

# Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries

Jennifer L. Harris\*<sup>†‡</sup>, Bradley J. Backes\*<sup>†§</sup>, Francesco Leonetti<sup>§</sup>, Sami Mahrus\*, Jonathan A. Ellman<sup>§</sup>, and Charles S. Craik\*

\*Department of Pharmaceutical Chemistry, Program in Chemistry and Chemical Biology, University of California, San Francisco, CA 94143; and <sup>§</sup>Chemistry Department, University of California, Berkeley, CA 94720

Edited by Christopher T. Walsh, Harvard Medical School, Boston, MA, and approved May 5, 2000 (received for review March 24, 2000)

A method is presented for the preparation and use of fluorogenic peptide substrates that allows for the configuration of general substrate libraries to rapidly identify the primary and extended specificity of proteases. The substrates contain the fluorogenic leaving group 7-amino-4-carbamoylmethylcoumarin (ACC). Substrates incorporating the ACC leaving group show kinetic profiles comparable to those with the traditionally used 7-amino-4-methylcoumarin (AMC) leaving group. The bifunctional nature of ACC allows for the efficient production of single substrates and substrate libraries by using 9-fluorenylmethoxycarbonyl (Fmoc)-based solid-phase synthesis techniques. The approximately 3-fold-increased quantum yield of ACC over AMC permits reduction in enzyme and substrate concentrations. As a consequence, a greater number of substrates can be tolerated in a single assay, thus enabling an increase in the diversity space of the library. Soluble positional protease substrate libraries of 137,180 and 6,859 members, possessing amino acid diversity at the P4-P3-P2-P1 and P4-P3-P2 positions, respectively, were constructed. Employing this screening method, we profiled the substrate specificities of a diverse array of proteases, including the serine proteases thrombin, plasmin, factor Xa, urokinase-type plasminogen activator, tissue plasminogen activator, granzyme B, trypsin, chymotrypsin, human neutrophil elastase, and the cysteine proteases papain and cruzain. The resulting profiles create a pharmacophoric portrayal of the proteases to aid in the design of selective substrates and potent inhibitors.

The ability of an enzyme to discriminate among many potential substrates is an important factor in maintaining the fidelity of most biological functions. While substrate selection can be regulated on many levels in a biological context, such as spatial and temporal localization of enzyme and substrate, concentrations of enzyme and substrate, and requirement of cofactors, the substrate specificity at the enzyme active site is the overriding principle that determines the turnover of a substrate. Characterization of the substrate specificity of an enzyme clearly provides invaluable information for the dissection of complex biological pathways. Definition of substrate specificity also provides the basis for the design of selective substrates and inhibitors to study enzyme activity.

Of the genomes that have been completely sequenced, 2% of the gene products encode proteases (1). This family of enzymes is crucial to every aspect of life and death of an organism. With the identification of new proteases, there is a need for the development of rapid and general methods to determine protease substrate specificity. While several biological methods, such as peptides displayed on filamentous phage (2, 3), and chemical methods, such as support-bound combinatorial libraries (4), have been developed to identify proteolytic substrate specificity, few offer the ability to rapidly and continuously monitor proteolytic activity against complex mixtures of substrates in solution.

The use of 7-amino-4-methylcoumarin (AMC) fluorogenic peptide substrates is a well-established method for the determination of protease specificity (5). Specific cleavage of the amide bond liberates the fluorogenic AMC leaving group, allowing for the simple determination of cleavage rates for individual substrates. More recently, arrays (6) and positional-scanning libraries (7) of AMC peptide substrate libraries have been used to rapidly profile the N-terminal specificity of proteases by sampling a wide range of

substrates in a single experiment. Each of these published efforts was designed for profiling caspases, cysteine proteases that require an Asp residue at the P1-position<sup>¶</sup> for substrate turnover. This requirement allows for the convenient attachment of the P1 Asp to the solid support through the carboxylic acid side chain. Because most proteases do not require P1 Asp/Glu for activity, libraries generated by these methods have limited applicability. Naturally, fluorogenic substrates that contain P1 amino acids that do not possess adequate side-chain functionality for attachment to a solid support in a straightforward manner (Gly, Leu, Val, Ile, Ala, Pro, Phe) will not be amenable to similar synthetic strategies.

We have recently developed 9-fluorenylmethoxycarbonyl (Fmoc)-based synthesis methods to displace support-bound peptides with nucleophiles in a final cleavage step to produce C-terminal-modified peptides (9). To prepare fluorogenic peptide substrates with any residue at the P1 position, AMC-amino acid derivatives are prepared and then used as nucleophiles to produce the AMC-peptide substrates (20). While this method has great utility, it became clear that a considerably more efficient and general strategy to prepare fluorogenic peptide substrate libraries could be achieved by meeting the following objectives: (i) The solid-phase synthesis method should enable *direct* incorporation of all 20 proteinogenic amino acids at every position, including the P1 position. (ii) The method should be compatible with Fmoc-based solid-phase peptide synthesis protocols that are amenable to automation. (iii) The method should be flexible enough to enable the rapid synthesis of any single substrate, substrate array, and positional scanning library.

Here we report a highly efficient method for the preparation of fluorogenic peptide substrate libraries based on the bifunctional fluorogenic leaving group 7-amino-4-carbamoylmethylcoumarin (ACC). By using Fmoc-synthesis protocols, all 20 proteinogenic amino acids can be *directly* coupled to the support-bound ACC leaving group to provide general sets of substrates for analyzing protease substrate specificity. The versatility of the solid-phase synthesis strategy allows for substrate arrays (6) and positional scanning libraries (7) of any configuration to be rapidly prepared. The substrate specificities of numerous representative serine and cysteine proteases are profiled in this paper to show the utility and generality of libraries generated by the ACC method.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: AMC, 7-amino-4-methylcoumarin; ACC, 7-amino-4-carbamoylmethylcoumarin; Nle, norleucine; PS-SCL, positional scanning-synthetic combinatorial library; DICl, diisopropylcarbodiimide; HOBT, 1-hydroxybenzotriazole; TFA, trifluoroacetic acid; Fmoc, 9-fluorenylmethoxycarbonyl; Pbf, 2,2,4,6,7-pentamethylidihydrobenzofuran-5-sulfonyl; Trt, trityl; Boc, *tert*-butoxycarbonyl; DMF, *N,N*-dimethylformamide; TIS, triisopropylsilane; HATU, *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; uPA, urokinase-type plasminogen activator; tPA, tissue plasminogen activator.

