoyl-Thr−Ile−Nle−Phe(NO_2)-Gln−Arg-NH_2, that allows the observation of the formation of a catalytically competent enzyme−substrate intermediate. The formation of the initial enzyme−substrate complex was observed by a small change in fluorescence of the 2-aminobenzoyl group arising from transfer from water to the enzyme surface. This was followed by the formation of a second enzyme−substrate complex at high enzyme concentration with enhanced fluorescence. Porter et al. attribute this to a conformational shift of the enzyme to produce a complex where all elements are poised for the chemical step. They observed that solvent isotope effects were present in two steps of the reaction, the step involving cleavage of the C−N bond and the isomerization step immediately prior to the rate-limiting cleavage.

The chemical mechanism proposed[220] for all aspartic peptidases (Figure 11) accounts for proton transfers to the substrate and is in general agreement with those suggested by others; however, intermediate F′T leaves the carbonyl oxygen of the substrate in the negatively charged state without hydrogen bonding

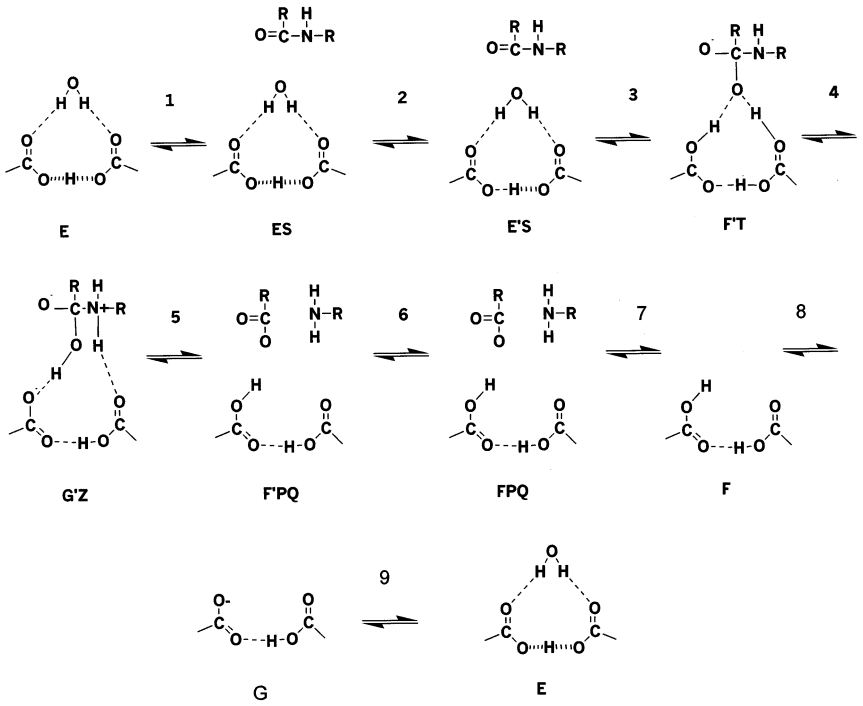

Figure 11. Catalytic mechanism proposed by Northrop.[220] In this mechanism, species E is the free enzyme poised for catalysis. Step 1 is the binding of substrate to form a loose complex (ES). Step 2 is the closing of the flap down upon the substrate to squeeze all components into the correct geometry and distances for the catalytic process to begin (E′S). Step 3 includes the removal of a proton from the bound water molecule to stimulate attack on the carbonyl carbon (F′T). Step 4 involves a proton transfer to the Nitrogen of the peptide bond (G′Z). Step 5 is the bond cleavage event (F′PQ). Step 6 is the opening of the flap to free the products (FPQ) and step 7 is release of the products (F). Step 8 includes a loss of one proton (G) and step 9 involves binding of a new water molecule and re-formation of the low-barrier hydrogen bond (E).

to stabilize this high energy form. In devising this scheme, it would have been better to take into account the structural work of the James and Blundell groups discussed above.

This discussion of mechanistic details comes down to the position of a proton: The symmetry proposal[220] would have one between the two catalytic Asp residues, while the structural analyses would suggest that the proton is on the outer oxygen of Asp 32 prior to substrate binding. Examining the structures of aspartic peptidases without bound ligands, one can get a good appreciation for the structural arrangement of the catalytic water molecule in relation to the two Asp side chains. The distances between the oxygen atoms of the Asps and the oxygen of the water molecule are between 2.621 and 3.147 Å, while the distances between the outer oxygen atoms and the oxygen of the water molecule are between 2.859 and 3.284 Å for structures 1PSN, 3APP, and 4APE. The average distances for the outer oxygen of the Asps to the water is 3.03 Å and for the inner is 2.89 Å, although this difference is not large enough to reach a definite conclusion. However, more important is that the angle $\gamma C'^{215}$-$\gamma O1^{215}$-$\gamma O1^{32}$ or $\gamma C'32$--$\gamma O1^{32}$-$\gamma O1^{215}$ is between 150 and 169°, while the angles for all combinations between the $\gamma C-O-O$[water] are between 90 and 104°. These angles are relevant to the orbitals on the oxygen atoms that are involved in forming hydrogen bonds. The wide angle between the two inner oxygen atoms suggests that a hydrogen bond is not formed between the inner oxygen atoms of the two Asp residues, while the more acute angle formed with the water molecule provides almost perfect alignment for hydrogen bonding to take place.

A further view on the catalytic mechanism was provided by Andreeva and Rumsh,[226] who have carefully analyzed the structures of 82 enzyme or enzyme−inhibitor complexes in detail. In their analysis, they have discovered a second conserved water molecule in the active site, designated W2. This takes up a position between Tyr 75, Asn 37, and Ser 35. In their mechanism, W2 is hydrogen bonded to the carbonyl oxygen of Asn 37 and to the OH of Ser 35 prior to binding of substrate. Upon substrate binding, the OH of Ser 35 rotates to provide a hydrogen bond to the outer oxygen of Asp 32 and the W2 water forms a new hydrogen bond to the OH of Tyr 75. These events are postulated to take place as the flap β-hairpin closes over the bound substrate (Figure 12). The new hydrogen bond to Asp 32 makes it easier to donate its proton to the carbonyl group of the scissile peptide bond, leading to the formation of an intermediate identical to that revealed in the crystallography of the difluoroketone derivatives by Veerapandian et al.[217] and James et al.[218] Conversely, Thr 218 on the other side of the flap plays a role in preventing the protonation of Asp 215, so that it is free to serve as a general base in the reaction. Upon substrate binding, the hydrogen bond between Thr 218 and Asp 215 is disrupted.

