BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

*Biochemistry* **2009,** *48,* 5753–5759 5753
DOI: 10.1021/bi900461e

# Profiling the Substrate Specificity of Viral Protease VP4 by a FRET-Based Peptide Library Approach[†]

Ozlem Dogan Ekici,[‡] Jinge Zhu,[‡] Ivy Yeuk Wah Chung,[§] Mark Paetzel,[§] Ross E. Dalbey,[‡] and Dehua Pei*[,‡]

[‡]*Department of Chemistry, The Ohio State University, 100 West 18th Avenue, Columbus, Ohio 43210, and* [§]*Department of Molecular Biology and Biochemistry, Simon Fraser University, South Science Building, 8888, University Drive, Burnaby, British Columbia, Canada V5A 1S6*

*Received March 17, 2009; Revised Manuscript Received May 12, 2009*

ABSTRACT: Knowing the substrate specificity of a protease is useful in determining its physiological substrates, developing robust assays, and designing specific inhibitors against the enzyme. In this work, we report the development of a combinatorial peptide library method for systematically profiling the substrate specificity of endopeptidases. A fluorescent donor (Edans) and quencher (Dabcyl) pair was added to the C- and N-termini of a support-bound peptide. Protease cleavage of the peptide removed the N-terminal quencher, resulting in fluorescent beads, which were isolated and individually sequenced by partial Edman degradation and mass spectrometry (PED−MS) to reveal the peptide sequence, as well as the site of proteolytic cleavage. The method was validated with bovine trypsin and *Escherichia coli* leader peptidase and subsequently applied to determine the substrate specificity of a viral protease, VP4, derived from the blotched snakehead virus (BSNV). The results show that VP4 cleaves peptides with a consensus sequence of (Abu/Ala/Pro)-X-Ala↓X, in agreement with the previously observed cleavage sites in its protein substrates. Resynthesis and a solution-phase assay of several representative sequences against VP4 confirmed the library screening results.

Proteases make up one of the most important groups of enzymes, comprising nearly 2% of the human proteome. To elucidate the function of a protease and/or design selective substrates and inhibitors, it is often necessary to define the substrate specificity of the enzyme. A variety of methods have been reported, ranging from kinetic assay of individual peptides to various library approaches such as phage display (*1*, *2*), position scanning (*3*−*5*), peptide microarrays (*6*), fluorescence resonance energy transfer (FRET)[1] assays (*7*−*9*), and activity-based probes (*10*, *11*). While each of these methods has provided valuable information about the substrate specificity of proteases and several of them remain popular choices of investigators, they each suffer from some drawbacks. For example, position scanning libraries are straightforward to synthesize and screen, do not require individual peptide sequencing, and are able to reveal the site of cleavage but cannot probe the prime side of the substrates or give individual sequences. As a variation of the position scanning method, peptide array does give individual sequences but is practically limited to testing $10^3$−$10^4$ sequences. Substrate phage display has the advantage of being able to generate and screen very large libraries, interrogate the specificity on both sides of the scissile bond, and give individual substrate sequences but is limited to 20 proteinogenic amino acids and does not reveal the actual site of cleavage. The FRET assay as originally reported by Meldal and co-workers (*7*) employs a fluorescent donor and quencher pair, placed on the two termini of a resin-bound substrate peptide. Cleavage of the peptide by an endopeptidase removes the N-terminal quencher, and the bead becomes fluorescent. The identity of the substrate peptide and the site of cleavage are determined by conventional Edman sequencing of the partially cleaved peptides on the positive bead. The FRET method has many advantages. It is compatible with both natural and unnatural building blocks, is capable of handling large libraries, and gives individual sequences. The main drawback is that Edman sequencing is very time-consuming (hours to a day per peptide) and expensive (more than $300 per peptide), making the method less ideal for high-throughput profiling of a large number of proteases. To overcome this drawback, we have recently developed an inexpensive, high-throughput peptide sequencing method, termed partial Edman degradation and mass spectrometry (PED−MS), which is ideally suited for sequencing support-bound peptides and peptoids [poly(*N*-alkylglycine)] (*12*−*14*).

The identity of protease substrates and information about the specificity can also be obtained by chemical labeling techniques to

[†]This work was supported by grants from the National Institutes of Health (GM62820 to D.P.), the National Science Foundation (MCB-0316670 to R.E.D.), and the Canadian Institute of Health Research (to M.P.).

*To whom correspondence should be addressed. Phone: (614) 688-4068. Fax: (614) 292-1532. E-mail: pei.3@osu.edu.

[1]Abbreviations: Abu, L-2-aminobutyric acid; BSNV, blotched snakehead virus; Bz-OSu, *N*-hydroxysuccinimidyl benzoate; Dabcyl, 4-{[4-(dimethylamino)phenyl]azo}benzoic acid; Edans, 5-[(2-aminoethyl) amino]naphthalene-1-sulfonic acid; FRET, fluorescence resonance energy transfer; IPNV, infectious pancreatic necrosis virus; Nic-OSu, *N*-hydroxysuccinimidyl nicotinate; PITC, phenylisothiocyanate; PED−MS, partial Edman degradation and mass spectrometry.

 Published on Web 05/12/2009 

identify protein fragments generated by the addition of a protease or activation of a protease pathway (*15*, *16*). Also, information about substrates of proteases can be determined within a proteome, with or without protease activation, by examining all the proteins on a one-dimensional sodium dodecyl sulfate (SDS)−polyacrylamide gel with tryptic digestion and shotgun LC−MS/MS methods (*17*).