<sup>†</sup>Present address: Chemistry Department, Genomics Institute of the Novartis Research Foundation, 3115 Merryfield Row, San Diego, CA 92121.

<sup>‡</sup>J.L.H. and B.J.B. contributed equally to this work.

<sup>¶</sup>Nomenclature for the substrate amino acid preference is P<sub>n</sub>, . . . , P<sub>2</sub>, P<sub>1</sub>, P<sub>2</sub>' . . . , P<sub>m</sub>'. Amide bond hydrolysis occurs between P<sub>1</sub> and P<sub>1</sub>' (8).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.140132697. Article and publication date are at [www.pnas.org/cgi/doi/10.1073/pnas.140132697](http://www.pnas.org/cgi/doi/10.1073/pnas.140132697)

## Materials and Methods

**Reagents and General Methods.** Rink Amide AM resin and Fmoc-amino acids were from Novabiochem (San Diego). The amine substitution level of the Rink resin (0.80 meq/g) was determined by a spectrophotometric Fmoc-quantitation assay (10). Anhydrous *N,N*-dimethylformamide (DMF) was from EM Science (Hawthorne, NY). *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) was from PerSeptive Biosystems (Foster City, CA). Diisopropylcarbodiimide (DICl), 1-hydroxybenzotriazole (HOBt), AcOH, Fmoc-Cl, trifluoroacetic acid (TFA), collidine, and triisopropylsilane (TIS) were from Aldrich. An Argonaut Quest 210 Organic Synthesizer was used to prepare Fmoc-P1-substituted ACC resins. Library synthesis was performed in 96-well plates by using the MultiChem synthesis apparatus of Robbins Scientific (Sunnyvale, CA). Human thrombin, plasmin, and factor Xa were used as received from Hematologic Technologies (Essex Junction, VT). Human light-chain urokinase-type plasminogen activator (uPA), tissue plasminogen activator (tPA), and neutrophil elastase were used as received from Calbiochem. Rat granzyme B was expressed and purified as described (11). Cruzain was expressed and purified as described (12). Rat trypsin was expressed and purified as described (13).

**ACC-Resin Synthesis.** 7-Fmoc-aminocoumarin-4-acetic acid was prepared by treating 7-aminocoumarin-4-acetic acid (14, 15) with Fmoc-Cl. 7-Aminocoumarin-4-acetic acid (10.0 g, 45.6 mmol) and H<sub>2</sub>O (228 ml) were mixed. NaHCO<sub>3</sub> (3.92 g, 45.6 mmol) was added in small portions followed by the addition of acetone (228 ml). The solution was cooled with an ice bath, and Fmoc-Cl (10.7 g, 41.5 mmol) was added with stirring over the course of 1 h. The ice bath was removed and the solution was stirred overnight. The acetone was removed with rotary evaporation and the resulting gummy solid was collected by filtration and washed with several portions of hexane. The material was dried over P<sub>2</sub>O<sub>5</sub> to give 14.6 g (80%) of cream-colored solid. <sup>1</sup>H NMR (400 MHz): δ 3.86 (s, 2), 4.33 (t, 1, *J* = 6.2), 4.55 (d, 2, *J* = 6.2), 6.34 (s, 1), 7.33–7.44 (m, 5), 7.56 (s, 1), 7.61 (d, 1, *J* = 8.6), 7.76 (d, 2, *J* = 7.3), 7.91 (d, 2, *J* = 7.4), 10.23 (s, 1), 12.84 (s, 1). <sup>13</sup>C NMR (101 MHz): δ 37.9, 47.4, 66.8, 67.2, 105.5, 114.6, 115.3, 121.1, 125.9, 126.9, 128.0, 128.6, 141.6, 143.6, 144.5, 150.7, 154.1, 154.8, 160.8, 171.4.

ACC-resin was prepared by condensation of Rink Amide AM resin with 7-Fmoc-aminocoumarin-4-acetic acid. Rink Amide AM resin (21 g, 17 mmol) was solvated with DMF (200 ml). The mixture was agitated for 30 min and filtered with a filter cannula, whereupon 20% piperidine in DMF (200 ml) was added. After agitation for 25 min, the resin was filtered and washed with DMF (3 times, 200 ml each). 7-Fmoc-aminocoumarin-4-acetic acid (15 g, 34 mmol), HOBt (4.6 g, 34 mmol), and DMF (150 ml) were added, followed by DICl (5.3 ml, 34 mmol). The mixture was agitated overnight, filtered, washed (DMF, three times with 200 ml; tetrahydrofuran, three times with 200 ml; MeOH, three times with 200 ml), and dried over P<sub>2</sub>O<sub>5</sub>. The substitution level of the resin was 0.58 mmol/g (>95%) as determined by Fmoc analysis (10).

**P1-Substituted ACC-Resin Synthesis.** Fmoc-ACC-resin (100 mg, 0.058 mmol) was added to 20 reaction vessels of an Argonaut Quest 210 Organic Synthesizer and solvated with DMF (2 ml). The resin was filtered and 20% piperidine in DMF (2 ml) was added to each vessel. After agitation for 25 min, the resin was filtered and washed with DMF (three times with 2 ml). An Fmoc-amino acid (0.29 mmol), DMF (0.7 ml), collidine (76 μl, 0.58 mmol), and HATU (110 mg, 0.29 mmol) were added to the designated reaction vessel, followed by agitation for 20 h. The resins were then filtered, washed with DMF (three times with 2 ml), and subjected a second time to the coupling conditions. A solution of AcOH (40 μl, 0.70 mmol), DICl (110 μl, 0.70 mmol), and nitrotriazole (80 mg, 0.70 mmol) in DMF (0.7 ml) was added to each of the reaction vessels, followed by agitation for 24 h. The resins were filtered, washed (DMF, three

