Analytical Biochemistry **288,** 156–167 (2001)
doi:10.1006/abio.2000.4886, available online at http://www.idealibrary.com on IDEAL®

# Specificity Assay of Serine Proteinases by Reverse-Phase High-Performance Liquid Chromatography Analysis of Competing Oligopeptide Substrate Library

József Antal,*[,1] Gábor Pál,† Bence Asbóth,* Zsuzsanna Buzás,* András Patthy,* and László Gráf*[,†]
**Analysis-Synthesis Center, Agricultural Biotechnology Center, Szent-Györgyi A. út 4, H-2100 Gödöllő, Hungary; and †Department of Biochemistry, Eötvös University, Puskin utca 3, H-1088 Budapest, Hungary*

Received June 20, 2000

**In this paper we present an HPLC method developed for quick activity and specificity analysis of serine proteinases. The method applies a carefully designed peptide library in which the individual components differ only at the potential cleavage site for enzymes. The library has seven members representing seven different cleavage sites and it offers substrates for both trypsin and chymotrypsin-like enzymes. The individual peptide substrates compete for the proteinase during the enzymatic reaction. The reaction is monitored by RP-HPLC separation of the components. We describe the systematic design of the competitive peptide substrate library and the test of the system with eight different serine proteinases. The specificity profiles of the investigated enzymes as determined by the new method were essentially identical to the ones reported in the literature, verifying the ability of the system to characterize substrate specificity. The tests also demonstrated that the system could detect even subtle specificity differences of two isoforms of an enzyme. In addition to recording qualitative specificity profiles, data provided by the system can be analyzed quantitatively, yielding specificity constant values. This method can be a useful tool for quick analysis of uncharacterized gene products as well as new forms of enzymes generated by protein engineering.** © 2001 Academic Press

***Key Words:*** **serine proteinases; substrate library; competing oligopeptides; RP-HPLC.**

Recent genome projects have been providing an enormous number of DNA sequences. One of the greatest challenges will be to sort out the potential genes and predict function for the encoded proteins by sequence analysis of the open-reading frames. However, even if these computational approaches fulfill the expectations, isolation and *in vitro* analysis of the gene products still remain indispensable. This is one of the major subjects of functional genomics that demands fast development of new high-throughput and preferably inexpensive test systems for various enzymes and other proteins. Here, we present an HPLC method that can be a prototype for such an assay system for quick activity and specificity analysis of serine proteinases. Such a method can be beneficial not only for the investigation of unknown gene products but also for protein engineering purposes. In particular, when alteration of enzyme specificity is the goal, large numbers of mutants are frequently generated. Combination of structural and functional analysis of these forms may reveal how subtle changes can successively lead to significant changes in specificity.

To better understand the structural basis for the characteristically different substrate specificity of trypsin and chymotrypsin, interconversion of these two enzymes was attempted. The only striking difference between their substrate binding pockets could be assigned to position 189 where there is an aspartate residue in all the trypsin sequences versus a serine residue in most chymotrypsins. This difference was for long accepted to be responsible for the fact that trypsin selectively cleaves after Arg and Lys, while chymotrypsin cleaves after Phe, Tyr, Trp, and Leu residues. However, a single Asp189Ser mutation did not transform trypsin to a chymotrypsin-like enzyme. For successful conversion complete loops had to be replaced, segments that are not involved in direct interaction with the substrate (1–4).

[1] To whom correspondence should be addressed. Fax: 36-28-430-338. E-mail: antal@abc.hu.



0003-2697/01 $35.00

This experimental approach generates an array of mutant enzymes that should be analyzed concerning their specificity changes. Enzyme specificity is usually defined by the ratios of individual $k_{cat}/K_M$ values determined in separate measurements with only a single substrate species present at a time. Instead, it should rather "be defined in terms of how well the enzyme can discriminate between substrates *present in the same reaction mixture*" (8). In addition to providing more exact information about substrate specificity, a competitive approach could also shorten the time of the assays. However, development of a competitive system on the basis of single traditional chromogenic or fluorogenic assays is a challenge. In these assays the cleavage site is an ester or amide bond formed between the carboxyl group of an amino acid residue and the hydroxyl or amine group of a chromogenic or fluorogenic leaving group. Application of these substrates presents two problems. One is that the detection of the different components in a competitive system requires different leaving groups that provide signals at distinct wavelengths so that parallel cleavages of the individual substrates can be followed simultaneously. This demand greatly restricts the variety of the different substrates that can be included in the library. The other problem is that the artificial leaving groups can interact with the enzyme in a way that might not mimic the normal peptide substrate interaction. The natural serine proteinase substrates make extensive contacts with the enzyme. The residue, after which the cleavage occurs ($P_1$) occupies the substrate binding pocket ($S_1$) of the enzyme. Substrate residues surrounding the cleavage site in both N-terminal ($P_n$–$P_2$) and C-terminal ($P_1'$–$P_m'$) directions also interact with appropriate subsites of the enzyme ($S_n$–$S_m'$) (14) (see Fig. 1). The short substrates built of only a few $P_n$–$P_1$ amino acid residue(s) and a chromogenic leaving group lack the native P′–S′ interactions. Moreover, different leaving groups can be engaged in different interactions with the enzyme. Thus, it is rather doubtful whether these compounds are fully adequate for the investigation of the $P_1$ specificity of serine proteinases.

The use of molecules better imitating the natural substrates, such as oligopeptides or proteins, would eliminate this problem, providing ample cleavage sites with different chemical character. However, hydrolysis of these compounds is more complicated to follow. In addition to the relatively low sensitivity, assays that intend to use large peptides or whole proteins are often unreliable because of the structural preferences and the possible additional specificity determinant regions at the polypeptide chain.

Methods that use mixtures of short synthetic peptides, such as when studying the specificity of collagenase (9) or the $P_1$–$P_1'$ site preference of several proteases (10), apply sophisticated detection systems. Usually, these approaches combine HPLC with mass spectrometry (MS) or automated peptide sequencing and the product analysis involves several steps, some of them slow and expensive.

The recently developed internally quenched fluorogenic (IQF)[2] peptide substrates that have reporter groups further away from the cleavage site might not suffer from artificial P′–S′ interactions and could represent a powerful competitive system when organized into libraries. However, the special substrates this method requires are not commercially available and the ultimate requirement for a sophisticated detection system renders this approach costly (5–7). Similarly to the classical chromogenic substrates the number of the different leaving groups that can be included in the library is limited. Perhaps due to these restrictions no competitive IQF assay system has been reported so far.

In the present paper we describe a systematic approach heading toward the design of a peptide library that allows routine assay of serine proteinase specificity.

