Part One
Overview of Aspartic Acid Proteases

Aspartic Acid Proteases as Therapeutic Targets. Edited by Arun K. Ghosh

ISBN: 978-3-527-31811-7

1
Introduction to the Aspartic Proteinase Family

Ben M. Dunn

Abbreviations

BACE	β-amyloid cleaving enzyme
C-terminal	carboxyl terminal
D-G-I-L-G-L	amino acid sequence using single-letter code
HIV	human immunodeficiency virus
Nle	norleucine
Nph	*para*-nitrophenylalanine
N-terminal	amino terminal
PDB	Protein Data Bank
PfPM2	*Plasmodium falciparum* plasmepsin 2

Symbols

k_{cat}/K_m	characterizes the kinetics of cleavage of an enzyme–substrate pair
K_{cat}	turnover number of an enzyme
K_m	Michaelis–Menten constant

1.1
Introduction

All cells, tissues, and organisms require proteolysis for the control of metabolism and growth. Even a virus, the smallest nucleic acid-based self-replicating organism, typically requires either host cell proteolysis or enzymes coded by its own genetic material to provide processing of initial viral gene products. Proteolysis is required to activate prohormones and other precursor molecules, to invade into cells and tissues, to release membrane-bound molecules, to provide a cascade effect in a variety of rapid response systems, to regulate cell growth, tissue homeostasis, remodeling, and

Aspartic Acid Proteases as Therapeutic Targets. Edited by Arun K. Ghosh

ISBN: 978-3-527-31811-7

renewal, and to stimulate cell division, to name but a few vital functions. This applies to pathogenic organisms as well as to normal cells and tissues.

Thus, it is not surprising that proteolytic enzymes from pathogens are targets for drug discovery or that creating molecules that will selectively block the activity of an enzyme from a pathogen, while not harming normal cellular function, is an ongoing endeavor in many pharmaceutical companies and academic labs around the world.

Classically, proteolytic enzymes have been divided into four groups based on their catalytic apparatus: aspartic, cysteine, metallo-, and serine proteases. However, recently, three new systems have been defined: the threonine-based proteosome system [1], the glutamate–glutamine system of eqolisin [2], and the serine–glutamate–aspartate system of sedolisin [3]. It remains to be seen if the proteases mentioned here represent the total spectrum of proteolytic mechanisms available in biology.

This book focuses on the aspartic proteinase family of proteolytic enzymes and specifically on several enzymes that are currently being pursued as drug targets for therapeutic intervention in humans. This chapter will describe several general features of these enzymes to provide a context in which to compare the specific examples found in the following chapters.

The aspartic proteinase family has a long and complicated history. In fact, one of the first processing involving enzyme action was discovered by accident. Legend has it that around 7000 BC, an Arabic traveler placed milk in a pouch made from the stomach of an animal, possibly a sheep or a cow, and set off on a journey across a desert. Different versions of the legend suggest that the traveler was either male or female. Upon reaching the destination, the traveler opened the pouch to find that the milk had coagulated and separated into curd and whey. Sampling the curd, the traveler decided that the process had been productive and it became possible to capture the essential nutrients of milk in a more concentrated and thus easier to carry form. Many centuries later, it was discovered that the enzyme renin, which is contained in the cells lining the stomach cavity of many ruminant animals, was responsible for the cleavage of kappa-casein proteins to cause precipitation that leads to the curd.

The modern history of the aspartic proteinase family can be noted by the publication of the amino acid sequence of porcine pepsin by Jordan Tang and his colleagues at the Oklahoma Medical Research Foundation [4]. This sequence determination was done totally by classical protein chemistry methods, which resulted in information that allowed the necessary primer design to enable the cloning and DNA sequencing of many other members of the family in recent years.

1.2
Sequence Alignment and Family Tree

The MEROPS database (http://merops.sanger.ac.uk/index.htm) contains a full listing of all sequences related to the aspartic proteinase family [5]. Specifically, the A1A family of pepsin-like enzymes contains 1167 entries as of 2009, most of which

have been derived from genome sequencing efforts over the past 10 years and some not even assigned a name yet. The A1B family of plant aspartic proteases contains another 406 entries, with each having a "plant-specific" insert of 55 amino acids. The A1B family will not be considered further, although the overall structure of these is very similar to the A1A family.

In addition to the A1A family, the "retropepsin" or A2A family contains another 228 sequences. This represents the retroviral enzymes related to HIV-1 protease as the archetypal member.

Due to the large number of sequences available, it is not feasible to show a sequence alignment or family tree for *all* the members of the pepsin and retropepsin families. Rather, selected members of the pepsin-like A1A family that are discussed in this book are aligned in Figure 1.1.

1.3 Three-Dimensional Structure

The second watershed event in the history of the aspartic proteinase family was the determination, nearly simultaneously, of the three-dimensional structures of three enzymes from fungi (James [6], Blundell [7], and Davies [8]) and that of porcine pepsin (Andreeva [9]). The structures of the fungal enzyme provided critical information that allowed the correct interpretation of the crystallographic data for porcine pepsin, yielding the structure solution. Remarkably, the overall structures of these enzymes were very similar, although important differences were immediately noted. In particular, by superimposing the N-terminal half of each enzyme, it was possible to see that the C-terminal halves all adopted significantly different orientations, while retaining the same internal architecture.

A figure showing the overall structure of the archetypical enzyme in this family, pepsin, is shown in Figure 1.2a, with several residues highlighted. This figure was prepared using PDB file 1QRP [10], a pepsin structure that contains an inhibitor molecule bound in the active site. This is seen in blue sticks in the top middle of the figure and this identifies the active site. Below the inhibitor can be seen the two catalytic aspartic acids in stick representation, which represent two of the regions of identity, one in the N-terminal domain of the enzyme and one in the C-terminal domain. In addition, in each domain, a strand of sequence (D-G-I-L-G-L in the N-terminal domain and I-L-G-D-V-F-I in the C-terminal domain) passes through a wide loop containing the catalytic domain. This is shown in detail in Figure 1.3 for the N-terminal domain. A conserved glycine must appear in the position where the strand passes through the loop due to steric constraints. Any larger residue would not fit in this crowded location. Figure 1.2b shows the structure of the *Candida albicans* homologue 1EAG [11] and this illustrates the overall secondary structure of the typical aspartic proteinase. Considerable segments of beta structure can be seen, with two orthogonally packed beta sheets in both the N- and C-terminal domains. In addition, at the "bottom" of the enzyme, a six-stranded beta sheet completes the structure. While there is some variation in the size of the beta strands/sheets, this basic

