Article

Fluorescence-Quenched Solid Phase Combinatorial Libraries in the Characterization of Cysteine Protease Substrate Specificity

Phaedria M. St. Hilaire, Marianne Willert, Maria Aparecida Juliano, Luiz Juliano, and Morten Meldal

J. Comb. Chem., **1999**, 1 (6), 509-523• DOI: 10.1021/cc990031u • Publication Date (Web): 21 September 1999

Fluorescence-Quenched Solid Phase Combinatorial Libraries in the Characterization of Cysteine Protease Substrate Specificity†

Phaedria M. St. Hilaire,*,‡ Marianne Willert,‡ Maria Aparecida Juliano,§ Luiz Juliano,§ and Morten Meldal*,‡

Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby-Copenhagen, Denmark, and Department of Biophysics, Escola Paulista de Medicina, Rua Tres de Maio 100, 04044-20 São Paulo, Brazil

Received June 16, 1999

To map the substrate specificity of cysteine proteases, two combinatorial peptide libraries were synthesized and screened using the archetypal protease, papain. The use of PEGA resin as the solid support for library synthesis facilitated the application of an on-resin fluorescence-quenched assay. Results from the screening of library 2 indicated a preference for Pro or Val in the S_3 subsite and hydrophobic residues in S_2; the most prevalent residue not being Phe but Val. The S_1 subsite exhibited a dual specificity for both small, nonpolar residues, Ala or Gly, as well as larger, Gln, and charged residues, Arg. Small residues predominated in the $S_1'-S_4'$ subsites. Active peptides from the libraries and variations thereof were resynthesized and their kinetics of hydrolysis by papain assessed in solution phase assays. Generally, there was a good correlation between the extent of substrate cleavage on solid phase and the k_{cat}/K_M's obtained in solution phase assays. Several good substrates for papain were obtained, the best substrates being Y(NO_2)PMPPLCTSMK(Abz) (k_{cat}/K_M = 2109 (mM s)$^{-1}$), Y(NO_2)PYAVQSPQK(Abz) (k_{cat}/K_M = 1524 (mM s)$^{-1}$), and Y(NO_2)PVLRQQRSK-(Abz) (k_{cat}/K_M = 1450 (mM s)$^{-1}$). These results were interpreted in structural terms by the use of molecular dynamics (MD). These MD calculations indicated two different modes for the binding of substrates in the narrow enzyme cleft.

Introduction

Combinatorial chemistry has revolutionized both drug discovery and fundamental approaches to the elucidation of various processes in the biological as well as the physical sciences. The development of the portion-mixing (split/mix or divide and combine) methodology[1,2] has made possible the rapid generation of millions of compounds for high throughput screening. An important advance in the rapid screening of these vast libraries has been the development of solid supports such as TentaGel,[3] PEGA,[4] POEPOP,[5] POEPS-3,[6] and SPOCC[7] that allow the screening of the libraries to be performed directly on resin-bound compounds. However, while solid phase assays are expedient, hits obtained from the screening process may not necessarily lead to highly active ligands for the particular receptor being studied. A thorough understanding of the system being studied as well as the advantages and limitations of the solid phase assay is necessary in order to correctly interpret the results.

One of our research goals is the application of the solid phase combinatorial methodology to characterize proteolytic enzymes, in particular cysteine proteases as a preface to the design of their specific inhibitors. To this end, we have developed an intramolecular fluorescence quenching assay for the determination of the substrate[8−11] as well as inhibitor specificity[12,13] of these enzymes. The conventional methods of characterizing the substrate specificity of proteolytic enzymes involve the systematic yet tedious synthesis of several substrates usually containing a chromophore (AMC or pNA) at one end. This approach is limited because it is practically impossible to synthesize all the possible different substrates for testing, and furthermore, the use of chromogenic substrates gives rise to information about the requirements of the nonprimed (S) or primed (S′) subsites independently. In addition, it has been demonstrated that subsite−substrate interaction is generally dependent on the substrate structure and is not necessarily additive.[14] The use of an internally quenched fluorescent substrate, on the other hand, gives information about both the nonprimed and primed subsites simultaneously, making the use of a fluorescence-quenched peptide library ideal for the investigation of the substrate specificity of proteolytic enzymes.[15,16]

† 1X, set as standard−100 times dilution of enzyme (8.34 μM); 20X, 20 times dilution of 1X enzyme solution; Abz, 2-amino-benzoyl; AMC, 7-amino-4-methylcoumarin; Boc, *tert*-butyloxycarbonyl; CHC, α-cyano-4-hydroxycinnamic acid; DhBt-OH, 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine; DMF, dimethylformamide; eddnp, E′, ethylenediamine-*N*-(2,4-dinitrophenyl); Fmoc, N^{α}-fluoren-9-ylmethyloxycarbonyl; HMBA, hydroxymethyl benzoic acid; MeIm, *N*-methyl imidazole; MSNT, 1-(mesitylene-2-sulfonyl)-3-nitro-*1H*-1,2,4-triazole; NEM, *N*-ethyl morpholine; PEGA, poly(ethylene glycol) acrylamide copolymer; Pfp, pentafluorophenyl; Pmc, 2,2,5,7,8-pentamethyl chroman-6-sulfonyl; PNA, *p*-nitroanilide; TBTU, *O*-(benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium tetrafluoroborate; TEA, triethyl amine; TFA, trifluoroacetic acid.

* Correspondence: Dr. Phaedria M. St. Hilaire or Dr. Morten Meldal, Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark. E-mail: pms@crc.dk or mpm@crc.dk.

‡ Carlsberg Laboratory.

§ Escola Paulista de Medicina.

10.1021/cc990031u CCC: $18.00
Published on Web 09/21/1999

Cysteine proteases comprise a wide class of proteolytic enzymes in plants and animals that serve important functions such as the degradation of muscular protein, the processing of propeptides, prohormones, and zymogens, and the processing of foreign antigens for immunological responses.[17] Defects in the regulation of cysteine protease activity have also been reported in connection with several diseased states including osteoporosis, cancer metastasis, muscular dystrophy, viral replication, and parasitic infection.[18] Cysteine proteases are divided into about 20 families; the best known family being that of papain. The papain family contains peptidases with a wide range of activities including endopeptidases with broad or narrow specificities, amino peptidases, dipeptidylpeptidases, and peptidases containing both endo- and exopeptidase activity. Although papain has been selected as the archetypal cysteine protease and other members of the family are described as papain-like based on their substrate specificity, there has been surprisingly little systematic characterization of the substrate specificity of papain. The specificity is still primarily based on the pioneering work of Schechter and Berger which utilized diastereomeric mixtures of D and L Ala and a Phe scan to determine the number of subsites required and the specificity of the enzyme.[19] Results from that study suggested that the subsite of papain, hence most cysteine proteases, is seven amino acids long and that there is a preference for an aromatic residue in the S_2 subsite. At least one other study examining the hydrolysis of tripeptide esters and amides also concluded that Phe is best in S_2 compared to Leu or Ala[20] while others have shown that AMC substrates containing Leu in S_2 are also fairly well hydrolyzed.[21] A few studies have attempted to explore the specificity of the S_1', S_2', and S_3' subsites through the synthesis of a limited number of fluorescent quenched substrates containing Phe at P_2.[22−25] Recently, however, studies utilizing internally quenched substrates based on a partial sequence of a protein inhibitor, cystatin, have shown that an extremely good substrate may be obtained with the general sequence: QxVxG.[17,26−28] These substrates do not contain the usual aromatic residue in the S_2 subsite, lending doubt to the suggested strict requirement of a bulky aromatic residue at S_2. More recently, there has been an attempt to investigate the substrate specificity of papain by parallel synthesis of different resin-bound (Dansyl-GGGFX_1X_2GGGG-linker resin) substrates having Phe in P_2 and varying the amino acids at P_1 and P_1' with 20 different amino acids.[29]

In the present work, we assess the effectiveness and limitations of the solid phase fluorescence-quenched assay as a tool for determining the substrate specificity of proteolytic enzymes. By screening combinatorial libraries, the enzyme itself is allowed to select the best substrate from thousands of peptides. Thus, through a combination of the library methodology and molecular dynamics calculations, we present herein the first systematic characterization of the substrate specificity of papain, the archetypal cysteine protease.

Results

Design and Construction of Solid Phase Libraries. The general structure of the internally quenched fluorescent library is shown in Figure 1. PEGA resin was the solid support of choice because of its excellent swelling in both organic and aqueous media, thus allowing library synthesis as well as on-bead screening. Furthermore, in contrast to other commercially available PEG-based resins, the gellike matrix of PEGA facilitates rapid diffusion of enzymes into the bead where they maintain optimal enzymatic activity.[8−15,29] The library was synthesized on $PEGA_{4000}$ resin[30] which contains a larger pore size than $PEGA_{1900}$[31] or TentaGel thereby allowing large macromolecules (50 000 to 90 000 Da) to freely diffuse in to the interior of the bead.[30] Seven positions (X_1-X_7) were randomized using all 20 genetically encoded amino acids. This length was selected in order to map the seven subsites of cysteine proteases as proposed by Schechter and Berger.[19] The peptide sequences were flanked by the fluorescence donor, 2-aminobenzoic acid (Abz), attached to the side chain of lysine (K(Abz)) and by the fluorescent quencher, 3-nitrotyrosine ($Y(NO_2)$). Two libraries were synthesized: the first contained the 20 amino acids in the seven randomized positions, while in the second, acidic residues (Asp and Glu) were excluded from the library. In the second library, a proline residue was inserted at position 8 at the C-terminal end of $Y(NO_2)$ in order to direct $Y(NO_2)$ away from the S_2 site of the enzyme. The Lys-Lys sequence was incorporated in order to double the loading of the resin; initial loading was 0.13 mmol/g and final loading was 0.21 mmol/g (loading was calculated by measuring the absorbance at 290 nm of the piperidene-dibenzofulvene adduct obtained from the cleavage of the Fmoc protecting group by piperidine). The base labile HMBA linker was included to facilitate substrate cleavage for analysis of the purity of the library.

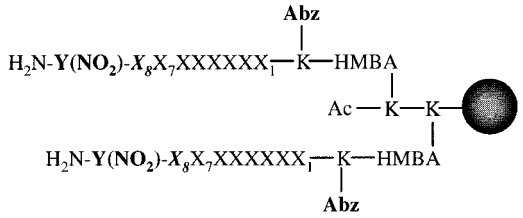

Figure 1. Peptide library construct. The libraries which contain free N-termini were synthesized on $PEGA_{4000}$ resin, and the loading was doubled by incorporation of the sequence of two Lys residues. X_1-X_7 are randomized positions while X_8 is omitted in library 1 and is Pro in library 2. Library 1 was synthesized without the HMBA linker.

