Communication

Rapid Optimization of Enzyme Substrates Using Defined Substrate Mixtures*

(Received for publication, August 7, 1991)

Judd Berman†, Michael Green‡, Elizabeth Sugg§, Rob Anderegg†, David S. Millington†,
Daniel L. Norwood†, Jerry McGeehan‡, and Jeffrey Wiseman§

From Glaxo Research Laboratories, Research Triangle Park, North Carolina 27709 and the §Division of Pediatric Genetics and Metabolism, Duke University Medical Center, Durham, North Carolina 27710

A strategy is described for the rapid optimization of \( k_{\text{cat}}/K_m \) for protease substrates. Selected positions of a given peptide substrate sequence are varied through synthesis with mixtures of amino acids. Incubation of the resulting peptide mixture with the enzyme of interest and analysis by high pressure liquid chromatography provides a direct measure of analogs with enhanced \( k_{\text{cat}}/K_m \). High performance liquid chromatography/continuous flow fast atom bombardment mass spectrometry is used to assign structure to each peak in the chromatogram. As an example of the utility and efficiency of “substrate mapping” we describe optimization of the collagenase substrate Dnp-Pro-Leu-Gly-Leu-Trp-Ala-d-Arg-NH₂ (where Dnp is dinitrophenyl) at the P₁ and P₂ positions. Six different mixtures were prepared for evaluation, representing the synthesis of 128 different synthetic substrates. “Substrate mapping” has led to Dnp-Pro-Leu-Gly-Cys(Me)-His-Ala-d-Arg-NH₂, a substrate that possesses a 10-fold better \( k_{\text{cat}}/K_m \) than Dnp-Pro-Leu-Gly-Leu-Trp-Ala-d-Arg-NH₂.

Substrate optimization is a critical step in providing a reliable and convenient enzyme assay and can provide information for the optimization of inhibitor structure. The optimization problem is pronounced when searching for appropriate substrates for proteolytic enzymes. Model substrates are usually shortened peptide segments derived from the physiological target sequences. Complications in substrate design arise when there are several target sequences which are cleaved by a single enzyme. A relevant example is the HIV\(^{+} \) protease, which processes a wide variety of differing proproteins essential for the maturation of HIV (1). The essential feature of optimal substrate design is rapid turnover (enhanced \( k_{\text{cat}}/K_m \)) of target peptides. To minimize the time required for the development of an optimal substrate, a method has been developed which allows the rapid mapping of substrate specificity. This method relies on the ability to synthesize defined mixtures of amino acids at each subsite of peptide substrates, readily assign structure to each component in the mixture, and subsequently carry out enzymatic screening of the mixtures. Structural assignment is greatly facilitated by the use of combined high performance liquid chromatography/continuous flow fast atom bombardment mass spectrometry (HPLC/CF-FAB-MS) (2-4).

As an example of the utility of the method, we describe optimization of the P₁ and P₂ positions of the collagenase substrate mimic Dnp-Pro-Leu-Gly-Leu-Trp-Ala-d-Arg-NH₂ (5-7). This substrate is cleaved specifically at the Gly-Leu linkage to yield Dnp-Pro-Leu-Gly-OH and H-Leu-Trp-Ala-d-Arg-NH₂. Using this template, three different substrate mixtures were generated at both the P₁ and P₂ positions, utilizing naturally occurring L-amino acids, D-amino acids, or miscellaneous unnatural amino acids, representing the synthesis of 128 different synthetic substrates. Results of the analyses of each of these substrates are reported below.