## MATERIALS AND METHODS

### *Chemicals*

All chemicals used were of analytical or sequencing grade as appropriate. HPLC-grade acetonitrile was obtained from Romil (Great Britain). *N*,*N*-Diisopropylethylamine, trifluoroacetic acid, and all chemicals for the peptide synthesis and sequencing were obtained from ABI–Perkin–Elmer (U.S.A.), except Fmoc-Arg-(Pmc)-OH, which was from Novabiochem (Switzerland). Bovine trypsin and chymotrypsin were from Sigma (Hungary), bovine plasmin, thrombin, and mouse endoproteinase Arg-C were from Boehringer-Mannheim (Germany), recombinant rat chymotrypsin (2) and a double mutant of rat trypsin (referred to as "Tyrpsin" in later parts of this paper because it showed increased selectivity against Tyr as measured with fluorogenic substrates (Pál *et al.*, manuscript in preparation)) were produced at the Department of Biochemistry, Eötvös University (Budapest, Hungary), and crayfish trypsin was isolated at ABC-ACE (Gödöllő, (Hungary).

### *Solid-Phase Synthesis of Oligopeptides*

The seven different His-Ala-Ala-Pro-Xxx-Ser-Ala-Asp-Ile-Gln-Ile-Asp-Ile oligopeptides with Lys, Arg, Tyr, Leu, Phe Trp, and Pro amino acid residues at the position signed Xxx were produced by ABI 431A automated peptide synthesizer with Fmoc chemistry.

[2] Abbreviation used: IQF, internally quenched fluorogenic.

**TABLE 1**

Substrate and Active Enzyme Concentrations in the Reaction Mixture

| Reaction components | Concentration [$\mu$M] |
|---|---|
| Substrate mixture components | 40 |
| Enzymes | |
| Rat chymotrypsin | 0.0150 |
| Bovine chymotrypsin | 0.0075 |
| Tyrpsin | 0.1500 |
| Bovine trypsin | 0.0012 |
| Crayfish trypsin | 0.0012 |
| Bovine plasmin | 0.0012 |
| Bovine thrombin | 0.0075 |
| Mouse endoproteinase Arg-C | 0.0075 |

*Purification of Oligopeptides*

Purification of oligopeptides was carried out by Waters 600 E preparative RP-HPLC system with Waters Delta Pak C18 300-Å 19-mm × 30-cm column.

*Amino Acid Analysis*

For quality control of the oligopeptides and identification of the individual substrate fragments separated on RP-HPLC, a Waters PICO-TAG amino acid analysis system was used with PITC derivatization

*Enzymatic Digestions*

All the reactions were carried out in 0.05 M Tris–HCl, pH 8.0, 18 mM $CaCl_2$ buffer at 25°C. Equimolar mixture of the components was made from 1 mg/ml solutions of oligopeptides dissolved in the reaction buffer. This mixture was digested with bovine and crayfish trypsins, bovine chymotrypsin A and rat chymotrypsin (that is similar to bovine chymotrypsin B), a mutant rat trypsin named Tyrpsin, bovine plasmin, bovine thrombin, and mouse endoproteinase Arg-C. The enzyme concentrations were optimized to ensure comparable enzyme reaction rates with the individual proteinases. The substrate and enzyme concentrations are shown in Table 1. The reaction volume was 700 $\mu$l. The digestion was terminated at different reaction times (0, 1, 4, 16, 32, and 64 min) by transferring 100-$\mu$l aliquots to the injector tubes containing 20 $\mu$l of 5 M acetic acid. The aliquots were analyzed on an automated HPLC system. In the case of bovine trypsin and chymotrypsin three independent digestions and measurements were carried out to check reproducibility of the assay.

*RP-HPLC Analysis*

A PE-ABI automated HPLC instrument with a 140 C Microgradient System, Series 200 Autosampler, 785 A Programmable Absorbance Detector, 900 Series Interface equipped with an Aquapore OD 300 reversed-phase column (2.1 × 220 mm) was used for the separations. Five-microliter aliquots of the terminated digestion mixture were loaded onto the column. Eluent A was water containing 0.1% (v/v) trifluoroacetic acid and eluent B was 80% acetonitrile and 20% water containing 0.1% (v/v) trifluoroacetic acid. Chromatographic conditions were set to acquire data only from retention range of the intact oligopeptides (see Results and Discussion, adjusting RP-HPLC conditions section). The column was equilibrated with 23% eluent B and the injection was followed by a 5-min hold of this composition. Then, a linear gradient from 23 to 35% of eluent B was applied in 26 min with 1 ml/min flow rate. This gradient profile ensured baseline separation of the oligopeptide components. Detection was carried out at 220 nm.

Chromatographic data obtained by RP-HPLC were collected and quantitated by the Perkin–Elmer Turbochrom 4 chromatography software.

## RESULTS AND DISCUSSION

*Substrate Design*

The major goal of this study was to construct a peptide substrate library that is applicable for routine chromatographic analysis of the $P_1$ specificity of various serine proteinases in a competing reaction system. Components of such a library should possess a common amino acid sequence that is varied at only a single position regarding the residue at the enzymatic cleavage site. The sequence of this special library must meet several theoretical and practical demands: (i) the common peptide backbone should be generally accepted by serine proteinases having either trypsin- or chymotrypsin-like specificity; (ii) it should not offer cleavage sites other than the one intended to be the $P_1$; (iii) all the initial components and the cleavage products should be easily separated from one another by a conventional RP-HPLC system; (iv) should be soluble under the conditions of the enzyme reaction; and finally, (v) the design should avoid sequences that can cause difficulties during the peptide synthesis and the subsequent purification steps.

According to the literature, only the $P_4$–$P_2'$ fragment of the substrate has significant effect in the midst of enzyme–substrate interaction (14). Thus, first a hexapeptide was designed corresponding to the region mentioned above. To cover both trypsin (Arg and Lys) and chymotrypsin-like (Tyr, Phe, Trp, and Leu) primary specificity, these six different amino acid residues had been chosen for the $P_1$ position. Since most serine proteinases are unable to cleave the peptide bond following proline, this residue was included to serve as an internal standard.

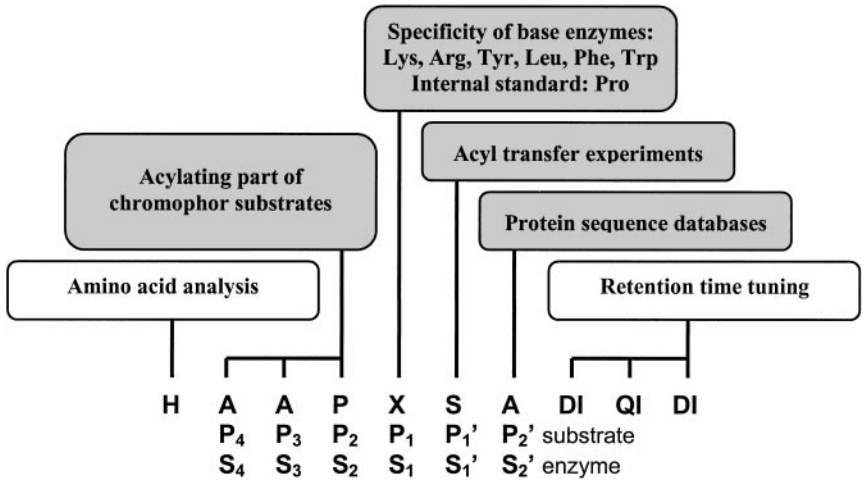


**FIG. 1.** Guide to substrate design. $P_4$–$P_2'$ are the substrate $S_4$–$S_2'$ are the enzyme subsites involved in enzyme–substrate interactions according to Schechter and Berger nomenclature. Lines from the text boxes point to the part of the sequence designed according to the respective information.