Viruses of the *Birnaviridae* family are characterized by their bisegmented double-stranded RNA genome (segments A and B) (*18*). The polyprotein generated by segment A includes proteins VP2, VP4, and VP3 in a linear arrangement. VP2 and VP3 are structural proteins, whereas VP4 is the protease responsible for releasing VP2, VP3, and itself from the polyprotein. Prototypical members of this virus family include the infectious pancreatic necrosis virus (IPNV) that infects fish and the infectious bursal disease virus (IBDV) that infects poultry. The aqua birnavirus blotched snakehead virus (BSNV) was first discovered by John and Richards in 1999 (*19*). The BSNV VP4 protease cleaves the BSNV segment A polyprotein at the VP2p−VP4 and VP4−VP3 junctions as well as several other sites located at the end of pVP2, producing the mature capsid protein VP2 along with the so-called X-peptide and a number of smaller peptides, some of which were found in the viral particle (*20*). Inspection of the cleavage sites in the BSNV segment A polyprotein suggested a cleavage preference for a Pro-X-Ala↓Ala motif, but there are some variations. The crystal structure of the BSNV VP4 protease (*21*) revealed that the shape and electrostatic nature of the VP4 binding site are generally consistent with the observed cleavage preferences. Given the very limited number of cleavage sites available and the spatial constraints (if the cleavage occurs intramolecularly), we felt that the above motif may not fully reflect the substrate specificity profile of VP4. In this work, we have combined the advantages of FRET-based combinatorial peptide libraries with PED−MS peptide sequencing to develop a powerful, high-throughput method for systematically profiling the substrate specificity of any endopeptidase. This method was first validated with trypsin and *Escherichia coli* leader peptidase, two enzymes whose substrate specificities had previously been well characterized (*3*, *9*, *22*). Subsequent application of the method to VP4 revealed its detailed substrate specificity profile.

## MATERIALS AND METHODS

*Materials*. PL-PEGA resin (0.2 mmol/g, 300−500 $\mu$m) was purchased from Polymer Laboratories Ltd. (Amherst, MA). All of the reagents for peptide synthesis were purchased from Advanced ChemTech (Louisville, KY), Novabiochem (San Diego, CA), or Bachem (Torrance, CA). Sodium 5-[(2-aminoethyl)amino]naphthalene-1-sulfonate (Edans) was purchased from Molecular Probes, Invitrogen Corp. (Carlsbad, CA). All other chemicals were purchased from Aldrich Chemical Co. (Milwaukee, WI) and Acros Organics. Bovine trypsin was purchased from Sigma (Milwaukee, WI). *E. coli* leader peptidase and VP4 protease were expressed and purified as previously described (*22*, *23*).

*Synthesis of the FRET-Based Peptide Library*. The peptide library was synthesized using the standard Fmoc/HBTU/HOBt synthesis protocol as previously reported (*22*) with the following modifications. The synthesis was conducted on 2.0 g of $PEGA_{1900}$ resin (300−500 $\mu$m, 0.2 mmol/g) in a homemade peptide synthesis apparatus (*24*). First, the seven-residue linker sequence, AAE(allyl)BBrM (where B is $\beta$-alanine and r is D-arginine), was synthesized using 4 equiv of reagents. Next, the random region was generated using the split-and-pool synthesis method (*25*−*27*). The resin was evenly divided into 19 aliquots and placed into 19 separate reaction vessels. Each aliquot was coupled with a different amino acid (4 equiv of reagent, 30 min), and the coupling reaction was repeated once. The resin from all the vessels was combined, mixed, and washed exhaustively with DMF, and the Fmoc group was removed with two treatments with 20% piperidine and DMF (5 + 15 min). The resin was then redistributed into 19 reaction vessels, and this process was repeated until all randomized positions were generated. The rest of the synthesis, including the addition of Fmoc-Lys(Dabcyl)-OH and the replacement of the allyl group from the C-terminal Glu with Edans and library deprotection, was conducted as previously described (*22*). Finally, the resin was washed with $CH_2Cl_2$ (5 × 10 mL) and stored in the same solvent at 4 °C.

*Library Screening*. A typical screening reaction involved ~30000 library beads, which were washed exhaustively with water and the proper protease reaction buffer [for trypsin, 0.1 M Hepes (pH 7.5) and 10 mM $CaCl_2$; for leader peptidase, 50 mM Tris-HCl (pH 8.0), 10 mM $CaCl_2$, and 1% Triton X-100; for VP4, 20 mM Tris-HCl, 100 mM NaCl, 0.1% Chaps, 1 mM EDTA, 10% glycerol, and 10 mM dithiothreitol (pH 7.4)]. The resin was suspended in 3 mL of the proper protease reaction buffer in a 60 mm × 15 mm Petri dish (Baxter Scientific Products) and treated with the desired protease (final concentrations of 0.063 $\mu$M, 0.083 mg/mL, and 0.059 mg/mL for trypsin, leader peptidase, and VP4, respectively) for 18 h at 37 °C. The dish was viewed under a fluorescence microscope (Olympus SZX12) using the appropriate filter set for the Edans group (exciter at 360 nm; emitter at 460 nm). Positive beads were identified by their intense fluorescence and removed from the library using a micropipet. A control screening was conducted under the same conditions with the exclusion of the proteases. This screening resulted in no fluorescent beads.

The positive beads from above were washed with water and pyridine, suspended in 160 $\mu$L of pyridine, and treated with 10% *N*-hydroxysuccinimidyl benzoate (Bz-OSu) for 12 min at room temperature. This treatment resulted in the benzoylation of any free N-terminus exposed by the enzymatic cleavage and the side chain of lysine. The beads were next treated with 20% piperidine in DMF to remove the N-terminal Fmoc group and individually sequenced by the PED−MS method using *N*-hydroxysuccinimidyl nicotinate (Nic-OSu) as the capping agent during PED reactions (*12*).