Each library was synthesized on 1.7 g of $PEGA_{4000}$ resin containing about 270 000 beads (200−800 μm; average bead size was 300−400 μm). The large variation in bead size is a consequence of the difficult polymerization of long PEG chains during resin synthesis and does not influence the screening process given the long time scale of the enzymatic reaction compared to differing rates of enzyme diffusion within beads of varying size. Since the theoretical number of compounds (20^7) exceeds the number of particles, all the possible compounds were not present. However, a large number (270 000) of compounds were obtained. The peptides were synthesized using the Fmoc/OPfp ester methodology[32] except for the incorporation of K(Abz) and $Y(NO_2)$ which were incorporated using TBTU/NEM activation.[33] Upon completion of synthesis and deprotection protocols, 35 randomly chosen beads were cleaved and the product purity

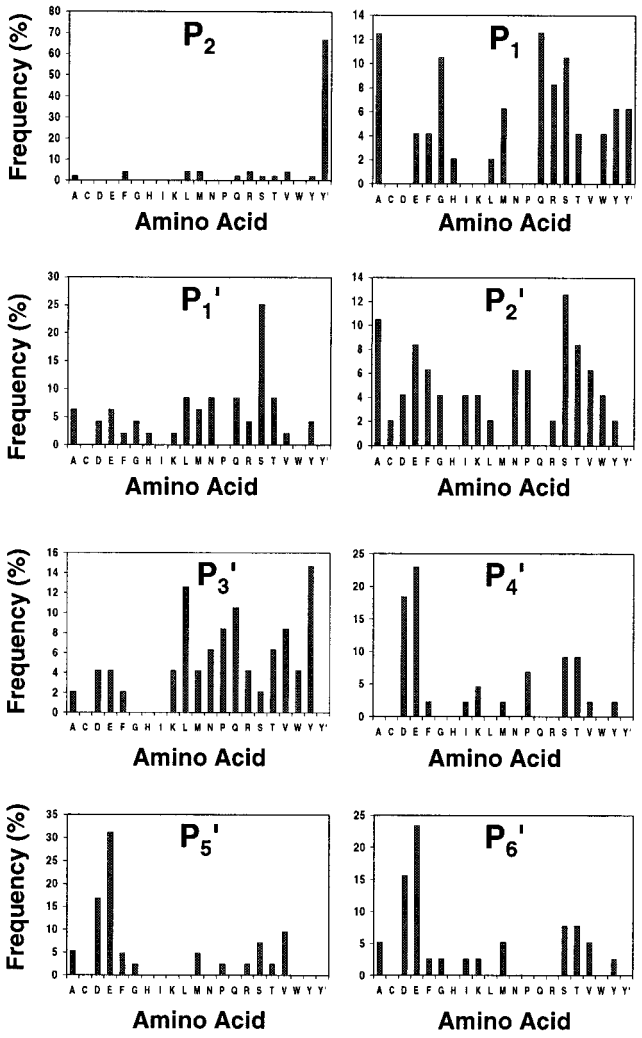

Figure 2. Amino acid frequency in enzyme subsites obtained from screening of library 1. It should be noted that an erroneously superficial impression of specificity can be obtained from this representation since there are three effects at once: the bias of $Y(NO_2)$ in S_2 subsite, the preference for peptides containing acidic residues, and to a much lesser extent, the specificity free from both biases.

was analyzed by MALDI-TOF-MS.[34] The mass spectra indicated that most of the peptides were at least 90−95% pure.

Screening of Peptide Libraries. After the synthesis and complete deprotection of library 1, 300 mg (ca. 50 000 beads) of the beads were incubated with papain (5.0 nM) for 2.5 h at 25 °C. Beads were then examined under UV light for the presence of fluorescent beads, which were transferred to TFA-treated cartridge filters for sequencing by Edman degradation. Because the enzyme reaction was terminated before complete hydrolysis of all the peptide attached to a single bead, each bead contained both the complete peptide sequence as well as the sequences of the remaining portion of the cleaved peptide (S′ amino acids). Thus, from Edman sequencing, it was possible to obtain the peptide sequence, the cleavage point and the approximate extent of cleavage. Beads (115) with varying degrees of brightness were isolated from the first library, and the amino acid subsite frequencies obtained from peptides attached to 40 of them are presented in Figure 2. The results from the sequencing showed that roughly 30% of the peptides were cleaved at two positions, and the alignments resulting from both cleavages were utilized. From these data, the nature of the amino acid in the S_2 subsite is important for directing substrate cleavage (consistent with literature results).[19,35] Surprisingly, by far the most preferred amino acid in S_2 was the fluorescence quencher $Y(NO_2)$, making it difficult to get reasonable data describing the residue specificity of subsites S_3 and S_4 which may also be important for substrate specificity. This result is in agreement with recent results which suggest that better binding of substrates of the type Dansyl-X-R-A-P-W is obtained when amino acids with aromatic side chains containing electron withdrawing groups occupy the S_2 subsite (X).[36] In the event that S_2 was not occupied by $Y(NO_2)$, hydrophobic residues, primarily Val, Leu, and Phe occupied that position. There was insufficient data to obtain a clear picture of the amino acid preference for positions S_4 and S_3. The S_1 subsite was occupied by small amino acids, Gln, Ala, Ser, or Gly and to a lesser extent, Arg. The S_1' subsite demonstrated a preference for Ser, and the S_2' was occupied by small amino acids Ser or Ala or by acidic residues Glu and Asp. The S_3' site demonstrated little specificity and was occupied by hydrophobic residues, Tyr, Leu, or Val as well as Pro and Gln. There was an exceedingly high preponderance of acidic residues (Glu and Asp) in the P_4', P_5', and P_6' positions, suggesting a substrate unrelated bias in the results. A series of test peptides were then synthesized to determine the best way to prevent substrate cleavage such that $Y(NO_2)$ occupied the S_2 position. It was found that placement of a Pro residue C-terminally to the $Y(NO_2)$ prevented such a cleavage (data not shown). Interestingly, placement of two proline residues N-terminally to the $Y(NO_2)$ enhanced and accelerated the cleavage (data not shown) in agreement with the substrate specificity found in library 2 (vide infra). A new library, library 2, was thus synthesized incorporating a Pro C-terminally to the $Y(NO_2)$ and excluding acidic residues.

Library 2, 300 mg (ca. 50 000 beads) of the beads, was incubated with papain (5.0 nM) for 3.5 h at 25 °C, and 34 bright beads were selected for Edman degradation. In two beads, the peptide was completely cleaved by the enzyme, providing the amino acids in only the primed subsites while in two other beads only K(Abz) was detected. In three cases, it was impossible to determine the sequence and cleavage point from the data. Consequently, the results from the Edman degradation of 29 beads are summarized in Table 1 and the amino acid subsite frequencies are shown in Figure 3. Again, about 30% of the peptides were cleaved in two positions. There was, in this case, a lack of preference for $Y(NO_2)$ in the S_2 subsite thus making it possible to obtain information about the occupancy of the S_4 and S_3 subsites. These subsites were occupied by small hydrophobic residues such as Pro, Val, Ala, and Gln, but with an overall preference for Pro (data does not include Pro residues fixed in the C-terminal position adjacent to $Y(NO_2)$). There was a clear preference for a hydrophobic residue in S_2 with Val being the preferred amino acid followed by Phe and Leu. S_1 was occupied by Arg, Gln, Ala, or Gly. When the small amino acids occupied S_1, Val occurred most often in S_2; whereas

Table 1. Substrates Obtained from Library 2 ([a]Y′-PX$_7$X$_6$X$_5$X$_4$X$_3$X$_2$X$_1$-[b]K′) after Incubation with Papain[c]

							⇓⇓								
bead[d]	P_8	P_7	P_6	P_5	P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'	P_5'	P_6'	P_7'	clv[e]
1					Y′	P	V	A	A	S	P	I	G	K′		31
2a					Y′	P	A	G	V	Q	Q	P	K	K′		15
2b						Y′	P	A	G	V	Q	Q	P	K	K′	24
3					Y′	P	-	A	W	P	P	G	G	K′		8
4		Y′	P	W	P	I	A	R	K	A	K′					22
5			Y′	P	Y	A	V	Q	S	P	Q	K′				28
6					Y′	P	F	K	S	R	K	Q	N	K′		35
7				Y′	P	V	L	R	Q	Q	R	S	K′			27
8		Y′	P	Y	V	P	M	R	Q	G	K′					17
9					Y′	P	V	S	G	Q	A	S	N	K′		15
10					Y′	P	L	Q	A	S	G	N	A	K′		26
11a		Y′	P	Q	Q	P	V	V	A	S	K′					35
11b	Y′	P	Q	Q	P	V	V	A	S	K′						24
12a	Y′	P	Q	Q	P	V	R	A	S	K′						17
12b		Y′	P	Q	Q	P	V	R	A	S	K′					23
13					Y′	P	F	K	L	G	I	S	N	K′		25
14					Y′	P	L	Q	A	S	G	N	A	K′		33
15					Y′	P	F	Q	A	R	R	F	R	K′		35
16					Y′	P	F	G	Q	R	T	P	A	K′		14
17[f]	P	V	W	P	Q	P	S	G	K′							-
18a[g]		Y′	P	R/C	Y	A	P	M	S	T	K′					5
18b[g]	Y′	P	R/C	Y	A	P	M	S	T	K′						8
19a					Y′	P	F	R	T	Q	T	R	P	K′		10
19b			Y′	P	F	R	T	Q	T	R	P	K′				30
20					Y′	P	V	A	T	A	G	P	V	K′		22
21					Y′	P	V	R	M	A	K	C	G	K′		8
22	Y′	P	L	P	P	A	M	Q	A	K′						9
23a					Y′	P	F	R	V	K	K	A	T	K′		13
23b			Y′	P	F	R	V	K	K	A	T	K′				12
24a					Y′	P	V	R	A	R	S	T	R	K′		21
24b			Y′	P	V	R	A	R	S	T	R	K′				20
25					Y′	P	V	G	T	S	I	Q	S	K′		25
26[g]		Y′	P	M	P	P	L	R/C	T	S	K′					52
27				Y′	P	P	L	Q	L	P	G	G	K′			40
28[h]									K	A	K′					
29[h]									G	W	Q	F	N	K′		

[a] Y′ = Y(NO$_2$). [b] K′ = Lys(Abz). [c] ⇓⇓ = Cleavage site as determined by Edman degradation. Substrates with more than one cleavage site are designated a and b. [d] Arbitrary number assigned to bead isolated from library. [e] Approximate percentage of peptide cleaved based on results of Edman degradation. [f] Y′ precedes P. [g] It was not possible to conclusively identify the amino acid as Cys or Arg. [h] Complete cleavage of peptide occurred: sequence of primed side only obtained.

when Arg or Lys occupied S_1, Phe or Val were present at S_2 with almost equal frequency. Ala, Thr, and Ser occupied S_1', while small amino acids Ser, Gln, and Ala occupied S_2'. Small amino acids particularly Pro, Gln, and Gly also predominated in S_3', S_4', and S_5' although in a less specific manner.

Kinetic Characterization of Selected Peptides. 1. Selection of Peptides. Lead peptides were selected for solution phase kinetics only from library 2 because the subsite occupancy trend obtained from the two libraries was similar and, additionally, peptides from library 2 were more centrally cleaved thus allowing amino acid substitutions from P_3 to P_4'. Peptide sequences with a high degree of cleavage as determined from amino acid sequencing data (Table 1: bead number 1, 26, 11, 5, 7, 10, 25, and 20 corresponding to compounds **1**, **18**, **19**, **20**, **21**, **23**, **24**, **26** in Table 2) and which contained only one cleavage point were selected as lead peptides. These and variations thereof were resynthesized and used for solution phase kinetic studies (Table 2). Peptide **1** was selected as the lead substrate since it was cleaved to a high degree (31%) and its sequence reflected the general enzyme subsite preferences determined from the library screen. Amino acid substitutions at each subsite were based on the statistical occurrences of particular amino acids in those subsites. To test the efficiency of the screening method, substrates with intermediate to low degree of cleavage (Table 1: bead number 9, 22, and 21 corresponding to compounds **29**, **30**, and **31** in Table 2) were also synthesized to see if they were indeed poor substrates. Because acidic residues had been excluded from the second library, these were incorporated into a few sequences (Table 2: **14**, **15**, and **16**) to investigate the effect they may have on the binding and catalysis of the substrate. Since the best known substrate for papain is the internally quenched substrate AbzQVVAGA-eddnp (**36**),[26] this substrate was also tested, and in order to directly compare with the current results, the substrate was also synthesized using K(Abz) and Y(NO$_2$) as the fluorescence donor/quencher pair (**32**, **33**, **34**, and **35**).