To ensure criteria i the following considerations were taken into account (Fig. 1). For the $P_4$–$P_2$ segment the Ala-Ala-Pro sequence was chosen. This motif is often used in chromogenic and fluorogenic substrates of trypsin and chymotrypsin (14). Incorporation of this sequence into the model peptide allows certain comparisons among the chromatographic and the spectroscopic results.

The acyl-transfer studies of Schellenberger and coworkers (11, 12) regarding the $P_1'$ site were considered. These results highlighted the importance of the $P_1'$ position and gave quantitative data about the differences in the contributions of the various amino acid residues to the enzyme–substrate interaction. Since our purpose was to construct a peptide library that, in addition to the $P_1$ site, does not differentiate between trypsin and chymotrypsin, we needed a $P_1'$ residue equally accepted by these two enzymes. After analysis of the available data, Ser residue seemed to be the most suitable candidate for the $P_1'$ position (11). To find an appropriate $P_2'$ residue, we compared sequences of substrate-like serine proteinase inhibitors. The result of the protein database search suggested a common preference of trypsin and chymotrypsin for hydrophobic amino acids at the $P_2'$ position. To prevent any possible steric hindrance caused by bulkier side chains, Ala residue was chosen. The choice of Ser-Ala motif as $P_1'$–$P_2'$ residues agrees with the results of a recent investigation that used a peptide library to find the optimal subsite occupancy of urokinase (15).

To satisfy criterion iii the retention times of the reactants and the products need to be significantly different. In RP-HPLC systems the relative retention times of short peptides can be calculated from their amino acid sequences by Rekker's method (16). As shown in Table 2 (Step 1), Rekker's calculated retention times of hexapeptides with the sequence Ala-Ala-Pro-Xxx-Ser-Ala and their expected cleaved fragments (Ala-Ala-Pro-Xxx and Ser-Ala respectively) predicted by this method are short and interspersed. The retention times can be adjusted by extending the oligopeptides to the N- and/or C-terminal directions. Separation of the intact peptides from their cleaved N-terminal fragments cannot be improved by N-terminal elongation since the respective retention times of these forms should change in the same degree. Nevertheless, for analytical purposes an extra His residue was added to the N-terminus. It was beneficial because His facilitates easier identification of the fragments by amino acid analysis. As expected, all the retention time values shifted upward to about the same extent, keeping the differences between the various forms almost unaltered (Table 2, Step 2), To achieve better separation, the peptides were lengthened at their C-termini. In addition to performing criterion ii, many different amino acid residues can ensure the same retention time shift when added to the peptides. Thus, the design of the extended forms gave us some freedom to consider certain practical aspects. To facilitate peptide synthesis and avoid hydrophobic or acidic cores in the peptide backbone, Asp-Ile and Gln-Ile motifs were chosen. Addition of a pair of these motifs results in an oligopeptide library with the overall His-Ala-Ala-Pro-Xxx-Ser-Ala-Asp-Ile-Gln-Ile sequence. As is presented in Table 2 (Step 4), the calculation predicts complete separation of the individual peptides from each other and from their expected proteolytic products in a way that the intact forms and the cleavage products segregate in two different groups.

**TABLE 2**

Substrate Design: Steps of Retention Time Adjusting

| Step | | Fragments | | Rekker's retention times if X is K | R | P | Y | L | F | W |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | N: | AAPX | | 9.6 | 12.7 | 20.6 | 33.7 | 39.1 | 44.1 | 48.5 |
| | FL: | AAPXSA | | 13.1 | 16.1 | 24.1 | 37.2 | 42.6 | 47.6 | 52 |
| | C: | SA | 3.5 | | | | | | | |
| 2 | N: | HAAPX | | 18.4 | 21.5 | 29.4 | 42.5 | 47.9 | 52.9 | 57.3 |
| | FL: | HAAPXSA | | 21.9 | 25.0 | 32.9 | 46.0 | 51.4 | 56.4 | 60.8 |
| | C: | SA | 3.5 | | | | | | | |
| 3 | N: | HAAPX | | 18.4 | 21.5 | 29.4 | 42.5 | 47.9 | 52.9 | 57.3 |
| | FL: | HAAPXSADI | | 49.2 | 52.3 | 60.2 | 73.3 | 78.7 | 83.7 | 88.1 |
| | C: | SADI | 30.8 | | | | | | | |
| 4 | N: | HAAPX | | 18.4 | 21.5 | 29.4 | 42.5 | 47.9 | 52.9 | 57.3 |
| | FL: | HAAPXSADIQI | | 79.8 | 82.9 | 90.8 | 104.0 | 109.3 | 114.0 | 119.0 |
| | C: | SADIQI | 61.4 | | | | | | | |
| 5 | N: | HAAPX | | 18.4 | 21.5 | 29.4 | 42.5 | 47.9 | 52.9 | 57.3 |
| | FL: | HAAPXSADIQIDI | | 107.1 | 110.0 | 118.1 | 131.0 | 136.6 | 142.0 | 146.0 |
| | C: | SADIQIDI | 88.7 | | | | | | | |

*Note.* N, N-terminal fragment after cleavage; C, C-terminal fragment after cleavage; FL, full-length peptide.

Although the sequence of the library with the seven chosen $P_1$ residues predicts fair separation for all the 15 individual molecular species (the initial components and the cleavage products), it does not allow a wider selection at the $P_1$ site. This is due to the fact that retention times of three components: the most retarded N-terminal cleavage product, the common C-terminal cleavage product, and the least retarded intact peptide are similar. However, as is demonstrated in Table 2 (Step 5), the His-Ala-Ala-Pro-Xxx-Ser-Ala-Asp-Ile-Gln-Ile-Asp-Ile oligopeptide library resulted by further elongation of the peptide library with another Asp-Ile segment would open a wide separation window between the intact and cleaved components.

### *Adjusting RP-HPLC Conditions*

Since the Rekker's method predicted complete segregation of the intact and the cleaved components in two groups, there were three possible choices how to follow the enzymatic reactions. One can analyze the consumption of the intact peptides, the appearance of the cleavage products, or both. Data collection for the retention range of both the intact and the cleaved peptides requires a very long running time. Since our goal was to develop a fast assay system, this collection mode was precluded. To decide whether to focus on the intact peptides or rather the cleavage products, we had to appreciate certain practical considerations such as the fact that sensitivity of the detection and the signal/noise ratio vary throughout different sections of the chromatogram.

The 220-nm signal of UV absorbance detection is a function of the absorbance of the peptide bonds and the aromatic amino acid residues. Thus in our system the UV signals produced by the intact peptides are higher by 15–50% than the signals of the shorter products. The small molar absorption coefficient of the pentapeptide products results in a relatively low signal/noise ratio. This makes the detection unreliable especially at the beginning of the reaction when only small amounts of cleaved product are generated. Therefore, accurate initial velocity determination becomes virtually impossible. The sensitivity can be raised by increasing the original peptide concentration but the applicable amount of the peptide is limited by the capacity of the analytical column.