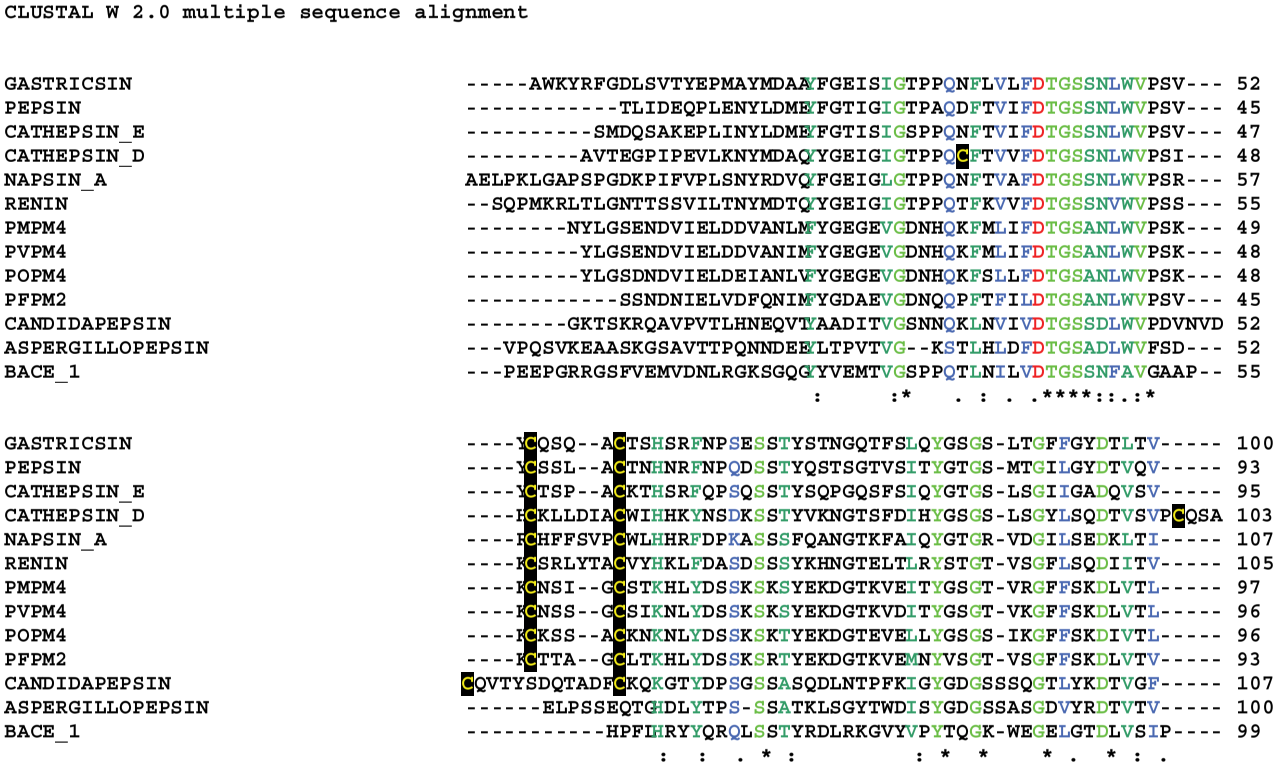

Figure 1.1 Alignment of amino acid sequences of aspartic proteases discussed in this book. The sequences were truncated at both the amino and carboxyl terminal ends to show the main catalytically active component of each protein. Thus, some membrane spanning segments at either end were deleted for this alignment. The color code is as follows: red font shows the two catalytic aspartic acid residues, one in each of the two structurally similar domains; light green font shows identical amino acids in all the sequences aligned here, also indicated below the alignment by an "*" symbol; sea green shows amino acids that are nearly identical in all sequences with very similar properties in all enzymes, also indicated by a ":" symbol below the alignment; and light blue shows amino acids that are similar in all sequences with similar properties, indicated below the alignment by a "." symbol. Although the cysteine residues are not completely conserved in this alignment, they are highlighted as C here due to their influence on the three-dimensional structure.

Figure 1.1 (*Continued*)