*Synthesis of Individual Peptides*. Individual peptide synthesis was performed on a 0.02 mmol scale on Rink amide resin using standard Fmoc chemistry. After side chain deprotection as described above, the peptides were purified by reversed-phase HPLC on a semipreparative C-18 column, eluted with a linear gradient from 25 to 50% $CH_3CN$ in $H_2O$ in the presence of 0.05% TFA over 26 min (monitored at 336 nm). Major fractions were subjected to ESI-MS analysis, and the fractions containing the desired peptides were evaporated in a SpeedVac concentrator to produce a reddish solid. All peptides were dissolved in $H_2O$, and their concentrations were determined on the basis of absorption at 495 nm ($\varepsilon_{495} = 2.07 \times 10^4\ M^{-1}\ cm^{-1}$).

*VP4 Protease Activity Assay*. The activity assay was conducted in a buffer (total volume of 0.5 mL) containing 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10% glycerol, 1% $\beta$-mercaptoethanol, and varying concentrations of peptide substrates (0−120 $\mu$M). The reaction was initiated by the addition of

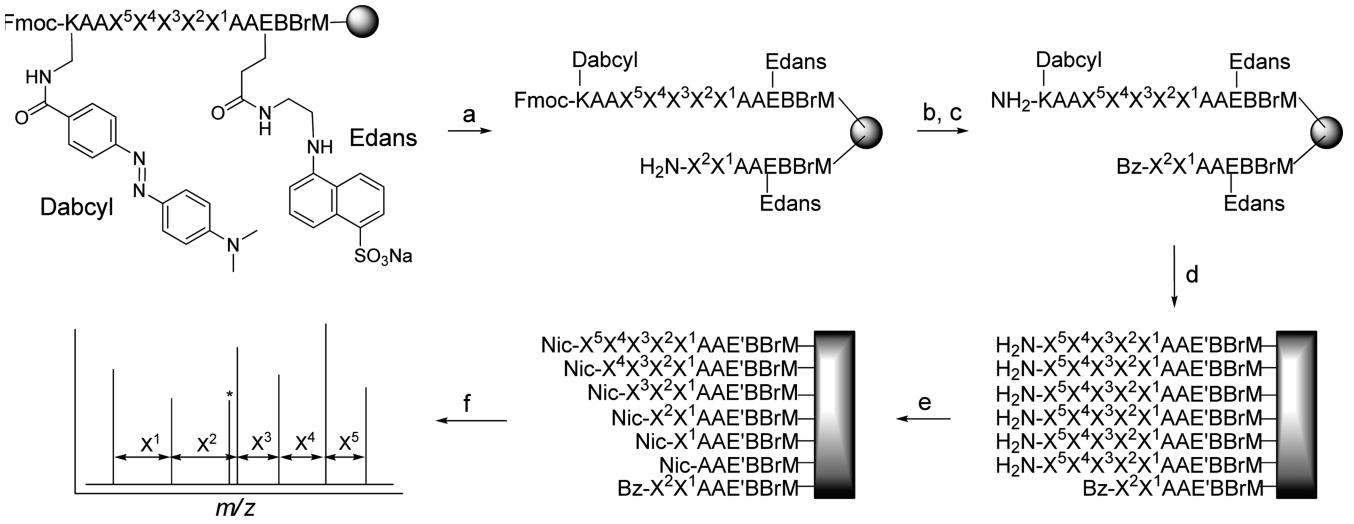


FIGURE 1: Design and screening of the FRET-based peptide library and postscreening sequence determination by PED−MS: (a) protease treatment, (b) Bz-OSu, (c) 20% piperidine in DMF, (d) PITC and then TFA (three times), (e) 5:1 PITC/Nic-OSu and then TFA (six times), and (f) CNBr. The example shown assumes protease cleavage between $X^2$ and $X^3$ residues. Bz, benzoyl; Nic, nicotinoyl; E′, Glu(Edans). The peak labeled with an asterisk is derived from the N-terminally benzoylated protease cleavage product; together with PED fragment Nic-$X^2X^1$AAE′BBrM, they form a doublet peak with $m/z$ ratios of $m$ and $m - 1$.

VP4 enzyme (final concentration of 0.01−0.02 mg/mL) and monitored continuously at 520 nm on an Aminco-Bowman Series 2 spectrometer (excitation at 340 nm) at room temperature. The initial rates calculated from the early regions of the progress curves (usually < 200 s) were fitted against the Michaelis−Menten equation [$V = k_{cat}[E][S]/(K_M + [S])$] using KaleidaGraph 3.6 to obtain the $k_{cat}$ and $K_M$ values. To define the relationship between fluorescence yield and product formation, one of the peptides (peptide III) was hydrolyzed to completion by incubation with excess VP4 overnight at 4 °C. Each unit of fluorescence increase at 520 nm under the experimental conditions described above corresponded to 32 $\mu$M of product formation.

*MS Analysis of VP4 Cleavage Products*. VP4 reactions were conducted as described above, except that the reactions were allowed to proceed overnight at 4 °C. The reaction mixtures were directly analyzed by ESI and/or MALDI mass spectrometry. The site of cleavage was identified on the basis of the masses of the reaction products. Control reactions in the absence of VP4 protease showed that the peptides were unaltered for several weeks at 4 °C.

## RESULTS AND DISCUSSION

*Design, Synthesis, and Screening of the FRET-Based Peptide Library*. A one-bead-one-compound (OBOC) library containing five random positions, Fmoc-K(Dabcyl)AA$X^5X^4X^3$-$X^2X^1$AAE(Edans)BBrM-resin [where $X^1$−$X^5$ represent L-2-aminobutyric acid (Abu or U) or any of the 18 proteinogenic amino acids except for Cys and Met, B is $\beta$-alanine, and r is D-arginine], was designed (Figure 1). The C-terminal sequence (BBrM) was added to provide a flexible linker (BB) and facilitate release of the peptide from the resin (CNBr cleavage after Met) and MS analysis (Arg) (*28*). D-Arginine was chosen to avoid undesired cleavage after the arginine by a trypsin-like activity. The N-terminal Lys and C-terminal Glu were included as handles, to which the fluorescence quencher (Dabcyl) and donor (Edans), respectively, were attached. The Ala residues before and after the random sequence were intended to move the bulky Dabcyl and Edans groups away from the protease cleavage site, which could potentially block access of the protease to the random region. The theoretical diversity of the library is $19^5$ or $2.48 \times 10^6$. The library was synthesized on 2 g of $PEGA_{1900}$ resin (~330000 beads, 0.2 mmol/g) using the split-and-pool synthesis method as previously described (*25*−*27*). $PEGA_{1900}$ resin was chosen as the solid support because of its ability to swell in the aqueous environment, permitting the permeation of relatively large biomolecules such as proteins into the beads (*29*).