Presence of any exopeptidase or non-serine-proteinase-like activities in the reaction mixture can result in unexpected cleavages of the substrate oligopeptides at positions other than the proposed $P_1$ site. Thus, by focusing only on the intact peptides can lead to false interpretation because it is indistinguishable whether the decrease of a peak area is a result of a predicted cleavage at the $P_1$ site or an unexpected one elsewhere.

However, detection of nonspecific cleavages becomes possible by measuring the amount of the common C-terminal fragment since it is generated by the desired reactions only. Any unexpected cleavage would de-

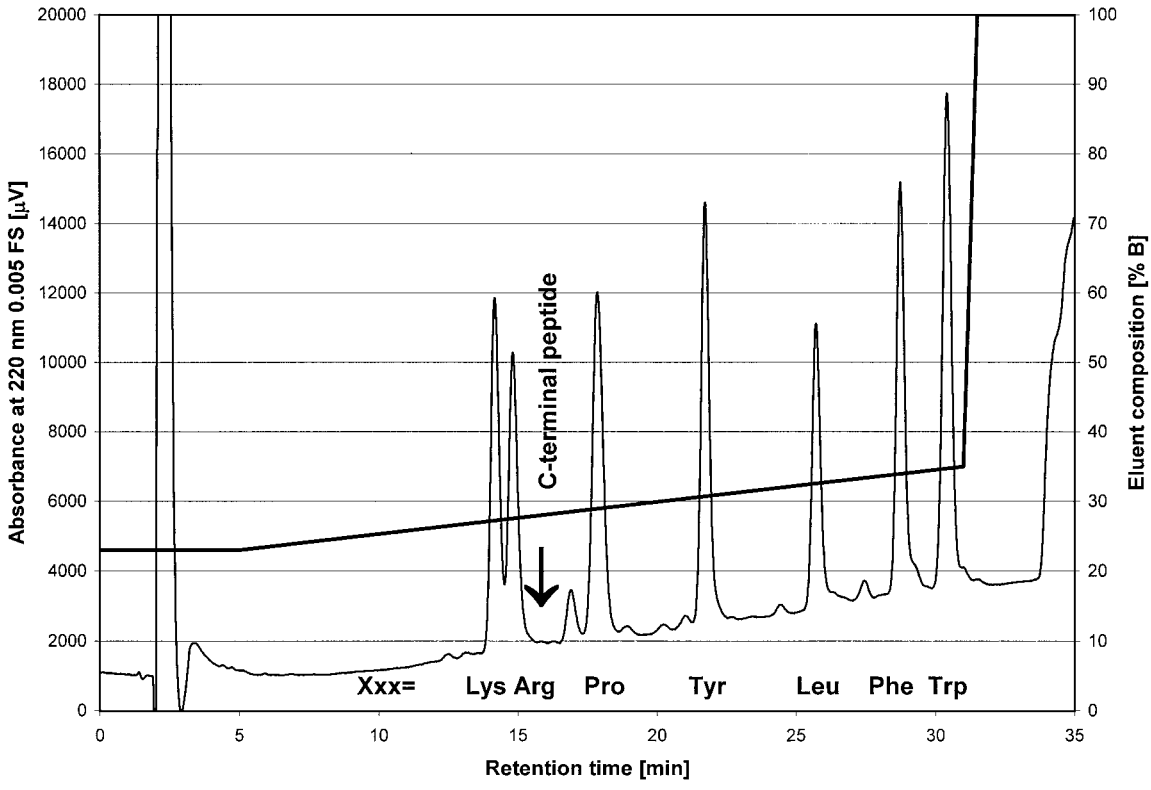


**FIG. 2.** RP-HPLC chromatogram of intact peptide mixture at the 0 min state. Five-microliter aliquots of the digestion mixture were loaded onto the Aquapore OD 300 reversed-phase column (2.1 × 220 mm) of the automated HPLC system. Eluent A was water containing 0.1% (v/v) trifluoroacetic acid and eluent B was 80% acetonitrile and 20% water containing 0.1% (v/v) trifluoroacetic acid. Chromatographic conditions were set to acquire data only from the retention range of the intact oligopeptides. The column was equilibrated with 23% eluent B and the injection was followed by a 5-min hold of this composition, after which a linear gradient from 23 to 35% eluent B was applied in 26 min, with a 1 ml/min flow rate. Detection was carried out at 220 nm.

crease the amount of the C-terminal fragment relative to the theoretical value. The above consideration led us to focus on the consumption of intact peptides with checking the amount of the common C-terminal product produced.

The chromatogram in Fig. 2 shows that the seven intact components of the substrate mixture were indeed separated from each other and were eluted in the expected order: Lys, Arg, Pro, Tyr, Leu, Phe, and Trp. The only discrepancy was that the C-terminal fragment eluted somewhat later than expected, between the Arg and Pro-containing peptides. Fortunately, neither the separation of the intact components nor the identification of the peaks was affected by this retention time shift.

*HPLC Analysis of the Enzymatic Digestion*

The seven-component peptide library was digested with bovine and rat chymotrypsin, bovine and crayfish trypsin, bovine plasmin and thrombin, mouse endoproteinase Arg-C, and Tyrpsin. Samples withdrawn after different time intervals from each enzyme digestion mixture were run on RP-HPLC giving sets of chromatograms. Figure 3 shows one set of such chromatograms in the case of rat chymotrypsin.

Chromatograms after 16-min digestion with different enzymes are presented on Fig. 4. Qualitative differences among the specificities of the individual enzymes can be easily observed at this time point by a simple visual inspection. Differences of the specificities can be more accurately demonstrated by using chromatography software to calculate the peak areas. The areas determined by the software were normalized with the zero minute value of the Pro-containing peptide used as internal standard (Table 3A). Parallel analyses in the case of bovine trypsin and chymotrypsin resulted in less than 2% average deviation in the respective peak areas.

The most accurate specificity data could be generated from the decrease of the peak areas at the 16-min stage compared to the corresponding 0-min values (Table 3B). At this time point none of the substrates had been completely consumed. The calculated specificity profiles are presented in Fig. 5. These data show that