The switch in orientation of the hydrogen bonding of W2 is proposed[226] to be triggered by the hydrogen bonding from the NH of the side chain of Trp 39 to the OH of Tyr 75. For the final steps of the reaction, the hydrogen taken from W1 by Asp 215 will be transferred to the nitrogen of the peptide bond. This will increase the electrophilicity of that atom, which in turn would draw electrons from the carbonyl

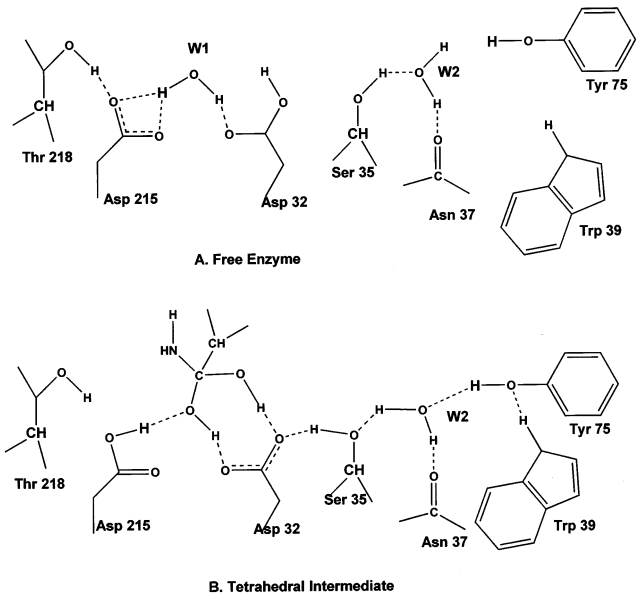

Figure 12. Mechanism proposed by Andreeva and Rumsh.[226] Part A represents the free enzyme, with hydrogen bonds from the Asp groups to the bound water molecule (W1) and no hydrogen bond between the two Asps. A hydrogen bond from Thr 218 to Asp 215 keeps the Asp in the negatively charged form and this residue acts as the general base to remove a proton from the water molecule (W1) during attack on a carbonyl group. Also, the second water molecule (W2) is hydrogen bonded to the -OH of Ser 35 and the carbonyl oxygen of Asn 37. Following approach of the peptide to be cleaved, the hydrogen bond from Thr 218 to Asp 215 is broken so that Asp 215 can take a proton from W1, which attacks the carbonyl (part B). The outer oxygen of Asp 32 donates a proton to the carbonyl group of the peptide bond, while the -OH of Ser 35 moves to form a new hydrogen bond to the oxygen of Asp 32. Water molecule W2 changes the hydrogen-bonding pattern to donate a hydrogen bond to the oxygen of the side chain of Ser 35 while accepting a hydrogen bond from Tyr 75. The formation of a new hydrogen bond from Trp 39 to the side chain oxygen of Tyr 75 facilitates the donation of the new hydrogen bond to W2. The new hydrogen bonds of Tyr 75 and Trp 39 are formed following the movement of the flap to bring Tyr 75 closer to both the W2 and the Trp. In the final steps of this mechanism, Asp 215 will give up a proton to the nitrogen of the leaving group and Asp 32 will take a proton away from the diol intermediate to create the electron push necessary to break the carbonyl carbon to nitrogen bond. Reformation of the hydrogen bonds with Thr 218 to Asp 215 and Ser 35 with W2 returns the structure to that seen in part A.

carbon and, hence the carbonyl oxygen. The reverse switch of Ser 35 to break the hydrogen bond to Asp 32 would increase the basicity of the Asp outer oxygen so that it will take back the proton it donated to the carbonyl oxygen, freeing electrons to re-form the carbonyl group. This would mean that the carboxyl product would leave the catalytic site with the proton from the inner oxygen of Asp 32. Re-formation of the Thr 218 hydrogen bond with Asp 215 and binding of a new W1 water molecule complete the catalytic cycle.

Marcinkeviciene et al.[227] have studied the binding of a substituted piperidine to porcine pepsin in relation to the binding of pepstatin. Their results provide strong support for the concept of an isomechanism where different conformational isomers or prototropic forms are in equilibrium. Several observations of the conformational flexibility of the aspartic family of enzymes were described in section III. Additional recent observations on the conformational flexibility of members of the aspartic peptidase class have been provided by Ullrich et al.[228] on conformational heterogeneity of the ground state of HIV-1 peptidase, by Porter's group on the binding of substrates and inhibitors to HIV-1 peptidase.[225,229] Whether these results support the low-barrier hydrogen bond proposal from Northrop, the hydrogen bonding network proposed by Andreeva and Rumsh, or a completely different scheme is not clear at present.

These speculations on the mechanism of the aspartic endopeptidases will not be settled before more structural work is done at higher resolution in both water and deuterium oxide. At the same time, some of the studies relating to kinetics and isotope effects should be repeated with better preparations of enzyme and more convenient substrates. Discoveries in these mechanistic details may provide new approaches to the design of inhibitors that could be of value in a number of disease states. Bursavich and Rich have provided a thorough analysis of inhibitor design with conformational ensembles in the forefront.[180]

IX. Summary

This review has described a variety of work conducted on the pepsin-like members of the aspartic endopeptidase family over the last dozen years. It is clear that the information base has expanded greatly and that we have new targets for drug discovery. It is also clear that we have new questions regarding the genetic relationships of these enzymes, the biosynthesis, folding, and targeting of the proteins, the mechanism of activation of the different types of proenzymes, and the mechanism of catalysis. The faster pace of discovery and analysis of the structure and function of enzymes of the pepsin-like family makes it easy to predict that the next dozen years will bring substantial progress in understanding the remaining questions.

X. Acknowledgment

The author would like to thank the members of his laboratory over the past 28 years for their efforts to advance our understanding of the structure−function relationships of aspartic peptidases. The work of this laboratory has been supported by NIH Grants DK18865, AI28571, and AI49063. The author also thanks John Kay, Alex Wlodawer, Michael James, Kohei Oda, and John Dame for their collaboration and insights into these enzymes.

XI. References

(1) Davies, D. R. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 189.
(2) Dunn, B. M. *Structure and Function of the Aspartic Peptidases*; Plenum Press: New York, 1991.
(3) Takahashi, K. *Aspartic Peptidases Structure, Function, Biology and Biomedical Implications*; Plenum Press: New York, 1995.
(4) James, M. N. G. *Aspartic Peptidases Retroviral and Cellular Enzymes*; Plenum Press: New York, 1998.