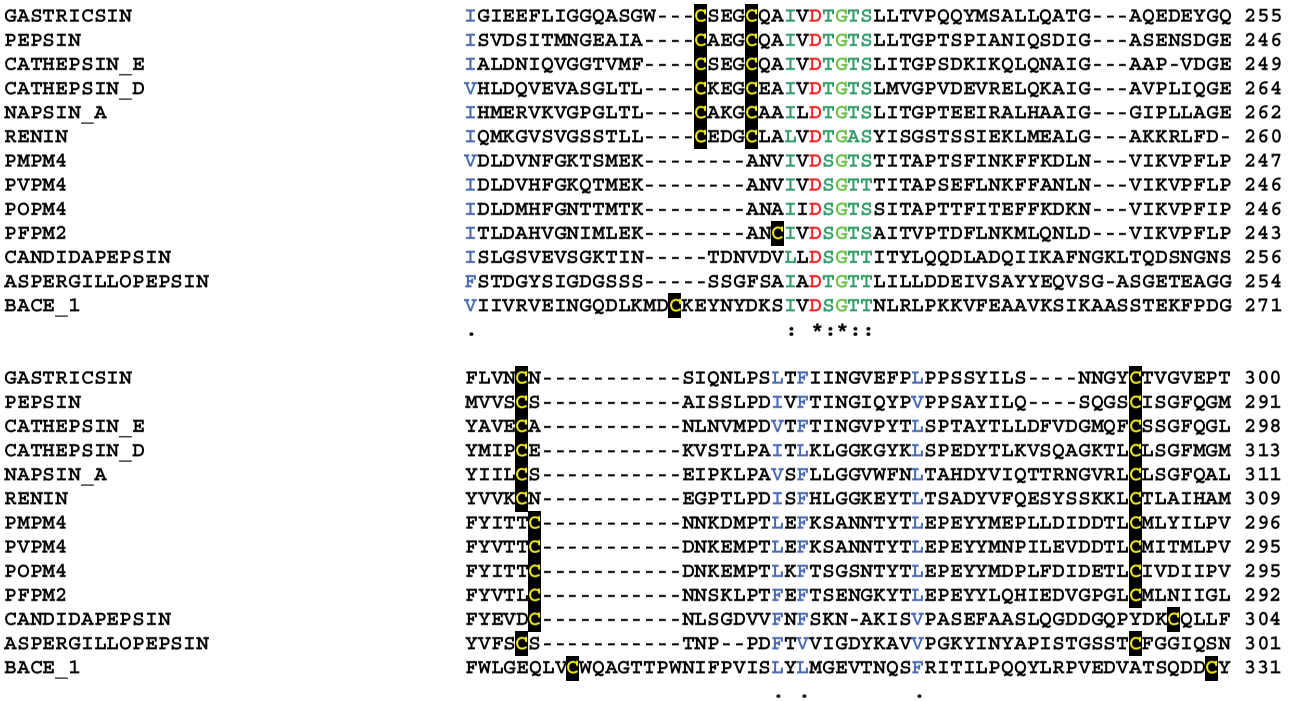

Figure 1.1 (*Continued*)

```
GASTRICSIN                YLSSQNGQPLWILGDVFLRSYYSVYDLG----NNRVGFATAA--- 338
PEPSIN                    DVPTESGE-LWILGDVFIRQYFTVFDRA----NNQVGLAPVA--- 328
CATHEPSIN_E               DIHPPAGP-LWILGDVFIRQFYSVFDRG----NNRVGLAPAVP-- 336
CATHEPSIN_D               DIPPPSGP-LWILGDVFIGRYYTVFDRD----NNRVGFAEAARL- 352
NAPSIN_A                  DVPPPAGP-FWILGDVFLGTYVAVFDRGDMKSSARVGLARARTR- 354
RENIN                     DIPPPTGP-TWALGATFIRKFYTEFDRR----NNRIGFALAR--- 346
PMPM4                     DID----KNTFILGDPFMRKYFTVFDYD----KESIGFAVAKN-- 331
PVPM4                     DID----SNTFILGDPFMRKYFTVFDYD----KESVGFAIAKN-- 330
POPM4                     DID----ENTFILGAPFMRKYFSVFDYD----NERVGFAVAKN-- 330
PFPM2                     DFP----VPTFILGDPFMRKYFTVFDYD----NHSVGIALAKKNL 329
CANDIDAPEPSIN             DVN-----DANILGDNFLRSAYIVYDLD----DNEISLAQVKYT- 339
ASPERGILLOPEPSIN          SGLG-----LSILGDVFLKSQYVVFNSE----GPKLGFAAQA--- 334
BACE_1                    KFAISQSSTGTVMGAVIMEGFYVVFDRA----RKR---------- 362
                                     :*  ::      ::
```

Figure 1.1 (*Continued*)

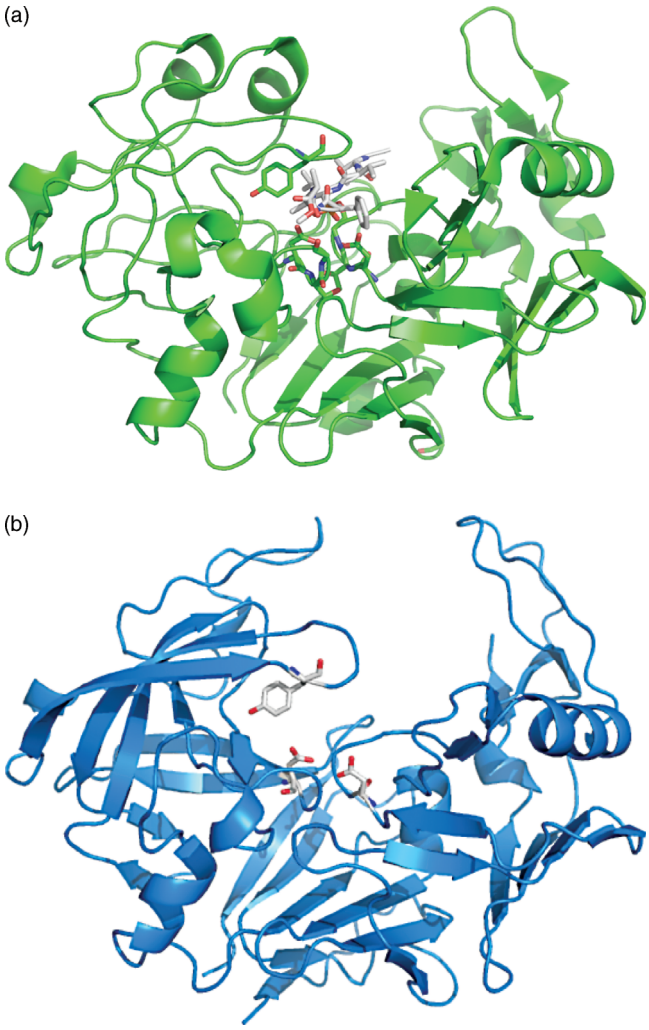

Figure 1.2 (a) Overall structure of pepsin generated by the Pymol program using PDB file 1QRP. A ribbon trace is used to show the backbone structure of the enzyme and a bound inhibitor is shown as sticks. Side chains of the enzyme are shown using stick representation with atoms colored according to the elemental identity (white = carbon; red = oxygen; blue = nitrogen). Only some conserved amino acids from Figure 1.1 are shown in this representation, as described in the text. (b) Figure showing the backbone structure of PDB file 1EAG for candidapepsin plus the two catalytic aspartic acids and Tyr75. This figure shows the orthogonal beta sheets in both the N-terminal domain, on the left-hand side, and the C-terminal domain, on the right-hand side. In addition, the six-stranded beta sheet that makes up the "bottom" of the protein is seen in the lower middle portion of the image. Three small sections of alpha helical structure are seen as well, one in the N-terminal domain and two in the C-terminal domain.