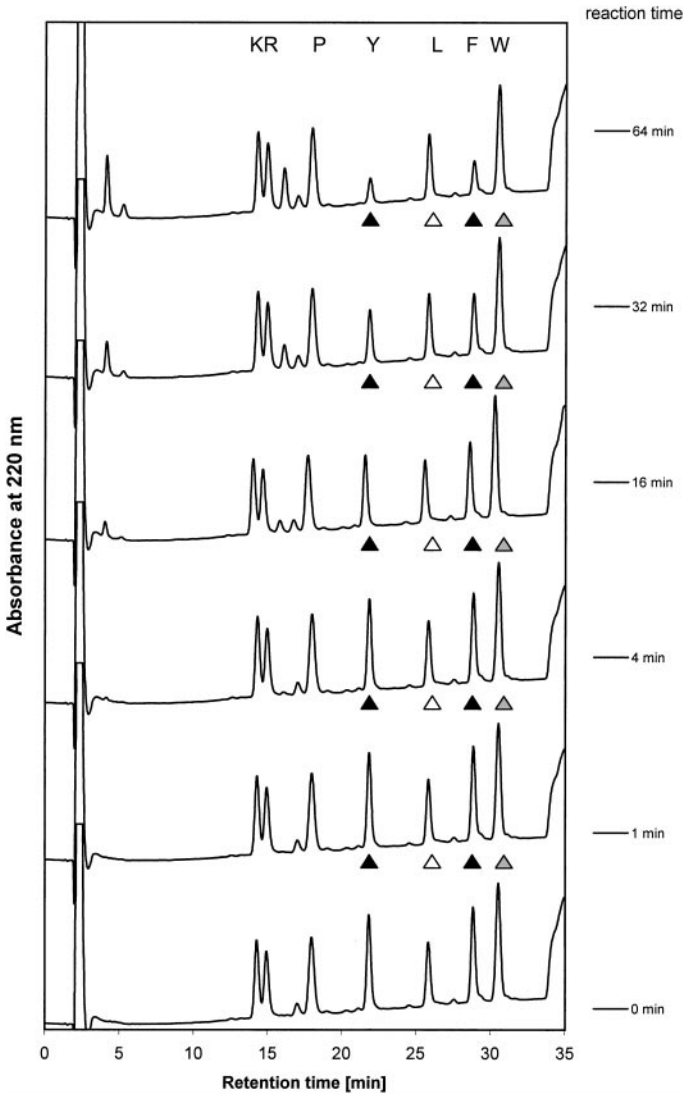


**FIG. 3.** RP-HPLC chromatogram set with different reaction times, in the case of rat chymotrypsin. HPLC conditions were the same as those described in the legend to Fig. 2. Arrowheads point to peaks belonging to peptides involved in the appropriate enzyme reaction. The darker the arrowhead, the higher the substrate acceptance of the enzyme is.

as long as none of the seven different substrates are completely consumed, only those are being cleaved that contain a $P_1$ side chain known to be preferred by the given enzyme. For example, while bovine trypsin cleaved 12 and 36% of the Lys- and Arg-containing peptides, respectively, none of the other substrates have been cleaved to any detectable extent (Fig. 5e). Similarly, while 8, 2, 4, and 17% of the Tyr, Leu, Phe, and Trp-containing peptides were digested by bovine chymotrypsin, respectively, no cleavage of the substrates with Lys and Arg $P_1$ residues could be detected (Fig. 5b).

In addition to clear demonstration of the well-known differences among several model enzymes, our HPLC system revealed significant substrate selectivity variation of two chymotrypsins from two different species. Specificity profiles of bovine chymotrypsin A and rat chymotrypsin (that can be regarded as the rat equivalent of bovine chymotrypsin B) were found to be markedly different. The rat enzyme showed high specificity toward Tyr and Phe residues compared to the bovine proteinase, which preferred Trp as a $P_1$ site residue (Table 3B, Figs. 5a and 5b). This finding is in good agreement with specificity differences of chymotrypsin A and B forms demonstrated recently by using fluorogenic substrates (17).

In the case of a mutant form of rat trypsin named Tyrpsin, our test system unambiguously demonstrated its high selectivity toward the Tyr substrate (the name Tyrpsin was given for this sharp tyrosine specificity). Quantitatively it means that more than 27% of the Tyr-containing peptide has been cleaved, whereas no measurable digestion of any other component occurred (Table 3B and Fig. 5c). Considering this outstanding selectivity compared to the two chymotrypsin isoforms, Tyrpsin might be used as a tool in protein sequencing. For this purpose, low frequency of the target residues is desired to obtain only a few long peptide fragments. Since tyrosine is a relatively rare residue Tyrpsin may be of analytical potential.

No considerable difference has been found between the specificity profiles of the bovine and the crayfish trypsins (Table 3B, Figs. 5e and 5f).

In our system bovine plasmin showed slight preference toward the Lys-containing peptide over the Arg-containing one (Table 3B and Fig. 5d). These data are in agreement with those presented with using chromogenic peptide substrates (18, 19). Although plasmin seems to prefer the lysine side chain when presented on a small peptide, all the known physiological plasmin cleavage sites contain Arg residue. This finding highlights the role of additional specificity determinant regions that often facilitate proper assembly of the enzyme–substrate complex that modulates the specificity of many selective enzymes of the serum such as plasmin (20). Since our substrate peptides lack any specific additional sites, slight basal $P_1$ Lys preference of plasmin could be detected.

Bovine thrombin and mouse endoproteinase Arg-C showed exclusive Arg specificity as expected (20–22) (Table 3B, Figs. 5g and 5h).

### *Quantitative Characterization of Enzyme Specificity*

Initial concentrations of the substrates were determined by amino acid analysis and set to be equal in the starting peptide mixture. Change of the individual substrate concentrations was calculated from the decrease