(5) Fusek, M.; Vetvicka, V. *Aspartic Peptidases Physiology and Pathology*; CRC Press: Boca Raton, 1995.
(6) Barrett, A. J.; Rawlings, N. D.; Woessner, J. F. *Handbook of Proteolytic Enzymes*; Academic Press: San Diego, 1998.
(7) Wlodawer, A.; Gustchina, A. *Biochim. Biophys. Acta* **2000**, *1477*, 16.
(8) Jean, L.; Long, M.; Young, J.; Pery, P.; Tomley, F. *Trends Parasitol.* **2001**, *17*, 491.
(9) Coombs, G. H.; Goldberg, D. E.; Klemba, M.; Berry, C.; Kay, J.; Mottram, J. C. *Trends Parasitol.* **2001**, *17*, 532.
(10) Banerjee, R.; Liu, J.; Beatty, W.; Pelosof, L.; Klemba, M.; Goldberg, D. E. *Proc. Natl. Acad. Sci U.S.A.* **2002**, *99*, 990.
(11) Francis, S. E.; Sullivan, D. J.; Goldberg, D. E.; *Annu. Rev. Microbiol.* **1997**, *51*, 97.
(12) Francis, S. E. Banerjee, R.; Goldberg, D. E. *J. Biol. Chem.* **1997**, *272*, 14961.
(13) Dame, J. B.; Reddy, G. R.; Yowell, C. A.; Dunn, B. M.; Kay, J.; Berry, C. *Mol. Biochem. Parasitol.* **1994**, *64*, 177.
(14) Moon, R. P.; Tyas, L.; Certa, U.; Rupp, K.; Bur, D.; Jacquet, C.; Matile, H.; Loetscher, H.; Grueninger-Leitch, F.; Kay, J.; Dunn, B. M.; Berry, C.; Ridley, R. G. *Eur. J. Biochem.* **1997**, *244*, 552.
(15) Hill, J.; Tyas, L.; Phylip, L. H.; Kay, J.; Dunn, B. M.; Berry, C. *FEBS Lett.* **1994**, *352*, 155.
(16) Westling, J.; Cipullo, P.; Hung, S.-H.; Saft, H.; Dame, J. B.; Dunn, B. M. *Protein Sci.* **1999**, *8*, 2001.
(17) Berry, C.; Humphreys, M. J.; Matharu, P.; Granger, R.; Horrocks, P.; Moon, R. P.; Certa, U.; Ridley, R. G.; Bur, D.; Kay, J. *FEBS Lett.* **1999**, *447*, 149.
(18) Wyatt, D.; Berry, C. *FEBS Lett.* **2002**, *513*, 159.
(19) Tyas, L.; Gluzman, I.; Moon, R. P.; Rupp, K.; Westling, J.; Ridley, R. G.; Kay, J.; Goldberg, D. E.; Berry, C. *FEBS Lett.* **1999**, *454*, 210.
(20) Le Bonniec, S.; Deregnaucourt, C.; Redeker, V.; Banerjee, R.; Grellier, P.; Goldberg, D. E.; Schrevel, J. *J. Biol. Chem.* **1999**, *274*, 14218.
(21) Humphreys, M.; Moon, R. P.; Klinder, A.; Fowler, S. D.; Rupp, K.; Bur, D.; Ridley, R. G.; Berry, C. *FEBS Lett.* **1999**, *463*, 43.
(22) Becker, M. M.; Harrop, S. A.; Dalton, J. P.; Kalinna, B. H.; McManus, D. P.; Brindley, P. J. *J. Biol. Chem.* **1995**, *270*, 24496.
(23) Brinkworth, R. I.; Harrop, S. A.; Prociv, P.; Brindley, P. J. *Inter. J. Parasitol.* **2000**, *30*, 785.
(24) Brindley, P. J.; Kalinna, B. H.; Wong, J. Y. M.; Bogitsh, B. J.; King, L. T.; Smyth, D. J.; Verity, C. K.; Abbenante, G.; Brinkworth, R. I.; Fairlie, D. P.; Smythe, M. L.; Milburn, P. J.; Bielefeldt-Ohmann, H.; Zheng, Y.; McManus, D. P. *Mol. Biochem. Parasitol.* **2001**, *112*, 103.
(25) Harrop, S. A.; Prociv, P.; Brindley, P. J. *Biochem. Biophys. Res. Commun.* **1996**, *227*, 294.
(26) Jolodar, A.; Miller, D. J. *Biochim. Biophys. Acta* **1998**, *1382*, 13.
(27) Geier, G.; Banaj, H.-J.; Heid, H.; Bini, L.; Pallini, V.; Zwilling, R. *Eur. J. Biochem.* **1999**, *264*, 872.
(28) Hashmi, S.; Tawe, W.; Lustigman, S. *Trends Parasitol.* **2001**, *17*, 387.
(29) Tcherepanova, I.; Bhattacharyya, L.; Rubin, C. S.; Freedman, J. H. *J. Biol. Chem.* **2000**, *275*, 26359.
(30) Gallego, S. G.; Slade, R. W.; Brindley, P. J. *Acta Trop.* **1998**, *71*, 17.
(31) Saftig, P.; Hetman, M.; Schmahl, W.; Weber, K.; Heine, L.; Mossman, H.; Koster, A.; Hess, B.; Evers, M.; von Figura, K.; Peters, C. *EMBO J.* **1995**, *14*, 3599.
(32) De Stasio, R.; Borrelli, L.; Kille, P. Parisi, E.; Filosa, S. *Mol. Reproduc. Dev.* **1999**, *52*, 126.
(33) Brooks, S.; Tyler, C. R.; Carnevali, O.; Coward, K.; Sumpter, J. P. *Gene* **1997**, *201*, 45.
(34) Capasso, C.; Lees, W. C.; Capasso, A. Scudiero, R.; Carginale, V.; Kille, P.; Kay, J.; Parisi, E. *Biochim. Biophys. Acta* **1999**, *1431*, 64.
(35) Capasso, C.; Riggio, M.; Scudiero, R.; Carginale, V.; di Prisco, G.; Kay, J.; Kille, P.; Parisi, E. *Biochim. Biophys. Acta* **1998**, *1387*, 457.
(36) Nielsen, L. B.; Nielsen, H. H. *Comp. Biochem. Physiol. Part B* **2001**, *128*, 351.
(37) Dragonetti, A.; Baldassarre, M.; Castino, R.; Demoz, M.; Luini, A.; Buccione, R.; Isodoro, C. *J. Cell Sci.* **2000**, *112*, 3289.
(38) Hussain, I.; Powell, D.; Howlett, D. R.; Tew, D. G.; Meek, T. D.; Chapman, C.; Gloger, I. S.; Murphy, K. E.; Southan, C. D.; Ryan, D. M.; Smith, T. S.; Simmons, D. L.; Walsh, F. S.; Dingwall, C.; Christie, G. *Mol. Cell Neurosci.* **1999**, *14*, 419.
(39) Vassar, R.; Bennett, B. D.; Babu-Kahn, S.; Kahn, S.; Mendiaz, E. A.; Denis, P.; Teplow, D. B.; Ross, S.; Amarante, P.; Loeloff, R.; Luo, Y.; Fisher, S.; Fuller, J.; Edenson, S.; Lile, J.; Jarosinski, M. A.; Biere, A. L.; Curran, E.; Burgess, T.; Louis, J. C.; Collins, F.; Treanor, J.; Rogers, G.; Citron, M. *Science* **1999**, *286*, 735.
(40) Yan, R.; Bienkowski, M. J.; Shuck, M. E.; Miao, H.; Tory, M. C.; Pauley, A. M.; Brashier, J. R.; Stratman, N. C.; Mathews, W. R.; Buhl, A. E.; Carter, D. B.; Tomasselli, A. G.; Parodi, L. A.; Heinrikson, R. L.; Gurney, M. E. *Nature* **1999**, *402*, 533.