2. Synthesis of Peptides. Peptides (except for AbzAVVAGAE′[26]) were synthesized on PEGA$_{800}$ resin (0.26 mmol/g; 4 μM peptide) using standard Fmoc/OPfp ester methodology except for the incorporation of linker, K(Abz) and Y(NO$_2$) were coupled using the Fmoc/TBTU/NEM methodology. After deprotection, peptides were purified using reverse

Table 2. Mass Characterization and Kinetic Parameters (k_{cat}/K_M) for Papain Hydrolysis of Internally Quenched Fluorogenic Substrates Derived from Library Results

no.	substrate[a]	mass exp	mass found[b]	clv (%)[c]	k_{cat}/K_M (mM s)$^{-1}$
1	[d]Y′PVA⇊A↓SPIGK′	1166.3	1166.7	31	760 ± 15
2	[e]**K**′PVA↓A⇊SPIG**Y**′	1166.3	1165.8		815 ± 38
3	**K**′VA⇊ASPIG**Y**′	1069.2	1069.6		83 ± 33
4	**Abz**VA⇊A↓SPIG**Y**′	940.0	940.9		190 ± 8
5	Y′**A**VA⇊A↓SPIGK′	1140.3	1140.3		735 ± 32
6	Y′P**F**A⇊A↓SPIGK′	1214.6	1214.1		620 ± 24
7	Y′P**L**A⇊A↓SPIGK′	1180.3	1180.8		540 ± 27
8	Y′P**Y**A⇊ASPIGK′	1230.6	1230.3		749 ± 7
9	Y′PV**G**⇊ASPIGK′	1152.3	1152.4		1825 ± 19
10	Y′PVA⇊**S**↓SPIGK′	1181.3	1181.9		469 ± 26
11	Y′PVA⇊A**P**PIGK′	1176.4	1177.0		259 ± 17
12	Y′PVA⇊A↓S**Q**IGK′	1197.3	1196.7		446 ± 7
13	Y′PVA⇊ASP**N**GK′	1167.3	1168.1		315 ± 7
14	Y′PVA⇊A↓SP**E**GK′	1182.3	1182.5		397 ± 22
15	Y′PV**G**⇊ASP**E**GK′	1168.2	1168.8[f]		1017 ± 44
16	Y′PV**E**⇊AS**G**IGK′	1184.3	1184.3		90 ± 7
17	Y′**S**PVA⇊ASPIGK′	1253.4	1255.5		1095 ± 53
18	Y′PMPPLC⇊TSMK′ [g]	1431.7	1431.6[f]	52	2109 ± 101
19	Y′PQQPVVA⇊SK′	1280.4	1280.3	35/24	190 ± 34
20	Y′PYAVQ⇊SPQK′	1344.5	1344.0	28	1524 ± 79
21	Y′PVL**R**⇊QQRSK′	1438.6	1437.7	27	1450 ± 41
22	Y′PV**G**⇊**R**QQRSK′	1382.5	1382.0		639 ± 60
23	Y′PLQ⇊ASGNAK′	1212.3	1212.7	26	242 ± 23
24	Y′PVG⇊TSIQSK′	1243.4	1243.7	25	800 ± 82
25	Y′PVG⇊**G**↓SIQSK′	1199.3	1199.2		254 ± 14
26	Y′PVA⇊TAGPVK′	1166.3	1167.4	22	920 ± 127
27	Y′PVA⇊**G**↓AGPVK′	1122.3	1122.8		674 ± 27
28	Y′PVA⇊TAGK′	970.1	970.0		168 ± 16
29	Y′PVS⇊GQASNK′	1214.3	1211.9[f]	15	79 ± 4
30	Y′PLPPA⇊M⇊QAK′	1279.5	1279.4	9	40 ± 3
31	Y′PVR⇊MAKCGK′	1316.6	1315.3	8	557 ± 45
32	**Y**′YQVVA⇊GA**K**′	1162.3	1163.8		460 ± 44
33	**Y**′PQVVA⇊GA**K**′	1092.2	1096.6		88 ± 17
34	**Y**′QVVA⇊GA**K**′	999.1	999.9[f]		149 ± 4
35	**Abz**QVVA⇊GAG**Y**′	928.0	928.3[f]		918 ± 127
36	**Abz**QVVA⇊GA**E**′ [h]	870.9	893.9		9351 ± 382
37	Y′FR⇊QQK′ [i]	1032.1	1032.6		84 ± 13
38	Y′**P**FR⇊QQK′	1129.2	1130.6		84 ± 4

[a] Amino acid substitutions are shown in bold; ⇊ and ↓ denote major and minor cleavage points of substrates in solution, respectively. [b] Mass determined by MALDI-TOF-MS unless otherwise noted. [c] Approximate cleavage of resin-bound peptide based on results of Edman degradation. [d] Y′ = Y(NO$_2$). [e] K′ = Lys(Abz). [f] Mass determined by ES-MS. [g] Original sequence obtained from library was Y′PMPPLC⇊TSK′. [h] E′ = eddnp. [i] Substrate for barley endoprotease A and B.

phase HPLC and characterized by the determination of the mass using MALDI-TOF-MS or ES-MS (Table 2). The synthesis of all peptides proceeded smoothly in high yield.

3. Kinetics of Hydrolysis. The kinetics of substrate hydrolysis using papain was first determined under pseudo-first-order conditions at 25 °C for the determination of the second-order rate constant, k_{cat}/K_M. The cleavage points were determined using MALDI-TOF-MS as described in the Experimental Section (Figure 4a,b). The results shown in Table 3 indicate that while most substrates had identical cleavage points both in solution and on solid phase, a few were cleaved in different or multiple positions compared to when they were resin-bound. For example, peptides **1** and **30** were cleaved in one position on solid phase but cleaved at two positions in the solution phase assay. Conversely, peptide **19** that showed two cleavage points on solid phase

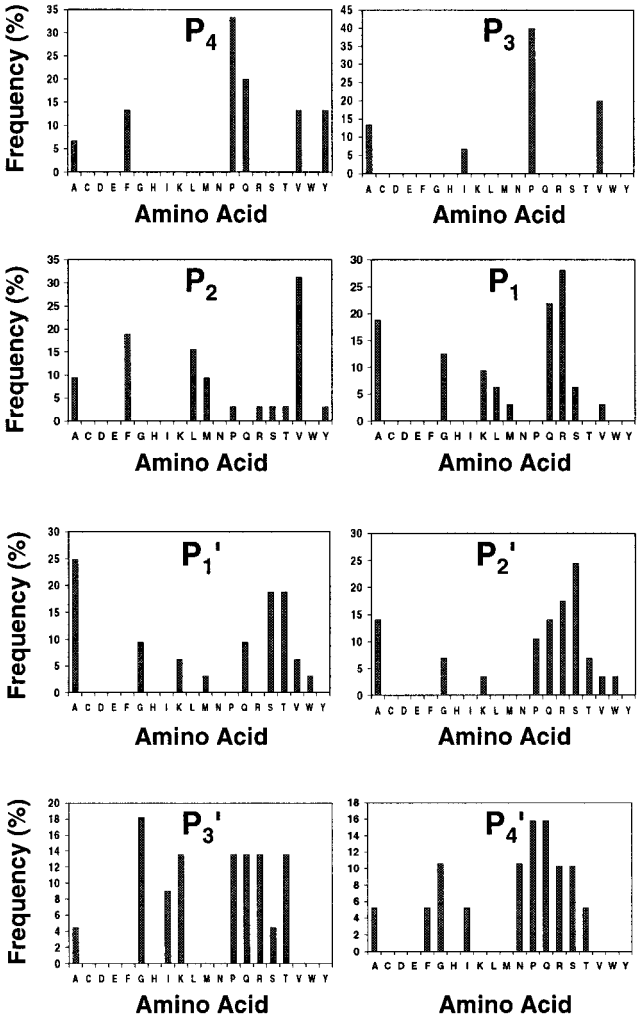

Figure 3. Amino acid frequency in enzyme subsites obtained from screening of library 2. Peptide sequences are shown in Table 1.

had only one cleavage point in solution. Substitution of lead peptides at various positions with different amino acids also lead to variations in the cleavage points (e.g. peptides **1** and **2**, **21** and **22**). For peptides showing significant hydrolysis at a second site, it can be shown that the measured k_{cat}/K_M corresponds to the sum of individual values of k_{cat}/K_M for each cleavage site and that the ratio of the resulting products from each site corresponds to the ratio of the individual k_{cat}/K_M values.[37] This ratio was determined from the MALDI-TOF-MS experiments and the measured k_{cat}/K_M corrected to reflect solely the specificity constant for hydrolysis at the P_1-P_1' position deemed the main cleavage point. MALDI-TOF-MS is not a generally quantitative technique; however, under certain circumstances the method can be used quantitatively.[38] Since the enzyme cleavage products have similar structures, it is expected that their ionization and flight properties may be similar.[39] Therefore, the relative product signal intensities may be a good measure of their relative quantities in the product mixture.

4. Effect of Substitution in the Various Subsites on Lead Peptide 1. The amino acids in subsites P_3 to P_4' were substituted by various amino acids primarily based on the library results (Table 2, **5**−**17**). Most of the modified peptides also manifested major and minor cleavage points although

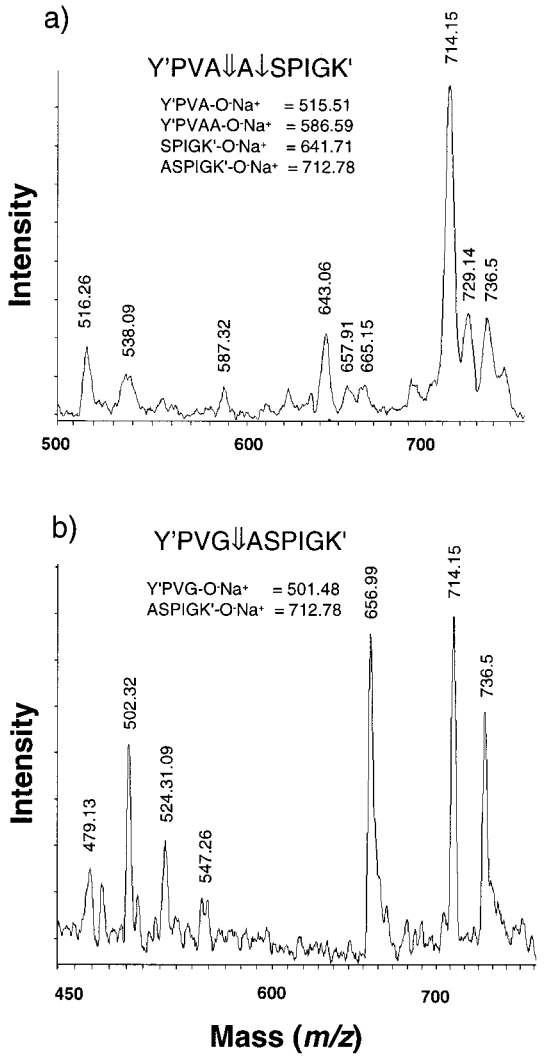

Figure 4. Representative mass spectra showing cleavage site of substrates. (a) Cleavage of substrate **1** in two positions. Often, di-sodiated species are also observed (538.09, 665.15, and 736.5). Peak at 657.91 arises from the buffer used in the assay. (b) Cleavage of substrate **9** at a single position. The peak at 479.13 represents the protonated species, those at 524.31 and 547.26 represent the di- and tri-sodiated species, respectively, while the peak at 656.99 comes from the buffer.

some had one cleavage point as in the original solid phase results and others had up to three (data not shown). Substrate **1** already possessed the preferred Pro residue at P_3, and substitution by another small hydrophobic residue, Val, resulted in cleavage of the substrate at three different sites which made direct comparison to **1** difficult (data not shown). Substitution of Ala (**5**) for Pro at P_3 had almost no effect on the specificity constant. Replacement of the hydrophobic residue in P_2 by other hydrophobic or aromatic residues Phe, Leu, or Tyr (**6**, **7**, **8**), led to a slight reduction in k_{cat}/K_M in the case of Phe and Leu but not for Tyr. This effect reflects the natural tendency of the enzyme to cleave substrates with $Y(NO_2)$ in the P_2 position (screening of library 1) and those with amino acids containing nitro- and chloro-substituted aromatic side chains in P_2.[36] Replacement of the Ala in the P_1 position by Gly almost doubled the k_{cat}/K_M, providing one of the best substrates (**9**) for papain. This large increase is probably due to the insertion of the more flexible Gly into the S_1 binding pocket in a manner that allows optimized contacts of the other residues with the enzyme. Substitution by Ser at P_1 led to a substrate with multiple cleavage points, making a direct comparison difficult (data not shown). Substitution of Ala in P_1' by Ser (**10**) resulted in a reduction of the specificity constant by about a half. The P_2' position was already occupied by the favored amino acid, Ser; thus a substitution by a Pro residue was performed to see the effect of Pro in P_2' since, in autohydrolysis of propapain, cleavage occurs such that Pro is present in the P_2' position (**11**).[40] The k_{cat}/K_M was reduced to about one-third its original value. Substitution of the Pro residue in P_3' with the most abundant amino acid Gln (**12**), resulted in a reduction of k_{cat}/K_M by about one-third. A substitution of Asn in P_4' (**13**) caused a large reduction of catalytic efficiency. Since acidic residues were excluded from the library, they were incorporated at sites where they had been observed in the first library far from (**14** and **15**) and near (**16**) the active site, in order to evaluate their contribution to substrate hydrolysis by papain. Insertion of Glu at P_1 or P_4' both resulted in a large reduction of the specificity constant.