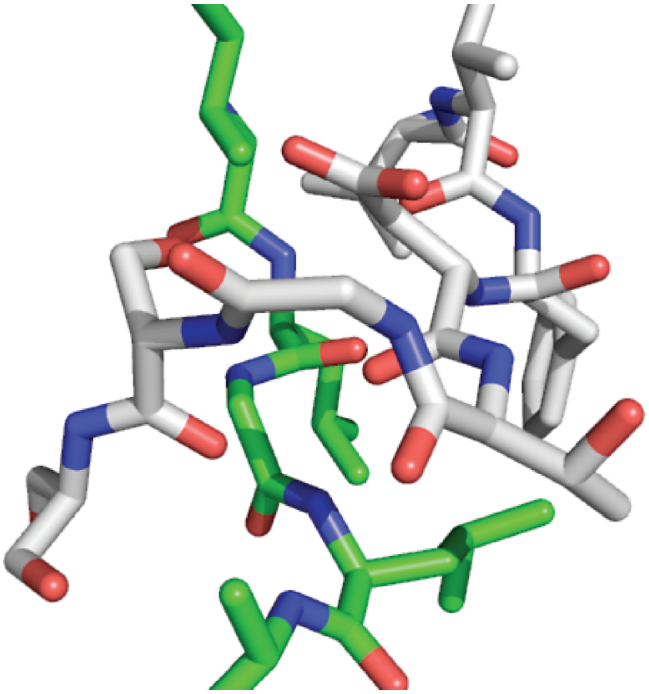

Figure 1.3 View of the N-terminal ψ-loop, where a strand of sequence (green = carbon atoms) passes through the wide loop (white = carbon atoms) that contains Asp32. A glycine must be conserved at the point where the strand passes through the loop for steric reasons.

structural arrangement is found in all the enzymes of the pepsin family, with a higher degree of variability found in the C-terminal domain.

An important conserved element of the structure is the "ψ-loop," one found in the N-terminal domain and one found in the C-terminal domain. In this loop, a strand of beta structure passes through a wide loop that contains the catalytic Asp residue. The two "ψ-loops" fix the central structure of the enzyme and thus define the aspartic proteinase catalytic machinery. An illustration is provided in Figure 1.3.

Two additional highly conserved residues are also highlighted in Figure 1.2a: first, Tyr75 of the pepsin sequence is shown. This residue influences substrate selectivity by interaction with amino acids in the P_1 and P_3 positions, according to the Schechter and Berger nomenclature [12]. In addition, Trp39, while not strictly conserved throughout the family (it is replaced by an Ala in BACE-1), plays an important role in stabilizing the internal core of the N-terminal domain and interacts with Tyr75 through a hydrogen bond in some structures, as shown in Figure 1.4.

In addition to the regions of identity or strong similarity, there are several places where gaps occur in the alignment shown in Figure 1.1. As can be seen in Figure 1.5, the points where gaps of more than two amino acids occur in comparison to the pepsin structure tend to occur on the surface of the proteins and are most likely

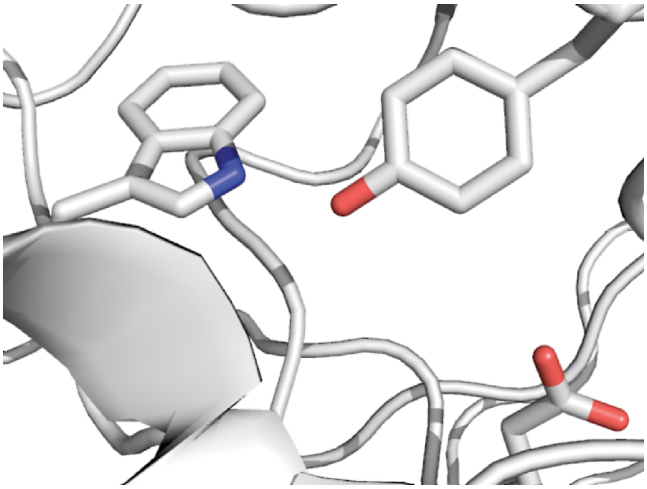

Figure 1.4 A view of Trp39 and Tyr75, showing that a hydrogen bond could form between the −OH of Tyr75 and the ring nitrogen of Trp39. For reference, Asp32 of the N-terminal domain is shown in the lower right-hand side of the figure.

insertions of sequences to expand or contract a loop to provide some specific property, such as interaction with a receptor or a binding partner. In terms of catalytic function, it is believed that these points do not influence either the rate of cleavage or the substrate specificity of the enzymes.

A superposition of several crystallographically determined structures is provided in Figure 1.6 to illustrate the similarities and also the differences in this family of

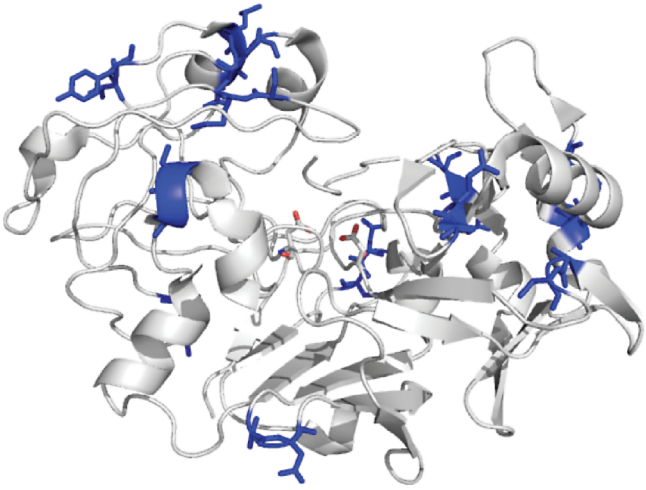

Figure 1.5 Pepsin structure with locations of significant (>2 amino acids) gaps for other enzymes in the alignment shown in Figure 1.1. All these insert locations occur on the surface of the pepsin molecule, indicating that the insertions will cause some expansion of a loop or turn.

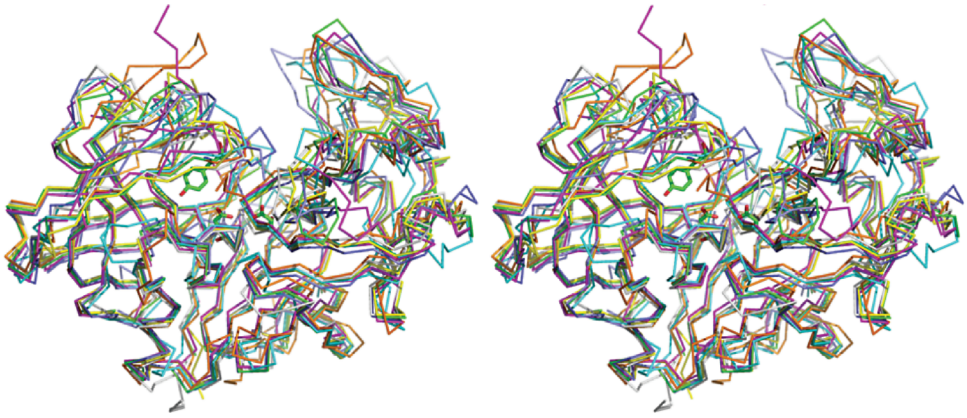

Figure 1.6 Superposition of several aspartic protease structures. This is a stereo view.

enzymes. While structures of all the proteins aligned in Figure 1.1 are not available at this time, enough are available to provide this comparison. It can be seen that the central region of all enzymes overlaps nicely, especially in the N-terminal domain. The regions of highest sequence and structural variability include several surface loops, again providing individual "character" to each member of the family.