(41) Sinha, S.; Anderson, J. P.; Barbour, R.; Basi, G. S.; Caccavello, R.; Davis, D.; Doan, M.; Dovey, H. F.; Frigon, N.; Hong, J.; Jacobson-Croak, K.; Jewett, N.; Keim, P.; Lieberburg, I.; Power, M.; Tan, H.; Tatsuno, G.; Tung, J.; Schenk, D.; Seubert, P.; Suomensaari, S. M.; Wang, S.; Walker, D.; Zhao, J.; McConlogue, L.; John, V. *Nature* **1999**, *402*, 537.
(42) Lin, X. L.; Koelsch, G.; Wu, S. L.; Downs, D.; Dashti, A.; Tang, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1456.
(43) Howlett, D. R.; Simmons, D. L.; Dingwall, C.; Christie, G. *Trends Neurosci.* **2000**, *23*, 565.
(44) Acquati, F.; Accarino, M.; Nucci, C.; Fumagalli, P.; Jovine, L.; Ottolenghi, S.; Taramelli, R. *FEBS Lett.* **2000**, *468*, 59.
(45) Hong, L.; Koelsch, G.; Lin, X. L.; Wu, S. L.; Terzyan, S.; Ghosh, A. K.; Zhang, X. C.; Tang, J. *Science* **2000**, *290*, 150.
(46) Yan, R. Q.; Han, P.; Miao, H.; Greengard, P.; Xu, H. X. *J. Biol. Chem.* **2001**, *276*, 36788.
(47) Haniu, M.; Denis, P.; Young, Y.; Mendiaz, E. A.; Fuller, J.; Hui, J. O.; Bennett, B. D.; Kahn, S.; Ross, S.; Burgess, T.; Katta, V.; Rogers, G.; Vassar, R.; Citron, M. *J. Biol. Chem.* **2000**, *275*, 21099.
(48) Evin, G.; Cappai, R.; Li, Q.-X.; Culvenor, J. G.; Small, D. H.; Beyreuther, K.; Masters, C. L. *Biochemistry* **1995**, *34*, 14185.
(49) Wolfe, M. S.; Xia, W.; Moore, C. L.; Leatherwood, D. D.; Ostazewski, B.; Rahmati, T.; Donkor, I. O.; Selkow, D. J. *Biochemistry* **1999**, *38*, 4720.
(50) Shearman, M. S.; Beher, D.; Clarke, E. E.; Lewis, H. E.; Harrison, T.; Hunt, T.; Nadin, A.; Smith, A. L.; Stevenson, G.; Castro, J. L. *Biochemistry* **2000**, *39*, 8698.
(51) Easler, W. P.; Kimberly, W. T.; Ostaszewski, B. L.; Diehl, T. S.; Moore, C. L.; Tsai, J.-Y.; Rahmati, T.; Xia, W.; Selkoe, D. J.; Wolfe, M. S. *Nat. Cell Biol.* **2000**, *2*, 428.
(52) Selkoe, D. J.; Wolfe, M. S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5690.
(53) Wolfe, M. S.; De Los Angeles, J.; Miller, D. D.; Xia, W.; Selkoe, D. J. *Biochemistry* **1999**, *38*, 11223.
(54) Wolfe, M. S.; Citron, M.; Diehl, T. S.; Xia, W.; Donkor, I. O.; Selkoe, D. J. *J. Med. Chem.* **1998**, *41*, 6.
(55) 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.
(56) Tatnell, P. J.; Powell, D. J.; Hill, J.; Smith, T. S.; Tew, D. G.; Kay, J. *FEBS Lett.* **1998**, *441*, 43.
(57) Schauer-Vukasinovic, V.; Bur, D.; Kling, D.; Gruninger, F.; Giller, T. *FEBS Lett.* **1999**, *462*, 135.
(58) Chuman, Y.; Bergman, A.-C.; Ueno, T.; Saito, S.; Sakaguchi, K.; Alaiya, A. A.; Franzen, B.; Bergman, T.; Arnott, D.; Auer, G.; Appella, E.; Jornvall, H.; Linder, S. *FEBS Lett.* **1999**, *462*, 129.
(59) Tatnell, P. J.; Cook, M.; Peters, C.; Kay, J. *Eur. J. Biochem.* **2000**, *267*, 6921.
(60) Schauer-Vukasinovic, V.; Wright, M. B.; Breu, V.; Giller, T. *Biochim. Biophys. Acta* **2000**, *1492*, 207.
(61) Azuma, T.; Pals, G.; Mohandas, T. K.; Couvreur, J. M.; Taggart, R. T. *J. Biol. Chem.* **1989**, *264*, 16748.
(62) Kageyama, T. *Eur. J. Biochem.* **1993**, *216*, 717.
(63) Kageyama, T.; Ichinose, M.; Tsukada, K.; Miki, K.; Kurokawa, K.; Okoiwai, O.; Tanji, M.; Yakabe, E.; Athauda, S. B. P.; Takahaski, K. *J. Biol. Chem.* **1992**, *267*, 16450.
(64) Okamoto, K.; Hao, Y.; Misumi, Y.; Ikehara, K.; Yamamoto, K. *Arch. Biochem. Biophys.* **1995**, *322*, 16450.
(65) Tatnell, P. J.; Lees, W. E.; Kay, J. *FEBS Lett.* **1997**, *408*, 62.
(66) Tatnell, P. J.; Roth, W.; Deussing, J.; Peters, C.; Kay, J. *Biochim. Biophys. Acta* **1998**, *1398*, 57.
(67) Kay, J.; Tatnell, P. J. In *Handbook of Proteolytic Enzymes*; Academic Press: London, 1998; p 819.
(68) Cook, M.; Caswell, R. C.; Richards, R. J.; Kay, J.; Tatnell, P. J. *Eur. J. Biochem.* **2001**, *268*, 2658.
(69) Bennett, K.; Levine, T.; Ellis, J. S.; Peanasky, R. J.; Samloff, I. M.; Kay, J.; Chain, B. M. *Eur. J. Immunol.* **1992**, *22*, 1519.
(70) Nishioku, T.; Kashimoto, K.; Yamashita, K.; Liou, S.-Y.; Kagamiishi, Y.; Maegawa, H.; Katsube, N.; Peters, C.; von Figura, K.; Saftig, P.; Katunuma, N.; Yamamoto, K.; Nakanishi, H. *J. Biol. Chem.* **2002**, *277*, 4816.
(71) Fowler, S. D.; Kay, J.; Dunn, B. M.; Tatnell, P. J. *FEBS Lett.* **1995**, *366*, 72.
(72) Kervinen, J.; Tormakangas, K.; Runeberg-Roos, P.; Guruprasad, K.; Blundell, T.; Teeri, T. H. In *Aspartic Peptidases: Structure, Function, Biology and Biomedical Implications*; Plenum Press: New York, 1995; p 241.
(73) Mutlu, A.; Gal, S. *Physiol. Plant.* **1999**, *105*, 569.
(74) Runeberg-Roos, P.; Saarma, M.; *Plant J.* **1995**, *15*, 139.
(75) Runeberg-Roos, P.; Kervinen, J.; Kovaleva, V.; Raikhel, N. V.; Gal, S. *Plant Physiol.* **1994**, *105*, 321.
(76) Verissimo, P.; Faro, C.; Moir, A. J. G.; Lin, Y.; Tang, J.; Pires, E. *Eur. J. Biochem.* **1996**, *235*, 762.
(77) Ramalho-Santos, M.; Verissimo, P.; Faro, C.; Pires, E. *Biochim. Biophys. Acta* **1996**, *1297*, 83.
(78) White, P. C.; Cordeiro, M. C.; Arnold, D.; Brodelius, P. E. Kay, J. *J. Biol. Chem.* **1999**, *274*, 16685.
(79) Faro, C.; Ramalho-Santos, M.; Vieira, M.; Mendes, A.; Simoes, I.; Andrade, R.; Verissimo, P.; Lin, X.-L.; Tang, J.; Pires, E. *J. Biol. Chem.* **1999**, *274*, 28724.

