The two sides of enzyme–substrate specificity: Lessons from the aspartic proteinases

ARTICLE *in* BIOCHIMICA ET BIOPHYSICA ACTA · APRIL 2000

Impact Factor: 4.66 · DOI: 10.1016/S0167-4838(99)00275-7 · Source: PubMed

2 AUTHORS, INCLUDING:

Ben M Dunn
University of Florida
251 PUBLICATIONS **5,298** CITATIONS

SEE PROFILE

ELSEVIER

Biochimica et Biophysica Acta 1477 (2000) 231–240

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.elsevier.com/locate/bba

Review

The two sides of enzyme–substrate specificity: lessons from the aspartic proteinases

Ben M. Dunn *, Su-Hwi Hung

Department of Biochemistry and Molecular Biology, University of Florida College of Medicine, P.O. Box 100245, Gainesville, FL 32610-0245, USA

Received 1 November 1999; accepted 1 December 1999

Abstract

Like most proteolytic enzymes, the aspartic proteinases bind substrates and most inhibitors within an extended active site cleft. Bound ligands typically adopt a β-strand conformation. Interactions with groups on both sides of the cleft determine the primary as well as secondary specificity of the enzymes. We have pursued the discovery of the sometimes subtle distinctions between members of the aspartic proteinase family by two routes. In the first case, we have constructed sets of oligopeptide substrates with systematic variation in each position to assess interactions at one position at a time. In the second type of experiment, we have altered residues of the enzymes in order to test theories of selectivity. The combination of the two approaches has provided a better understanding of the forces involved in determining specificity of enzyme action.

Keywords: Aspartic proteinase; Substrate specificity; Enzyme active site; Enzyme kinetics; Chromogenic substrate; Mutagenesis of active site residues

1. Introduction

From the earliest days of the modern era of biochemistry, a major thrust has been to understand how proteins of similar function isolated from different organisms compare in their structure and enzymatic activities [1]. In recent years, we have enjoyed the luxury not only of having access to total sequence and structural information, but also the methodology to vary primary structure at will. Thus, we can now vigorously pursue solutions to questions that were posed 30 or more years ago.

With respect to sequence and structure analyses, the aspartic proteinase family was a late arrival among the proteases, even though pepsin was the first enzyme to be crystallized [2]. While trypsin, chymotrypsin, papain, and carboxypeptidase were sequenced and characterized structurally in the 1960s, the first aspartic proteinase sequence was not available until 1973 [3], and the first structures were reported starting in the late 1970s [4–6]. In the years since, however, a considerable body of information regarding sequence and structure has been developed for new members of this family [7].

While comparisons of amino acid sequence and three-dimensional structure are straightforward, describing the functional relationships of proteases is fraught with peril. Classically, digestion of a standard sequence, such as the oxidized B-chain of insulin or hemoglobin, has been used to define the substrate specificity of a new protease [8]. This offers

* Corresponding author.

0167-4838/00/$ – see front matter
PII: S0167-4838(99)00275-7

several advantages, including low cost, ready availability of reagents, and well defined products, but suffers from some disadvantages, including limited sequence variation, inability to quantitate the kinetics of cleavage of an individual bond, and possible product inhibition. With a continuously expanding substrate population following the first cut in the peptide chain, the test protease is presented with a variety of choices. Nonetheless, this method has served to categorize proteases into broad groups, and is still recommended for the purposes of initial classification.

2. The use of designed sets of synthetic peptides for discovery of peptidase specificity

With the advent of solid phase synthesis, it became feasible to prepare peptides with longer and more complex sequences. In early studies of the specificity of pepsin, for example, di- and tri-peptides were prepared for kinetic studies [9]. We have learned from extensive structural studies that most endopeptidases have an elongated active site cleft capable of interacting with multiple residues of a substrate protein (Fig. 1). This extensive interaction has one very practical consequence: it forces the substrate sequence to adopt an extended β-strand conformation. This stretches the polypeptide to the maximum extent, and, in effect, 'pins' the substrate down so that it may be attacked at the sensitive bond. The benefit of holding the substrate in this way is difficult to measure, and the interpretation is controversial. However, the catalytic benefit has been quantitated by the work of Hofmann and colleagues [10] in the case of penicillopepsin by comparing the cleavage of substrates of increasing size. When a substrate is increased in size from a dipeptide to about five residues (P3-to-P2′), the k_{cat} value increases by several hundred fold.

Thus it seemed that comparisons would be facilitated if peptides were prepared that were long enough to fill up the active site cleft completely. This would permit positioning of the peptides within the extended cleft to take advantage of all interactions that could hold down the sequence. It is known that hydrogen bonding between carbonyl oxygen and amide nitrogens of the substrate peptide and corresponding groups of the enzyme is very extensive. However, any peptide sequence, except those involving proline, will present the same set of backbone groups for such interactions. The side chains of the residues comprising the peptide substrate and the active site of the enzyme also must make contacts that both strengthen the interactions and dictate the selectivity of the enzyme–substrate combination. Thus, a sequence that presents the optimal amino acid residues for a given active site cleft will achieve the best combination of binding and catalysis, leading to the highest values of k_{cat}/K_m, the specificity constant.