Prior to protease treatment, the library beads were nonfluorescent because the fluorescence of the Edans group is efficiently quenched by the N-terminal Dabcyl group. Protease cleavage at the random region would remove the quencher, and the C-terminal Edans group, which remains bound to the bead, would render the bead intensely fluorescent. The fluorescent beads are readily detected under a fluorescence microscope and manually removed from the library. The peptide sequences on the positive beads were determined by the PED−MS method (*12*). To determine the site of protease cleavage, the original PED−MS method was slightly modified as follows. A positive bead selected from the library was treated with *N*-hydroxysuccinimidyl benzoate (Bz-OSu); this resulted in the benzoylation of the free N-terminus exposed by the protease action and lysine side chains (Figure 1). Next, the N-terminal Fmoc group of any unreacted full-length peptide was removed by treatment with 20% piperidine in DMF. The resulting bead was subjected to three cycles of Edman degradation to remove the N-terminal KAA sequence, followed by five cycles (or more) of PED using a 5:1 (molar) mixture of phenyl isothiocyanate (PITC) and Nic-OSu (*12*). Finally, the PED products were released from the bead with cyanogen bromide and analyzed by MALDI-MS. The sequence of the original peptide on the bead was inferred from the peptide fragment pattern in the mass spectrum as previously described (*12*−*14*). A pair of doublet peaks at $m/z$ $m$ and $m/z$ $m - 1$ indicates that the protease cleavage site is between the two residues surrounding the doublet peaks (Figure 1). Note that benzoylation of the lysine side chain has the beneficial effect of differentiating lysine from glutamine, which both have the same nominal residual mass of 128. Benzoylation of the lysine side chain increases its residual mass to 232, while that of glutamine

Table 1: Most Preferred Peptide Substrates of Bovine Trypsin[a]

| bead no. | peptide sequence | bead no. | peptide sequence |
|---|---|---|---|
| 1 | AAKRKHR↓AA | 20 | AAKKER↓QAA |
| 2 | AAXXXXR↓AA | 21 | AAKVAR↓KAA |
| 3 | AAKHHRR↓AA | 22 | AAGKRR↓QAA |
| 4 | AAXSFGR↓AA | 23 | AATQPR↓KAA |
| 5 | AAQXXXR↓AA | 24 | AAEGDR↓QAA |
| 6 | AAPRAR↓NAA | 25 | AAEHQR↓AAA |
| 7 | AAXXQR↓UAA | 26 | AAHUKQR↓AA |
| 8 | AARVGR↓HAA | 27 | AAKSR↓HLAA |
| 9 | AAQTKUR↓AA | 28 | AAXXSGR↓AA |
| 10 | AARRKQR↓AA | 29 | AAXXK↓GRAA |
| 11 | AAERGAR↓AA | 30 | AATHKUK↓AA |
| 12 | AAKSR↓LHAA | 31 | AATGKK↓GAA |
| 13 | AAGKR↓QRAA | 32 | AAETLK↓UAA |
| 14 | AAKXXR↓KAA | 33 | AAKAKK↓UAA |
| 15 | AAXXRKR↓AA | 34 | AAEEHSK↓AA |
| 16 | AAEVLPR↓AA | 35 | AAXXKQK↓AA |
| 17 | AAXXKGR↓AA | 36 | AAETTK↓UAA |
| 18 | AAEGR↓ULAA | 37 | AAKRK↓HRAA |
| 19 | AAXXXR↓NAA | 38 | AARRK↓QRAA |

[a] U, 2-aminobutyric acid; X, unidentified amino acid. The down arrow denotes the site of cleavage. The random region of the FRET library is underlined.

remains 128. Leucine and isoleucine, two other amino acids of degenerate residual mass (113), were not differentiated in this library.

*Evaluation of the FRET Library against Trypsin and Leader Peptidase*. The FRET library was first tested on bovine trypsin and *E. coli* leader peptidase, two enzymes whose substrate specificities had previously been well-defined (*3*, *9*, *22*). Approximately 180 mg of the library (~30000 beads) was incubated with 0.063 $\mu$M bovine trypsin for 18 h at pH 7.5. A total of 55 highly fluorescent beads were collected and sequenced by the PED−MS method. The site of cleavage was determined unambiguously for 38 of the 55 beads. Among the 38 sequences, 28 were cleaved after an Arg residue and the other 10 were digested after a Lys (Table 1). Thus, consistent with previous reports (*3*), trypsin has a strong preference for a positively charged residue at the $P_1$ position. It does not have any obvious preference at other positions. Approximately 180 mg of the library was also screened against *E. coli* leader peptidase, resulting in 51 fluorescent beads. Sequencing by PED−MS produced 50 unambiguous sequences, all of which contained an Ala-X-Ala general motif, known to be the recognition motif of leader peptidase (Table 2). The cleavage site was determined for 26 of the 50 beads, and all 26 peptides were cleaved between two Ala residues (Ala as $P_1$ and $P_1'$ residues). Among them, 25 peptides were cleaved between the two fixed Ala-Ala residues just C-terminal to the random region, where the remaining one (bead 1) was cleaved immediately N-terminal to the Ala-Ala motif. At the $P_3$ site, all 26 peptides contained a small residue, including 15 Ala residues, eight Abu residues, two Val residues, and one Ser residue. On the other hand, a variety of residues were found at the $P_2$ position. Interestingly, the leader peptidase appears to prefer a proline at the $P_6$ position (13 of the 26 sequences had Pro at the $P_6$ position). This trend was not observed in the earlier studies, because the previous libraries contained only four random positions (*9*, *22*). A proline at the $P_6$ (or $P_5$) position may help break the transmembrane α-helix of signal peptides and facilitate the interaction between a peptide substrate and the protease active site (*30*).