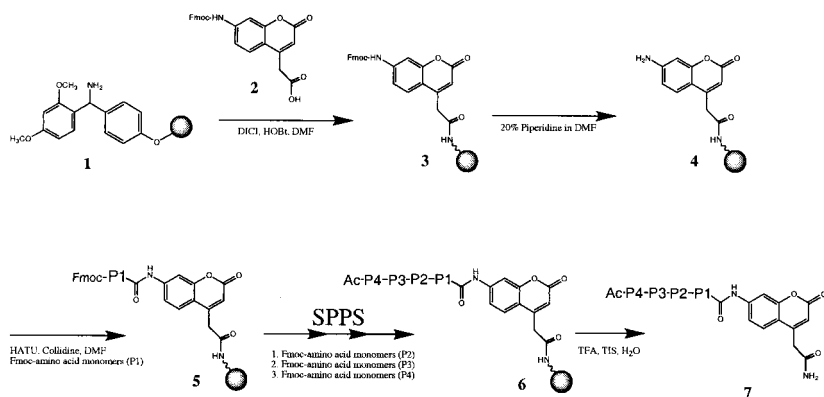
times with 2 ml; tetrahydrofuran, three times with 2 ml; MeOH, three times with 2 ml), and dried over P<sub>2</sub>O<sub>5</sub>. The substitution level of each resin<sup>||</sup> was determined by Fmoc analysis (10).

**P1-Diverse Library Synthesis.** Individual P1-substituted Fmoc-amino acid ACC-resin (≈25 mg, 0.013 mmol) was added to wells of a MultiChem 96-well reaction apparatus. The resin-containing wells were solvated with DMF (0.5 ml). After filtration, a 20% piperidine in DMF solution (0.5 ml) was added, followed by agitation for 30 min. The wells of the reaction block were filtered and washed with DMF (three times with 0.5 ml). To introduce the randomized P2 position, an isokinetic mixture (16) of Fmoc-amino acids [4.8 mmol, 10 eq per well; Fmoc-Ala-OH, 3.4; Fmoc-Arg(Pbf)-OH, 6.5; Fmoc-Asn(Trt)-OH, 5.3; Fmoc-Asp(*O*-*t*-Bu)-OH, 3.5; Fmoc-Glu(*O*-*t*-Bu)-OH, 3.6; Fmoc-Gln(Trt)-OH, 5.3; Fmoc-Gly-OH, 2.9; Fmoc-His(Trt)-OH, 3.5; Fmoc-Ile-OH, 17.4; Fmoc-Leu-OH, 4.9; Fmoc-Lys(Boc)-OH, 6.2; Fmoc-Nle-OH, 3.8; Fmoc-Phe-OH, 2.5; Fmoc-Pro-OH, 4.3; Fmoc-Ser(*O*-*t*-Bu)-OH, 2.8; Fmoc-Thr(*O*-*t*-Bu)-OH, 4.8; Fmoc-Trp(Boc)-OH, 3.8; Fmoc-Tyr(*O*-*t*-Bu)-OH, 4.1; Fmoc-Val-OH, 11.3] was preactivated with DICl (390 μl, 2.5 mmol), and HOBt (340 mg, 2.5 mmol) in DMF (10 ml). The solution (0.5 ml) was added to each of the wells. The reaction block was agitated for 3 h, filtered, and washed with DMF (three times with 0.5 ml). The randomized P3 and P4 positions were incorporated in the same manner. The Fmoc of the P4 amino acid was removed and the resin was washed with DMF (three times with 0.5 ml) and treated with 0.5 ml of a capping solution of AcOH (150 μl, 2.5 mmol), HOBt (340 mg, 2.5 mmol), and DICl (390 μl, 2.5 mmol) in DMF (10 ml). After 4 h of agitation, the resin was washed with DMF (three times with 0.5 ml) and CH<sub>2</sub>Cl<sub>2</sub> (three times with 0.5 ml), and treated with a solution of 95:2.5:2.5 TFA/TIS/H<sub>2</sub>O. After incubation for 1 h the reaction block was opened and placed on a 96-deep-well titer plate and the wells were washed with additional cleavage solution (twice with 0.5 ml). The collection plate was concentrated, and the material in the substrate-containing wells was diluted with EtOH (0.5 ml) and concentrated twice. The contents of the individual wells were lyophilized from CH<sub>3</sub>CN/H<sub>2</sub>O mixtures. The total amount of substrate in each well was conservatively estimated to be 0.0063 mmol (50%) on the basis of yields of single substrates.

**P1-Fixed Library Synthesis.** Multigram quantities of P1-substituted ACC-resin could be synthesized by the methods described. Three libraries with the P1 position fixed as Lys, Arg, or Leu were prepared. Fmoc-amino acid-substituted ACC resin (≈25 mg, 0.013 mmol, of Lys, Arg, or Leu) was placed in 57 wells of a 96-well reaction block: three sublibraries denoted by the second fixed position (P4, P3, P2) of 19 amino acids (cysteine was omitted and norleucine was substituted for methionine). Synthesis, capping, and cleavage of the substrates were identical to those described in the previous section, with the exception that for P2, P3, and P4 sublibraries, individual amino acids (5 eq of Fmoc-amino acid monomer, 5 eq of DICl, and 5 eq of HOBt in DMF), rather than isokinetic mixtures, were incorporated in the spatially addressed P2, P3, or P4 positions.

**Synthesis of Single Substrates.** Single substrates for kinetic analysis were prepared by the methods described above. The unpurified

<sup>||</sup>Fmoc-amino acid, coupling efficiency (double coupling): Fmoc-Ala-OH, >95%; Fmoc-Arg(Pbf)-OH, 73% (80%); Fmoc-Asn(Trt)-OH, >95%; Fmoc-Asp(*O*-*t*-Bu)-OH, >95%; Fmoc-Glu(*O*-*t*-Bu)-OH, 77% (>95%); Fmoc-Gln(Trt)-OH, 73% (>95%); Fmoc-Gly-OH, >95%; Fmoc-His(Trt)-OH, 72% (>95%); Fmoc-Ile-OH, 57% (60%); Fmoc-Leu-OH, 86% (>95%); Fmoc-Lys(Boc)-OH, 75% (>95%); Fmoc-Met-OH, 94% (>95%); Fmoc-Nle-OH, 83% (>95%); Fmoc-Phe-OH, >95%; Fmoc-Pro-OH, 63% (70%); Fmoc-Ser(*O*-*t*-Bu)-OH, 85% (>95%); Fmoc-Thr(*O*-*t*-Bu)-OH, 73% (84%); Fmoc-Trp(Boc)-OH, 77% (>95%); Fmoc-Tyr(*O*-*t*-Bu)-OH, 86% (>95%); Fmoc-Val-OH, 69% (80%). Pbf, 2,2,4,6,7-pentamethylidihydrobenzofuran-5-sulfonyl; Trt, trityl; Boc, *tert*-butoxycarbonyl; Nle, norleucine.