In addition, making the peptide substrates a length that should fill up the active site cleft may have the added advantage of maintaining a fixed cleavage point, which will facilitate comparisons between peptides. In the case of pepsin, it was known, from studies with the oxidized B-chain of insulin [11], that the primary specificity (the two amino acids between which cleavage is favored) is for two aromatic amino acids. By incorporating -Phe-Phe- into oligopeptides, it is possible to 'lock' the cleavage site of a series of oligopeptides to the bond between the two aromatic [12]. This provides a convenient template for studies of the 'secondary' specificity, meaning the preferences of a given enzyme for the amino acid sequence surrounding the primary specificity site, -Phe-Phe-. In the work discussed below, we will show that the effects of variation in the flanking amino acids can be substantial, leading to quantifiable differences in specificity.

In considering which sequence to prepare to examine the specificity of pepsin, one can choose to prepare a peptide with the simplest possible sequence, Gly-Gly-Gly-Phe-Phe-Gly-Gly, as it was known that the active site cleft appeared capable of binding about seven amino acids. Variation could then be achieved by replacing each glycine with the other 19 amino acids. However, we took advantage of an earlier report [13] that detailed the results of a survey of pepsin cleavage sites in 177 proteins studied by sequence analysis between 1967 and 1972. This yielded a probability analysis describing which residues will appear in various positions on either side of the cleavage site. We chose the residue with the highest probability at each positions to design a sequence, Pro-Thr-Glu-Phe-Phe-Arg-Leu. In the case of the P3

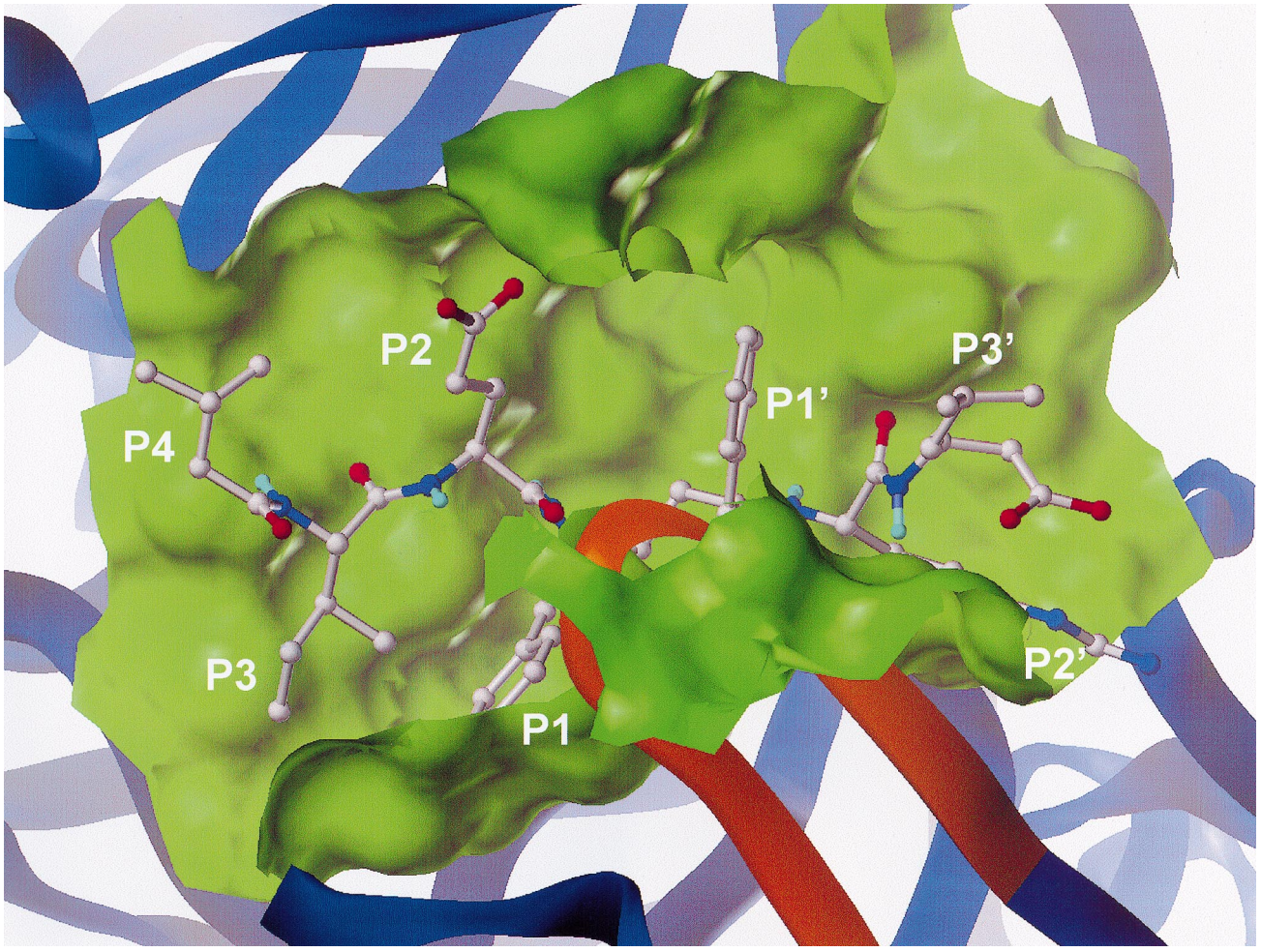

Fig. 1. Model of a heptapeptide substrate bound in the active site cleft of an aspartic proteinase. The enzyme backbone is shown as a shaded blue ribbon, except for the residues of the 'flap', a β-hairpin comprising residues 75–82 that partially occludes the cleft, given in orange. The Connelly solvent accessible surface of the active site (green) was calculated after removing bound ligand and water molecules. In addition, residues 78 and 79 at the tip of the 'flap' were not used in calculating the surface, so that the P1 side chain of the substrate can be viewed in this presentation. If included, the surface would obscure the P1Phe residue. The side chains of the substrate alternate along the cleft due to the β-strand conformation of bound ligands. Residues P4, P2, P1′ and P3′ are directed toward the C-terminal domain, which contributes one side of the active site cleft (top half of this view). Residues P3, P1, and P2′ are directed toward the N-terminal domain that contributes the other side of the cleft (bottom half of this view).