5. Other Trends. Other sequences obtained from library 2 and variations thereof (**18**−**31**) were synthesized in an attempt to investigate the correlation between solid phase and solution phase hydrolysis. The results suggest that there is not necessarily a direct correlation between the solid phase and solution phase enzymatic hydrolysis of the same substrate. Some sequences that should result in good substrates according to the degree of cleavage on solid phase do not (e.g. **19** and **23**) while others perform as expected (e.g. **18**, **20**, and **21**). Most sequences that were expected to be poor substrates (**30** and **31**) had low k_{cat}/K_M's. To investigate the number of subsites required, peptide **26** was shortened by two amino acids (**28**) on the primed side and peptide **1** was shortened by one amino acid (**3** and **4**) and lengthened by one amino acid on the nonprimed side (**17**). Results indicate that lengthening the peptide on the non-primed side slightly increases the specificity constant whereas shortening the peptide on either side lowers k_{cat}/K_M. Short peptides (**37** and **38**) that were good substrates for barley endoprotease A and B[41] were poorly hydrolyzed by papain. These results suggest that seven to eight subsites are indeed the minimum required for optimal catalytic activity of papain-like cysteine proteases.

6. Influence of Donor/Quencher Pair. The influence of the position and nature of the donor/quencher pair was investigated using two series of peptides: **1**−**4** and **32**−**36**. The results suggest that the placement and nature of the groups can greatly influence k_{cat}/K_M and the cleavage point of the substrate, particularly if the cleavage point is relatively close to the reporter group. Changing the position of K(Abz) and $Y(NO_2)$ in **1** and **2** resulted in a slight change in k_{cat}/K_m, probably due to a change in cleavage site. The effect is more pronounced in **3** and **4** when substitution of K(Abz) with Abz in P_3 halves the specificity constant. It appears that in certain substrates using eddnp as the quencher instead of $Y(NO_2)$ greatly enhances the k_{cat}/K_M of the substrate (**35** and **36**), probably due to interaction of the eddnp group with aromatic residues W177 and W181 (Figure 6f).

Table 3. Kinetic Parameters for the Hydrolysis of Selected Internally Quenched Fluorogenic Substrates by Papain

	$[S] \ll K_M$		$^1/_8K_m < [S] < 8\ K_M$		
no.	substrate[a]	k_{cat}/K_M (mM s)$^{-1}$	K_M (μM)	k_{cat} (s)$^{-1}$	k_{cat}/K_M (mM s)$^{-1}$
3	[b]**K′**VA⇊ASPIG**Y′**	583 ± 33	35.8 ± 8.2	23.8 ± 2.4	665
9	[c]Y′PV**G**⇊ASPIGK′	1825 ± 19	37.6 ± 8.6	21.6 ± 1.6	575
11	Y′PVA⇊A**P**PIGK′	259 ± 17	40 ± 3.6	11 ± 0.42	275
15	Y′PV**G**⇊ASP**E**GK′	1017 ± 44	22.9 ± 11.8	19.1 ± 1.6	834
16	Y′PV**E**⇊AS**G**IGK′	90 ± 7	50.5 ± 47.1	5 ± 1.8	99
17	Y′**S**PVA⇊ASPIGK′	1095 ± 53	11.4 ± 2.9	10.6 ± 0.77	930
18	Y′PMPPLC⇊TSMK′	2109 ± 101	6.8 ± 1.9	10.1 ± 0.25	1472
21	Y′PVL**R**⇊QQRSK′	1450 ± 41	22.6 ± 10.3	33.0 ± 4.6	1460
26	Y′PVA⇊TAGPVK′	920 ± 127	16.9 ± 10.9	13 ± 1.1	769
28	Y′PVA⇊TAGK′	168 ± 16	52.9 ± 20.3	10.2 ± 1.5	193
34	**Y′**QVVA⇊GA**K′**	149 ± 4	37.9 ± 14.7	6.4 ± 1.2	169
35	**Abz**QVVA⇊GAG**Y′**	918 ± 127	17.4 ± 4.4	12.7 ± 0.60	730
36	**Abz**QVVA⇊GA**E′**[d]	9351 ± 382			
		31 000[e]	1.6[e]	46[e]	29 000[e]

[a] Amino acid substitutions are shown in bold; ⇊ denotes cleavage point of substrates in solution. [b] K′ = Lys(Abz). [c] Y′ = Y(NO$_2$). [d] E′ = eddnp. [e] Kinetic parameters determined at 30 °C.[26]

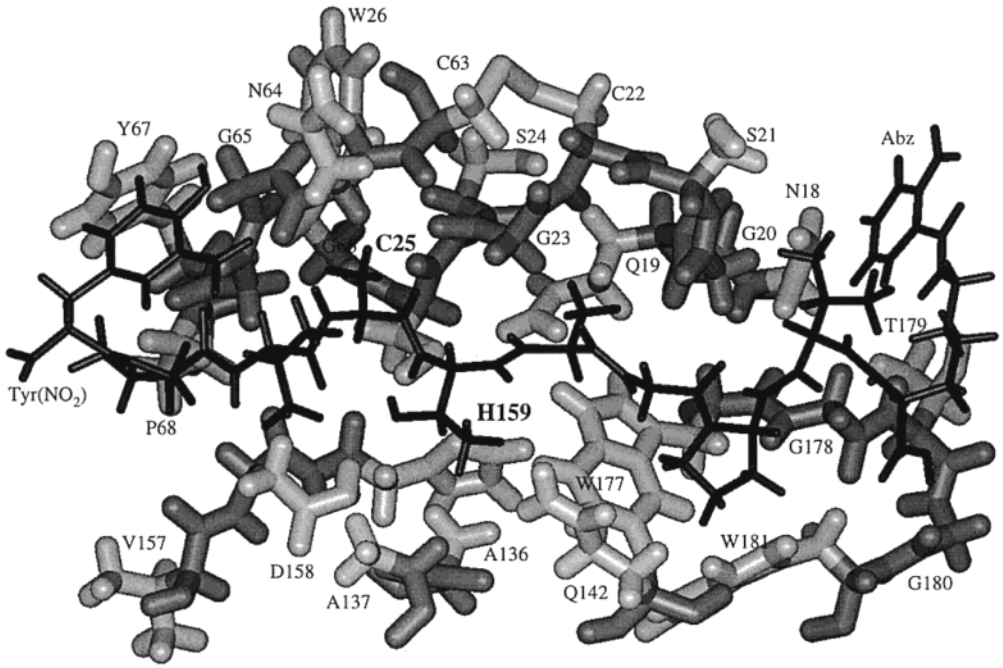

Figure 5. The detailed structure of the papain binding site with peptide **26** displayed as a thin black stick model. The structure is in the same orientation as shown with the Connolly surfaces in Figure 6a and Figure 6b−f. Only residues forming the active site are shown as a thick stick model with light gray side chains and darker gray backbone atoms. The residues are indexed as numbered from the N-terminus of papain. Residues C25 and H159 form part of the catalytic triad.

To determine whether a high or low specificity constant was a result of changes in binding or in catalytic rate, K_M's and k_{cat}'s for selected substrates were determined from Hanes plots using substrate concentrations ranging from 2 to 150 μM (Table 3). Examination of the K_M of the best substrates obtained from the library revealed that a good substrate was obtained either by tight binding to the enzyme (**18**, k_{cat} = 10.1 s^{-1}; K_M = 6.8 μM) or by having a high catalytic rate compensating for less effective binding (**21,** k_{cat} = 33.0 s^{-1}; K_M = 22.6 μM). On shortening the substrate at the primed side, the rate of catalysis remained the same but the binding was reduced by a factor of 3 (**26** and **28**). In the series of peptides **3**, **9**, **11**, **15**, **16** and **17**, the rate of hydrolysis of the G−A or A−A amide bond was essentially the same except in peptides **11** and **16** where having two prolines in P_2' and P_3' or a Glu in P_1 results in marked reductions of the catalytic rate. All these substrates had roughly the same K_M except for **17** which was 3−4-fold better and for **16** which was reduced by about 25%. Peptides **34**−**36** explored the effect of the donor and quencher groups and it is clear that a higher specificity constant arose from lower K_M's.

7. Modeling. The crystal structure of papain complexed with the inhibitor Suc-QVVAA-pNA with a resolution of 1.7 Å was used as the starting point for all calculations.[42] To rationalize the results obtained from screening of the substrate libraries, seven representative substrates (**1**, **18**, **21**, **26**, **28**, **35**, **36**) were docked into the active site of papain to investigate the following: multiple cleavages (**1**), the binding of amino acids with large side chains (**21**), additional, less specific interaction with P_4 and P_5 (**18**), the importance of P_3' and P_4' (**26** and **28**), and the importance of fluorescent quencher for binding (**35** and **36**). The peptide substrates

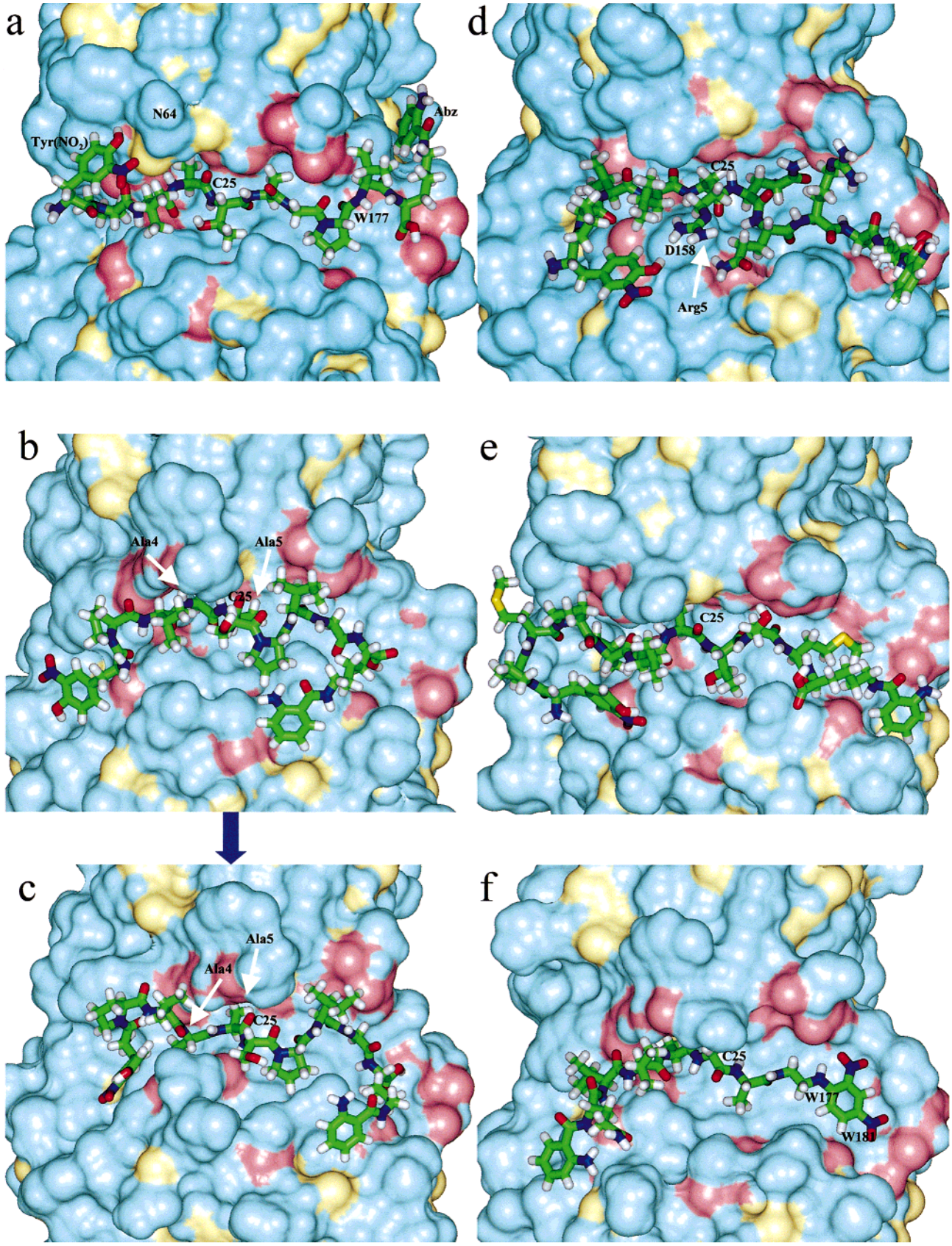

Figure 6. Molecular dynamics calculations of the enzyme−substrate complex using papain coordinates determined at 1.7 Å resolution and the most representative substrates found in the library. (a) The typical binding substrate **26** (Y(NO_2)PVATAGPVK(Abz)). Pro2 is buried in P_3 at the bottom of the cleft, Val3 is protruding at P_2, Ala4 is buried under G65 in P_1 close to C25, Thr5 and Ala6 are in small but obvious pockets S_1' and S_2', respectively, and Pro8 is in contact with W177. (b and c) The sliding of substrate **1** is possible due to the presence of small and flexible amino acids. (d) Interaction of papain with substrate **21** which contains Arg at P_1. The Arg was observed to interact with D158. (e) The increased number of contacts between **18** and the S-region lead to increased substrate binding and rate of hydrolysis. (f) In **36**, the distance between P_1 and the dinitrophenyl group is perfect for stacking on the W177 lining the bottom of the P_3' and P_4' subsites.