**Fig. 1.** Synthesis of 7-amino-4-carbamoylmethylcoumarin (ACC) substrates. SPPS, solid-phase peptide synthesis using standard Fmoc protocols.

products were subjected to reversed-phase preparatory HPLC followed by lyophilization.

**Fluorescence Properties of ACC.** The fluorescence of free ACC and peptidyl-derivatized ACC was detected on a Spex fluorimeter thermostated to 25°C. Excitation wavelengths of 300–410 nm, 5-nm intervals, were used with emission wavelengths of 410–500 nm, 5-nm intervals, to determine optimal excitation and emission parameters.

**Enzymatic Assay of Library.** The concentration of proteolytic enzymes was determined by absorbance measured at 280 nm (17). The proportion of catalytically active thrombin, plasmin, trypsin, uPA, tPA, and chymotrypsin was quantitated by active-site titration with 4-methylumbelliferyl *p*-guanidinobenzoate (MUGB) or methylumbelliferyl *p*-trimethylammoniumcinnamate chloride (MUTMAC) (18).

Substrates from the positional scanning–synthetic combinatorial libraries (PS-SCLs) were dissolved in DMSO. Approximately  $1.0 \times 10^{-9}$  mol of each P1-Lys, P1-Arg, or P1-Leu sublibrary (361 compounds) was added to 57 wells of a 96-well Microfluor plate (Dyner Technologies, Chantilly, VA) for a final concentration of 0.1  $\mu$ M. Approximately  $1.0 \times 10^{-10}$  mol of each P1-diverse sublibrary (6,859 compounds) was added to 20 wells of a 96-well plate for a final concentration of 0.01  $\mu$ M in each compound. Hydrolysis reactions were initiated by the addition of enzyme (0.02–100 nM) and monitored fluorimetrically with a Perkin–Elmer LS50B luminescence spectrometer, with excitation at 380 nm and emission at 450 nm or 460 nm. Assays of the serine proteases were performed at 25°C in a buffer containing 50 mM Tris at pH 8.0, 100 mM NaCl, 0–5 mM CaCl<sub>2</sub>, 0.01% Tween-20, and 1% DMSO (from substrates). Assay of the cysteine proteases, papain and cruzain, was performed at 25°C in a buffer containing 100 mM sodium acetate at pH 5.5, 100 mM NaCl, 5 mM DTT, 1 mM EDTA, 0.01% Brij-35, and 1% DMSO (from substrates).

**Single Substrate Kinetic Assays.** Thrombin concentration ranged from 5 to 20 nM. The final concentration of substrate ranged from 0.005 to 2 mM; the concentration of DMSO in the assay was less than 5%. Hydrolysis of AMC and ACC substrates was monitored fluorimetrically with an excitation wavelength of 380 nm and an emission wavelength of 460 nm on a Fluoromax-2 spectrofluorimeter. Cruzain (10 nM) was incubated with 600  $\mu$ M Ac-Leu-Thr-Phe-Lys-ACC substrate. Aliquots were removed at various time points and applied to a C<sub>18</sub> reverse-phase HPLC column with a 10–40% gradient of 95:4.9:0.1 CH<sub>3</sub>CN/H<sub>2</sub>O/TFA. Matrix-assisted laser desorption ionization (PE Biosystems Voyager) mass spectrometry data were collected on the HPLC fractions.

## Results

**Synthesis of ACC Substrates.** The fluorogenic substrates were prepared by using a bifunctional fluorophore that incorporates a site for peptide synthesis and a second site for attachment to a solid support (Fig. 1). *N*-Fmoc-protected bifunctional coumarin **2** is attached to acid-labile Rink linker **1** by using standard coupling conditions (Fig. 1). The Fmoc-protecting group of the coumarin **3** is removed by brief treatment with 20% piperidine in DMF to provide support-bound coumarin **4**. Because of the very poor nucleophilicity of the coumarin amine, modified coupling conditions of Carpino *et al.* (19) are used to provide P1-substituted ACC-resin **5** in good yields (57–>95%)<sup>||</sup>. For amino acids that typically provide lower coupling yields, such as Pro (63%), a second coupling provides increased substitution levels (70%). Remaining free aniline is efficiently capped by using the nitrotriazole active ester of AcOH prepared *in situ*. The substitution levels of the P1-ACC resins can be accurately assessed by the spectrophotometric quantitation of the fulvene-piperidine adduct resulting from piperidine deprotection of the Fmoc protecting group (10). Standard Fmoc-based synthesis methods provide support-bound ACC substrates **6**, and cleavage from the support by treatment with acid then produces the fluorogenic peptide substrates **7**.

**Fluorescence Properties of ACC.** The excitation and emission maxima of the amino-conjugated ACC substrates are 325 nm and 400 nm, respectively (Table 1). Cleavage of the substrate by a protease to release the free ACC results in a shift of the excitation and emission maxima to 350 nm and 450 nm, respectively (Table 1). The ACC fluorophore has an approximately 2.8-fold higher fluorescence yield than AMC at the excitation and emission wavelengths of 380 nm and 460 nm (Table 1). The enhanced fluorescence of the ACC group allows for the more sensitive detection of proteolytic activity.