Figure 1.7a and b provides some details of the interactions that occur within the active site of the aspartic proteases when peptide-like inhibitors bind. Specifically, Figure 1.7a shows the N-terminal half of the binding cleft of PfPM2. This is described in more detail in Section 1.5.

1.4 Conferences and Progress

An additional piece of history to recount here is to cite the series of books that have been produced following conferences devoted to the topic of the aspartic proteinase family. This series began with the conference organized in Norman, OK, by Jordan Tang in 1976 [13]. It was at this conference that the initial three-dimensional structures were described and it was noticed by Tang that the N- and C-terminal halves of the proteins seem to have repeating sequences [14]. This has been born out by the sequence analyses of multiple enzymes derived from DNA sequence analysis. The Norman meeting was followed by a workshop in London in 1982, organized by Tom Blundell and John Kay. This short but highly useful conference did not, regrettably, produce a book, although many publications resulted from collaborations initiated there. The second full meeting of scientists engaged in studying the aspartic proteinase family took place in 1984 in Prague, organized by Vladimir Kostka [15]. This meeting was notable for the switch from the "Acid Protease" nomenclature to "Aspartic Proteinases," induced mainly by the efforts of John Kay. Another important feature of this meeting was the addition of "and Their Inhibitors" to the title. This resulted from the realization that inhibitors of this class of enzyme could have

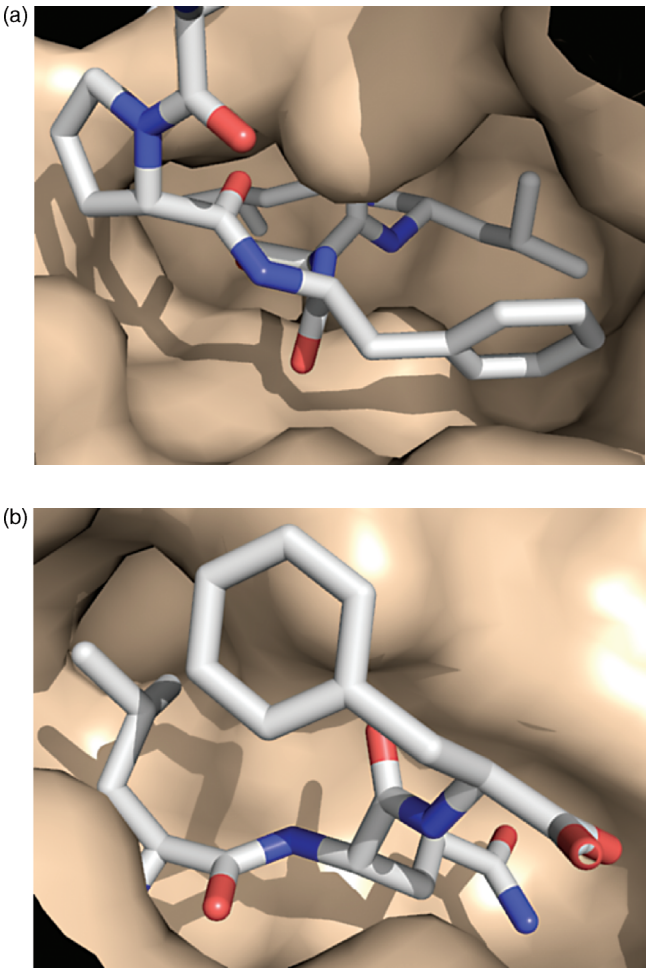

Figure 1.7 (a) View of the N-terminal half of the binding cleft of *P. falciparum* plasmepsin 2 with a bound peptidomimetic inhibitor. The figure was created using the Pymol program and the PDB file 2r9b. On the left of the figure, a proline residue in the P_4 position of the inhibitor makes contact with the C-terminal side of the active site cleft. On the right-hand side of the figure, a phenylalanine in P_3 and a leucine in P_1 make contact with the N-terminal side of the active site cleft. (b) View of the C-terminal half of the binding cleft of *P. falciparum* plasmepsin 2 with a bound peptidomimetic inhibitor. Thefigure was created using the Pymol program and the PDB file 2r9b. On the left of the figure, a leucine side chain in the P'_1 position makes contact with the S'_1 pocket in the enzyme. On the right-hand side of the figure, a glutamine side chain in the P'_2 position interacts with a deep pocket in the N-terminal half of the enzyme. The C-terminal residue is a phenylalanine, seen at the top of the figure. It makes contact with a gap in the C-terminal side of the cleft.

important impacts on treatment of an increasing number of diseases. It also gave further impetus to the exploding area of X-ray crystallography as a tool to develop important information about binding specificity. Another workshop was held in Tokyo, organized by Michael Samloff and attended by about 50 scientists. The next

full meeting of this group was organized by Bent Foltmann and held in Elsinor, Denmark, in 1988. This conference was marked by the consumption of enormous amounts of fine Danish beer, and, more important, saw the beginning of the realization that the retroviral systems produced enzymes with significant sequence and, presumably at that time, mechanistic identities to the classical systems studied before. At that time it was decided to hold another conference 3 years later and to switch the venue to the United States. I accepted the job of organizing this conference to be held in 1991; however, the incredibly rapid progress on the developing story of HIV protease led to the decision to move this meeting up by 1 full year and it took place in 1990 in Sonoma County, CA. This conference resulted in another book [16] that attempted to capture the content of the conference. Again, due to the rapidly increasing pace of discovery, Kenji Takahashi organized the next conference and it was held in Gifu, Japan, in 1993 and followed rapidly by publication of a fourth book from this conference [17]. In addition to many refinements in our understanding of the structure and function of the enzymes of this class, the Gifu conference featured many new biological systems where activity was shown to cause pathology, thus generating new targets for drug discovery. Michael James organized the next meeting in the series in 1996 in Banff, Canada. This conference revealed many new areas of investigation, so the traditional sequence of topics was modified significantly. The emphasis was on new enzymes from new areas of biology, illustrated by the emergence of parasitic and plant systems, in addition to the retroviral, mammalian, and microbial proteases, as can be seen in the book that resulted [18]. A very stimulating session dealt with issues related to the conversion of the proenzyme form of many of the aspartic proteinases into their mature and more active forms. In 1999, Carlos Faro and Euclides Pires organized a conference in Funchal, on the island of Madeira, Portugal. At this conference, the new methods of molecular biology, functional genomics, and gene manipulation were highlighted. An additional feature of this meeting was a strong emphasis upon featuring contributions from new investigators; unfortunately, this conference did not result in a book. Another conference was organized as a satellite meeting of the International Proteolysis Society in Nagoya in 2003. Following the IPS meeting, a dedicated group of scientists traveled by bus to Kyoto, where Yoshiaki Kiso and Kohei Oda arranged a very packed program that was notable for the high quality of the presentations. Among the advances discussed was the new area of membrane-bound proteases and the relationship with Alzheimer's disease, thus expanding further the scope of this enzyme family.