were aligned in the direction inferred from most available crystal structures of papain complexed with peptide inhibitors; that is, the orientation with S_3 between V133 and Y67 forward to $S_3'-S_5'$ at W181 and W177 (Figure 5). The possibility that peptide substrates can be aligned in the reverse direction has been discussed.[43] Our attempts, however, to also model substrates in that orientation were unsuccessful since the cooperative orientation of the substrate side chains in the S and S′ sites did not allow close proximity of the active sulfhydryl to the scissile bond (i.e. <4 Å).

The results of the calculations are best illustrated with substrate **26** ($k_{cat}/K_M = 920$ (mM s)$^{-1}$; intermediate binding energy $K_M = 17$ μM and intermediate catalytic efficiency $k_{cat} = 13$ s^{-1}) which binds in the classical way predicted from small peptide-like inhibitor complexes with papain, i.e. in an extended conformation and with very close contacts to the enzyme active site (Figure 5). Several starting conformations particularly with different orientations of the P_1 side chains were attempted, and for the most part, the same ending conformation with the Ala methyl group bound in a small pocket under N64 and G65 and pointing toward W26 was attained at equilibrium. Other contact residues for P_1 are G66, C25, and Ala 160. The P_2 Val side chain is oriented away from the cleft and interacts with residues Y61, G65, and D158. The orientation of Pro in P_3 was with the pyrrolidine ring pointing toward the cavity with P68 at the bottom of the cleft. Other P_3 contact residues are V150, V133, and Y67. P_1' is to the left of the binding cleft contacting D158, A136, H159, A137, and Q142, while P_2' points to the right to interact with G23, C22, G20, and Q19 (when viewing along the substrate cleft in the N → C terminal direction). The positions of S_3' (W177 and Q147), S_4' (W177, W181, Q147, and L143), and S_5' (G20, N18, and S21) were less obvious, and various binding modes were obtained with different substrates (vide infra). According to the current model, there may even be an S_5' binding subsite comprising residues W181, G178, T179, and G180. Interestingly, there was a stacking of the Pro ring at P_4' onto the aromatic side chain of W177.

The docking of substrate **1** (Y(NO_2)PVA⇓⇓A⇓SPIGK(Abz)) with a major and a minor cleavage site examined the ability of papain to cleave substrates with small amino acids such as Ala-Ala or Gly-Ala at P_1-P_1' or Ser-Thr at $P_1'-P_2'$ in two different positions (Table 2). The substrate bound in a conformation which allowed "sliding" of the amino acid in P_1-P_1' along the cleft to optimize interactions either at S_3, S_2, or at S_1', S_2', S_3'. The "sliding" actually occurred after release of constraints to the sulfhydryl group during the calculation going from the complex in Figure 6b to the one in Figure 6c and was possible because in papain there is a small cavity beneath the S_2 subsite that can accommodate the small Ala methyl group. When the equilibrium conformation obtained for **26** was used as the starting point for **1** and the constraints on the sulfhydryl group were maintained, the calculation converged to a single equilibrium conformation similar to that in Figure 6b.

Other substrates obtained from the library screening were those with Arg in P_1 and with relatively large amino acid side chains in P_2-P_3' as typified by **21** (Y(NO_2)PVLR⇓⇓Q-QRSK(Abz)). Modeling of complexes with these types of substrates was considerably more difficult since the motion of the more bulky peptides in the narrow cleft was limited. The result was an orientation of the P_3 and P_2 side chains as for **26**; however, the P_1 Arg side chain bent upward to interact with the carboxylate of D158 (Figure 6d). During the calculation, the distance between the guanidyl nitrogen and the carboxyl oxygen varied from 2 to 5 Å. The larger residues in P_1' and P_2' (QQ) were accommodated by the backbone forming a kink in the chain allowing the P_2' Gln to interact with S_1' and the P_1' Gln to interact with S_2' with subsites defined as previously reported[43] and by the docking of substrate **26**.

The addition of residues in P_4 and P_5 as in substrate **18** (Y(NO_2)PMPPLC⇓⇓TSMK(Abz)) gave an increased binding to the enzyme. From the molecular modeling calculations, the Pro rich sequence results in a turn structure protruding from the Pro buried in P_3 that allows all three amino acids in P_4-P_6 to interact in a less specific manner with the shallow nonprimed part of the substrate cleft (Figure 6e). This increased enzyme substrate interaction leads to lower K_M's.

Modeling of **28**, the shortened form of **26** lacking Gly and Pro in P_3' and P_4', demonstrated the absence of stacking interactions between the Pro at P_4' on the aromatic side chain of W177 (data not shown). A similar stacking with $\pi-\pi$ interactions was observed with substrate **36** (Figure 6f), which had an almost ideal orientation and distance between the Ala in P_1 and the dinitrophenyl group of eddnp to facilitate stacking on W177 (and W181). Interestingly, when a change in χ_1 of W177 occurred during one annealing, perfect intercalation of the dinitrophenyl group between the indole rings of W177 and W181 was obtained, suggesting that such an intercalation may also be at play during enzyme cleavage.

Discussion

Library Screening. The combinatorial library approach encompasses three steps: the design and construction of libraries of chemical diversity, followed by the application of screening and assay techniques, and finally the identification of the active compounds. Screening of libraries and identification of active compounds depend on the mode of library construction and the type of receptor being investigated. Synthetic peptide libraries have been used to characterize the substrate specificity of various proteolytic enzymes both in solution[44−46] and on solid phase.[8,10,30] But how effective are these methods? Do the hits obtained from the library screen truly represent the best possible substrate or inhibitor for that enzyme? In the screening of solution phase libraries (i.e. screening of mixtures of compounds), that question cannot be answered since the precise identity of the active compounds in the mixture is usually not known and the identity of the substrate or inhibitor is obtained from a statistical analysis of the preferred residue at various positions. Despite the tedium of the process, good substrates/inhibitors are nonetheless obtained.[45,47] In the one-bead-one-compound libraries, good lead compounds can be obtained both by direct identification of a single active compound as well as from a statistical analysis of favored residues in the

various enzyme subsites. Furthermore, different families of substrates may be identified and it is possible to correlate the results of the solid phase screening to the subsequent solution phase assays.

In the present work, we endeavor to assess the efficiency of the solid phase fluorescence-quenched library screening as a means of characterizing the substrate specificity of proteolytic enzymes, in particular cysteine proteases. To this end, we have used papain, the archetypal cysteine protease, as the model enzyme and have attempted to explain the substrate specificity using molecular modeling. Screening of library 1 randomly generated from all 20 natural amino acids yielded substrates with a preponderance of acidic residues primarily in subsites away from the cleavage site, i.e. S_4', S_5', and S_6'. These results are artifactual and do not necessarily arise from preferential selection of these substrates by the enzyme. The apparent preference for acidic residues is most likely due to an increase in the effective local concentration of the enzyme in beads with a high concentration of negative charges (acidic residues) because at pH 6.8 papain is an overall basic molecule with a pI of 8.75. Replacement of particular residues with acidic residues in substrates found from the screening of library 2 (without acidic residues) did not yield better substrates (Table 2: **14** and **15**) with increased k_{cat}/K_M values, suggesting that they are not preferred in those subsites. This occurrence also suggests that the nature of the residue $P_4'-P_6'$ is less important for the cleavage of the substrate. Interestingly, the amino acid subsite preferences for the P_2-P_2' obtained from screening library 2 were similar to the results obtained from screening library 1, after disregarding the preponderance of acidic residues and the preference for $Y(NO_2)$ in S_2. This finding suggests that in the event that the enzyme concentration is locally increased due to additional unrelated affinity, the substrate preference must still be satisfied in order to achieve enzymatic hydrolysis.

The intramolecular fluorescence-quenched assay has the advantage that information about the amino acid preference in both the nonprimed and the primed subsites can be obtained in a single experiment. However, the method demands that the peptide be flanked by fluorescent donor and quencher groups that inevitably interact with the enzyme, thus influencing the cleavage points and the absolute value of the kinetic parameters. In the present library, 2-amino benzoic acid (Abz), attached to the side chain of Lys, was used as the fluorescent donor and 3-nitrotyrosine ($Y(NO_2)$) as the fluorescent quencher. These residues are relatively small compared to other fluorescent donor/quencher pairs that are considerably more hydrophobic (e.g. DABCYL/EDANS[23−25,48]) and have the additional advantage that they are readily incorporated into the peptide during synthesis. The results obtained from the screening of library 1 (Figure 2) illustrate the situation wherein the fluorescence quencher ($Y(NO_2)$) strongly influences the cleavage site of the substrates and subsequently biases the screening results. However, the results from the kinetics of hydrolysis of lead substrates (Table 2) obtained from library 2 suggest that while the positioning of $Y(NO_2)$ and K(Abz) at either end of the peptide does not significantly influence the k_{cat}/K_M values, they may influence the cleavage point of the substrate (Table 2: **1**−**4**). In contrast, other fluorescent donors and quenchers may heavily influence the value of the specificity constant in a way that depends on the nature of the substrate and the distance of the reporter molecule from the cleavage point. This influence is particularly evident when the reporter group is situated at important subsites, e.g. $Y(NO_2)$ in P_2 or eddnp in $P_3'-P_4'$. It can be seen that the use of eddnp instead of $Y(NO_2)$ as the quencher residue enhances 1 order of magnitude the k_{cat}/K_M value in the shorter substrates analyzed (Table 2: **35** and **36**).

Of fundamental importance is the question of correlation between enzymatic activity on solid phase compared to that in solution. Is the best substrate obtained from the solid phase assay the best in solution? Based on the data in Table 2, the answer is not so straightforward and there does not appear to be a direct, linear correlation between activity on solid phase compared to that in solution, particularly in terms of ranking of the substrates. Generally, a substrate that was greater than 20% cleaved on solid phase was a good or excellent substrate in solution phase assays (**1**, **18**, **20**, **21**, **24**, **26**) with two exceptions (**19** and **23**). The converse is also true; substrates with a low degree (10−15%) of solid phase cleavage were generally poor substrates in solution phase assays (**29**, **20**, **31**). The ranking of the substrates in increasing order of degree of cleavage on solid phase, was not the same as the ranking based on increasing k_{cat}/K_M obtained in the solution phase assays. Furthermore, the number and position of the cleavage site of the peptide differed on solid phase compared to in solution. The molecular modeling and the amino acid substitutions shown in Table 2 suggest that subtle changes in the substrate are required for changes in the cleavage site in certain substrates. The reasons for these differences are not immediately obvious but are related to the microenvironment and dynamics of the substrate on the resin. Furthermore, the active site of papain is a narrow cleft without significant subsite "binding pockets" and the preference for small amino acids in most subsites (Figures 2 and 3) allows the substrate to easily slide thus resulting in alternative modes of binding because of minor changes in the substrate. The resin linker (HMBA) is situated in the extended primed subsites and may thus influence the binding conformation of the substrate. Another possibility for lack of correlation between degree of cleavage on solid phase and solution phase k_{cat}/K_M values could be due to higher local enzyme concentration in the beads which contain peptides that may bind tightly to a noncatalytic binding site of the enzyme. Thus, a peptide may be cleaved to greater extent than another because there was a higher concentration of enzyme present. Such local effects are not at play in the solution phase assay. These liabilities notwithstanding, overall there is a high probability that a good substrate on solid phase will be a good one in solution and vice versa: the peptide with the highest degree of cleavage on solid phase was one of the best cleaved in solution (**18**) while the one with one of the lowest degrees of cleavage was the poorest substrate (**30**).