position (Thr), we chose a residue that was more hydrophilic than the other best residues, Ile and Ala. In P2′, we chose Arg for similar reasons over Val. We chose to put a *p*-NO_2Phe in P1′, to replace Tyr, which was the highest amino acid listed, as Hofmann and Hodges had reported [14] that cleavage of the peptide bond next to the substituted phenylalanine residue would yield a shift in absorbance. The resulting peptide was a good pepsin substrate and provided a convenient assay [15].

We subsequently added an extra Lys residue at the amino-terminal end to improve solubility and varied the residue in the P3 position [16]. In this study, we found that the sequence Lys-Pro-Ala-Glu-Phe-Nph-Arg-Leu gave superior kinetic parameters with porcine pepsin ($k_{cat}/K_m = 2.77 \times 10^6$ M^{-1} s^{-1}). Another peptide, Lys-Pro-Ile-Glu-Phe-Nph-Arg-Leu, proved to be an excellent substrate for several enzymes of the aspartic proteinase family, and was adopted as the standard substrate for assay of those enzymes in several laboratories. Subsequently, Pohl and Dunn [17] reported on the pH dependence of kinetic pa-

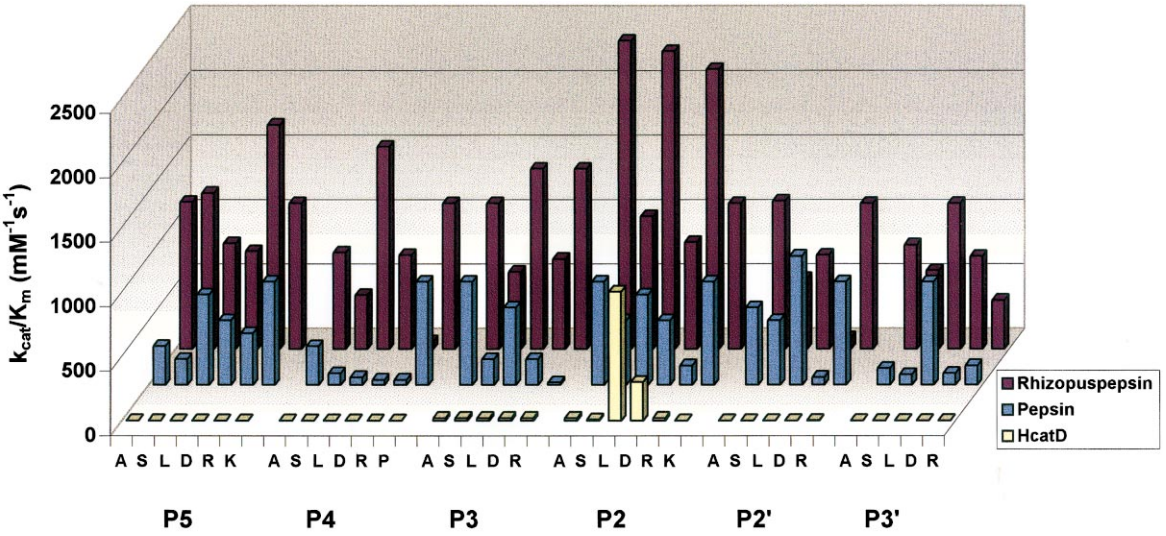

Fig. 2. Plot of the specificity constant, k_{cat}/K_m, for cleavage of a series of peptides based on Lys-Pro-Ala-Lys-Phe-Nph-Arg-Leu, where the bond between Phe and Nph is cleaved in all cases. Substitution in each position is indicated on the *x*-axis. Three enzymes were evaluated for each peptide, Rhizopuspepsin, pepsin, and human cathepsin D.

rameters for derivatives of Lys-Pro-Ala-Glu-Phe-Nph-Arg-Leu where Lys residues were substituted into various positions. Dramatic pH dependence was seen for several substrates, with higher pH favoring binding and cleavage. These effects arise from specific interactions within the enzyme subsites. However, a general electrostatic attraction could also account for the observed effects. This question was explored by site-directed mutagenesis, as discussed below.

Following these initial explorations, we decided to construct a set of peptides based on our discoveries. We first made the series of peptides based on Lys-Pro-Ala-Lys-Phe-Nph-Arg-Leu, with the substitution of Ala, Ser, Asp, Arg, and Leu in positions P5, P4, P3, P2, P2′, and P3′. We retained the -Phe-Nph- cleavage site in these peptides, and all cleavages were shown to occur between the two aromatic amino acids. The substitutions explored six 'secondary' subsites of the active site cleft with respect to tolerance for small, hydrogen-bonding, negatively or positively charged, or larger side chains. Each peptide produces a similar change in absorbance upon cleavage, thus we could compare rates of catalysis for different enzymes. This series of peptides taught us some valuable lessons (Fig. 2). First of all, some members of the aspartic proteinase family, such as porcine pepsin [18] and *Rhizopus chinensis* aspartic proteinase [19], have very broad substrate preferences, and can cleave nearly all of the 28 unique peptides of this set. For pepsin, the only peptide that was better than the parent sequence was Lys-Pro-Ala-Lys-Phe-Nph-Leu-Leu, where Leu replaces Ala in P2′. In the P4 position, the Pro of the parent peptide was clearly preferred, with little cleavage seen when Arg, Asp Leu or Ser were substituted into the parent sequence. In addition, an Arg residue in P3 yielded a peptide that was not cleaved by porcine pepsin.