1.5
Exploration of the Active Sites of Aspartic Proteinase and Relation to Drug Discovery

To learn about the requirements for strong and productive binding in the active sites of enzymes, numerous studies have examined the binding of peptides of various types. In one example, peptides derived from the prosegment of porcine pepsinogen [19] were synthesized and tested [20] as inhibitors of the mature enzyme pepsin. Considerable inhibitory potential was found and with the preparation of derivatives

by substitution of several amino acids, it was found that a central hydrophobic stretch of the peptide was critical to the inhibitory function [21]. However, years later it was found that the actual binding site for the peptide was not at the active site, but was at a separate location near the bottom of the enzyme [22]. This binding displaced the N-terminal beta strand of the mature enzyme, allowing it to move into the active site and cause the observed inhibition.

One point learned from this study is that it is important to have a good assay for inhibition that truly reflects the occupation of the active site cleft and thus yielding true competitive inhibition. Accordingly, some efforts were invested in developing an assay for pepsin that would allow variation of the substrate concentration that would provide a mechanism for testing the type of inhibition. After checking out the assay developed in Fruton's laboratory [23] that utilized fluorescence, we decided to prepare a chromogenic substrate due to low cost of synthesis and far greater stability. Powers *et al.* [24] had published a compilation of the peptide bonds cleaved by porcine pepsin in protein sequencing studies, which also provided the amino acid sequence surrounding the cleavage site. The data from that study allowed Dunn *et al.* to design a substrate sequence of Pro-Thr-Glu-Phe*Phe-Arg-Leu, where the asterisk indicates the predicted point of cleavage [25]. In this sequence, each amino acid was chosen from either the top or the second-best amino acid in the Powers *et al.*'s report. Why was the second-best amino acid chosen in some cases? This permitted the use of some hydrophilic amino acids to add solubility to the peptide sequence. Although this peptide was shown to be cleaved by porcine pepsin, it was impossible to make a solution of high enough concentration to permit determination of a K_m value. To address this problem, Dunn *et al.* added an eighth amino acid on the amino terminus, Lys, to increase the solubility. The resulting peptide has turned out to be a useful substrate for many aspartic proteinases of the A1 pepsin-like enzyme family. Only the retroviral enzymes were not able to cleave the peptide.

The sequence Lys-Pro-Thr-Glu-Phe-Phe-Arg-Leu was further modified to place a *para*-nitrophenylalanine to the right of the cleavage site, yielding Lys-Pro-Thr-Glu-Phe*Nph-Arg-Leu. Cleavage of the peptide bond between -Phe*Nph- results in a decrease in absorbance in the range of 300–310 nm that can be correlated with the conversion of the substrate to the products: a pentapeptide and a tripeptide.

Following the discovery that the synthetic peptide substrate was useful for analysis of the activity of other aspartic proteinases, including chymosin and cathepsin D, a series of substitutions in the peptide were made to probe the preferences for amino acids in the positions flanking the cleavage site. This work originally began with the efforts of Joseph Fruton and his colleagues at Yale University in the 1960s. However, his work was hampered by a lack of solubility and by using mostly tetrapeptides, which did not pick up all the binding interactions made possible by the elongated active site cleft of this class of enzyme. Once the crystallographic work showed the extended nature of the active site, the Dunn lab was able to exploit this to conduct studies of the various subsites along the cleft according to the principles first described by Schechter and Berger [12].

In addition, the efforts of Theo Hofmann in Toronto [26] revealed the value of adding additional amino acids to the right and to the left of the cleavage point. By

filling the P_3 and the P'_2 subsites, the activity of enzymatic cleavage could be optimized, with little additional advantage gained by adding additional amino acids. However, extending the substrate to the octapeptide described above brought advantages in solubility that more than made up for the additional cost of synthesizing the longer peptides.

In a series of studies, the pH dependence of several aspartic proteinases was detailed, along with the "secondary" specificity [27, 28]. Revealing the optimal substrate for various enzymes permitted the use of these substrates for characterization of the binding of competitive inhibitors [29]. From these studies, it was discovered that the most general substrate was Lys-Pro-Ile-Glu-Phe-Nph-Arg-Leu, and this has been proven of great value in studies of at least 20 aspartic proteinases.

At that point, a number of groups were interested in developing inhibitors of human renin in an attempt to contribute to the control of blood pressure, as renin initiates the angiotensinogen to angiotensin conversion. Unfortunately, at that time, producing renin in pure form in reasonable quantities was very difficult and the limited selectivity of the enzyme precluded applying the approach of using the synthetic peptides described above.

Around this time, several groups began to produce recombinant enzymes from this family and again the substrates described above proved valuable in establishing the quality of the recombinant forms. If an enzyme produced by expression in bacteria, yeast, or insect cells was shown to have kinetic parameters including K_m, k_{cat}, and k_{cat}/K_m with the peptide Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu that were equivalent to those shown by the purified natural enzymes, then one could have confidence that the recombinant enzyme was identical in properties to the native enzyme. If one could also study the binding of several standard inhibitors and again find agreement, within experimental error, with values obtained previously with the native material, this would further provide proof of the quality of the protein. In cases where the values did not agree in a satisfactory way, it was most often due to poor folding of the recombinant material.

Further improvements in peptide synthesis allowed Beyer *et al.* [30] to use a combinatorial approach to explore the active site specificity of the aspartic proteinase family. Two series of peptides were designed with the general structures as follows:

1) Series 1 = Lys-Pro-X-Glu-P_1*Nph-Y-Leu

   Series 1 contains 19 separate pools where the P_1 amino acid is one of the following: Ala, Arg, Gly, Ser, Thr, His, Asp, Asn, Lys, Met, Leu, Ile, Phe, Tyr, Trp, Pro, Glu, Gln, or Nle, where Nle is norleucine.