Table 2: Preferred Peptide Substrates of *E. coli* Leader Peptidase[a]

| bead no. | peptide sequence | bead no. | peptide sequence |
|---|---|---|---|
| 1 | AAQRALA↓AA | 14 | AAPGYAEA↓A |
| 2 | AAUQEAQA↓A | 15 | AAWUFAQA↓A |
| 3 | AAEKSAHA↓A | 16 | AAPNKUUA↓A |
| 4 | AAPGLAEA↓A | 17 | AAHEDULA↓A |
| 5 | AAPRLAUA↓A | 18 | AAPRNUFA↓A |
| 6 | AAAUPAKA↓A | 19 | AAPKRUAA↓A |
| 7 | AAPGVALA↓A | 20 | AAARLURA↓A |
| 8 | AAXXEANA↓A | 21 | AAPKGUKA↓A |
| 9 | AAVXXAKA↓A | 22 | AAPKUUQA↓A |
| 10 | AANRNAQA↓A | 23 | AAGKTUKA↓A |
| 11 | AARATAWA↓A | 24 | AAPKRVEA↓A |
| 12 | AAPRKATA↓A | 25 | AAPRAVEA↓A |
| 13 | AAQKRAQA↓A | 26 | AAPKKSLA↓A |

[a] U, 2-aminobutyric acid; X, unidentified amino acid. The down arrow denotes the site of cleavage. The random region of the FRET library is underlined.

We have noticed that sequence determination of positive hits from the FRET library by PED−MS has a somewhat lower success rate than other libraries (non-FRET-based), which are typically > 90% (*12*, *13*). This is at least partially due to excessive protease reaction, leaving an insufficient amount of full-length peptide on the bead for PED−MS analysis. Trypsin, which is a very active protease, had a higher percentage of undefined sequences at the N-terminal side of the scissile bond as compared to the leader peptidase (which is less active against peptide substrates) (Tables 1 and 2). Another contributing factor is the fact that the Edans group at the C-terminus undergoes photolytic degradation during MALDI-TOF analysis, resulting in more complex MS spectra (vide infra). Finally, the use of Bz-OSu to block the protease-exposed N-terminus produces doublet peaks separated by only 1 amu; it is sometimes difficult to differentiate the $m/z$ $m - 1$ peak (especially when its intensity is low) from noise peaks. The latter has undoubtedly contributed to the less-than-desirable successive rate in identifying the sites of cleavage (~61%). These problems can be avoided by controlling the extent of protease cleavage (a few percent of cleavage is usually sufficient to render a bead intensely fluorescent), removing the Edans group prior to MS analysis, and/or using a different acylating agent (other than Bz-OSu). These studies are already underway in our laboratories.

*Substrate Specificity of Viral Protease VP4*. Approximately 180 mg of the FRET library (~30000 beads) was screened against protease VP4 (0.059 mg/mL) in four equal aliquots to give a total of ~50 fluorescent beads. PED−MS sequencing of these beads produced 28 unambiguous sequences, and the protease cleavage site(s) was determined for 22 of the peptides (Table 3). The peptide sequences and/or the site of cleavage for the other beads could not be determined due to poor spectral quality. Because eight of the 22 peptides were cleaved at two different sites, a total of 30 cleavage sites were identified. Like the *E. coli* leader peptidase, most of the VP4 cleavage reactions occurred after the invariant C-terminal Ala-Ala motif instead of the random region. Inspection of the cleavage sites reveals a consensus recognition motif of (Abu/Ala/Pro)-X-Ala↓X (Figure 2). Five of the six peptides whose site(s) of cleavage was not identified also contain at least one such motif in each sequence. Presumably, these peptides were also cleaved at the consensus motifs. Thus, VP4 strongly prefers an alanine at the $P_1$ position, with the sterically similar Abu as the only other accepted residue.

Table 3: Selected VP4 Substrates[a]

| bead no. | peptide sequence | bead no. | peptide sequence |
|---|---|---|---|
| 1 | RWURU↓AA↓E | 15 | KKKVPAA↓E |
| 2 | RKUKU↓AA↓E | 16 | HRKHVAA↓E |
| 3 | KRURA↓AA↓E | 17 | TLKHUAA↓E |
| 4 | RAURA↓AA↓E | 18 | KNKUUAA↓E |
| 5 | RUTRU↓AA↓E | 19 | RKRUUA↓A↓E |
| 6 | KKURAAA↓E | 20 | TRRULA↓AE |
| 7 | SRPRPAA↓E | 21 | SRRAUA↓AE |
| 8 | SVGRGAAE[b] | 22 | RRKULA↓AE |
| 9 | RSKLU↓AA↓E | 23 | KRRPLA↓AE |
| 10 | RRYLA↓AA↓E | 24 | GVAARAAE[b] |
| 11 | KTRHUAA↓E | 25 | DEETVAAE[b] |
| 12 | RSRVUAAE[b] | 26 | RRLRPAA↓E |
| 13 | DRRVVAA↓E | 27 | RTHKUAAE[b] |
| 14 | RPYRAAA↓E | 28 | SURRKAAE[b] |

[a] U, 2-aminobutyric acid. The down arrow denotes the site of cleavage. [b] The cleavage sites of these cleaved peptides were not identified due to poor MS spectral quality. The random region of the FRET library is underlined.