Characterization of the Substrate Specificity of Papain. A secondary goal of this investigation was to characterize

the substrate specificity of papain by implementation of a methodology that provides information on the specificity of both the nonprimed and primed sites simultaneously. Using an intramolecular fluorescence-quenched solid phase combinatorial library approach, we were able to obtain very good substrates for papain and to characterize the subsite specificity from S_4 to S_4'. The papain active site, a narrow cleft between the two tightly interacting domains, is lined predominantly by protein backbone of limited flexibility, and most of the 16 contact residues are small amino acids, i.e. five Gly, three Ala, two Cys, and one Ser. Interaction of the substrate with the active cleft of the enzyme is therefore dominated by contacts to the protein backbone. In order for the substrate to arrive at the short distance required for cleavage, the rigid cavity demands flexibility of the substrate during binding and this was clearly reflected by the relative ease of docking substrates with small amino acid side chain groups into the active site and by the preponderance of substrates composed of small amino acids naturally selected from the library screening. There is a general absence of distinct binding pockets that can accommodate large amino acid side chains in the active site.

Subsites S_4 and S_3 showed a preference for the small nonpolar amino acids Val or Ala or Pro. While there is some debate over the actual existence of well-defined subsites beyond S_3,[49] results from the modeling support the lack of well-defined subsites. The active site broadens into a hydrophobic patch from the P_3 subsite onward, making it relatively easy for hydrophobic residues of some substrates to orient themselves in ways that optimize enzyme−substrate interactions. For example, Y(NO_2)PMPPLCTSMK(Abz) (**18**) with Met and Pro in P_5 and P_4 has the best specificity constant of all the identified substrates due to increased binding energy (Table 3).

In substrates **1**, **26**, and **28**, there is a small, but significant preference for Pro in P_3 due to the perfect fit of the Cβ−Cδ of the pyrrolidine ring into a small cavity with P68 at the bottom. A similar interaction is observed with **21** which has the other preferred residue, Val, in P_3. In fact, all substrates had the same orientation at P_3/P_2. Consistent with literature results, a hydrophobic residue was best at S_2; however, that residue could be Val, Phe, Leu, Tyr, or Y(NO_2). From both the statistical analysis of the library results (Figure 3) and the specificity constants for selected substrates (Table 2), it appears that Val is slightly favored over Phe, Tyr, and Leu in substrates containing a small residue at P_1 but an equal frequency in those containing Arg or Lys at P_1. The P_2 Val side chain is oriented away from the binding cleft. The apparent lack of structural requirement in S_2 and strict requirement for a hydrophobic residue at P_2 indicate that the side chain does not fit into a specific pocket but rather interacts with side chains of hydrophobic residues (e.g. V150, Y61) at the rim of the cleft. Interestingly, the S_1 subsite exhibited a dual specificity for small, nonpolar amino acids as well as the larger, charged Arg. For small residues, the orientation of the P_1 side chain was into a small pocket under G65 at the left side of the cleft. Larger substrates clearly bind in a different mode since there is no space for a large side chain in the small P_1 pocket, and in the case of substrate **21** the salt bridge formed between the Arg in P_1 and D158 may result in the extra force needed to get sufficient proximity of the carbonyl and the sulfhydryl group to give fast rate of hydrolysis (k_{cat} = 33.0 s^{-1}) and poorer binding (K_M = 22.6 μM) (Table 3).

The presence of small residues at P_1-P_2' causes some substrates to be cleaved in more than one position because of "sliding" of the small amino acid side chains into small cavities in the enzyme at S_1-S_2'. The remaining subsites ($S_1'-S_4'$) all preferred small nonpolar residues with a preference for Ala and Ser. These results are consistent with previous studies in solution[23,24] and on solid phase[29] which determined that small and/or hydrophobic residues such as Ser, Val, Ala, and Leu, were best at $P_1'-P_2'$. Most good substrates, i.e. those with a k_{cat}/K_M greater than 1000 (Table 2: **9**, **15**, **17**, **18**, and **20**) contained primarily small amino acids in P_3-P_3' although exceptional substrates (Tables 2 and 3: **21**) also contained larger residues in these positions. Although S_1' and S_2' are well defined for substrates with small amino acids, other substrates with larger amino acids in those positions bind with the side chains away from the cleft toward solvent. Substrate **28** which lacks residues Gly and Pro in $P_3'$and P_4' had a similar k_{cat} and a higher K_M (Table 3) than **26**, underscoring the importance of the stacking of Pro at P_4' on the aromatic side chain of W177. This stacking interaction plays a significant role in the specificity of the $S_3'-S_4'$ subsite of papain and could be the reason for the very high K_M (Table 3) found for **36** in contrast to substrate **35** for which stacking interactions of the Y(NO_2) with W177 (and W181) are not optimal.

Several good substrates for papain were obtained directly from the library screening and from modification of lead peptides. The sequences of the substrates vary significantly but can be loosely classified into two groups: those containing part of the cystatin-like sequence xVxA(T)x[26] or the C hordein-like xQQx sequence (C hordein is a substrate for endoproteases from barley).[41] The best substrates obtained were Y′PMPPLCTSMK′ (k_{cat}/K_M = 2109 (mM s)$^{-1}$), Y′PYAVQSPQK′ (k_{cat}/K_M = 1524 (mM s)$^{-1}$), and Y′PVLRQQRSK′ (k_{cat}/K_M = 1450 (mM s)$^{-1}$) and not those containing cystatin-like sequences. These substrates had higher k_{cat}/K_M values than Y′QVVA⇓⇓GAK′ (k_{cat}/K_M = 149 (mM s)$^{-1}$), the equivalent of AbzQVVA⇓⇓GAeddnp (k_{cat}/K_M = 9351 (mM s)$^{-1}$). Interestingly, a substrate Y′PQQPVVA⇓⇓SK′ (k_{cat}/K_M = 190 (mM s)$^{-1}$ very similar to the QVVAGA substrate was obtained directly from the library screen, and the k_{cat}/K_M is on the same order of magnitude as Y′QVVA⇓⇓GAK′. The QVVA motif was only superior in the case where the eddnp moiety is stacked above W177/W181, creating additional favorable binding interactions.

Conclusions

A new methodology which combines the screening of combinatorial libraries of substrates with molecular modeling of hits from the library has been developed for the complete characterization of proteolytic enzyme specificity. The approach combines the natural selection of substrates by the enzyme from a large library with the modeling of the interaction of the enzyme with these substrates to allow

rational design of ligands with desired properties. Of key importance is the ability to perform high throughput screening of the resin-bound substrates using a solid support that does not greatly influence the outcome of the screen. An awareness of factors such as the nature and placement of reporter groups and linker and properties of the receptor that can skew the screening results should be maintained. This methodology once properly applied is quite effective, providing a fairly good correlation between activity on solid phase and in solution and has generated, in this case, very good substrates for papain. In addition, we have mapped the subsite specificities of papain and in contrast to previous results, it appears that the nature of the residue in S_2 alone does not dominate the specificity requirements. Rather, it is the synergistic relationship between the residues in subsites S_2, S_1, and S_1' that govern the specificity of the enzyme. We also suggest that although the S′ interactions are very important for the substrate binding and cleavage, multiple binding modes in S′ subsites are possible. Furthermore, comparison of the calculations performed on the various substrates demonstrated that the enzyme, even with the quite restricted motion in the active site observed during calculations, reaches complex structures with quite different topologies to accommodate the different substrates. This is probably a prerequisite for the binding and cleavage of such diverse families of substrates by general proteases such as papain.

Experimental Section

Materials And Methods. All solvents were purchased from Labscan Ltd. (Dublin, Ireland). Dichloromethane was distilled from P_2O_5 and was stored over 3 Å molecular sieves under argon. Fluoren-9-ylmethoxycarbonyl (Fmoc) amino acids and their pentafluorophenyl (Pfp) ester derivatives were purchased from Bachem and NovaBiochem. Fmoc-Lys(Boc-Abz)-OH and Fmoc-Tyr(NO_2)-OH were prepared as previously described.[50] The substitution of the resins was determined by spectrophotometric analysis at 290 nm of the dibenzofulvene-piperidine adduct formed upon deprotection of the amino terminal using a Perkin-Elmer Lambda 7 UV/vis spectrophotometer. Purification of peptides was performed by preparative reverse phase HPLC on a Waters HPLC system with a delta pak C-18 column (200 × 25 mm) and a linear gradient of A (0.1% TFA in water) and B (0.1% TFA in 90% aqueous MeCN) at a flow rate of 20 mL/min. Amino acid analyses were carried out in a Beckman 600 amino acid analyzer following hydrolysis with 6 M HCl with 5% (v/v) phenol at 110 °C for 48 h. Amino acid sequencing was performed on resin-bound substrates using an Applied Biosystems Sequencer models 477A or 470A equipped with an on-line phenylthiohydantoin analyzer (model 120A) according to the protocol of the manufacturer. MALDI-TOF-MS of synthetic peptides was performed on a Finnigan Lasermat 2000 with a matrix of α-cyano-4-hydroxycinnamic acid while peptides from library beads were analyzed in the reflectron mode on a Bruker Reflex III using the same matrix. ES mass spectra were recorded with a VG-Quatro instrument from Fisons. Enzyme kinetics were performed using a temperature-controlled Perkin-Elmer luminescence Spectrometer (LS 50B).

Enzyme. Papain, twice recrystallized from papaya latex, was purchased from Sigma (Lot # 84H7220) and used without further purification. The molar concentration was determined by active site titration with E-64 [*trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane] (Sigma; Lot # 77H0328) as previously described using Abz-CRQQY-(NO_2)D-OH as the substrate.[41] The concentration of enzyme in a 100× diluted solution (1X) was 8.34 μM. For assays, the 1X solution was diluted 20-fold giving rise to the 20X solution. Papain was activated as the 1X solution by incubation in activation buffer (50 mM phosphate buffer, pH 7.2, augmented with 10 mM cysteine, 1 mM EDTA, and 0.08% Brij 35) for 10 min at room temperature. A fresh solution of enzyme was activated every 3 h for use in solution phase assays.

General Methods for Solid Phase Peptide Synthesis. Syntheses of peptides and libraries were carried out manually by MCPS.[51,52] on PEGA resin respectively in a 20 column Teflon synthesis block. In general, N^{α}-Fmoc amino acid OPfp esters with the following side chain protecting groups—*O*-*t*Bu for Asp and Glu, *t*Bu for Tyr, Ser and Thr, Trt for Cys, Asn and Gln, Boc for His, Lys and Trp and Pmc for Arg—were used for the synthesis. Each coupling step was carried out using the amino acid (3 equiv) in DMF along with Dhbt-OH (1 equiv) as an acylation catalyst as well as an indicator of the reaction completeness.[53] In certain cases, reaction completion was also assessed using the Kaiser test. To coupling reactions that were incomplete was added an additional 1−2 equiv of the amino acid ester. Each coupling step was followed by washing with DMF (6×), removal of the Fmoc group by treatment with 20% piperidine in DMF (4 + 16 min), and then another DMF washing step (6×). At the end of the synthesis, the resin was washed with CH_2Cl_2 (6×) and dried by air suction over a period of 1 h. The side chain protecting groups were removed by treatment with a mixture of TFA:thioanisole:ethanedithiol:water (87.5:5:2.5:5) initially for 10 min and then for 2.5 h. The resin was then washed with 95% acetic acid (4×), DMF (2×), 5% DIPEA (2×), DMF (2×), and CH_2Cl_2 (6×) then dried in vacuo. Cleavage of peptides from the resin was effected by treatment with 0.1 M NaOH for 2 h followed by washing with water (7×). The combined filtrate was neutralized with 0.1 N HCl, and the crude peptides were purified by preparative HPLC.