Rhizopus chinensis aspartic proteinase illustrated even broader selectivity, cleaving all 28 peptides at measurable rates, with only Arg in P4 and Asp in P2′ producing peptides that were cleaved at slower rates. In addition, this study has discovered several residues that were preferred for the *Rhizopus* enzyme; for example, Arg in P5, Leu in P4, Leu/Arg in P3, Ala/Leu/Arg in P2 all produced rates higher than seen with the parent peptide. This information may be used to design a more selective substrate with the highest possible specificity constant.

In contrast to the broad specificity exhibited by porcine pepsin and the fungal enzyme, some enzymes, such as human cathepsin D [20], are unable to handle the Lys residue in the P2 position of this series. As 24 of the peptides of the set have the Lys substitution at this point, little of value could be

gained by studying enzymes of this type with this set of peptides. As seen in Fig. 2, only the substrates with Leu or Asp substituted in P2 gave measurable cleavage with human cathepsin D.

We knew that the substitution of a Glu into P2 would convert a poor substrate for cathepsin D to a good one. Thus, in a second series of syntheses, we prepared a new set of peptides based on Lys-Pro-Ile-Glu-Phe-Nph-Arg-Leu. This set provided the versatility needed to study the preferences of many members of the aspartic proteinase family. In addition, we have substituted a wider variety of amino acids in positions P3, P2, P2′, and P3′, as that is where the largest definition of specificity differences occurred. Data obtained for this series of peptides are presented in Fig. 3 for the enzymes pepsin, cathepsin D, and plasmepsin II [21] from the malaria parasite, *Plasmodium falciparum*. We see a wide variety of values of the specificity constant, k_{cat}/K_m, for these substitutions. In each case, these reflect the preference of the individual subsites for interaction with the amino acid residue in the substrate peptide at the corresponding position. The range of values exhibited for the series of peptide substrates can be as large as 500-fold or more for the various substitutions. Some of the 'enhancements' are difficult to measure, as some substituents yield peptides that are cleaved at such slow rates that the value of the specificity constant is difficult to quantitate. It should be noted here that the *lack* of cleavage can also provide useful information in a drug discovery project, as one may wish to prevent binding to a human enzyme, for example, while promoting binding to an enzyme from a pathogen.

Another advantage of this analysis is that we can compare enzyme efficiency across the different proteinases. If we discover the optimal substrate for each enzyme among this series of peptides, we can relate the catalytic potency of the enzymes to each other. As can be observed in Figs. 2 and 3, the values of k_{cat}/K_m are in the range of 5×10^6 M^{-1} s^{-1} for each enzyme with the best substrate. As the active sites of the different enzymes are all similar and utilize equivalent catalytic machinery, therefore, this is not surprising.

In the case of plasmepsin II in Fig. 3, the values of the specificity constant were multiplied by a factor 10 to make them comparable to those of human cathepsin D and porcine pepsin. In plasmepsin II, we have found that the replacement of the Arg in the P2′ position of the parent peptide by Ala produces a peptide with superior kinetic properties. Thus, all of the other peptides tested in this series are not good substrates as they all have Arg in P2′. For