EXPERIMENTAL PROCEDURES

Peptide Synthesis — The protected peptides at the P₁ and P₂ positions were prepared starting with 0.5 mmol of \( p \)-methylbenzhydrylamine resin (1.1 mmol/g) (Peptides International, Louisville, KY). The residues with fixed assignments were double coupled (prepared hydroxybenzotriazole esters in N-methylpyrrolidone) followed by capping with acetic anhydride. For coupling the mixed residues, an equimolar amount of each of the amino acids (0.25 mmol total) was converted to the hydroxybenzotriazole ester and coupled using an extended cycle (1 h) followed by exhaustive capping with acetic anhydride. Trifunctional amino acids were protected as follows: N"-Boc-Cys(MeBzl)-OH, N"-Boc-Asp(Xhp)-OH, N"-Boc-Glu(Xhp)-OH, N"-Boc-His(Dnp)-OH, N"-Boc-Lys(2-Cl-Z)-OH, N"-Boc-Arg(Tos)-OH, N"-Boc-Ser(Bzl)-OH, N"-Boc-Thr(Bzl)-OH, N"-Boc-Tyr(2-Br-Z)-OH, N"-Boc-Orn(Xhp)-OH, N"-Boc-Cys(p-amino-Cbz)-Phe-OH, N"-Boc-Asn was used in the mixture as its preactivated p-nitrophenyl ester derivative. After the final coupling and N"-Boc deprotection, the residues were neutralized and treated with 2,4-dinitrofluorobenzene (0.55 mmol) and disopropylethylamine (0.55 mmol) in N-methylpyrrolidone (10 ml) for 3 h. The residues were then treated with thioenol (His(Dnp) removal), and the resulting residues were cleaved using anisole (2 ml) in anhydrous HF (10 ml) at -10 °C for 40 min. The HF was removed in vacuo, and the residues were diluted with trifluoroacetic acid (50 ml) and filtered. The residues were then additionally rinsed with trifluoroacetic acid (2 X 50 ml). The trifluoroacetic acid extracts were concentrated in vacuo, diluted with aqueous 60% CH₂CN (150 ml), and lyophilized to yellow solids. Enzyme Digests—Peptide mixtures were dissolved by warming in dimethyl sulfoxide. Insoluble material was pelleted by centrifugation. Concentrations of Dnp-containing peptides were determined by measuring absorbance at 375 nm (Dnp \( \epsilon_{375} = 16,000 \, \text{M}^{-1} \, \text{cm}^{-1} \)). Substrate mixtures were diluted to 250 \( \mu \text{M} \) final concentration in assay buffer (200 mM NaCl, 30 mM Tris, 5 mM CaCl₂, 0.05% Brij, pH 7.6). Standard substrate, Dnp-Pro-Leu-Gly-Leu-Trp-Ala-d-Arg-NH₂, was added to the d and miscellaneous amino acid mixtures to a final concentration of 15 \( \mu \text{M} \). The digestes were initiated by adding recombinant human fibroblast collagenase or collagenase/EDTA (control) to each mixture (250 \( \mu \text{M} \), 10-15 \( \mu \text{M} \) per substrate). The reactions were followed by monitoring the increases in tryptophan fluorescence (excitation, 280/emission, 346) of the mixtures, and they were quenched with EDTA after 30-50% hydrolysis of Dnp-Pro-Leu-Gly-Leu-Trp-Ala-d-Arg-NH₂.

HPLC/CF-FAB-MS—Chromatography was performed using a Waters 600 solvent delivery system and two Waters Delta Pak C₁₈ columns connected in series (150 × 3.9 mm, inner diameter). Solvent A consisted of aqueous 0.1% trifluoroacetic acid with 1% glycerol, and solvent B consisted of aqueous 0.1% trifluoroacetic acid, 1% acetonitrile. The gradient was as follows: 60% A (3 min), 10% A (17 min), 90% A (21 min), and 60% A (25 min). The flow rate was 1 ml/min, and the UV detector was operated at 375 nm. The peptides were analyzed by fast atom bombardment mass spectrometry on a 720C mass spectrometer (Fisons Instruments).
glycerol, and 60% CH₂CN. Chromatograms were developed over 90 min using either linear or concave gradients. Flow rates for all analyses were 1 ml/min. The liquid chromatography effluent was split, and about 10 μl/min was directed into the probe of a VG 70S double focusing mass spectrometer. The magnetic field was scanned repetitively once every 5 s from m/z 500 to m/z 1500. The ion source temperature was maintained at 50 °C, and the accelerating voltage was 5 kV. Alternatively, the chromatography was as above except 1% thioglycerol was used in place of glycerol in both solvent A and B. The liquid chromatography effluent was split, and about 10 μl/min was directed into the frit-FAB probe of a JEOL SX-102 mass spectrometer. The magnetic field was scanned repetitively once every 6 s from m/z 100 to m/z 1000. The ion source temperature was maintained at 55 °C, and the accelerating voltage was 10 kV.