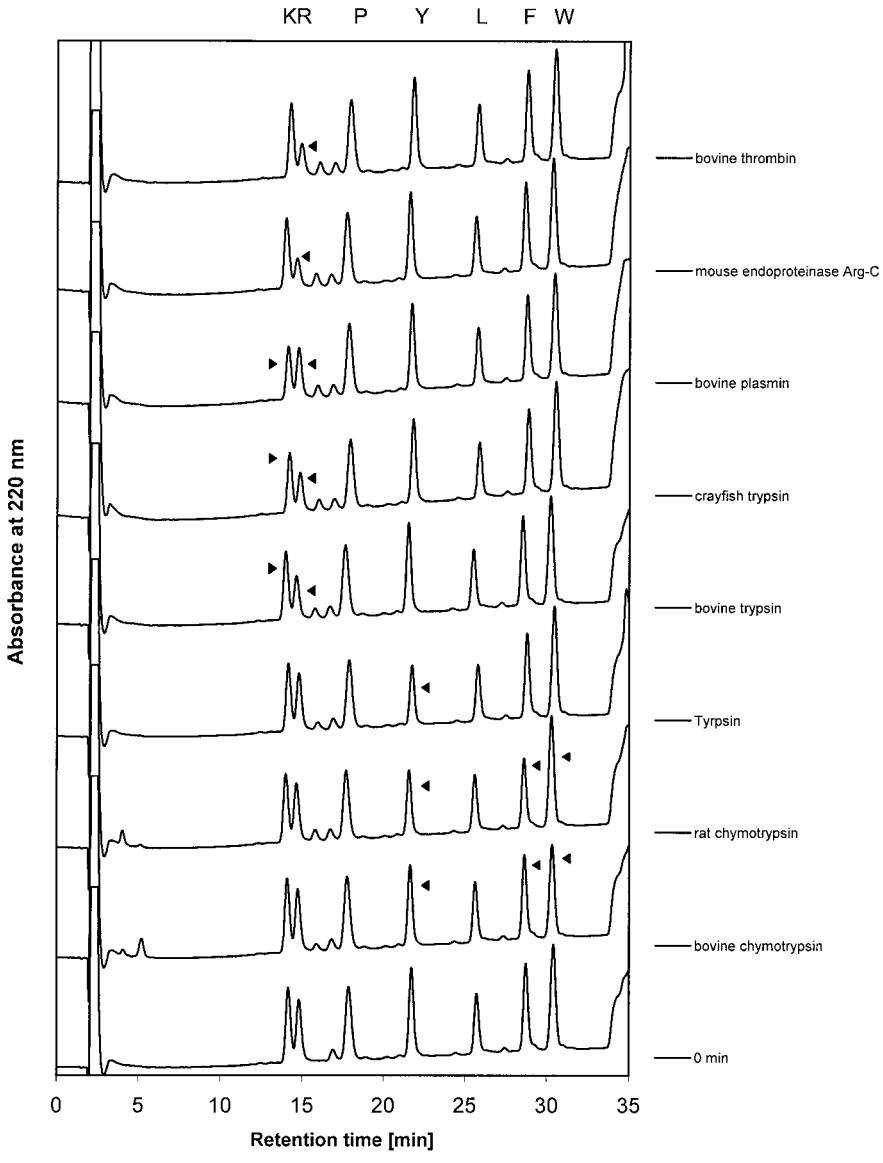


**FIG. 4.** RP-HPLC chromatograms after 16-min digestion with different enzymes. HPLC conditions were the same as those described in the legend to Fig. 2. Arrowheads point to peaks belonging to peptides involved in the appropriate enzyme reaction.

of the corresponding substrate peak areas after separation of the components by RP-HPLC. Quantitative evaluation of the enzyme specificity profiles was based on the equation that describes the enzymatic reaction when competing substrates are present (8).

In such a system, the rates of the enzymatic reactions for the individual substrates $i$ and $j$ can be expressed as

$$v_i = \frac{d[\mathrm{P}_i]}{dt} = \frac{[\mathrm{S}_i] \cdot V_i/K_{\mathrm{M}i}}{1 + [\mathrm{S}_j]/K_{\mathrm{M}j} + [\mathrm{S}_i]/K_{\mathrm{M}i}} \quad [1]$$

$$v_j = \frac{d[\mathrm{P}_j]}{dt} = \frac{[\mathrm{S}_j] \cdot V_j/K_{\mathrm{M}j}}{1 + [\mathrm{S}_j]/K_{\mathrm{M}j} + [\mathrm{S}_i]/K_{\mathrm{M}i}}, \quad [2]$$

where $v$ is the actual velocity of the reaction (or initial velocity at the zero time point), [P] is the product concentration, $t$ is time, $V$ is the maximal velocity, [S] is the substrate concentration, and $K_{\mathrm{M}}$ is the Michaelis constant. Ratios of the individual rates will be then

$$\frac{v_i}{v_j} = \frac{[\mathrm{S}_i] \cdot V_i/K_{\mathrm{M}i}}{[\mathrm{S}_j] \cdot \mathrm{V}_j/\mathrm{K}_{\mathrm{M}j}} = \frac{[\mathrm{S}_i] \cdot [\mathrm{E}_\mathrm{T}] \cdot (k_{\mathrm{cat}i}/K_{\mathrm{M}i})}{[\mathrm{S}_j] \cdot [\mathrm{E}_\mathrm{T}] \cdot (k_{\mathrm{cat}j}/K_{\mathrm{M}j})}. \quad [3]$$

That is, because $V = [\mathrm{E}_\mathrm{T}] \cdot k_{\mathrm{cat}}$ where $[\mathrm{E}_\mathrm{T}]$ is the total enzyme concentration and $k_{\mathrm{cat}}$ is the catalytic constant. The $k_{\mathrm{cat}}/K_{\mathrm{M}}$ ratio often referred to as catalytic efficiency is the most informative parameter for an enzyme-catalyzed reaction. As Eq. [3] shows, the ratio of the indi-

**TABLE 3**

| Peptide of the substrate mixture containing | 0 min | Rat chymotrypsin | Bovine chymotrypsin | Tyrpsin | Bovine trypsin | Crayfish trypsin | Bovine plasmin | Bovine thrombin | Mouse endoproteinase Arg-C |
|---|---|---|---|---|---|---|---|---|---|
| A. Normalized peak areas after a 16 min digestion [$\mu$V · s] | | | | | | | | | |
| Lys | 202697 | 202376 | 203525 | 202561 | 178924 | 164850 | 151605 | 202686 | 203130 |
| Arg | 187078 | 186926 | 187667 | 186351 | 120620 | 118399 | 159092 | 100473 | 94545 |
| Tyr | 233602 | 168465 | 214921 | 168993 | 233635 | 233997 | 233472 | 233183 | 233593 |
| Leu | 156656 | 153680 | 153210 | 156826 | 156404 | 156545 | 156756 | 156676 | 156721 |
| Phe | 230251 | 187899 | 220686 | 230590 | 230475 | 230295 | 230432 | 230289 | 230340 |
| Trp | 307810 | 303338 | 255391 | 306997 | 306437 | 307459 | 307295 | 307659 | 308003 |
| B. Decrease of peak areas after a 16-min digestion with respective enzyme [%] | | | | | | | | | |
| Lys | | 0.0 | 0.0 | 0.0 | 11.7 | 18.7 | 25.2 | 0.0 | 0.0 |
| Arg | | 0.0 | 0.0 | 0.0 | 35.5 | 36.7 | 15.0 | 46.3 | 49.5 |
| Tyr | | 27.9 | 8.0 | 27.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Leu | | 1.9 | 2.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Phe | | 18.4 | 4.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Trp | | 1.5 | 17.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| C. Decrease of peak areas after a 1-min digestion with respective enzyme [%] | | | | | | | | | |
| Lys | | 0.0 | 0.0 | 0.0 | 2.9 | 1.8 | 3.9 | 0.0 | 0.0 |
| Arg | | 0.0 | 0.0 | 0.0 | 3.8 | 2.6 | 2.1 | 4.2 | 2.7 |
| Tyr | | 2.4 | 1.0 | 3.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Leu | | 0.6 | 0.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Phe | | 2.0 | 0.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Trp | | 0.4 | 2.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

vidual reaction rates depends on the ratios of the actual substrate concentrations and the ratios of the respective $k_{cati}/K_{Mi}$ values. Rearranging the equation yields:

$$\frac{v_i/[S_i]}{v_j/[S_j]} = \frac{k_{cati}/K_{Mi}}{k_{catj}/K_{Mj}}. \quad [4]$$

This value quantifies how the enzyme discriminates between pairs of individual substrates. Indeed, it is the most relevant parameter that can characterize the selectivity of an enzyme. Ratios of the catalytic efficiency values when one of these values is used as a reference describe a specificity profile that is characteristic for the enzyme.