   In each pool, the synthetic step to add "Y" used a mixture of all the 19 amino acids listed above. Likewise, the synthetic step to add amino acid "X" also used the same mixture of the 19 amino acids. In these mixtures, the amino acids that reacted slower were supplemented in concentration to make the reaction rate approximately equal.

   Thus, in each of the 19 pools, there were 361 total peptides (19×19) and each had a different P_1 amino acid.

2) Series 2 = Lys-Pro-[best P_3]-X-Nph-P'_1-[best P'_2]-Y

Series 2 is prepared in the same way as series 1, but the amino acid in P'_1 is varied over the 19 amino acids listed above. At synthetic steps adding the X and Y amino acids, the mixture of the 19 amino acids was used. In addition, the Nph group was moved to the P_1 position. In this case, the chromogenic change on cleavage to the "right" of the Nph amino acid yielded an increase in absorbance, rather than a decrease as seen for series 1.

The 19 pools in series 1 and the 19 in series 2 were used to screen for the P_1 and P'_1 specificity, respectively. A wide variety of aspartic proteinases were used and the resulting changes in absorbance determined. A simple plot of the rate of absorbance change against the P_1 or P'_1 amino acid showed the optimal amino acid for those two positions. While Phe was the most common optimal residue for both the P_1 and P'_1 positions, a number of enzymes preferred other amino acids. For example, plasmepsin 2 from the malarial parasite *Plasmodium falciparum* prefers Leu in P_1, while human gastricsin prefers Trp in P_1. These differences probably reflect the size of the S_1 pocket in the enzyme's surface.

Following incubation of a given enzyme with the best P_1 or P'_1 pool, the resulting penta- and tripeptide products could be separated by HPLC and identified by mass spectrometry. The quantity of the products revealed which peptides out of the 361 peptides in the mixture were most rapidly cleaved, as long as the analysis was done using only about 10% cleavage of the pool. The identity of the penta- and tripeptide products report on the specific amino acid that was more favorable. In this fashion, it was possible to identify the preferences for the P_3-P_2-P_1-P'_1-P'_2-P'_3 positions in the substrate. As all peptides had the Lys-Pro dipeptide on the amino terminal, the study did not probe those sites. Surprisingly, the information determined could then be used to design inhibitors that turned out to be excellent inhibitors when the scissile peptide bond was replaced with a peptidomimetic, such as $-CH_2-NH-$, or reduced peptide bond. This converts a good substrate into a good inhibitor, with some peptides designed in this way yielding nanomolar inhibitors [31].

Confirmation of the binding was provided by the structure of one of the peptidomimetic inhibitors bound into the active site of plasmepsin 2. Each of the amino acids from the P_4 position to the P'_3 position can be seen to fit nicely into the active site cleft into the "subsites" or pockets that are associated with binding of substrates and inhibitors. Figure 1.7a shows the interactions of the P_4 Pro residue, the P_3 Phe residue, and the P_1 Leu residue of the inhibitor Lys-Pro-Phe-Ser-Leu-CH_2-NH-Leu-Gln-Phe. The P_4 Pro residue makes a nice interaction with a small depression on the enzyme surface, helping to explain the success of the whole series of peptides. The P_3 Phe residue points into the S_3 pocket and makes strong interactions, while the Leu at P_1 fits nicely in the S_1 subsite, but does not fill up all the space available. The P_1 Leu residue is better for the plasmepsins than for other enzymes in the pepsin class, perhaps because the S_1 pocket in the plasmepsins is slightly smaller than in the others. Figure 1.7b shows interactions of the P'_1, P'_2, and P'_3 amino acids with the active site of plasmepsin 2. The P'_1 Leu makes a nice contact with the S'_1 pocket on the C-terminal domain of the enzyme, while the Gln at P'_2 fits nicely in the S'_2 pocket.

The side chain of the Gln makes a good hydrogen bonding contact with a carbonyl oxygen of the enzyme surface. Finally, the P'_3 Phe interacts with a gap in the enzyme structure. This is not part of the extended active site cleft, as the inhibitor C-terminus curls back over the P'_1 residue. This is also seen in the case of structures of pepstatin binding to aspartic proteinases. The interactions of these amino acid side chains, combined with the hydrogen bonding of the backbone carbonyl (C=O) and amide nitrogens (N−H) with complementary groups of the enzyme backbone, provide the binding energy necessary to yield tight and specific binding. In the future, such structures can be used to fine-tune the binding interactions by placing other specific groups into the peptide sequence.

This narrative has emphasized one approach to developing inhibitors for the aspartic proteinase family of enzymes, which can be termed the "substrate-based inhibitor design." Other approaches are described by various authors in this book. In any event, the development of sensitive assays for studying the inhibitory effect of new compounds is a critical element in successful inhibitor design and analysis.

References

1 Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science*, **268**, 533–539.

2 Fujinaga, M., Cherney, M.M., Oyama, H., Oda, K., and James, M.N.G. (2004) The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 3364–3369.

3 Wlodawer, A., Li, M., Gustchina, A., Oyama, H., Dunn, B.M., and Oda, K. (2003) Structural and enzymatic properties of the sedolisin family of serine-carboxyl peptidases. *Acta Biochimica Polonica*, **50**, 81–102.

4 Tang, J., Sepulveda, P., Marciniszyn, J., Chen, K.C.S., Huang, W.Y., Tao, N., Liu, D., and Lanier, J.P. (1973) Amino-acid sequence of porcine pepsin. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3437–3439.

5 Rawlings, N.D., Morton, F.R., Kok, C.Y., Kong, J., and Barrett, A.J. (2008) *MEROPS*: the peptidase database. *Nucleic Acids Research*, **36**, D320–D325.

6 James, M.N.G. and Sielecki, A.R. (1983) Structure and refinement of penicillopepsin at 1.8 Å resolution. *Journal of Molecular Biology*, **163**, 299–361.

7 Blundell, T., Jenkins, J.A., Sewell, B.T., Pearl, L.H., Cooper, J.B., Tickle, I.J., Veerapandian, B., and Wood, S.P. (1990) X-ray analysies of aspartic proteinases II. three-dimensional structure of the hexagonal crystal from of porcine pepsin at 2.3 Å resolution. *Journal of Molecular Biology*, **214**, 199–222.

8 Suguna, K., Bott, R.R., Padlan, E.A., Subramanian, E., Sheriff, S., Cohen, G.H., and Davies, D.R. (1987) Structure and refinement at 1.8 Å resolution of the aspartic proteinase from *Rhizopus chinensis*. *Journal of Molecular Biology*, **196**, 877–900.