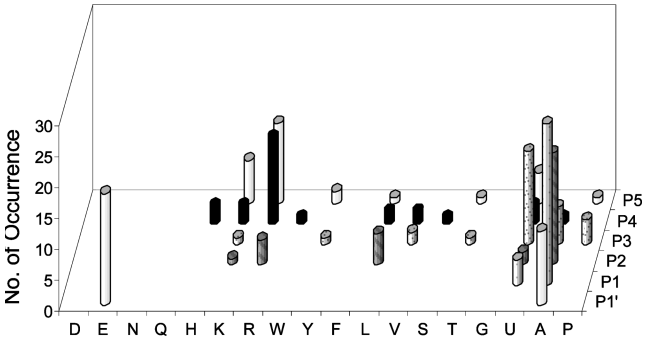


FIGURE 2: Substrate specificity of VP4 protease. The histograms display the amino acids identified at each position ($P_1'$ to $P_5$). The number of occurrences on the *y*-axis represents the number of selected peptide sequences that contained a particular amino acid at a certain position. U, 2-aminobutyric acid; L, leucine or isoleucine.

It also requires a small hydrophobic residue at the $P_3$ position, with a preference for Abu (15 of 30), Ala (six of 30), Pro (four of 30), Val (two of 30), and Thr (one of 30). VP4 appears to have broad specificity at the $P_1'$ position, since both Ala and the bulky Glu(Edans) are accepted. It also tolerates a variety of residues of different physicochemical properties at the $P_2$ position (e.g., Ala, Abu, Arg, Leu/Ile, and Lys). At the $P_4$ and $P_5$ positions, the selected peptides show an overrepresentation of positively charged and hydrophobic residues. Some preference for positively charged residues also extends to the $P_6$ and $P_7$ positions.

*Kinetic Properties of Selected VP4 Substrates and Confirmation of the Cleavage Site*. Three representative peptides selected from the library, K(Dabcyl)AKRURA↓AA↓E(Edans)G, K(Dabcyl)ARSKLU↓AA↓E(Edans)G, and K(Dabcyl)ARRLRPAA↓E(Edans)G (peptides I−III in Table 4), were individually synthesized to confirm the screening results. A fourth peptide [K(Dabcyl)ARSKLAAAE(Edans)G], which is analogous to peptide II, was designed to test whether protease VP4 indeed prefers an Abu over Ala at the $P_3$ position. The four peptides were first treated with excess VP4 overnight, and the cleavage products were analyzed by both ESI and MALDI-TOF mass spectrometry to determine the site of cleavage. Figure 3 shows the MALDI-TOF mass spectra of untreated peptide III (panel A) and peptide III after treatment with VP4 (panel B). Peptide III gave a major peak at *m*/*z* 1724.1, consistent with the calculated mass of the peptide (M + 1). The other peaks of lower *m*/*z* values were derived from peptide III, due to breakdown of the Edans and/or Dabcyl groups during the MS analysis. After overnight treatment with VP4, the peak at *m*/*z* 1724.2 (and those of its breakdown products) exhibited much reduced abundance, while an intense new peak appeared at *m*/*z* 1289.9. The loss of 434.3 amu indicates that VP4 cleaved peptide III between Ala and Glu(Edans) residues, consistent with the site predicted by PED−MS analysis. No signal that corresponds to any cleavage at other positions was observed (the *m*/*z* 1156.8 peak is likely a breakdown product derived from the *m*/*z* 1289.9 species). Similar analyses showed that VP4 also cleaved peptides II and IV exclusively at the same site. For peptide I, VP4 cleavage occurred predominantly between Ala and Glu(Edans), but it also resulted in minor cleavage after the URA motif (data not shown). Therefore, the solution-phase analyses indicate that the major cleavage site for these four peptides (and likely all of the selected peptides in Table 3) was at the Ala-Glu(Edans) peptide bond. During library screening, it is possible that restricted accessibility to the major cleavage site (the peptides were attached to the solid support via their C-termini) allowed the enzyme to generate a small amount of cleavage at the more N-terminal sites (which were presumably more accessible).

Peptides I−IV were next assayed against VP4, and their kinetic constants were determined at pH 8.0 (Table 4). The three peptides selected from the library had similar kinetic constants, with their $k_{cat}$, $K_M$, and $k_{cat}/K_M$ values of 0.075−0.35 $min^{-1}$, 30−71 $\mu$M, and 41−166 $M^{-1}s^{-1}$, respectively. Substitution of Ala for Abu at the $P_3$ position reduced the VP4 activity by ~3-fold, consistent with the observation that Abu was the most frequently selected by VP4 at this position. The activities of VP4 against peptide substrates are low as compared to that of classical serine proteases (e.g., trypsin) but are not unusual among the subfamily of serine proteases that utilize a Ser-Lys dyad as the principal catalytic machinery (*31*, *32*).

*Comparison with Known Cleavage Sites of VP4*. BSNV VP4 is synthesized as a polyprotein ($NH_2$-pVP2-X-VP4-VP3-COOH) that is processed through the proteolytic activity of VP4. The polyprotein is cleaved at four different sites (between residues 417 and 418, 486 and 487, 557 and 558, and 791 and 792) to release VP2, peptide X, VP4, and VP3 (*20*). Inspection of the four cleavage sites (Table 5) shows that the $P_1$ residue in all four sites is alanine. The $P_3$ residues are proline (*2*), cysteine (*1*), and alanine (*1*), while the $P_1'$ and $P_2$ residues are highly variable. The $P_4$ and $P_5$ residues are positively charged (Arg) or hydrophobic (Ile and Tyr). Thus, the specificity profile revealed by our library approach is in excellent agreement with the in vivo cleavage sites. A related protease, the VP4 of infectious pancreatic necrosis virus (IPNV), cleaves its own polyprotein precursor at six different sites and has a very similar recognition motif, (Ser/Thr)-X-Ala↓(Ser/Ala)-Gly (*33*).