Since less than 10% of the substrate is converted to product, the measured apparent velocity approximates the theoretical initial velocity value and in the denominator of the equation the mean substrate concentration can be applied.

$$v_i/[S_i] = \frac{\frac{[S_{0i}] - [S_{ti}]}{t}}{\frac{[S_{0i}] + [S_{ti}]}{2}}. \quad [5]$$

In terms of the data derived from the chromatograms, $[S_{0i}]$ and $[S_{ti}]$ can be calculated from the corresponding substrate peak areas. For convenience, the substrate concentrations $[S_{0i}]$ and $[S_{ti}]$ are defined as percentage values with $[S_{0i}]$ being 100%. As long as the initial assumption for the limited substrate conversion applies, the $v_i/[S_i]$ values can be considered as pseudo-first-order rate constants. A short 1-min digestion of the peptide substrate library with the eight different enzymes individually fulfilled this assumption. The corresponding pseudo-first-order rate constants were calculated by Eq. [5] using peak area data from Table 3C. The pseudo-first-order rate constant values are presented in Table 4.

As was shown in Eq. [4], the ratio of the pseudo-first-order rate constant values equal to the ratio of the catalytic efficiency. Therefore, the catalytic constant values ($k_{cat}$:$K_M$) themselves cannot be extracted directly from our data. However, once a $k_{cat}/K_M$ value for any of the components is determined in an independent assay, the corresponding values for all the other peptides of library can be calculated from their catalytic efficiency values determined with our technique.

## CONCLUSIONS

A carefully designed oligopeptide substrate library in combination with a simple RP-HPLC separation

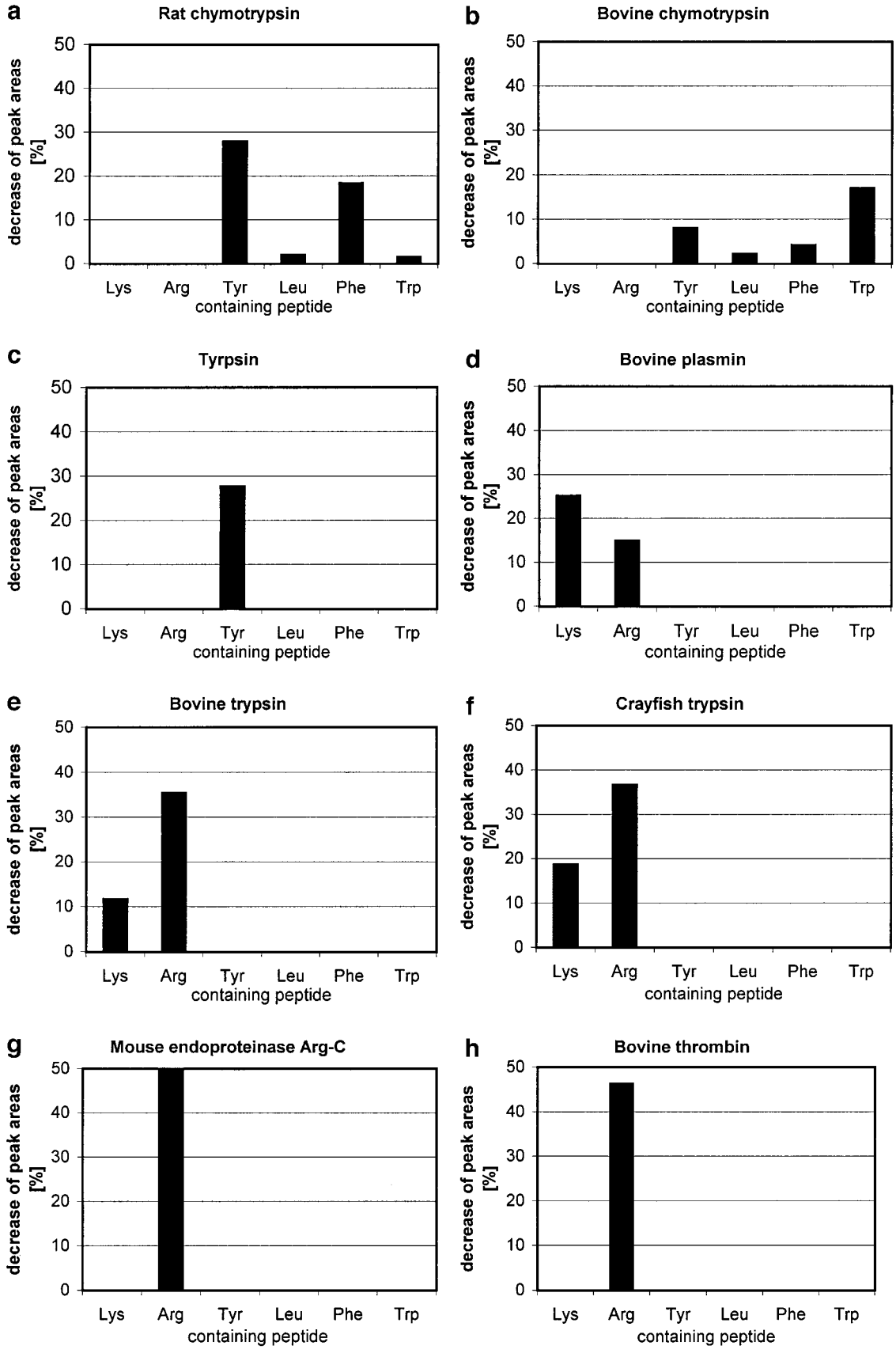


**FIG. 5.** Graphical presentation of enzyme specificities. Diagrams represent the data of Table 3B.

**TABLE 4**

$v_i/[S_i]$ Values at a 1-min Digestion with Different Enzymes

| Peptide of the substrate mixture containing | $v_i/[S_i]$ values, 1 min digestion with | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Rat chymotrypsin | Bovine chymotrypsin | Tyrpsin | Bovine trypsin | Crayfish trypsin | Bovine plasmin | Bovine thrombin | Mouse endoproteinase Arg-C |
| Lys | — | — | — | 0.0294 | 0.0182 | 0.0398 | — | — |
| Arg | — | — | — | 0.0387 | 0.0263 | 0.0212 | 0.0429 | 0.0274 |
| Tyr | 0.0243 | 0.0101 | 0.0315 | — | — | — | — | — |
| Leu | 0.0060 | 0.0070 | — | — | — | — | — | — |
| Phe | 0.0202 | 0.0090 | — | — | — | — | — | — |
| Trp | 0.0040 | 0.0243 | — | — | — | — | — | — |

*Note.* Data were calculated by Eq. [5] using data from Table 3C.

described here was capable of providing information on $P_1$ specificity of eight different serine proteinases (Table 5). Using this assay we could distinguish even between two isotypes of chymotrypsin revealing their different substrate specificities and the system could also demonstrate outstanding tyrosine specificity of a mutant trypsin, called Tyrpsin. Longer reaction times yielded qualitative information that was easy to interpret, not only by the specificity profiles formed by rendering data of peak areas into block diagrams, but also even by simple visual inspection of the results. The substrate preference profiles that we determined here for each enzyme are essentially identical to the ones provided by assays with only a single chromogenic substrate analyzed at a time (2, 17). Shorter incubation times when less than 10% of the substrate molecules was cleaved allowed calculation of specificity constants from data resulting from chromatography software, thus provided quantitative description of enzyme specificity. Even the specificity constants as calculated for individual substrate pairs by our method are close to those determined with the corresponding chromogenic substrates (1, 17).