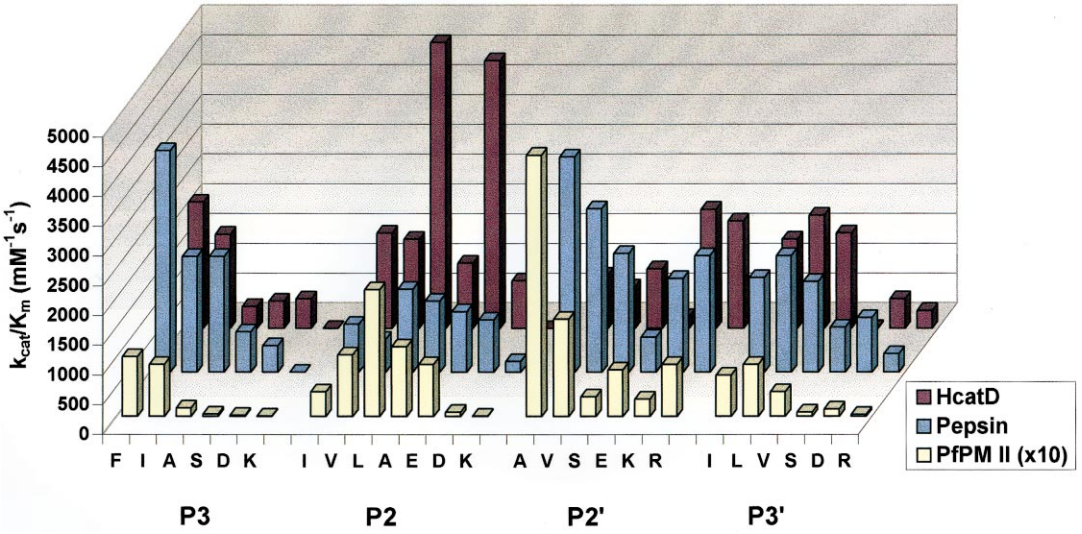

Fig. 3. Plot of the specificity constant, k_{cat}/K_m, for cleavage of a series of peptides based on Lys-Pro-Ile-Glu-Phe-Nph-Arg-Leu. Substitution in each of four positions are indicated on the *x*-axis. Three enzymes were evaluated for each peptide, human cathepsin D, pepsin and plasmepsin II (PfPM-II). The values for PfPM-II were multiplied by 10 in order to make the numbers visible on the same scale for the other two enzymes, as most of these peptides contain Arg in P2′, which is less favorable for PfPM-II.

plasmepsin II, then, a better analysis of the subsite preferences would be achieved by comparing peptides based on Lys-Pro-Ile-Glu-Phe-Nph-Ala-Leu as the parent peptide. The peptide series we have prepared do not cover all of 'sequence space', thus we may not have optimized the kinetics for these enzyme–substrate pairs. However, we have learned a great deal about the active site preferences of the enzymes, and have placed the different enzymes into separate groups based on these observations. In addition, we have found important clues to absolute specificity of the various enzymes.

From Fig. 3, we can see that the P3 and P3′ subsites of all three enzymes tested prefer hydrophobic amino acids. We know that the P3 position is relatively deep and hydrophobic from the structural analyses, thus this finding is in agreement. In P2, however, we see a considerable divergence for human cathepsin D, with both Leu and Glu providing the optimal substrates for that enzyme. In P2′, another surprise was observed with human cathepsin D: Arg and Lys were the preferred substitutions [22]. In contrast, pepsin and plasmepsin II clearly preferred Ala or Val at that position.

An alternative approach to the discovery of proteinase specificity would be to construct combinatorial libraries of peptides with all possible amino acids at each position. For just the four positions shown in Fig. 3 (P3, P2, P2′, and P3′) and just considering the natural 20 amino acids, this would be a library of 160 000 peptides. With a strategy of using quenched fluorescent substrates, such libraries can be constructed and analyzed to discover the preferences of new proteinases.

3. The use of site-specific mutagenesis to explore aspartic proteinase specificity

In parallel with our work on sets of peptide substrates, the results from extensive crystallographic investigations have provided clues to factors involved in peptidase active site specificity. Subsequently, we have utilized mutagenesis to test hypotheses developed from both approaches.

Lowther et al. [23] studied the S1 specificity of Rhizopuspepsin. Many fungal enzymes had been shown to have a trypsin-like specificity in that they would cleave a sequence following Lys. Lowther et al. postulated that this was due to a unique Asp residue in the Rhizopuspepsin sequence. They prepared mutants in which the Asp residues at position 30 and 77 were changed to Ile and Ser, respectively. They analyzed the cleavage of a peptide in which the Phe residue at position P1 in our normal series was replaced by a Lys residue. Whereas the change of Asp 30 to Ile was worth a factor of 10 in the specificity constant for cleavage of the Lys-containing peptide, k_{cat}/K_m, the change of Asp 77 was worth a factor of 10^5. Asp 77 occurs at the tip of the 'flap', the β-hairpin that is positioned over the active site cleft in aspartic proteinases. It is unique to the fungal enzymes, including those from various *Candida* species.

Westling et al. [21] have done the reverse experiment, putting an Asp residue into the sequence of the enzyme discussed above, plasmepsin II, at the corresponding position on its 'flap'. Using the same substrate containing Lys at the P1 position, an increase was observed in the kinetic parameters, indicating that a trypsin-like function had been added to plasmepsin-II.

Rao studied the specificity of porcine pepsin, focusing on the unique Glu residues at positions 13 and 287 [18]. These two positions are on opposite sides of the binding cleft. Glu 13 is located on one side of the S3 specificity pocket, while Glu 287 is found in the S2 specificity pocket. These two residues are not found together in any other aspartic proteinase. In exploring the S3 specificity, it was observed that a peptide substrate containing an Arg residue was not cleaved by pepsin under the normal conditions of assay, namely pH 3.0. However, if the pH is increased from 3.0 to 5.5, a steady increase in the rate of cleavage of the ArgP3 peptide was observed. This was attributed to titration of Glu 13, with the resulting negative charge providing an electrostatic complement to the positively charged Arg side chain. In order to investigate this hypothesis, Rao prepared the Glu13Ala variant by site-directed mutagenesis. That derivative had normal activity on substrate having hydrophobic amino acids in P3, but exhibited no pH dependence for cleavage of the substrate with Arg in P3. The cleavage rate was very low at pH 3.0, and approximately equal to that shown by the enzyme of wild-type sequence.