9 Andreeva, N.S., Zdanov, A.S., Gustchina, A.E., and Federov, A.A. (1984) *Journal of Molecular Catalysis A: Chemical*, **259** 11353–11365; Sielecki, A.R., Federov, A.A., Boodhoo, A., Andreeva, N.S., and James, M.N.G. (1983) Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *Journal of Molecular Biology*, **214** 143–170.

10 Fujinaga, M., Cherney, M.M., Tarasova, N.I., Bartlett, P.A., Hanson, J.E.,

and James, M.N.G. (2000) Structural study of the complex between human pepsin and a phosphorus-containing peptidic transition-state analog. *Acta Crystallographica Section D: Biological Crystallography*, **56**, 272–279.

11 Cutfield, S.M., Dodson, E.J., Anderson, B.F., Moody, P.C., Marshall, C.J., Sullivan, P.A., and Cutfield, J.F. (1995) The crystal structure of a major secreted aspartic proteinase from Candida albicans in complexes with two inhibitors. *Structure*, **3**, 1261–1271.

12 Schechter, I.and Berger, A. (1967) On size of active site of proteases. I. Papain. *Biochemical and Biophysical Research Communications*, **27**, 157–163.

13 Tang, J. (ed.) (1977) *Acid Proteases: Structure, Function, and Biology, Advances in Experimental Medicine and Biology*, vol. **95**, Plenum Press, New York.

14 Tang, J., James, M.N.G., Hsu, N., Jenkins, J.A., and Blundell, T.L. (1978) Structural evidence for gene duplication in the evolution of the acid proteases. *Nature*, **271**, 618–621.

15 Kostka, V. (ed.) (1985) *Aspartic Proteinases and Their Inhibitors*, Walter de Gruyter, Berlin.

16 Dunn, B.M. (ed.) (1991) *Structure and Function of the Aspartic Proteinases: Genetics, Structures, and Mechanisms, Advances in Experimental Medicine and Biology*, vol. **306**, Plenum Press, New York.

17 Takahashi, K. (ed.) (1993) *Aspartic Proteinases: Structure, Function, Biology, and Biomedical Implications, Advances in Experimental Medicine and Biology*, vol. **362**, Plenum Press, New York.

18 James, M.N.G. (ed.) (1998) *Aspartic Proteinases: Retroviral and Cellular Systems, Advances in Experimental Medicine and Biology*, vol. **436**, Plenum Press.

19 Herriott, R.M. (1938) Kinetics of the formation of pepsin from swine pepsinogen and identification of an intermediate compound. *The Journal of General Physiology*, **22**, 65–78.

20 Dunn, B.M., Deyrup, C., Moesching, W.G., Gilbert, W.A., Nolan, R.J., and Trach, M.L. (1978) Inhibition of pepsin by zymogen activation fragments. *The Journal of Biological Chemistry*, **253** 7269–7275.

21 Dunn, B.M., Lewitt, M., and Pham, C. (1983) Inhibition of pepsin by analogs of pepsinogen (1–12) with substitution in the 4–7 region. *The Biochemical Journal*, **209**, 355–362.

22 Masa, M., Maresova, L., Vondrasek, J., Horn, M., Jezek, J., and Mare, M. (2006) Cathepsin D propeptide: mechanism and regulation of its interaction with the catalytic core. *The Biochemical Journal*, **45**, 15474–15482.

23 Sachdev, G.P., Johnston, M.A., and Fruton, J.S. (1972) Fluorescence studies on interaction of pepsin with its substrates. *The Biochemical Journal*, **11**, 1080–1086.

24 Powers, J.C., Harley, A.D., and Myers, D.V. (1977) Subsite specificity of porcine pepsin, in *Acid Proteases: Structure, Function, and Biology* (ed. J. Tang), Plenum Press, New York, pp. 141–157.

25 Dunn, B.M., Kammermann, B., and McCurry, K.R. (1984) The synthesis, purification, and evaluation of a new chromophoric substrate for pepsin and other aspartyl proteases. *Analytical Biochemistry*, **138**, 68–73.

26 Balbaa, M., Cunningham, A., and Hofmann, T. (1993) Secondary substrate binding in aspartic proteinases: contributions of subsites S(3) and S′(2) to k_{cat}. *Archives of Biochemistry and Biophysics*, **306**, 297–303.

27 Dunn, B.M., Jimenez, M., Parten, B.F., Valler, M.J., Rolph, C.E., and Kay, J. (1986) Systematic series of synthetic chromophoric substrates for aspartic proteinases. *The Biochemical Journal*, **237**, 899–906.

28 Dunn, B.M., Valler, M.J., Rolph, C.E., Jimenez, M., and Kay, J. (1987) The pH dependence of the hydrolysis of chromogenic substrates of the type, Lys-Pro-Xaa-Yaa-Phe-(NO_2)Phe-Arg-Leu, by selected aspartic proteinases: evidence for specific interactions in subsites S_3 and S_2. *Biochimica et Biophysica Acta*, **913**, 122–130.

29 Rao, C.M., Scarborough, P.E., Kay, J., Batley, B., Rapundalo, S., Klutchko, S., Taylor, M.D., Lunney, E.A., Humblet, C.C., and Dunn, B.M. (1993) Specificity in the binding of inhibitors to the active site of human/primate aspartic proteinases: analysis of P_2-P_1-P'_1-P'_2 variation. *Journal of Medicinal Chemistry*, **36**, 2614–2620.

30 Beyer, B.B., Johnson, J.V., Chung, A.Y., Li, T., Madabushi, A., Agbandje-McKenna, M., McKenna, R., Dame, J.B., and Dunn, B.M. (2005) Active-site specificity of digestive aspartic peptidases from the four species of *Plasmodium* that infect humans using chromogenic combinatorial peptide libraries. *Biochemistry*, **44**, 1768–1779.

31 Liu, P., Marzahn, M.R., Robbins, A.H., Gutiérrez-de-Terán, H., Rodríguez, D., McClung, S., Stevens, S.M., Jr., Yowell, C.A., Dame, J.B., McKenna, R., and Dunn, B.M. (2009) Recombinant plasmepsin 1 from the human malaria parasite *Plasmodium falciparum*: expression of an engineered variant, enzymatic characterization, combinatorial chemistry-based peptide inhibitor design and crystallographic analysis, and modeling of active site interactions of three plasmepsins. *The Biochemical Journal*, **48**, 4086–4099.