Solid Phase Substrate Library Synthesis. The libraries of the structure Y(NO_2)$X_8X_7X_6X_5X_4X_3X_2X_1$K(Abz) (Figure 1) containing all 20 genetically encoded amino acids were prepared on $PEGA_{4000}$ resin.[11] The capacity of the resin was doubled (final loading 0.21 mmol/g) by synthesizing the peptide on the amino group of the side chains of two lysines (Figure 1). Using the syringe technology[54] two FmocLys-(Boc)OPfp were coupled to the $PEGA_{4000}$ resin (1,7 g, 0.12 mmol/g; 200−800 μm beads), and after removal of the Fmoc group, the N-terminus was acetylated and the Boc side chain protecting groups were removed by treatment with 50% TFA in CH_2Cl_2 for 30 min. After the appropriate washing protocol, HMBA (3 equiv) in DMF was coupled under TBTU (2.9 equiv)−NEM (4.5 equiv) activation. The resin was washed and dried by lyophilization before the coupling of Fmoc-

Lys(BocAbz)-OH (2.5 equiv) in dry CH_2Cl_2 under the agency of MSNT (2 equiv) and MeIm (2.5 equiv).[55] The reaction was allowed to proceed for 45 min, after which the resin was washed with DMF (2×) and CH_2Cl_2 (1×). The coupling procedure was repeated for another 45 min and the resin washed with DMF (7×). The resin was evenly distributed in the 20 wells of a Teflon synthesis block, and the Fmoc group was removed. An Fmoc amino acid OPfp ester was coupled into each of the 20 wells. After completion of the coupling, the block was filled with DMF up to 1 cm above the top of the wells and inverted, and the resin was mixed vigorously for 30 min in the mixing chamber. The block was once more inverted, evenly distributing the resins once more into the wells for washing and Fmoc deprotection protocols. This procedure was repeated for the incorporation of the first seven amino acids of the library in the case of library 1. In the case of library 2, after the coupling of the seven randomized positions, FmocProOPfp was added to all the wells. In both libraries, Fmoc(Y(NO_2))-OH was incorporated using TBTU/NEM preactivation for 7 min. The library was deprotected and washed as described in general SPPS methods. The mass and purity of a random sampling of compounds attached to 36 beads (ca. 100 pmol/bead) were analyzed by high resolution MALDI-TOF-MS. For this purpose, 36 resin beads were randomly chosen and the peptides cleaved off by treatment with 10% TEA in MeOH. For some of those compounds, the sequences were also determined by Edman degradation and correlated with the masses obtained.

Solid Phase Multiple Column Peptide Synthesis. Lead peptides from the library screen were synthesized on $PEGA_{800}$ resin (0.21 mmol/g, 4 μM/well, 150−300 μm beads). The HMBA linker was attached as described for the library synthesis with the exception that a glycine was first coupled to the resin before coupling of the linker. Peptide synthesis was carried out as described for library synthesis with the omission of the mixing step. After deprotection and washing, the peptides were cleaved with 0.1 M NaOH (350 μL/well) for 2 h and the products collected in tubes placed beneath the wells. The solutions were neutralized with 0.1 M HCl, and the peptides were purified by reverse phase HPLC. The peptides were characterized by MALDI-TOF-MS or ES-MS, and their purity was assessed by analytical HPLC at 215 and 320 nm.

Solid Phase Library Screening. The library beads (300 mg) were washed with assay buffer (50 mM phosphate buffer, pH 7.2, augmented with 2 mM cysteine, 1 mM EDTA, and 0.08% Brij 35; 3×) and then incubated with activated papain (135 μL of 20X in 10 mL buffer; 5 nM) at 25 °C in a small glass Petri dish (4 cm diameter). The fluorescence intensity of the beads was monitored with a fluorescent microscope every 30 min for indications of hydrolysis. After 2.5 h for library 1 and 3.5 h for library 2, several beads showed a fluorescent "ring", indicating hydrolysis of some of the peptides on the beads. The reaction was then stopped by treatment with 2% aqueous TFA solution and washed with water (2×), 2% $NaHCO_3$ (2×), and then water (3×). The fluorescence intensity of the beads was assessed by inspection with a fluorescent microscope, and bright beads were collected and transferred to a TFA-treated cartridge filter for on-resin sequence analysis. The amino acid sequence and the cleavage point of the peptide substrates were determined by Edman degradation. The extent of cleavage was determined by a comparison of the picomoles of an amino acid in both the noncleaved and cleaved peptide in different cycles of the degradation.

Enzymatic Hydrolysis of Fluorescence-Quenched Substrates in Solution. Substrates were dissolved in water, DMF, or combinations thereof to a final concentration of approximately 5 mM. Hydrolysis of substrates was carried out at 25 °C in 50 mM phosphate buffer, pH 6.8, augmented with 2 mM cysteine, 1 mM EDTA, and 0.08% Brij 35. Hydrolysis was followed by measurement of the increase in intensity of the Abz fluorescence (λ_{EX} = 320 nm, λ_{EM} = 420 nm, slit width: 10 nm). Under pseudo first-order conditions, the rate of initial hydrolysis was measured continuously for 5 min for four concentrations of each substrate. Total hydrolysis was effected by the addition of 50 μL of 1X enzyme (final concentration of 0.2 μM), and final fluorescence was measured after 24 h. k_{cat}/K_M values were calculated from the following equation: k_{cat}/K_M = slope/($[E]_o(I_f - I_o)$) where slope = $\nu = \Delta I/\Delta t$, $[E]_o$ is initial enzyme concentration, and $(I_f - I_o) \equiv [S]$ assuming $[S] \ll K_M$. For determinations of k_{cat} and K_M, assays were carried out under similar conditions in a 100 μL flow cell using six different substrate concentrations ranging from 2 to 150 μM. The slit width was reduced to 2.5 nm to compensate for the higher fluorescent intensities of more concentrated solutions. The slope was converted into moles of substrate hydrolyzed per second by use of a standard curve. Substrate concentrations were also correct using the standard curve. The standard curve was obtained from measurement of the fluorescence intensity of AbzGAGAAF-OH derived from the total hydrolysis of AbzGAGAAFFA-Y(NO_2)D-OH by subtilisin at different concentrations (20−180 μM). The kinetic parameters k_{cat} and K_M for each substrate were then determined from Hanes plots ($[S]/\nu$ vs $[S]$) of the measured values.

Determination of Cleavage Point in Solution Phase Assays. The cleavage point was determined in separate microassays on more concentrated solutions but maintaining a similar enzyme-to-substrate ratio as in the solution assays above. Substrates (6 μL of 5 mM stock) were incubated with 2 μL of 1X enzyme (activated with activation buffer without Brij 35 since peaks from Brij occur in the mass range of interest) in 12 μL assay buffer (without Brij 35) for 30 min. The hydrolysis mixture (2 μL) was mixed with CHC matrix (1 μL), and the MALDI-TOF spectra were acquired and analyzed (Figure 4).

Modeling. Molecular dynamics calculations were carried out on a Silicon Graphics Octane workstation using the InsightII/Discover program. The coordinates from the crystal structure of papain complexed with peptide inhibitor (Suc-QVVAA-pNA) determined at 1.7 Å resolution[42] were used as input for the calculations after removal of the inhibitor. The calculations were performed with seven different substrates **1**, **18**, **21**, **26**, **28**, **35**, and **36** in order to explain variations in the substrate preferences found in the two "one-bead-one-substrate" libraries. During all calculations, most

of the amide backbone of papain was fixed, while all the side chains were allowed to move. Each calculation was carried out as annealing at decreasing temperatures, 650 and then 500 and 300 K. At 500 and 300 K, the papain residues in contact with the substrate were allowed to move freely in sequence ranges of five successive amino acids or less. The substrates were initially energy-minimized and added in extended conformations approximately 15 Å away from the binding site. A weak constraint of ~3 kcal/mol (boundaries 3 and 4 Å) was added between the scissile bond and the active thiol, and this was maintained throughout all calculations. Initially, in the first calculations a weak constraint of ~3 kcal/mol (boundaries 4 and 8 Å) between the terminal residues and papain residues at either end of the active cleft were added; however, this constraint was later omitted since no significantly different results were obtained with these extra constraints. Additional constraints of ~3 kcal/mol (boundaries 2 and 3 Å) were applied for distances between the sulfhydryl hydrogen and π-nitrogen of H159 and between the τ-NH of the H159 and the N175 side chain carbonyl. The calculations were initiated with 10 steps of minimization and then 30 000 to 60 000 steps were calculated at each temperature at 1 fs intervals. The development of each calculation was monitored, and when persistent obstructions to the progress (i.e. distance of >5 Å between the sulfur atom and the scissile carbonyl carbon) of the calculation were observed, a new starting substrate conformation was used. Only the φ-ψ angles allowed in the φ-ψ space were used as starting conformations. As many as five independent successful calculations were performed for each substrate, and when the calculations were allowed to reach equilibrium, the results obtained with different starting conformations converged to one or two related bound conformations. Constraints were completely removed from the final structure and the complex was subjected to extended calculations, in order to determine whether the substrate remained bound to the papain molecule in a stable conformation.

Acknowledgment. This work was supported by the INCO-DC program (EU contract number ERBIC18CT97-0225), the Danish National Research Foundation, and the Brazilian Research Foundations FAPESP and PADCT. Dr. Ib Svendsen, Bodil Corneliussen, and Lone Sørensen are acknowledged for peptide sequencing. Dr. David Simpson is thanked for many helpful discussions about enzyme kinetics and Hanne Christiansen for help with peptide synthesis. High resolution MALDI-TOF-MS and ES-MS were carried out by Pia Breddam and Anita Jansson.

References and Notes

(1) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−493.

(2) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, *354*, 82−84.

(3) Rapp, W.; Zhang, L.; Häbish, R.; Bayer, E. Polystyrene-Polyoxyethylene graftcopolymers for high-speed peptide synthesis. In *Peptides 1988, Proc. Eur. Pept. Symp.*, 1 ed.; Jung, G., Bayer, E., Eds.; Walter de Gruyter: Berlin, 1989; pp 199−201.

(4) Meldal, M. PEGA: A flow stable poly(ethylene glycol) dimethyl acrylamide copolymer for solid phase synthesis. *Tetrahedron Lett.* **1992**, *33*, 3077−3080.

(5) Renil, M.; Meldal, M. POEPOP and POEPS: Inert poly(ethylene glycol) cross-linked polymeric supports for solid phase synthesis. *Tetrahedron Lett.* **1996**, *37*, 6185−6188.

(6) Buchardt, J.; Meldal, M. A chemically inert hydrophilic resin for solid phase organic synthesis. *Tetrahedron Lett.* **1998**, *39*, 8695−8698.

(7) Rademann, J.; Grötli, M.; Meldal, M.; Bock, K. SPOCC: Resin for solid phase organic chemistry and enzyme reactions. *J. Am. Chem. Soc.* **1999**, *121*, 5459−5466.

(8) Breddam, K.; Meldal, M. Substrate preferences of glutamic-acid-specific endopeptidases assessed by peptide substrates based on intramolecular fluorescence quenching. *Eur. J. Biochem.* **1992**, *206*, 103−107.

(9) Grøn, H.; Meldal, M.; Breddam, K. Extensive comparison of substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching. *Biochemistry* **1992**, *31*, 6011−6018.

(10) Juliano, M. A.; Nery, E. D.; Scharfstein, J.; Meldal, M.; Svendsen, I.; Walmsley, A.; Juliano, L. Characterization of substrate specificity of the major cysteine protease (Cruzipain) from *Trypanosoma cruzi*. *Biochem. J.* **1997**, *323*, 427−433.

(11) Renil, M.; Meldal, M.; Delaisse, J.-M.; Foged, N. T. Fluorescent quenched peptide libraries as a tool for identification of enzyme substrates for matrix metalloproteinase (MMP)-9 from osteoclasts. In *Peptides 1996, Proc. Eur. Pept. Symp.* 1 ed.; Ramage, R.; Epton, R., Eds.; Mayflower Scientific: Kingswindford, 1997; pp 753−754.