**Proteolytic Comparison of ACC and AMC.** To evaluate ACC as a proteolytic leaving group, matched tetrapeptide substrates were made that differed only in the leaving group, ACC or the tradi-

**Table 1.** Fluorescence properties of AMC and ACC and their substrate conjugates

Compound	$\lambda_{\text{max}}$ , nm		RFU/nM*	
	ex	em	$\lambda_{\text{em}} = 450$ nm	$\lambda_{\text{em}} = 460$ nm
ACC	350	450	5,750	4,390
7-Nle-Thr-Pro-Lys-ACC	325	400	6.4	4.6
AMC	340	440	2,600	1,550
7-Nle-Thr-Pro-Lys-AMC	330	390	3.3	2.2

\*RFU, relative fluorescence units; for both columns,  $\lambda_{\text{ex}} = 380$  nm.

**Table 2. Steady-state kinetic constants for thrombin acting on AMC and ACC substrates**

Substrate	$k_{cat}$ , $s^{-1}$	$K_m$ , $\mu M$	$k_{cat}/K_m$ , $\mu M^{-1}s^{-1}$
Ac-Nle-Thr-Pro-Lys-AMC	$31.0 \pm 0.9$	$115 \pm 10$	$0.26 \pm 0.03$
Ac-Nle-Thr-Pro-Lys-ACC	$33.7 \pm 2.7$	$125 \pm 13$	$0.28 \pm 0.05$
Ac-Leu-Gly-Pro-Lys-AMC	$2.3 \pm 0.2$	$160 \pm 25$	$0.015 \pm 0.002$
Ac-Leu-Gly-Pro-Lys-ACC	$3.2 \pm 0.4$	$195 \pm 30$	$0.018 \pm 0.003$

tionally used AMC. The two thrombin-susceptible sequences with ACC or AMC, P4-Nle-P3-Thr-P2-Pro-P1-Lys and P4-Leu-P3-Gly-P2-Pro-P1-Lys, showed comparable kinetic constants as substrates of thrombin (Table 2). A significant advantage of ACC substrates over AMC substrates is the ease of synthesizing ACC substrates compared with AMC substrates. By employing the synthesis methods described, any amino acid ACC substrate can be prepared rapidly with Fmoc-based synthesis protocols.

#### Profiling Proteases with a P1-Diverse Library of 137,180 Substrates.

To test the possibility of attaching all amino acids to the P1 site in the substrate a P1-diverse tetrapeptide library was created. The P1-diverse library consists of 20 wells in which only the P1 position is systematically held constant as each one of the proteinogenic amino acids, excluding cysteine and including norleucine. The P2, P3, and P4 positions consist of an equimolar mixture of the 19 amino acids, for a total of 6,859 substrate sequences per well. Several serine and cysteine proteases were profiled to test the applicability of this library for the identification of the optimal P1 amino acid. Chymotrypsin showed the expected specificity for large hydrophobic amino acids (Fig. 2A). Trypsin and thrombin showed preference for P1 basic amino acids (Arg > Lys) (Fig. 2B and C). Plasmin also showed a

preference for basic amino acids (Lys > Arg) (Fig. 2D). Granzyme B, the only known mammalian serine protease to have P1-Asp specificity, showed a distinct preference for aspartic acid over all other amino acids, including the other acidic amino acid, Glu (Fig. 2E). The P1 profile for human neutrophil elastase has the canonical preference for alanine and valine (Fig. 2F). The cysteine proteases, papain (Fig. 2G) and cruzain (Fig. 2H) showed the broad P1-substrate specificity that is known for these enzymes, although there is a modest preference for arginine.

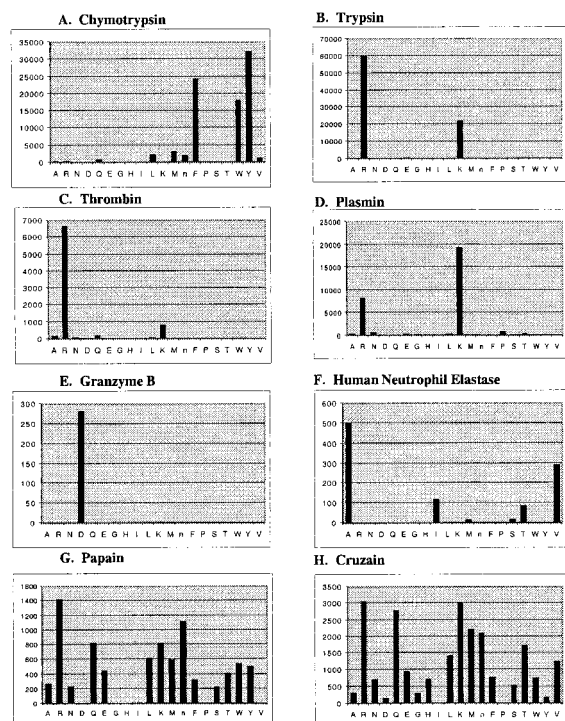
**Profiling of Serine Proteases with P1-Fixed Positional Libraries.** The extended P4–P2 substrate specificity of several serine proteases was profiled with tetrapeptide libraries in which the P1 position was held constant. Three sublibraries denoting the second fixed position (P4, P3, P2) and consisting of 19 wells addressing a fixed amino acid (Cys was omitted and Nle was substituted for Met) were screened (361 compounds per well and 6,859 compounds per library). Because of the enhanced fluorescence properties of the ACC fluorophore, the concentration of each substrate could be reduced to  $0.1 \mu M$ , versus  $0.25 \mu M$  for the AMC substrates (20).

Plasmin, a protease involved in fibrinolysis, has a P1 preference for lysine. Recently, we have shown plasmin to have a distinct preference for aromatic amino acids at the P2 position and lysine at P4 (20). As is consistent with those data, the substrate specificity profile of plasmin in the ACC P1-fixed lysine library is for P4-lysine, broad P3-specificity, and P2-aromatic amino acids (Fig. 3A). This sequence correlates with the sequences found in plasmin's physiological substrates such as vitronectin (21) and factor X (22), which are cleaved by plasmin at Lys-Gly-Tyr-Arg and Ile-Thr-Phe-Arg, respectively.