The crystal structures of BSNV and IPNV VP4 with and without bound substrates or intermediates have previously been determined (*21*, *33*). The substrate is bound to the enzyme active site in an extended conformation, forming an antiparallel $\beta$-sheet with the $\beta$-strands that line the VP4 binding groove (Val546−Val549 and Gly627−Ser633). The VP4 $S_1$ and $S_3$ subsites are formed by primarily hydrophobic residues, and the $P_1$ and $P_3$ side chains fit snugly into the small pockets, providing an explanation for why small, hydrophobic residues are strongly preferred at these two positions. The side chains of $P_2$ and $P_1'$ residues point into the solvent, suggesting that any amino acid (except for proline) should be acceptable at these positions. The $P_4$−$P_6$ side

Table 4: Kinetic Constants of Selected VP4 Substrates[a]

| entry no. | peptide sequence | $k_{cat}$ ($min^{-1}$) | $K_M$ ($\mu M$) | $k_{cat}/K_M$ ($M^{-1}$ $s^{-1}$) |
|---|---|---|---|---|
| I | K(Dabcyl)AKRURAAA↓E(Edans)G-OH | $0.17 \pm 0.02$ | $71 \pm 19$ | 41 |
| II | K(Dabcyl)ARSKLUAA↓E(Edans)G-OH | $0.075 \pm 0.010$ | $30 \pm 2$ | 43 |
| III | K(Dabcyl)ARRLRPAA↓E(Edans)G-OH | $0.35 \pm 0.02$ | $35 \pm 4$ | 166 |
| IV | K(Dabcyl)ARSKLAAA↓E(Edans)G-OH | $0.035 \pm 0.005$ | $43 \pm 13$ | 14 |

[a] U, 2-aminobutyric acid. The down arrow denotes the experimentally observed site of cleavage.

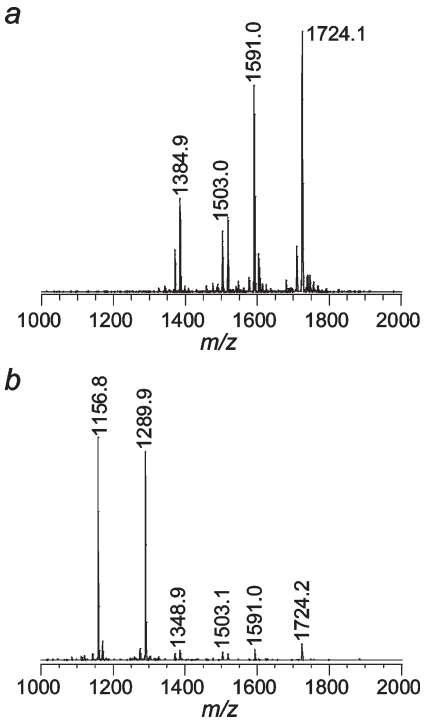


FIGURE 3: MALDI-TOF mass spectra of peptide III before (a) and after (b) overnight treatment with protease VP4.

Table 5: Known in Vivo Cleavage Sites of BSNV VP4

| BSNV cleavage sites | peptide sequence |
|---|---|
| VP2−pVP2 (residues 417 and 418) | KIAGA↓FGWG |
| pVP2−X (residues 486 and 487) | RIPLA↓SSDE |
| X−VP4 (residues 557 and 558) | LRPQA↓ADLP |
| VP4−VP3 (residues 791 and 792) | FYCGA↓ADEE |

chains make hydrophobic contacts with shallow pockets on the VP4 surface. Positively charged residues such as Arg and Lys are selected from the library probably because the hydrocarbon portions of their side chains can engage in hydrophobic interactions.

*Conclusion*. We have developed a new combinatorial library method for systematically and rapidly profiling the substrate specificity of proteases and endopeptidases by coupling the previously reported FRET-based assays with our high-throughput peptide sequencing method (PED−MS). The effectiveness of the current method was demonstrated with three different proteases, including the viral protease VP4. VP4 cleaves after (Abu/Ala/Pro)-X-Ala motifs, a specificity profile similar to that of bacterial leader peptidase. Our method has several advantages over the previously reported methods. It is capable of identifying individual peptide sequences and the actual sites of proteolytic cleavage. It is compatible with peptide libraries containing nonproteinogenic amino acids. Furthermore, hit identification by PED−MS is much faster and less expensive than traditional Edman sequencing. Compared to some of the other methods (e.g., phage display), a drawback of the current method is that our library is presently limited to $10^7$ sequences or five completely random amino acid residues. Also, in carrying out our method, one must take care to limit the extent of protease reaction (preferably $<50\%$ substrate-to-product conversion) so that sufficient full-length peptides are left for PED−MS sequencing.