(80) D'Hondt, K.; Stack, S.; Gutteridge, S.; Van dekerckhove, J.; Krebbers, E.; and Gal, S. *Plant Mol. Biol.* **1997**, *33*, 187.
(81) Mutlu, A.; Chen, X.; Reddy, S. M.; Gal, S. *Seed Sci. Res.* **1999**, *9*, 75.
(82) Domingos, A.; Cardoso, P. C.; Xue, Z.-T.; Clemente, A.; Brodelius, P. E.; Pais, M. S. *Eur. J. Biochem.* **2000**, *267*, 6824.
(83) Cruz de Caravalho, M. H.; d'Arcy-Lameta, A.; Roy-Macauley, H.; Gareil, M.; El Maarouf, H.; Pham-Thi, A.-T.; Zuily-Fodil, Y. *FEBS Lett.* **2001**, *492*, 242.
(84) Asakura, T.; Watanabe, H.; Abe, K.; Arai, S. *Eur. J. Biochem.* **1995**, *232*, 77.
(85) Asakura, T.; Watanabe, H.; Abe, K.; Arai, S. *J. Agric. Food Chem.* **1997**, *45*, 1070.
(86) Asakura, T.; Matsumoto, I.; Funaki, J.; Arai, S.; Abe, K. *Eur. J. Biochem.* **2000**, *267*, 5115.
(87) Payie, K. G.; Weadge, J. T.; Tanaka, T.; Yada, R. Y. *Biotechnol. Lett.* **2000**, *22*, 1515.
(88) Ramalho-Santos, M.; Verissimio, P.; Cortes, L.; Samyn, B.; van Beeumen, J.; Pires, E.; Faro, C. *Eur. J. Biochem.* **1998**, *255*, 133.
(89) Egas, C.; Lavoura, N.; Resende, R.; Brito, R. M. M.; Pires, E.; de Lima, M. C. P.; Faro, C. *J. Biol. Chem.* **2000**, *275*, 38190.
(90) Sali, A.; Veerapandian, B.; Cooper, J. B.; Moss, D. S.; Hofmann, T.; Bundell, T. L. *Proteins: Struct. Funct. Genet.* **1992**, *12*, 158.
(91) Sielecki, A. R.; Fujinaga, M.; Read, R. J.; James, M. N. G. *J. Mol. Biol.* **1991**, *219*, 671.
(92) Hartsuck, J. A.; Koelsch, G.; Remington, S. J. *Proteins: Struct. Funct. Genet.* **1992**, *13*, 1.
(93) Kageyama, T. *Cell. Mol. Life Sci.* **2002**, *59*, 288.
(94) Abad-Zapatero, C.; Rydel, T. J.; Erickson, J. *Proteins: Struct. Funct. Genet.* **1990**, *8*, 62.
(95) Sielecki, A. R.; Fedorov, A. A.; Boodhoo, A.; Andreeva, N. S.; James, M. N. G. *J. Mol. Biol.* **1990**, *214*, 143.
(96) Cooper, J. B.; Khan, G.; Taylor, G.; Tickle, I. J.; Blundell, T. L. *J. Mol. Biol.* **1990**, *214*, 199.
(97) Koelsch, G.; Mares, M.; Metcalf, P.; Fusek, M. *FEBS Lett.* **1994**, *343*, 6.
(98) Khazanovich-Bernstein, N.; Cherney, M. M.; Loetscher, H.; Ridley, R. G.; James, M. N. G. *Nat. Struct. Biol.* **1999**, *6*, 32.
(99) Shi, Y.-P.; Chen, E.; Yin, K.-C.; Na, S.; Garsky, V. M.; Lai, M.-T.; Li, Y.-M.; Platchek, M.; Register, R. B.; Sardana, M. K.; Tang, M.-J.; Thiebeau, J.; Wood, T.; Schafer, J. A.; Gardell, S. J. *J. Biol. Chem.* **2001**, *276*, 10356.
(100) Moore, S. A.; Sielecki, A. R.; Chernaia, M. M.; Tarasova, N. I.; James, M. N. G. *J. Mol. Biol.* **1995**, *247*, 466.
(101) Khan, A. R.; Cherney, M. M.; Tarasova, N. I.; James, M. N. G. *Nat. Struct. Biol.* **1997**, *4*, 1010.
(102) Dunn, B. M. *Nat. Struct. Biol.* **1997**, *4*, 969.
(103) Lee, A. Y.; Gulnik, S. V.; Erickson, J. W. *Nat. Struct. Biol.* **1998**, *5*, 866.
(104) Metcalf, P.; Fusek, M. *EMBO J.* **1993**, *12*, 1293.
(105) Baldwin, E. T.; Narayana Bhat, T.; Gulnik, S.; Hosur, M. V.; Sowder, R. C. III; Cachau, R. E.; Collins, J.; Silva, A. M.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, 6796.
(106) Kornfeld, S.; Mellman, I. *Annu. Rev. Cell Biol.* **1989**, *5*, 483.
(107) Suguna, K.; Padlan, E. A.; Bott, R.; Boger, J.; Parris, K. D.; Davies, D. R. *Proteins: Struct. Funct. Genet.* **1992**, *13*, 195.
(108) Canduri, F.; Teodoro, L. G. V. L.; Fadel, V.; Lorenzi, C. C. B.; Hial, V.; Gomes, R. A. S.; Neto, J. R.; de Avezado, W. F. *Acta Crystallogr. D*, **2001**, *57*, 1560.
(109) Frazao, C.; Bento, I.; Costa, J.; Soares, C. M.; Verissimo, P.; Faro, C.; Pires, E.; Cooper, J.; Carrondo, M. A. *J. Biol. Chem.* **1999**, *274*, 27694.
(110) Kervinen, J.; Tobin, G. J.; Costa, J.; Waugh, D. S.; Wlodawer, A.; Zdanov, A. *EMBO J.* **1999**, *18*, 3947.
(111) Groves, M. R.; Dhanaraj, V.; Badasso, M.; Nugent, P.; Pitts, J. E.; Hoover, D. J.; Blundell, T. L. *Protein Eng.* **1998**, *11*, 833.
(112) Gilliland, G. L.; Toner Olivia, M.; Dill, J. In *Structure and Function of the Aspartic Peptidases*; Dunn, B. M., Ed.; Plenum Press: New York, 1991; p 23.
(113) Newman, M.; Safro, M.; Frazao, C.; Khan, G.; Zdohav, G.; Tickle, I. J.; Blundell, T. L.; Andreeva, N. *J. Mol. Biol.* **1991**, *221*, 1295.
(114) Andreeva, N.; Dill, J.; Gilliland, G. L. *Biochem. Biophys. Res. Commun.* **1992**, *184*, 1074.
(115) Silva, A. M.; Lee, A. Y.; Gulnik, S. V.; Majer, 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.
(116) Khan, A. R.; Parrish, J. C.; Fraser, M. E.; Smith, W. W.; Bartlett, P. A.; *James*, M. N. G. *Biochemistry* **1998**, *37*, 16839.
(117) Fujinaga, M.; Chernaia, M. M.; Tarasova, N. I.; Mosimann, S. C.; James, M. N. G. *Protein Sci.* **1995**, *4*, 960.
(118) Foundling, S. I.; Cooper, J.; Watson, F. E.; Cleasby, A.; Pearl, L. H.; Sibanda, B. L.; Hemmings, A.; Wood, S. P.; Blundell, T. L.; Valler, M. J.; Norey, C. G.; Kay, J.; Boger, J.; Dunn, B. M.; Leckie, B. J.; Jones, D. M.; Atrash, B.; Hallet, A.; Szelke, M. *Nature* **1987**, *327*, 349.
(119) Monod, M.; Togni, G.; Hube, B.; Sanglard, D. *Mol. Microbiol.* **1994**, *13*, 357.