Thrombin prefers cleavage after P1 arginine to cleavage after P1 lysine. However, the specificity preference of thrombin, when profiled with both the P1-Arg and P1-Lys libraries, shows little difference in the extended subsites (Fig. 3B and ref. 20). Thrombin has a preference for aliphatic amino acids at the P4 position, little preference at P3, and strict preference for proline at the P2 position. There is a strong correlation of thrombin's preferred substrate sequence determined by the synthetic libraries to the sequences found in its physiological substrates (20).

Two enzymes that have been extensively characterized for their extended specificity are tPA (3, 23) and uPA (24, 25). Both tPA and uPA are responsible for converting plasminogen into active plasmin, and both show high specificity for cleavage after P1 Arg. We observe that both enzymes also show similar preference for small amino acids at P2 (Gly/Ala/Ser) and no significant preference at P4, except for the low activity of acidic amino acids (Fig. 3C and D). In contrast, their P3 preferences are quite disparate, with tPA showing preference for aromatic amino acids (Phe and Tyr) and uPA, for small polar amino acids (Thr and Ser). This difference in P3 specificity was also identified through substrate–phage display and noted by Ke *et al.* (24) to be a major distinction between the two plasminogen activators.

Factor Xa is an enzyme that fulfills the critical physiological functions of activating prothrombin and factor VII in the blood coagulation cascade (26). Through profiling with the P1-Arg library, we find factor Xa to show a minor preference for P4 aliphatic amino acids, broad substrate specificity in P3, with the absence of P3-proline activity, and a P2 preference for glycine (Fig. 3E). This quantitative information agrees with the qualitative sequences that are efficiently hydrolyzed by factor Xa in a substrate–phage system (2) as well as kinetic studies on tripeptide *para*-nitroanilide (27) and AMC substrates (27, 28). Furthermore, the factor Xa P4-P1 cleavage sequence determined here is found in physiologically relevant substrates: the cleavage sequences in prothrombin are Ile-Glu-Gly-Arg and Ile-Asp-Gly-Arg; cleavage sequence in factor VII is Pro-Gln-Gly-Arg; and the cleavage sequence in the autolysis loop of factor Xa is Glu-Lys-Gly-Arg (29).



**Fig. 2.** ACC P1-diverse library. The library consists of 20 wells with 6,859 compounds per well (137,180 compounds total). P1 profiles of several serine and cysteine proteases are shown. The y axis is the picomolar fluorophore produced per second. The x axis provides the spatial address of the amino acid as represented by the one-letter code (with "n" representing norleucine).

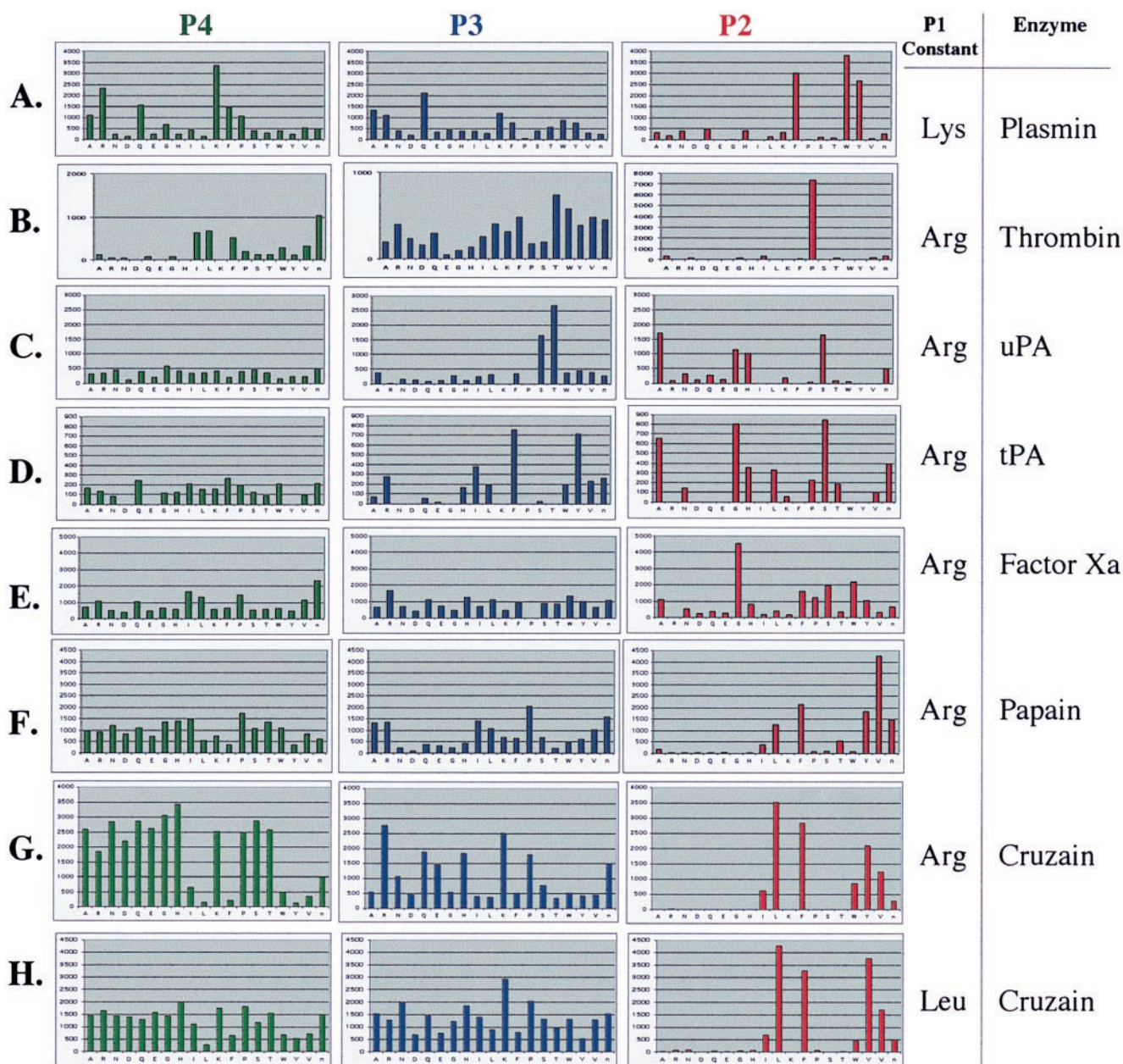


Fig. 3. Profiles of serine and cysteine proteases acting on a P1-fixed ACC P5-SCL. The y axis is the picomolar fluorophore produced per second. The x axis provides the spatial address of the amino acid as represented by the one-letter code (with "n" representing norleucine).