## REFERENCES

1. Matthews, D. J., and Wells, J. A. (1993) Substrate phage: Selection of protease substrates by monovalent phage display. *Science 260*, 1113–1117.
2. Smith, M. M., Shi, L., and Navre, M. (1995) Rapid identification of highly active and selective substrates for stromelysin and matrilysin using bacteriophage display libraries. *J. Biol. Chem. 270*, 6440–6449.
3. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc. Natl. Acad. Sci. U.S.A. 97*, 7754–7759.
4. Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S., and Ellman, J. E. (2000) Synthesis of position-scanning libraries of fluorogenic peptide substrates to define the extended substrate specificity of plasmin and thrombin. *Nat. Biotechnol. 18*, 187–193.
5. Turk, B. E., Huang, L. L., Piro, E. T., and Cantley, L. C. (2001) Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. *Nat. Biotechnol. 19*, 661–665.
6. Salisbury, C. M., Maly, D. J., and Ellman, J. E. (2002) Peptide microarrays for the determination of protease substrate specificity. *J. Am. Chem. Soc. 124*, 14868–14870.
7. Meldal, M., Svendsen, I., Breddam, K., and Auzanneau, F.-I. (1994) Portion-mixing peptide libraries of quenched fluorogenic substrates for complete subsite mapping of endoprotease specificity. *Proc. Natl. Acad. Sci. U.S.A. 91*, 3314–3318.
8. Hilaire, P. M. S., Willert, M., Juliano, M. A., Juliano, L., and Meldal, M. (1999) Fluorescence-quenched solid phase combinatorial libraries in the characterization of cysteine protease substrate specificity. *J. Comb. Chem. 1*, 509–523.
9. Rosse, G., Kueng, E., Page, M. G. P., Schauer-Vukasinovic, V., Giller, T., Lahm, H.-W., Hunziker, P., and Schlatter, D. (2000) Rapid identification of substrates of novel proteases using a combinatorial peptide library. *J. Comb. Chem. 2*, 461–466.
10. Liu, Y., Patricelli, M. P., and Cravatt, B. F. (1999) Activity-based protein profiling: The serine hydrolases. *Proc. Natl. Acad. Sci. U.S.A. 96*, 14694–14699.
11. Bogyo, M., Berhelst, S., Bellingard-Dubouchaud, V., Toba, S., and Greenbaum, D. (2000) Selective targeting of lysosomal cysteine proteases with radiolabelled electrophilic substrate analogs. *Chem. Biol. 7*, 27–38.
12. Sweeney, M. C., and Pei, D. (2003) An improved method for rapid sequencing of support-bound peptides by partial Edman degradation and mass spectrometry. *J. Comb. Chem. 5*, 218–222.
13. Thakkar, A., Wavreille, A. S., and Pei, D. (2006) Traceless capping agent for peptide sequencing by partial Edman degradation and mass spectrometry. *Anal. Chem. 78*, 5935–5939.
14. Thakkar, A., Cohen, A. S., Connolly, M. D., Zuckermann, R. N., and Pei, D. (2009) High-throughput sequencing of peptoids and peptide-peptoid hybrids by partial Edman degradation and mass spectrometry. *J. Comb. Chem. 11*, 294–302.
15. Schilling, O., and Overall, C. M. (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat. Biotechnol. 26*, 685–694.

16. Mahrus, S., Trinidad, J. C., Barkan, D. T., Sali, A., Burlingame, A. L., and Wells, J. A. (2008) Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell 134*, 866–876.
17. Dix, M. M., Simon, G. M., and Cravatt, B. F. (2008) Global mapping of the topography and magnitude of proteolytic events in apoptosis. *Cell 134*, 679–691.
18. Dobos, P., Hill, B. J., Hallett, R., Kells, D. T., Becht, H., and Teninges, D. (1979) Biophysical and biochemical characterization of five animal viruses with bisegmented double-stranded RNA genomes. *J. Virol. 32*, 593–605.
19. John, K. R., and Richards, R. H. (1999) Characteristics of a new birnavirus associated with a warm-water fish cell line. *J. Gen. Virol. 80*, 2061–2065.
20. Da Costa, B., Soignier, S., Chevalier, C., Henry, C., Thory, C., Huet, J. C., and Delmas, B. (2003) Blotched snakehead virus is a new aquatic birnavirus that is slightly more related to avibirnavirus than to aquabirnavirus. *J. Virol. 77*, 719–725.
21. Feldman, A. R., Lee, J., Delmas, B., and Paetzel, M. (2006) Crystal structure of a novel viral protease with a serine/lysine catalytic dyad mechanism. *J. Mol. Biol. 358*, 1378–1389.
22. Ekici, O. D., Karla, A., Paetzel, M., Lively, M. O., Pei, D., and Dalbey, R. E. (2007) Altered −3 substrate specificity of *Escherichia coli* signal peptidase I mutants as revealed by screening a combinatorial peptide library. *J. Biol. Chem. 282*, 417–425.
23. Lee, J., Feldman, A. R., Delmas, B., and Paetzel, M. (2006) Expression, purification and crystallization of a birnavirus-encoded protease, VP4, from blotched snakehead virus (BSNV). *Acta Crystallogr. F62*, 353–356.
24. Wavreille, A. S., Garaud, M., Zhang, Y., and Pei, D. (2007) Defining SH2 domain and PTP specificity by screening combinatorial peptide libraries. *Methods 42*, 207–219.
25. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., and Knapp, R. J. (1991) A new type of synthetic peptide library for identifying ligand-binding activity. *Nature 354*, 82–84.
26. Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T., and Cuervo, J. H. (1991) Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature 354*, 84–86.
27. Furka, A., Sebestyen, F., Asgedom, M., and Dibo, G. (1991) General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res. 37*, 487–493.
28. Sweeney, M. C., Wavreille, A. S., Park, J., Butchar, J. P., Tridandapani, S., and Pei, D. (2005) Decoding protein-protein interactions through combinatorial chemistry: Sequence specificity of SHP-1, SHP-2, and SHIP SH2 domains. *Biochemistry 44*, 14932–14947.
29. Kress, J., Zanaletti, R., Amour, A., Ladlow, M., Frey, J. G., and Bradley, M. (2002) Enzyme accessibility and solid supports: Which molecular weight enzymes can be used on solid supports? An investigation using confocal Raman microscopy. *Chem.—Eur. J. 8*, 3769–3772.
30. von Heijne, G. (1983) Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem. 133*, 17–21.
31. Beebe, K. D., Shin, J., Peng, J., Chaudhury, C., Khera, J., and Pei, D. (2000) Substrate recognition through a PDZ domain in tail-specific protease. *Biochemistry 39*, 3149–3155.
32. Dev, I. K., Ray, P. H., and Novak, P. (1990) Minimum substrate sequence for signal peptidase I of *Escherichia coli*. *J. Biol. Chem. 265*, 20069–20072.
33. Lee, J., Feldman, A. R., Delmas, B., and Paetzel, M. (2007) Crystal structure of the VP4 protease from infectious pancreatic necrosis virus reveals the acyl-enzyme complex for an intermolecular self-cleavage reaction. *J. Biol. Chem. 282*, 24928–24937.