Scarborough and Dunn explored the specificity of

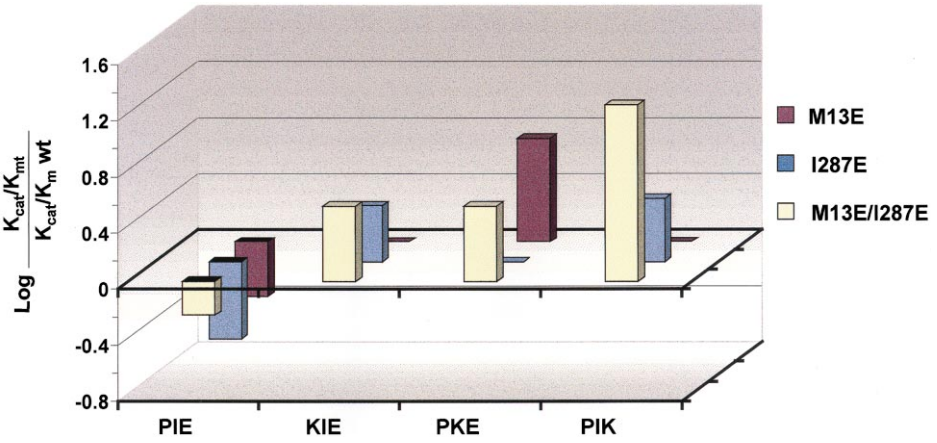

Fig. 4. Plot of the logarithm of the ratio of the specificity constant for three mutant enzymes over that for wild-type PfPM-II. Values are plotted for four substrates, the parent peptide, Lys-Pro-Ile-Glu-Phe-Nph-Arg-Leu, and three peptides where Lys is substituted in P4, P3, and P2. Negative values indicate that the wild-type enzyme cleaves the peptide more efficiently, whereas positive values indicate that the mutant enzyme is more efficient.

cathepsin D in the S2 position [24]. As discussed above, cathepsin D was unable to cleave a peptide substituted with Lys in the P2 position. Cathepsin D has a Met residue at the equivalent position to Glu 287 in porcine pepsin. Therefore Scarborough and Dunn created the Met287Glu mutant in cathepsin D and saw an increase in cleavage of the LysP2 substrate. It was not possible to quantitate the magnitude of the enhancement, because the native enzyme did not cleave the peptide at measurable rates, even with prolonged incubation.

A very thorough dissection of these effects is presented in Fig. 4, using data obtained by Westling et al. [21] in the plasmepsin-II system. In this figure three mutant enzymes are compared to the enzyme of wild-type sequence. By calculating the ratio of the specificity constant for the mutant form to the specificity constant of the wild-type enzyme and taking the log to the base 10, Fig. 4 represents a relative energy diagram. Four substrates were used for comparison. In the left side, the 'standard' substrate, Lys-Pro-Ile-Glu-Phe-Nph-Arg-Leu, has been evaluated. It can be seen that all three mutant enzymes are slightly less active against the standard peptide (negative values of log specificity constant–mutant/specificity constant–wild-type). The other peptides evaluated place a Lys residue in P4, P3, or P2. The mutants include the Met13Glu variant, where the Met is in the S3 subsite, and the Ile287Glu variant, where 287 is located on the other side of the cleft, approximately mid-way between the S2 and S4, based on pepstatin binding [25]. Considering the Met13Glu mutant first, when a Lys residue is placed in the P3 position, the specificity constant for the mutant enzyme is increased relative to that of the wild-type enzyme. Substrates with Lys placed in P4 or P2 were cleaved at about the same rate as by wild-type enzyme. On the other hand, when evaluating the Ile287Glu variant, both the peptides with Lys substituted in P2 and P4 show enhanced cleavage by the mutant enzyme, again relative to the wild-type enzyme. This is in agreement with the positioning of Ile 287 *between* subsites S2 and S4, so that a Glu in that position could interact with a Lys residue in either P2 or P4. Finally, the double mutant, Met13Glu/Ile287Glu, shows enhanced cleavage of all three Lys-containing substrates.

The results in Fig. 4 demonstrate convincingly that the efficiency of cleavage of a peptide substrate can be influenced by substitution in both the peptide and the enzyme, and that the effects are specific to the position of substitution. Furthermore, the effects can be observed in positions that are peripheral to the actual site of bond scission.

Beyer and Dunn have reported on binding in the S2′ subsite of human cathepsin D [22]. As mentioned earlier, the P2′ specificity of human cathepsin D was notable in that the basic amino acids Lys and Arg were preferred. Inspection of the S2′ subsite from the crystal structure of Baldwin et al. [26] revealed a

largely hydrophobic cavity. However, it was clear that the S2′ subsite was open to the aqueous surrounding. Thus, Beyer and Dunn have postulated that a Lys or Arg side chain can place its methylene groups in contact with the hydrophobic S2′ site, and still allow the positively charged end to erupt out of the site in order to make contact with solvent.

Another observation with human cathepsin D was that an Asp residue, when substituted into P2′, produced a substrate that was not cleaved at measurable rates. Again, the hydrophobic nature of the subsite was consistent with that observation. Beyer and Dunn altered the nature of S2′ by replacing Ile 128 with an Arg residue. This variation of the enzyme had reduced cleavage activity on substrates possessing hydrophobic amino acids in P2′. However, the Ile128Arg mutant showed a significant enhancement in the cleavage of the peptide with Asp substituted in P2′. This represents an 'addition of function' that is in accord with the principles discussed in this review.

4. Other examples of alterations in specificity derived from electrostatic interactions discovered by mutation of active site amino acids or variation in substrate sequence

The classic study of Wells et al. [26] on subtilisin demonstrated that engineering of charged groups into enzymes can have up to a 2000-fold effect on k_{cat}/K_m. They used both variation in substrate sequence and mutagenesis of enzyme groups to established critical positions where the effects could be observed.