**Profiling of Cysteine Proteases with P1-Fixed Positional Libraries.** The positional substrate libraries with the ACC fluorogenic leaving group are also conducive for defining cysteine protease specificity. The P4–P2 extended substrate specificities for papain and cruzain were defined by using the ACC P1-fixed arginine or leucine library. Cysteine proteases of the papain-like class have been shown to have primary substrate specificity at the P2 position (30) rather than the P1 position as is seen in the chymotrypsin-like class of serine proteases. The P2 position usually shows a preference for hydrophobic amino acids. Indeed, we observe papain to have a preference for P2 Val > Phe > Tyr > Nle (Fig. 3F) and cruzain to have a P2 preference for Leu > Tyr > Phe > Val (Fig. 3G). While the P3 specificity is rather broad, papain does show a preference for Pro, whereas cruzain has a preference for the basic amino acids, arginine and lysine. The P4 position is very broad for both enzymes, but interesting observations arise from testing all possible substrates. There is a lack of activity for large aliphatic and aromatic amino

acids, the exact amino acids that are preferred in the P2 library. This absence is also seen in a P4 library in which the P1 position is held constant as leucine (Fig. 3H). One possible reason for the observations in the P4 library is that the tetrapeptide substrates are out of register. Cleavage is not occurring at the P1-amidocarbonylmethylcoumarin bond, but rather, at the P3–P2 amide bond because the large hydrophobic P4 amino acid binds to the S2 pocket of the enzyme. Incubation of the single substrate Ac-Leu-Thr-Phe-Lys-ACC with cruzain and analysis of the cleavage products confirmed this observation. Product fragments corresponding to cleavage between Thr-Phe were observed (data not shown).

### Discussion

A fluorogenic leaving group has been designed that can be attached to acid-labile Rink linker to provide ACC-resin 3. After Fmoc removal, all 20 amino acids can be coupled to the aniline efficiently. For amino acids that are traditionally difficult to

couple (Ile, Val, etc), free, unreacted aniline may remain on the support and complicate subsequent synthesis and assay operations. A specialized capping step employing the 3-nitrotriazole active ester of acetic acid in DMF efficiently acylates the remaining aniline. The resulting acetic acid-capped coumarin that may be present in unpurified substrate solutions was shown not to be a substrate for all proteases tested (data not shown). P1-substituted resins that are provided by these methods can be used to prepare any ACC-fluorogenic substrate and consequently any library configuration.

To assess the performance of the ACC leaving group, an ACC-P1-fixed lysine PS-SCL was prepared for comparison to an AMC-P1-fixed lysine PS-SCL prepared by previously developed methods (20). Three libraries, denoting the second fixed position (P2, P3, P4) and consisting of 19 wells addressing a fixed amino acid (cysteine was omitted and norleucine was substituted for methionine) were prepared (361 compounds per well and 6,869 compounds per library). Plasmin and thrombin showed comparable kinetic profiles in both libraries, demonstrating the equivalency of ACC to AMC as a leaving group for the determination of substrate specificity. The major difference between the libraries was the amount of enzyme and substrate required for sufficient fluorescence signal. The substrate concentration for the ACC library was reduced to 0.1  $\mu\text{M}$  per substrate per well, compared with 0.25  $\mu\text{M}$  for the AMC library. The enzyme concentration was also reduced. The increased fluorescence sensitivity of the ACC group will be very important for assaying proteases that are available only in limited amounts. For additional validation, specific substrates that differed only in fluorogenic leaving groups, ACC or AMC, were synthesized. Steady-state kinetic constants of thrombin were measured for these substrates and shown to be similar for the ACC- and AMC-containing substrates (Table 2).

We tested the ability to introduce an additional element of diversity in the PS-SCLs through the preparation of a tetrapeptide library consisting of 20 wells (cysteine was omitted and norleucine was included) addressing a fixed P1 amino acid. The P4-P3-P2 positions in this library consisted of an equimolar mixture of 19 amino acids (cysteine was omitted and norleucine was substituted for methionine) for a total of 6,858 substrates per well and 137,180 substrates per library. To avoid insolubility of the substrates as well as to maintain  $k_{\text{cat}}/K_{\text{m}}$  conditions, the concentration for each individual substrate was decreased to approximately 0.01  $\mu\text{M}$ . The effectiveness of the library was validated through the profiling of various serine and cysteine proteases. The library was able to distinguish proteases that had specificity for P1 acidic amino acids (granzyme B), P1 large hydrophobic amino acids (chymotrypsin), P1 small hydrophobic amino acids (human neutrophil elastase), P1 basic amino acids (trypsin, thrombin, plasmin), and P1 multiple amino acids (papain and cruzain) (Fig. 2).

The extended substrate specificities of several serine proteases involved in blood coagulation were determined by libraries in

which the P1 position was held constant as either Lys or Arg, depending on the preferred P1 specificity of the protease. Thrombin, plasmin, uPA, tPA, and factor Xa (Fig. 3 A–E) displayed profiles that are consistent with what is known about their specificity, thus demonstrating that the libraries are functioning as designed. The extended substrate specificity of the cysteine proteases, papain and cruzain, were also profiled with P1-positioned libraries and indicate a strong preference for hydrophobic amino acids in the P2 position.

The presented PS-SCL strategy allows for the rapid and facile determination of proteolytic substrate specificity. However, the method will not be applicable to all proteases. Requirement of some proteases, such as metalloproteinases and aspartyl proteases, for interactions C-terminal to the cleavage site may limit the usefulness of these substrates. However, we have evidence that some metalloproteinases and aspartyl proteases have the ability to hydrolyze ACC substrates (data not shown), and the bifunctionality of the ACC fluorophore may allow for the incorporation of prime side substituents. We have also demonstrated, with the papain-fold protease cruzain, that correct register of the substrate is important. Alternative library formats may solve the problems of substrate register for particular proteases. For example, fixing the P2 position as a large hydrophobic amino acid may circumvent preferential internal cleavage by papain-fold proteases and lead to proper register of the substrate. These limitations are present for any method of substrate specificity determination and must be considered in such experiments.