(120) Hube, B. *Curr. Top. Med. Mycol.* **1996**, *7*, 55.
(121) Shaller, M.; Korting, H. C.; Schafer, W.; Bastert, J.; Chen, W.; Hube, B. *Mol. Microbiol.* **1999**, *34*, 169.
(122) De Bernardis, F.; Mondello, F.; Scaravelli, G.; Pachi, A.; Giorlamo, A.; Agatensi, L.; Cassone, A. *J. Clin. Microbiol.* **1999**, *5*, 1376.
(123) Hube, B.; Sanglard, D.; Odds, F. C.; Hess, D.; Monod, M.; Schafer, W.; Brown, A. J. P.; Gow, N. A. R. *Infect. Immun.* **1997**, *65*, 3529.
(124) Abad-Zapatero, C.; Goldman, R.; Muchmore, S. W.; Hutchins, C.; Stewart, K.; Navaza, J.; Payne, C. D.; Ray, T. L. *Protein Sci.* **1996**, *5*, 640.
(125) Cutfield, S. M.; Dodson, E. J.; Anderson, B. F.; Moody, P. C. E.; Marshall, C. J.; Sullivan, P. A.; Cutfield, J. F. *Structure* **1995**, *3*, 1261.
(126) Fusek, M.; Smith, E. A.; Monod, M.; Dunn, B. M.; Foundling, S. I. *Biochemistry* **1994**, *33*, 9791.
(127) Symersky, J.; Monod, M.; Foundling, S. I. *Biochemistry* **1997**, *36*, 12700.
(128) Koelsch, G.; Tang, J.; Loy, J. A.; Monod, M.; Jackson, K.; Foundling, S. I.; Lin, X. *Biochim. Biophys. Acta* **2000**, *1480*, 117.
(129) Zaugg, C.; Borg-van-Zepelin, M.; Reichard, U.; Sanglard, D.; Monod, M. *Infect. Immun.* **2001**, *69*, 405.
(130) Pichova, I.; Pavlickova, L.; Dostal, J. Dolesjse, Hruskova-Heidlingsfeldova, O.; Weber, J.; Rumi, T.; Soucek, M. *Eur. J. Biochem.* **2001**, *268*, 2669.
(131) Khan, A. R.; Khazanovich-Bernstein, N.; Bergmann, E. M.; James, M. N. G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10968.
(132) Tanaka, T.; Yada, R. Y. *Protein Eng.* **2001**, *14*, 669.
(133) Bohak, Z. *J. Biol. Chem.* **1969**, *244*, 4638.
(134) van den Hazel, H. B.; Kielland-Brandt, M. C.; Winther, J. R. *J. Biol. Chem.* **1993**, *268*, 18002.
(135) van den Hazel, H. G.; Kielland-Brandt, M. C.; Winther, J. R. *J. Biol. Chem.* **1995**, *270*, 8602.
(136) Richter, C.; Tanaka, T.; Yada, R. Y. *Biochem. J.* **1998**, *335*, 481.
(137) Fukuda, R.; Umebayashi, K.; Horiuchi, H.; Ohta, A.; Takagi, M. *J. Biol. Chem.* **1996**, *271*, 14252.
(138) Richter, C.; Tanaka, T.; Koseki, T.; Yada, R. Y. *Eur. J. Biochem.* **1999**, *261*, 746.
(139) Fortenberry, S. C.; Chirgwin, J. M. *J. Biol. Chem.* **1995**, *270*, 9778.
(140) Harrison, T. M.; Chidgey, M. A. J.; Brammar, W. J.; Adams, G. J. *Proteins: Struct. Funct. Genet.* **1989**, *5*, 259.
(141) Baxter, J. D.; Duncan, K.; Chu, W.; James, M. N. G.; Russell, R. B.; Haidar, M. A.; Denoto, F. M.; Hsueh, W.; Reudelhuber, T. L. *Recent Prog. Horm. Res.* **1991**, *47*, 211.
(142) Norman, J. A.; Hadjilambris, O.; Baska, R.; Sharp, D. Y.; Kumar, R. *Mol. Pharmacol.* **1992**, *41*, 53.
(143) Rothwell, V.; Kosowski, S.; Hadjilambris, O.; Baska, R.; Norman, J. *DNA Cell Biol.* **1993**, *12*, 291.
(144) Beyer, B. M.; Dunn, B. M. *J. Biol. Chem.* **1996**, *271*, 15590.
(145) Azim, M. K.; Zaidt, Z. H. *Biochem. Biophys. Res. Commun.* **1999**, *264*, 825.
(146) Francky, A.; Francky, B. M.; Strukelj, B.; Gruden, K.; Ritonja, A.; Krizaj, I.; Kregar, I.; Pain, R. H.; Pungercar, J. *Eur. J. Biochem.* **2001**, *268*, 2362.
(147) Dunn, B. M.; Hung, S.-H. *Biochim. Biophys. Acta* **2000**, *1477*, 231.
(148) Abita, J. P.; Delaage, M.; Lazdunski, M.; Savrda, J. *Eur. J. Biochem.* **1969**, *8*, 314.
(149) James, M. N. G.; Sielecki, A. R.; and Hofmann, T. In *Aspartic Peptidases and Their Inhibitors*; Kostka, V., Ed.; Walter de Gruyter: Berlin, 1985; p 163.
(150) Lowther, W. T.; Majer, P.; Dunn, B. M. *Protein Sci.* **1995**, *4*, 689.
(151) Shintani, T.; Kobayashi, M.; Ichishima, E. *J. Biochem.* **1996,** *120*, 974.
(152) Shintani, T.; Nomura, K.; Ichishima, E. *J. Biol. Chem.* **1997**, *272*, 18855.
(153) Okoniewska, M.; Tanaka, T.; Yada, R. Y. *Biochem. J.* **2000**, *349*, 169.
(154) Okoniewska, M.; Tanaka, T.; Yada, R. Y. *Protein Eng.* **1999**, *12*, 55.
(155) Andreeva, N.; Bochkarev, A.; Pechik, I. In *Aspartic Peptidases*; Takahashi K., Ed.; Plenum Press: New York, 1995; p 19.
(156) Tanaka, T.; Teo, K. S.; Lamb, K. M.; Harris, L. J.; Yada, R. Y. *Protein Pept. Lett.* **1998**, *5*, 19.
(157) Suzuki, J.; Sasaki, K.; Saso, Y.; Hamu, A.; Kawasaki, H.; Nishiyama, H.; Horinouchi, S.; Beppu, T. *Protein Eng.* **1989**, 563.
(158) Beppu, T.; Park, Y.-N.; Aikawa, J.-I.; Nishiyama, M.; Horinouchi, S. In *The Aspartic Peptidases: Structure, Function, Biology, and Biomedical Implications*; Takahashi, K., Ed.; Plenum Press: New York, 1995; p 501.
(159) Park, Y.-N.; Aikawa, J.-I.; Nishiyama, M.; Horinouchi, S.; Beppu, T. *Protein Eng.* **1996**, *9*, 869.
(160) Tang, J.; Koelsch, G. *Protein Pept. Lett.* **1995**, *2*, 257.
(161) Gilliland, G. L.; Winborne, E. L.; Nachman, J.; Wlodawer, A. *Proteins: Struct. Funct. Genet.* **1990**, *8*, 82.
(162) Scarborough, P. E.; Dunn, B. M. *Protein Eng.* **1994**, *7*, 495.
(163) Scarborough, P. E.; Guruprasad, K.; Topham, C.; Richo, G. R.; Conner, G. E.; Blundell, T. L.; Dunn, B. M. *Protein Sci.* **1993**, *2*, 264.