Caputo et al. [27] studied the unique cleavage specificity of granzyme B, which is involved in apoptosis. This enzyme cleaves at sites with Asp in P1, a preference that is unique among mammalian serine proteinases. By replacing Arg 226 with Glu, they were able to create a variant that had the ability to cleave thiolesters of the general structure Boc-Ala-Ala-Xaa-SBzl, where Xaa was Met, Phe, or Arg, but not Asp. This single substitution, however, was not sufficient to permit cleavage of amide substrates.

Olsen et al. [28] studied the cleavage efficiency of the novel aspartic proteinase yaspin (Yap3p). Molecular modeling had suggested that the active site cleft of this enzyme contained multiple acidic residues within or close to S6, S3, S2, S1, S1′, S2′ and S3′. Substitution of several residues by Arg in the substrate CCK_{13-33} resulted in cleavage enhancements. One result suggested that binding of an Arg into the S2 site was so favorable that it forced a change in the primary specificity. These observations supported the model developed that attributes cleavage specificity to electrostatic interactions.

The study of the cysteine proteinases have provided many examples of electrostatic effects on catalysis, especially as the catalytic groups are known to be the ion pair between a Cys^- and His^+. Taylor et al. [29] explored the pH dependence of caricain, formerly known as papaya proteinase A, and the pH dependence of several mutants where titratable groups in the enzyme were substituted. This study compared the efficiency of catalysis of both neutral and charged substrates in an effort to identify residues that perturbed the pH profile for enzyme action. An Asp residue at sequence position 158 was proposed as a likely candidate, but eliminated based on the observation of a pH dependence in the Asp-158Asn mutant. Later work from Brocklehurst's lab [30] suggested that the catalytic system of cysteine proteinases was inactive until modulated by electrostatic effects. Pinitglang et al. [30] proposed Glu 50 as the residue responsible for a pK_a of around 4 that provides this modulation to the Cys/His catalytic pair. Ikeuchi et al. [31] have studied variants of caricain where Glu 50 is altered to Ala, and observed an *increase* in catalytic competence. This result challenges the concept that an electrostatic switch is required for cysteine proteinases to function, and this controversy is likely to continue until additional studies explore other potential residues.

5. Summary

Due to the extended binding mode of peptides in the active site of aspartic proteinases, it has been possible to study interactions at positions flanking the 'primary' specificity sites, P1 and P1′. The use of two aromatic amino acids, which fit the primary specificity, to anchor the peptide ligands into the proper alignment within the active site has permitted variation in the amino acids in as many as six other sites in the substrates. These approaches can be used

to determine the space available for amino acids, by comparing the efficiency of catalysis for peptides with amino acids of varying size. In addition, the effects of electrostatic interactions have been studied in four different aspartic proteinases. By changing amino acids in selected locations in the enzyme active site and using sets of oligopeptide substrates for analysis, it has been possible to quantitate the enhancements provided by complementary charges in both components. These findings also have value in the design of active site-directed inhibitors, in that additional specificity can be obtained by optimizing interactions within the active site. The observations made in our studies are in agreement with findings within other families of proteolytic enzymes.

Acknowledgements

The authors would like to acknowledge the efforts of other members of this laboratory on these projects over the past 10 years. We thank the Protein Chemistry Core at the University of Florida for synthesis of peptides and analysis of recombinant proteins. We are grateful to Professor Kohei Oda for stimulating discussions regarding selectivity in protease function. The figures were prepared with the assistance of Mr. Mihn Lam.

References

[1] H. Neurath, K.A. Walsh, W.P. Winter, Evolution of structure and function of proteases, Science 158 (1967) 1638–1644.

[2] J.H. Northrop, Crystalline pepsin. I. Isolation and tests of purity, J. Gen. Physiol. 13 (1930) 739–766.

[3] J. Tang, P. Sepulveda, J. Marciniszyn Jr., K.C.S. Chen, W.-Y. Huang, D. Liu, J.P. Lanier, Amino acid sequence of porcine pepsin, Proc. Natl. Acad. Sci. USA 70 (1973) 3437–3439.

[4] T.L. Blundell, J.A. Jenkins, B.T. Sewell, L.H. Pearl, J.B. Cooper, I.J. Tickle, B. Veerapandian, S.P. Wood, X-ray analyses of aspartic proteinases. The three dimensional structure at 2.1 Å resolution of endothiapepsin, J. Mol. Biol. 211 (1990) 919–941.

[5] K. Suguna, R.R. Bott, E.A. Padlan, E. Subramanian, S. Sheriff, G.H. Cohen, D.R. Davies, Structure and refinement at 1.8 Å resolution of the aspartic proteinase from *Rhizopus chinensis*, J. Mol. Biol. 196 (1987) 877–900.

[6] I.N. Hsu, L.T.J. Delbaere, M.N.G. James, T. Hofmann, Penicillopepsin from *Penicillium janthinellum* crystal structue at 2.8 Å and sequence homology with porcine pepsin, Nature 266 (1983) 140–145.

[7] A.J. Barrett, N.D. Rawlings, J.F. Woessner, Aspartic peptidases, in: Handbook of Proteolytic Enzymes, Academic Press, New York, 1998, p. 801, and following chapters.

[8] D.G. Smyth, Techniques in enzymic hydrolysis, Methods Enzymol. XI (1967) 214–231.

[9] G.P. Sachdev, J.S. Fruton, Pyridyl esters of peptides as synthetic substrates of pepsin, Biochemistry 8 (1969) 4231–4238.