(12) Meldal, M.; Svendsen, I. Direct visualization of enzyme inhibitors using a portion mixing inhibitor library containing a quenched fluorogenic peptide substrate. 1: Inhibitors for subtilisin Carlsberg. *J. Chem. Soc., Perkin Trans. 1* **1995**, 1591−1596.

(13) Meldal, M.; Svendsen, I.; Juliano, L.; Juliano, M. A.; Del Nery, E.; Scharfstein, J. Inhibition of Cruzipain visualized in a fluorescence quenched solid phase inhibitor library. D-Amino acid inhibitors for Cruzipain, cathepsin B and cathepsin L. *J. Pept. Sci.* **1998**, *4*, 83−91.

(14) Grøn, H.; Breddam, K. Interdependency of the binding subsites in subtilisin. *Biochemistry* **1992**, *31*, 8967−8971.

(15) Meldal, M.; Svendsen, I.; Breddam, K.; Auzanneau, F.-I. Portion-mixing peptide libraries of quenched fluorogenic substrates for complete subsite mapping of endoprotease specificity. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3314−3318.

(16) Meldal, M. Combinatorial solid phase assay for enzyme activity and inhibition. In *Combinatorial Peptide Libraries*, 1 ed.; Schmuel, C., Ed.; Humana Press: Totowa, NJ, 1998; pp 51−82.

(17) Serveau, C.; Juliano, L.; Bernard, P.; Moreau, T.; Mayer, R.; Gauthier, F. New substrates of papain based on the conserved sequence of natural inhibitors of the cystatin family. *Biochemie* **1994**, *76*, 153−158.

(18) Harrison, M. J.; Burton, N. A.; Hillier, I. H. Catalytic mechanism of the enzyme papain: predictions with a hybrid quantum mechanical/molecular mechanical potential. *J. Am. Chem. Soc.* **1997**, *119*, 12285−12291.

(19) Schechter, I.; Berger, A. On the size of the active site of proteases I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157−162.

(20) Brubacher, L. J.; Zaher, M. R. A kinetic study of hydrophobic interactions at the S_1 and S_2 sites of papain. *Can. J. Biochem.* **1979**, *57*, 1064−1072.

(21) Alves, L. C.; Almeida, P. C.; Franzoni, L.; Juliano, L.; Juliano, M. A. Synthesis of N^{α}-protected aminoacyl 7-amino-4-methylcoumarin amide by phosphorous oxychloride and preparation of specific fluorogenic substrates for papain. *Pept. Res.* **1996**, *9*, 92−96.

(22) Menard, M.; Carmona, E.; Plouffe, C.; Bromme, D.; Konishi, Y.; Lefebvre, J.; Storer, A. C. The specificity of the S_1' subsite of cysteine proteases. *FEBS Lett.* **1993**, *328*, 107−110.

(23) Garcia-Echeverria, C.; Rich, D. H. New intramolecularly quenched fluorogenic peptide substrates for the kinetic specificity of papain. *FEBS Lett.* **1992**, *297*, 100−102.

(24) Garcia-Echeverria, C.; Rich, D. H. Effect of P_2' substituents on kinetic constants for hydrolysis by cysteine proteinases. *Biochem. Biophys. Res. Commun.* **1993**, *187*, 615−619.

(25) Garcia-Echeverria, C.; Rich, D. H. Kinetic studies of papain: Effect of P_3' substituents and donor/acceptor pairs of intramolecularly quenched fluorogenic substrates. *Lett. Pept. Sci.* **1995**, *2*, 77−82.

(26) Gauthier, F.; Moreau, T.; Lalmanach, G.; Brillard-Bourdet, M.; Ferrer-Di Martino, M.; Juliano, L. A new, sensitive fluorogenic substrate for papain based on the sequence of the cystatin inhibitory site. *Arch. Biochem. Biophys.* **1993**, *306*, 304−308.

(27) Lalmanach, G.; Hoebeke, J.; Moreau, T.; Brillard-Bourdet, M.; Ferrer-Di Martino, M.; Borras-Cuesta, F.; Gauthier, F. Interaction between cystatin-derived peptides and papain. *J. Protein Chem.* **1993**, *12*, 23−31.

(28) Lalmanach, G.; Serveau, C.; Brillard-Bourdet, M.; Chagas, J. R.; Mayer, R.; Juliano, L.; Gauthier, F. Conserved cystatin segments as models for designing specific substrates and inhibitors of cysteine proteinases. *J. Protein Chem.* **1995**, *14*, 645−653.

(29) Leon, S.; Quarrell, R.; Lowe, G. Evaluation of resins for on-bead screening: a study of papain and chymotrypsin specificity using pega-bound combinatorial peptide libraries. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2997−3002.

(30) Renil, M.; Ferreras, M.; Delaisse, J. M.; Foged, N. T.; Meldal, M. PEGA supports for combinatorial peptide synthesis and solid phase enzymatic library assays. *J. Pept. Sci.* **1998**, *4*, 195−210.

(31) Auzanneau, F. I.; Meldal, M.; Bock, K. Synthesis, characterization and biocompatibility of PEGA resins. *J. Pept. Sci.* **1994**, *1*, 31−44.

(32) Atherton, E.; Sheppard, R. C. Solid phase peptide synthesis using N-α-fluorenylmethyloxycarbonylamino acid pentafluorophenyl esters. *J. Chem. Soc., Chem. Commun.* **1985**, 165−166.

(33) Knorr, R.; Trzeciak, A.; Bannwarth, W.; Gillessen, D. New coupling reagents in peptide synthesis. *Tetrahedron Lett.* **1989**, *30*, 1927−1930.

(34) Smart, S. S.; Mason, T. J.; Bennell, P. S.; Maeji, N. J.; Geysen, H. M. High-throughput purity estimation and characterisation of synthetic peptides by electrospray mass spectrometry. *Int. J. Pept. Protein Res.* **1998**, *47*, 47−55.

(35) Schechter, I.; Berger, A. On the active site of proteases. III. Mapping the active site of papain and specific peptide inhibitors of papain. *Biochem. Biophys. Res. Commun.* **1968**, *32*, 898−902.

(36) Lecaille, F.; Serveau, C.; Gauthier, F.; Lalmanach, G. Revisiting the S_2 specificity of papain by structural analogues of Phe. *FEBS Lett.* **1999**, *445*, 311−314.

(37) Berti, P. J.; Faerman, C. H.; Storer, A. C. Cooperativity of papain-substrate interaction energies in the S_2 to S_2' subsites. *Biochemistry* **1991**, *30*, 1394−1402.

(38) Schmidt, J.; Wermann, M.; Rosche, F.; Demuth, H.-U. The use of MALDI-TOF mass spectrometry in quantification of the stability of prolyl endopeptidase. *Protein Pept. Lett.* **1996**, *3*, 385−392.

(39) Valero, M.-L.; Giralt, E.; Andreu, D. An investigation of residue-specific contributions to peptide desorption in MALDI-TOF mass spectrometry. *Lett. Pept. Sci.* **1999**, *6*, 109−115.

(40) Groves, M.; Coulombe, R.; Jenkins, J.; Cygler, M. Structural basis for specificity of papain-like cysteine protease proregions towards their cognate enzymes. *Proteins: Struct. Funct. Genet.* **1998**, *32*, 504−514.

(41) Davy, A.; Svendsen, I.; Sørensen, S. O.; Sørensen, M. B.; Rouster, J.; Meldal, M.; Simpson, D. J.; Cameron-Mills, V. Substrate specificity of two barley cysteine endoproteases EP-A and EP-B. *Plant Physiol.* **1998**, *117*, 255−261.

(42) Yamamoto, A.; Tomoo, K.; Doi, M.; Ohishi, H.; Inoue, M.; Ishida, T.; Yamamoto, D.; Tsuboi, S.; Okamoto, H.; Okada, Y. Crystal structure of papain-succinyl-Gln-Val-Val-Ala-Ala-*p*-nitroanilide complex at 1.7-Å resolution: noncovalent binding mode of a common sequence of endogenous thiol protease inhibitors. *Biochemistry* **1992**, *31*, 11305−11309.

(43) Lalonde, J. M.; Zhao, B.; Smith, W. W.; Janson, C. A.; DesJarlais, R. L.; Tomaszek, T. A.; Carr, T. J.; Thompson, S. K.; Oh, H.-J.; Yamashita, D. S.; Veber, D. F.; Abdel-Meguid, S. Use of papain as a model for the structure-based design of Cathepsin K inhibitors: crystal structures of two papain-inhibitor complexes demonstrate binding to S′-subsites. *J. Med. Chem.* **1998**, *41*, 4567−4576.

(44) Thornberry, N. A.; Rano, T. A.; Peterson, E. P.; Rasper, D. M.; Timkey, T.; Garcia-Calco, M.; Houtzager, V. M.; Nordstrom, P. A.; Roy, S.; Vaillancourt, J. P.; Chapman, K. T.; Nicholson, D. W. A combinatorial approach defines specificities of members of the capase family and granzyme B. *J. Biol. Chem* **1997**, *272*, 17907−17911.

(45) Rano, T. A.; Timkey, T.; Peterson, E. P.; Rotonda, J.; Nicholson, D. W.; Becker, J. W.; Chapman, K. T.; Thornberry, N. A. A combinatorial approach for determining protease specificities: application to interleukin-1β converting enzyme (ICE). *Chem. Biol.* **1997**, *4*, 149−155.

(46) Arnold, D.; Keilholz, W.; Schild, H.; Dumrese, T.; Stevanovic, S.; Rammensee, H.-G. Substrate specificity of cathepsins D and E determined by N-terminal and C-terminal sequencing of peptide pools. *Eur. J. Biochem.* **1997**, *249*, 171−179.

(47) Eichler, J.; Lucka, A. W.; Houghten, R. A. Cyclic peptide template combinatorial libraries: synthesis and identification of chymotrypsin inhibitors. *Pept. Res.* **1994**, *7*, 300−307.

(48) Wang, G. T.; Matayoshi, E.; Huffaker, H. J.; Krafft, G. A. Design and synthesis of new fluorogenic HIV protease substrates based on resonance energy transfer. *Tetrahedron Lett.* **1990**, *31*, 6493−6496.

(49) Turk, D.; Guncar, G.; Podobnik, M.; Turk, B. Revised definition of Substrate binding sites of papain-like cysteine proteases. *Biol. Chem.* **1998**, *379*, 137−147.

(50) Meldal, M.; Breddam, K. Anthranilamide and nitrotyrosine as a donor acceptor pair in internally quenched fluorescent substrates for endopeptidases − Multicolumn peptide synthesis of enzyme substrates for subtilisin carlsberg and pepsin. *Anal. Biochem.* **1991**, *195*, 141−147.

(51) Holm, A.; Meldal, M. Multiple Column Peptide Synthesis. In *Peptides 1988, Proc. Eur. Pept. Symp., 20th*, 1 ed.; Jung, G., Bayer, E., Eds.; Berlin: Walter de Gruyter, 1989; pp 208−210.

(52) Meldal, M.; Holm, C. B.; Bogginess, G.; Jacobean, M. H.; Holm, A. Multiple column peptide synthesis, Part 2(1,2). *Int. J. Pept. Protein Res.* **1993**, *41*, 250−260.

(53) Atherton, E.; Cameron, L.; Meldal, M.; Sheppard, R. C. Self-indicating activated esters for use in solid phase peptide synthesis. Fluorenylmethoxycarbonyl amino acid derivatives of 3- hydroxy-4-oxodihydrobenzotriazine. *J. Chem. Soc., Chem. Commun.* **1986**, 1763−1765.

(54) Christensen, M. K.; Meldal, M.; Bock, K. Synthesis of mannose 6-phosphate-containing disaccharide threonine building blocks and their use in solid phase glycopeptide synthesis. *J. Chem. Soc., Perkin Trans. 1* **1993**, 1453−1460.

(55) Blankemeyer-Menge, B.; Nimtz, M.; Frank, R. An Efficient method for anchoring Fmoc-amino acids to hydroxyl-functionalised solid supports. *Tetrahedron Lett.* **1990**, *31*, 1701−1704.

CC990031U