In conclusion, we have developed a fluorogenic leaving group that allows for the preparation of general fluorogenic substrate libraries. Due to the bifunctional nature of the fluorophore, libraries can be rapidly prepared by standard Fmoc-based synthesis strategies. Multiple serine and cysteine proteases with a broad spectrum of specificities have been profiled through libraries generated in the described manner. The substrate specificity profiles gathered are in agreement with known specificities and provide a “fingerprint” of the protease. Unlike other currently used combinatorial methods to determine substrate specificity, PS-SCLs are screens of substrate space, not selections for the optimal substrates. The ability to screen all possible substrates yields not only efficient substrate sequences but also suboptimal substrate sequences. Thus, this pharmacophoric information allows for the design of not only sensitive substrates and inhibitors but, equally important, selective substrates and inhibitors. Likewise, this information will be useful for dissecting the physiological pathways in which proteases operate, through identification of downstream macromolecular substrates.

We thank Toshihiko Takeuchi, Sherin Halfon, Robert Maeda, and Keith Burdick for insightful discussions and careful reading of the manuscript. This work was supported in part by National Institutes of Health Grants CA72006 and AI35707 (to C.S.C.) and GM54051 (to J.A.E.) and a National Institutes of Health Biotechnology Training Grant Fellowship (to J.L.H.).

- Barrett, A. J., Rawlings, N. D. & Woessner, J. F. (1998) *Handbook of Proteolytic Enzymes* (Academic, London).
- Matthews, D. J. & Wells, J. A. (1993) *Science* **260**, 1113–1117.
- Ding, L., Coombs, G. S., Strandberg, L., Navre, M., Corey, D. R. & Madison, E. L. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7627–7631.
- Lam, K. S. & Lebl, M. (1998) *Methods Mol. Biol.* **87**, 1–6.
- Zimmerman, M., Ashe, B., Yurewicz, E. & Patel, G. (1977) *Anal. Biochem.* **78**, 47–51.
- Lee, D., Adams, J. L., Brandt, M., DeWolf, W. E., Jr., Keller, P. M. & Levy, M. A. (1999) *Bioorg. Med. Chem. Lett.* **9**, 1667–1672.
- Rano, T. A., Timkey, T., Peterson, E. P., Rotonda, J., Nicholson, D. W., Becker, J. W., Chapman, K. T. & Thornberry, N. A. (1997) *Chem. Biol.* **4**, 149–155.
- Schechter, I. & Berger, A. (1968) *Biochem. Biophys. Res. Commun.* **27**, 157–162.
- Backes, B. J. & Ellman, J. A. (1999) *J. Org. Chem.* **64**, 2322–2330.
- Bunin, B. A. (1998) *The Combinatorial Index* (Academic, San Diego).
- Harris, J. L., Peterson, E. P., Hudig, D., Thornberry, N. A. & Craik, C. S. (1998) *J. Biol. Chem.* **273**, 27364–73.
- Eakin, A. E., Mills, A. A., Harth, G., McKerrow, J. H. & Craik, C. S. (1992) *J. Biol. Chem.* **267**, 7411–7420.
- Halfon, S. & Craik, C. S. (1996) *J. Am. Chem. Soc.* **118**, 1227–1228.
- Kanaoka, Y., Kobayashi, A., Sato, E., Nakayama, H., Ueno, T., Muno, D. & Sekine, T. (1984) *Chem. Pharm. Bull.* **32**, 3926–3933.
- Besson, T., Joseph, B., Moreau, P., Viaud, M. C., Coudert, G. & Guillaumet, G. (1992) *Heterocycles* **34**, 273–291.
- Ostresh, J. M., Winkle, J. H., Hamashin, V. T. & Houghten, R. A. (1994) *Biopolymers* **34**, 1681–1689.
- Gill, S. C. & von Hippel, P. H. (1989) *Anal. Biochem.* **182**, 319–326.
- Jameson, G. W., Roberts, D. V., Adams, R. W., Kyle, W. S. A. & Elmore, D. T. (1973) *Biochem. J.* **131**, 107–117.
- Carpino, L. A., Ionescu, D. & Elfaham, A. (1996) *J. Org. Chem.* **61**, 2460–2465.
- Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S. & Ellman, J. A. (2000) *Nat. Biotechnol.* **18**, 187–193.
- Chain, D., Kreizman, T., Shapira, H. & Shaltiel, S. (1991) *FEBS Lett.* **285**, 251–256.
- Prydzial, E. L., Lavigne, N., Dupuis, N. & Kessler, G. E. (1999) *J. Biol. Chem.* **274**, 8500–8505.
- Coombs, G. S., Dang, A. T., Madison, E. L. & Corey, D. R. (1996) *J. Biol. Chem.* **271**, 4461–4467.
- Ke, S. H., Coombs, G. S., Tachias, K., Navre, M., Corey, D. R. & Madison, E. L. (1997) *J. Biol. Chem.* **272**, 16603–16609.
- Ke, S. H., Coombs, G. S., Tachias, K., Corey, D. R. & Madison, E. L. (1997) *J. Biol. Chem.* **272**, 20456–20462.
- Davie, E. W., Fujikawa, K. & Kisiel, W. (1991) *Biochemistry* **30**, 10363–10370.
- Cho, K., Tanaka, T., Cook, R. R., Kisiel, W., Fujikawa, K., Kurachi, K. & Powers, J. C. (1984) *Biochemistry* **23**, 644–650.
- Lottenberg, R., Christensen, U., Jackson, C. M. & Coleman, P. L. (1981) *Methods Enzymol.* **80**, 341–361.
- Brandstetter, H., Keuhne, A., Bode, W., Huber, R., von der Saal, W., Wirthensohn, K. & Engh, R. A. (1996) *J. Biol. Chem.* **271**, 29988–29992.
- Rawlings, N. D. & Barrett, A. J. (1994) *Methods Enzymol.* **244**, 461–486.