(164) Beyer, B. M.; Dunn, B. M. *Protein Sci.* **1998**, *7*, 88.
(165) Westling, J.; Yowell, C. A.; Majer, P.; Erickson, J. W.; Dame, J. B.; Dunn, B. M. *Exp. Parasitol.* **1997**, *87*, 185.
(166) Cao, Q.-N.; Stubbs, M.; Ngo, K. Q. P.; Ward, M.; Cunningham, A.; Pai, E. F.; Tu, G.-C.; Hofmann, T. *Protein Sci.* **2000**, *9*, 991.
(167) Hofmann, T.; Allen, B.; Bendiner, M.; Blum, M.; Cunningham, A. *Biochemistry* **1988**, *27*, 1140.
(168) Balbaa, M.; Cunningham, A.; Hofmann, T. *Arch. Biochem. Biophys.* **1993**, *306*, 297.
(169) Williams, M. G.; Wilsher, J.; Nugent, P.; Mills, A.; Dhanaraj, V.; Fabry, M.; Sedlacek, J.; Uusitalo, J. M.; Penttila, M. E.; Pitts, J. E.; Blundell, T. L. *Protein Eng.* **1997**, *9*, 991.
(170) Bhatt, D.; Dunn, B. M. *Bioorg. Chem.* **2000**, *28*, 374.
(171) Cottrell, T. J.; Harris, L. J.; Tanaka, T.; Yada, R. Y. *J. Biol. Chem.* **1995**, *270*, 19974.
(172) Gulnik, S. V.; Suvorov, L. I.; Majer, P.; Collins, J.; Kane, B. P.; Johnson, D. G.; Erickson, J. W. *FEBS Lett.* **1997**, *413*, 379.
(173) Pimenta, D. C.; Oliveira, A.; Juliano, M. A.; Juliano, L. *Biochim. Biophys. Acta* **2001**, *1544*, 113.
(174) Gulnik, S. V.; Suvorov, L. I.; Majer, P.; Erickson, J. W. *Protein Pept. Lett.* **2000**, *7*, 219.
(175) Turner, R. T., III; Koelsch, G.; Hong, L.; Castenheira, P.; Ghosh, A.; Tang, J. *Biochemistry* **2001**, *40*, 12230.
(176) Ghosh, A. K.; Bilcer, G.; Harwood: C.; Kawahama, R.; Shin, D.; Hussain, K. A.; Hong, L.; Loy, J. A.; Nguyen, C.; Koelsch, G.; Ermolieff, J.; Tang, J. *J. Med. Chem.* **2001**, *44*, 2865.
(177) Gustchina, E.; Rumsh, L.; Ginodman, L.; Majer, P.; Andreeva, N. *FEBS Lett.* **1996**, *379*, 60.
(178) Visser, S.; Slangen, C. J. van Rooijen, P. J. *Biochem. J.* **1987**, *244*, 553.
(179) Kuzmic, P.; Sun, C. Q.; Zhao, Z. C.; Rich, D. H. *Tetrahedron* **1991**, *47*, 2519.
(180) Bursavich, M. G.; Rich, D. G. *J. Med. Chem.* **2002**, *45*, 541.
(181) Cronin, N. B.; Badasso, M. O.; Tickle, I. J.; Dreyer, T.; Hoover, D. J.; Rosati, R. L.; Humblet, C. C.; Lunney, E. A.; Cooper, J. B. *J. Mol. Biol.* **2000**, *303*, 745.
(182) Aguilar, C. F.; Cronin, N. B.; Badasso, M.; Dreyer, T.; Newman, M. P.; Cooper, J. B.; Hoover, D. J.; Wood, S. P.; Johnson, M. S.; Blundell, T. L. *J. Mol. Biol.* **1997**, *267*, 899.
(183) Bailey, D.; and Cooper, J. B. *Protein Sci.* **1994**, *3*, 2129.
(184) Fraser, M. E.; Strynadka, N. C. J.; Bartlett, P. A.; Hanson, J. E.; James, M. N. G. *Biochemistry* **1992**, *31*, 5201.
(185) Majer, P.; Collins, J. R.; Gulnik, S. V.; Erickson, J. W. *Protein Sci.* **1997**, *6*, 1458.
(186) Ripka, A. S.; Satyshur, K. A.; Bohacek, R. S.; Rich, D. H. *Org. Lett.* **2001**, *3*, 2309.
(187) Dales, N. A.; Bohancek, R.; Satyshur, K. A.; Rich, D. H. *Org. Lett.* **2001**, *3*, 2313.
(188) Bursasvich, M. G.; West, C. W.; Rich, D. H. *Org. Lett.* **2001**, *3*, 2317.
(189) Oefner, C.; Binggeli, A.; Breu, V.; Bur, D.; Clozel, J. P.; D'arcy, A.; Dorn, A.; Fischli, W.; Gruninger, F.; Guller, R.; Hirth, G.; Marki, H.; Mathews, S.; Miller, M.; Ridley, R. G.; Stadler, H.; Viera, E.; Wilhelm, M.; Winkler, F.; Wostl, W. *Chem. Biol.* **1999**, *6*, 127.
(190) Jiang, S.; Prigge, S. T.; Wei. L.; Gao, Y.-E.; Hudson, T. H.; Gerena, L.; Dame, J. B.; Kyle, D. E. *Antimicrob. Agents Chemother.* **2001**, *45*, 2577.
(191) Nezami, A.; Luque, I.; Kimura, T.; Kiso, Y.; Freire, E. *Biochemistry* **2002**, *41*, 2273.
(192) Haque, T. S.; Skillman, A. G.; Lee, C. E.; Habashita, H.; Gluzman, I. Y.; Ewing, T. J.; Goldberg, D. E.; Kuntz, I. D.; Ellman, J. A. *J. Med. Chem.* **1999**, *42*, 1428.
(193) Caroll, C. D.; Johnson, T. O.; Tao, S.; Lauri, G.; Orlowski, M.; Gluzman, I. Y.; Goldberg, D. E.; Dolle, R. E. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3203.
(194) Caroll, C. D.; Patel, H.; Johnson, T. O.; Guo, T.; Orlowski, M.; He, Z. M.; Cavallaro, C. L.; Guo, J.; Oksman, A.; Gluzman, I. Y.; Connelly, J.; Chelsky, D.; Goldberg, D. E.; Dolle, R. E. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2315.