[10] T. Hofmann, B. Allen, M. Bendiner, M. Blum, A. Cunningham, The effect of secondary substrate binding in penicillopepsin the contributions of subsites S2′ and S3 to k_{cat}, Biochemistry 27 (1988) 1140–1147.

[11] A.P. Ryle, J. Leclerc, F. Falla, The substrate specificity of pepsin C, Biochem. J. 110 (1968) p4.

[12] J.S. Fruton, Specificity and mechanism of pepsin action on synthetic substrates, Adv. Exp. Med. Biol. 95 (1977) 131–140.

[13] J.C. Powers, A.D. Harley, D.V. Myers, Subsite specificity of porcine pepsin, Adv. Exp. Med. Biol. 95 (1977) 141–157.

[14] T. Hofmann, R.S. Hodges, A new chromogenic substrate for penicillopepsin and other fungal aspartic proteinases, Biochem. J. 203 (1982) 603–610.

[15] B.M. Dunn, B. Kammermann, K.R. McCurry, The synthesis, purification, and evaluation of a new chromophoric substrate for pepsin and other aspartyl proteases, Anal. Biochem. 138 (1084) 68–73.

[16] B.M. Dunn, M. Jimenez, B.F. Parten, M.J. Valler, C.E. Rolph, J. Kay, A systematic series of synthetic chromogenic substrates for aspartic proteinases, Biochem. J. 237 (1986) 899–906.

[17] J. Pohl, B.M. Dunn, Secondary enzyme-substrate interactions use of kinetic evidence for ionic interaction between substrate side chains and the pepsin active site, Biochemistry 27 (1988) 4827–4834.

[18] C. Rao, A Detailed Investigation into the Enzymatic Specificity of Porcine Pepsin, Ph.D. Thesis, University of Florida, Gainesville, FL, 1994.

[19] W.T. Lowther, B.M. Dunn, The promiscuous active site specificity/binding preferences of the fungal aspartic proteinase, Rhizopuspepsin, Protein Peptide Lett. 5 (1998) 303–316.

[20] P.E. Scarborough, K. Guruprasad, T.L. Blundell, G.R. Richo, G.E. Conner, B.M. Dunn, Exploration of subsite binding specificity of human cathepsin D through kinetics and knowledge-based molecular modeling, Protein Sci. 2 (1993) 264–276.

[21] J. Westling, P. Cipullo, S.-H. Hung, H. Saft, J.B. Dame, B.M. Dunn, Active site specificity of slasmepsin II, Protein Sci. 8 (1999) 2001–2009.

[22] B.M. Beyer, B.M. Dunn, Prime region subsite specificity characterizaion of human cathepsin D. The dominant role of position 128, Protein Sci. 7 (1997) 88–95.

[23] W.T. Lowther, P. Majer, B.M. Dunn, Engineering the sub-

strate specificity of Rhizopuspepsin: the role of Asp 77 of fungal aspartic proteinases in facilitating the cleavage of oligopeptide substrates with lysine in P_1, Protein Sci. 4 (1995) 689–702.

[24] P.E. Scarborough, B.M. Dunn, Redesign of the substrate specificity of human cathepsin D: the dominant role of position 287 in the S_2 subsite, Protein Eng. 7 (1994) 495–502.

[25] E.T. Baldwin, T.N. Bhat, S. Gulnik, M.V. Hosur, R.C. Sowder, R.E. Cachau, J. Collins, A.M. Silva, J.W. Erickson, Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design, Proc. Natl. Acad. Sci. USA 90 (1993) 6796–6800.

[26] J.A. Wells, D.B. Powers, R.R. Bott, T.P. Graycar, D.A. Estell, Designing substrate specificity by protein engineering of electrostatic interactions, Proc. Natl. Acad. Sci. USA 84 (1987) 1219–1223.

[27] A. Caputo, J.C. Parrish, M.N.G. James, J.C. Powers, R.C. Bleackley, Electrostatic reversal of serine proteinase substrate specificity, Proteins Struct. Funct. Genet. 35 (1999) 415–424.

[28] V. Olsen, K. Guruprasad, N.X. Cawley, H.-C. Chen, T.L. Blundell, Y.P. Loh, Cleavage efficiency of the novel aspartic protease yapsin 1 (Yap3p) enhanced for substrates with arginine residues flanking the P1 site: correlation with electronegative active site pockets predicted by molecular modeling, Biochemistry 37 (1998) 2768–2777.

[29] M.A. Taylor, K.C. Baker, I.F. Connerton, N.J. Cummings, G.W. Harris, I.M.J. Henderson, S.T. Jones, R.W. Pickersgill, I.G. Sumner, J. Warwicker, P.W. Boodenough, An unequivoval example of cysteine proteinase activity affected by multiple electrostatic interactions, Protein Eng. 7 (1994) 1267–1276.

[30] S. Pinitglang, A.B. Watts, M. Patel, J.D. Reid, M.A. Noble, S. Gul, A. Bokth, A. Naeem, H. Patel, E.W. Thomas, S.K. Sreedharan, C. Verma, A classical enzyme active center motif lacks catalytic competence until modulated electrostatically, Biochemistry 36 (1997) 9968–9982.

[31] Y. Ikeuchi, N.A. Katerelos, P.W. Goodenough, The enhancing of a cysteine proteinase activity at acidic pH by protein engineering, the role of glutamic 50 in the enzyme mechanism of caricain, FEBS Lett. 437 (1998) 91–96.