Although the system we developed can certainly be used for quantitative analyses, we would emphasize its great potentials as a fast qualitative specificity test for uncharacterized serine proteinases. In addition to high reproducibility, an important advantage of this assay is its ability to be used in the case of colorful or even unpurified samples where the classic photometric assays are restricted; therefore, our assay will be a useful test for high-throughput analysis of natural raw samples.

Oligopeptide libraries overlapping with some $P_1$ amino acid residues and organized on the basis of the designed oligopeptide backbone are predicted to allow good separation of the components with as many as 14 different amino acid residues incorporated at the $P_1$ site (His, Ala, Ser, Asp, and Gln should be excluded because they are involved in the common peptide sequence and Cys also is undesirable because of possible intermolecular disulfide bridge formation). It provides an almost complete screen for analysis of serine proteinase $P_1$ specificity.

**TABLE 5**

Relative Specificities of the Investigated Enzymes

| Peptide of the substrate mixture containing | Relative $v_i/[S_i]$ values, 1 min digestion with | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Leu at $P_1$ as the basis | | | Lys at $P_1$ as the basis | | | | |
| | Rat chymotrypsin | Bovine chymotrypsin | Tyrpsin | Bovine trypsin | Crayfish trypsin | Bovine plasmin | Bovine thrombin | Mouse endoproteinase Arg-C |
| Lys | — | — | — | 1.0 | 1.0 | 1.0 | — | — |
| Arg | — | — | — | 1.3 | 1.5 | 0.5 | — | — |
| Tyr | 4.0 | 1.4 | — | — | — | — | — | — |
| Leu | 1.0 | 1.0 | — | — | — | — | — | — |
| Phe | 3.4 | 1.3 | — | — | — | — | — | — |
| Trp | 0.7 | 3.5 | — | — | — | — | — | — |

*Note.* Data were calculated with data from Table 4.

## ACKNOWLEDGMENT

The technical assistance of Ms. E. Karácsony is appreciated.

## REFERENCES

1. Gráf, L., Jancsó, Á., Szilágyi, L., Hegyi, Gy., Pintér, K., Náray-Szabó, G., Hepp, J., Medzihdradszky, K., and Rutter, W. J. (1988) Electrostatic complementarity within the substrate-binding pocket of trypsin. *Proc. Natl. Acad. Sci. USA* **85,** 4961–4965.
2. Venekei, I., Szilágyi, L., Gráf, L., and Rutter, W. J. (1996) Attempts to convert chymotrypsin to trypsin. *FEBS Lett.* **379,** 143–147.
3. Hedstrom, L., Szilagyi, L., and Rutter, W. J. (1992) Converting trypsin to chymotrypsin: The role of surface loops. *Science* **255,** 1249–1253.
4. Hedstrom, L. (1996) Trypsin: A case study in the structural determinants of enzyme specificity. *Biol. Chem.* **377,** 465–470. [Review]
5. Chagas, J. R., Juliano, L., and Prado, E. S. (1991) Intramolecularly quenched fluorogenic tetrapeptide substrates for tissue and plasma kallikreins. *Anal. Biochem.* **192,** 419–425.
6. Pimenta, D. C., Chao, J., Chao, L., Juliano, M. A., and Juliano, L. (1999) Specificity of human tissue kallikrein toward substrates containing Phe–Phe pair of amino acids. *Biochem. J.* **339,** 473–479.
7. Portaro, F. C., Santos, A. B., Cezari, M. H., Juliano, M. A., Juliano, L., and Carmona, E. (2000) Probing the specificity of cysteine proteinases at subsites remote from the active site: Analysis of P4, P3, P2′ and P3′ variations in extended substrates. *Biochem. J.* **347,** 123–129.
8. Cornish-Bowden, A. (1995) Fundamentals of Enzyme Kinetics, Portland Press, London.
9. Berman, J., Green, M., Sugg, E., Anderegg, R., Millington, D. S., Norwood, D. L., McGeehan, J., and Wiseman J. (1992) Rapid optimization of enzyme substrates using defined substrate mixtures. *J. Biol. Chem.* **267,** 1434–1437.
10. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27,** 157–162.
11. Schellenberger, V., Turck, Ch. W., Hendstrom, L., and Rutter, W. J. (1993) Mapping the S′ subsites of serine proteases using acyl transfer to mixtures of peptide nucleophiles. *Biochemistry* **32,** 4349–4353.
12. Schellenberger, V., Turck, Ch. W., and Rutter, W. J. (1994) Role of the S′ subsites in serine protease catalysis. Active-site mapping of rat chymotrypsin, rat trypsin, alpha-lytic protease, and cercarial protease from *Schistosoma mansoni. Biochemistry* **33,** 4251–4257.
13. Birkett, A. J., Soler, D. F., Woltz, R. L., Bond, J. S., Wiseman, J., Berman, J., and Harris, R. B. (1991) Determination of enzyme specificity in a complex mixture of peptide substrates by N-terminal analysis. *Anal. Biochem.* **196,** 137–143.
14. Thompson, R. C., and Blout, E. R. (1973) Dependence of the kinetic parameters for elastase-catalyzed amide hydrolysis on the length of peptide substrates. *Biochemistry* **12,** 57–65.
15. Ke, S. H., Coombs, G. S., Tachias, K., Corey, D. R., and Madison, E. L. (1997) Optimal subsite occupancy and design of a selective inhibitor of urokinase. *J. Biol. Chem* **272** (33), 20456–20462.
16. Sasagawa, T., and Teller, D. C. (1987) *in* Handbook of HPLC for the Separation of Amino Acids, Peptides and Proteins (Hancock, W. S., Ed.), Vol 2, pp. 53–65, CRC Press, Boca Raton, FL.
17. Hudáky, P., Kaslik, Gy., Venekei, I., and Gráf, L. (1999) The differential specificity of chymotrypsin A and B is determined by amino acid 226. *Eur. J. Biochem.* **259,** 528–533.
18. Pierzchala, P. A., Dorn, C. P., and Zimmermann, M. (1979) A new fluorogenic substrate for plasmin. *Biochem. J.* **183,** 555–559.
19. Clavin, S. A., Bobbitt, J. L., Shuman, R. T., and Smithwick, E. L., Jr. (1977) Use of peptidyl-4-methoxy-2-naphthylamides to assay plasmin. *Anal. Biochem.* **80,** 355–365.
20. Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S., and Ellman J. A. (2000) Synthesis of positional-scanning libraries of fluorogenic peptide substrates to define the extended substrate specificity of plasmin and thrombin. *Nat. Biotechnol.* **18,** 187–193.
21. Chang, J. Y. (1985) Thrombin specificity. Requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate. *Eur. J. Biochem.* **151,** 217–224.
22. Schenkein, I., Levy, M., Franklin, E. C., and Frangione, B. (1977) Proteolytic enzymes from the mouse submaxillary gland. Specificity restricted to arginine residues. *Arch. Biochem. Biophys.* **182,** 64–70.