(195) Saheki, T.; Matsuda, Y.; Holzer, H. *Eur. J. Biochem.* **1074**, *47*, 325.
(196) Dreyer, T.; Valler, M. J.; Kay, J.; Charlton, P.; Dunn, B. M. *Biochem. J.* **1985**, *231*, 777.
(197) Phylip, L. H.; Lees, W. E.; Brownsey, B. G.; Bur, D.; Dunn, B. M.; Winther, J. R.; Gustchina, A.; Li, M.; Copeland, T.; Wlodawer, A.; Kay, J. *J. Biol. Chem.* **2001**, *276*, 2023.
(198) Abu-Erreish, G. M.; Peanasky, R. J. *J. Biol. Chem.* **1974**, *249*, 1558.
(199) Martzen, M. R.; McMullen, B. A.; Smith, N. E.; Fujikawa, K.; Peanasky, R. J. *Biochemistry* **1990**, *29*, 7366.
(200) Kageyama, T. *Eur. J. Biochem.* **1998**, *253*, 804.
(201) Abu-Erreish, G. M.; Peanasky, R. J. *J. Biol. Chem.* **1974**, *249*, 1566.
(202) Valler, M. J.; Kay, J.; Aoyagi, T.; Dunn, B. M. *J. Enzymol. Inhib.* **1985**, 77.
(203) Ng, K. K. S.; Petersen, J. F. W.; Cherney, M. M.; Garen, C.; Zalatoris, J. J.; Rao-Naik, C.; Dunn, B. M.; Martzen, M. R.; Peansaky, R. J.; James, M. N. G. *Nat. Struct. Biol.* **2000**, *7*, 653.
(204) Lenarcic, B.; Ritonja, A.; Strukelj, B.; Turk, B.; Turk, V. *J. Biol. Chem.* **1997**, *272*, 13899.
(205) Lenarcic, B.; Turk, V. *J. Biol. Chem.* **1999**, *274*, 563.
(206) Strukelj, B.; Lenarcic, B.; Gruden, K.; Pungercar, J.; Rogelj, B.; Turk, V.; Borsch, D.; Jongsma, M. A. *Biochem. Biophys. Res. Commun.* **2000**, *269*, 732.
(207) Christeller, J. T.; Farley, P. C.; Ramsay, R. J.; Sullivan, P. A.; Laing, W. A. *Eur. J. Biochem.* **1998**, *254*, 160.
(208) Dash, C.; Ahmad, A.; Nath, D.; Rao, M. *Antimicrob. Agents Chemother.* **2001**, *45*, 2008.
(209) Dash, C.; Phadtare, S.; Deshpande, V.; Rao, M. *Biochemistry* **2001**, *40*, 11525.
(210) Bott, R.; Subramanian, E.; Davies, D. *Biochemistry* **1982**, *21*, 6956.
(211) James, M. N. G.; Sielecki, A. R.; Salituro, F.; Rich, D. H.; Hofmann, T. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 6137.
(212) Rich, D. H.; Bernatowicz, M. S.; Agarwal, N. S.; Kawai, M.; Salituro, F. G.; Schmidt, P. G. *Biochemistry* **1985**, *24*, 3165.
(213) Gelb, M. H.; Svaren, J. P.; Abeles, R. H. *Biochemistry* **1985**, *24*, 1813.
(214) Thairivongs, S.; Pals, D. T.; Kati, W. M.; Turner, S. R.; Thomasco, L. M.; Watt, W. *J. Med. Chem.* **1985**, *28*, 1553.
(215) Thairivongs, S.; Pals, D. T.; Kati, W. M.; Turner, S. R.; Thomasco, L. M.; Watt, W. *J. Med. Chem.* **1986**, *29*, 2080.
(216) Fearon, K.; Spaltenstein, A.; Hopkins, P. B.; Gelb, M. H. *J. Med. Chem.* **1987**, *30*, 1617.
(217) Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, T. L.; Rosati, R. L.; Dominy, B. W.; Damon, D. B.; Hoover, D. J. *Protein Sci.* **1992**, *1*, 322.
(218) James, M. N. G.; Sielecki, A. R.; Hayakawa, K.; Gelb, M. H. *Biochem.* **1992**, *31*, 3872.
(219) Coates, L.; Erskine, P. T.; Wood, S. P.; Myles, D. A. A.; Cooper, J. B. *Biochem.* **2001**, *40*, 13159.
(220) Northrop, D. *Acc. Chem. Res.* **2001**, *34*, 790.
(221) Piana, S.; Carloni, P. *Proteins: Struct. Funct. Genet.* **2001**, *39*, 26.
(222) Hunkpillar, M. W.; Richards, J. H. *Biochemistry* **1972**, *11*, 2829.
(223) Meek, T. D.; Rodriguez, E. J.; Angeles, T. S. *Methods Enzymol.* **1994**, *241*, 127.
(224) Polgar, L.; Szeltner, Z.; Boros, I. *Biochemistry* **1994**, *33*, 9351.
(225) Porter, D. J. T.; Hanlon, M. H.; Furfine, E. S. *Biochemistry* **2002**, *41*, 1302.
(226) Andreeva, N. S.; Rumsh, L. D. *Protein Sci.* **2001**, *10*, 2439.
(227) Marcinkeviciene, J.; Kopcho, L. M.; Yang, T.; Copeland, R. A.; Glass, B. M.; Combs, A. P.; Falahatpisheh, N.; Thompson, L. *J. Biol. Chem.* **2002**, *32*, 28677.
(228) Ullrich, B.; Laberge, M.; Tolgyesi, F.; Szeltner, Z.; Polgar, L.; Fidy, J. *Protein Sci.* **2000**, *9*, 2232.
(229) Furfine, E. S.; D'Souza, E.; Ingold, K. J.; Leban, J. J.; Spector, T.; Porter, D. J. T. *Biochemistry* **1992**, *31*, 7886.

CR010167Q