RESULTS AND DISCUSSION

The syntheses of substrate mixtures were accomplished using routine solid-phase peptide synthesis techniques except at positions substituted with equimolar mixtures of amino acids. At these mapping sites, the chemistry was modified and a 2-fold excess of resin was used. The use of the activated amino acid mixture as the limiting reagent produced an approximately equimolar ratio of the final peptides despite the different coupling rates of the various amino acids (8). Peptides were cleaved from the solid-phase support using anhydrous HF followed by exhaustive extraction with trifluoroacetic acid. The desired stoichiometry at the mapping positions was verified by amino acid analysis (data not shown).

Six different mixtures were prepared for evaluation, representing the synthesis of 126 different synthetic substrates. Each of these substrate stock mixtures was hydrolyzed with recombinant human fibroblast collagenase. The reactions were initiated by adding collagenase or collagenase/EDTA (control) to each mixture and quenched with EDTA after 30–50% hydrolysis of Dnp-Pro-Leu-Gly-Leu-Trp-Ala-D-Arg-NH₂. Each digest was chromatographed by HPLC using reversed-phase separation conditions optimized for the particular substrate mixture. Eluting peaks were integrated by peak area or peak height (Dnp [C] = 16,000 M⁻¹ cm⁻¹).

Under the competitive conditions employed in these digests, the relative specificity for any two competing substrates (i.e. kcat/Km value) is given by the ratios of the values of ln(S/So) for each substrate at the time of quenching (9). This kinetic analysis is general and can be applied to two or more substrates that are competing for flux through the same enzyme regardless of substrate concentration. The presence of inhibitors in a mixture will decrease the absolute rate of turnover of all substrates but will not affect the substrate specificity, kcat/Km. Potent inhibitors, therefore, can be detected although the identity of the inhibitory peptide(s) will not be readily discernible. The important point is that substrate specificity is determined by the ratios of kcat/Km alone (10). The integrated peak areas were used to calculate (ln(S/So)) for each substrate from the digests versus controls. Comparison of these values to Dnp-Pro-Leu-Gly-Leu-Trp-Ala-D-Arg-NH₂ as standard was used to translate relative kcat/Km to absolute values and allowed comparison of results across all six digests.

The composition of each peak in the chromatogram was determined by HPLC/CF-FAB-MS. The FAB spectrum of each synthetic peptide contained molecular weight information (in the form of (M + H)⁺) and structurally informative fragmentation. Fig. 1, top, shows the HPLC trace (UV = 375 nm) of the L-amino acid Pᵢ₁-substituted heptapeptide mixture. A reconstructed total ion chromatogram (m/z 800–1000) under identical conditions is shown in Fig. 1, bottom. In favorable cases, fragmentation patterns allow differentiation of isobaric compounds. As an example, Fig. 2 shows the mass chromatograms of the protonated parent ions (m/z 904, Fig. 2, bottom) along with those of two sequence-specific ions m/z 299 and m/z 313 representing ions unique to the leucine and isoleucine Pᵢ₁-substituted peptides, respectively (11). This approach allowed assignment of most of the individual components in the complex mixtures of peptides without the necessity of tedious and time-consuming isolation procedures.

Analysis of the results from the six enzyme digests gave...
both expected and unexpected results. Incubations of the D-amino acid P; and P;'-substituted heptapeptide mixtures with collagenase and examination of the resulting hydrolysates by HPLC revealed little change in the chromatograms over the time required for 50% hydrolysis of Dnp-Pro-Leu-Gly-Leu-Trp-Ala-D-Arg-NH$_2$ (data not shown). These data indicate that D-amino acids are not tolerated by collagenase at the P; or P;'-positions.

A variety of structurally diverse L-amino acid replacements is well accommodated at the P; site (Table I). The normal amino acid at P; in human skin a1 chain of collagen is Ala. However, the S; site of collagenase apparently is able to adjust to many naturally occurring L-amino acids with the notable exceptions of the acidic residues (Asp and Glu), the sterically mobile residue Gly, and the imino acid Pro. The fastest turnover was obtained with the substrate containing His at P; and the basic residue Arg also gave significant enhancement. The map containing miscellaneous amino acid replacements at P; showed a similar allowance of a wide variety of amino acids as seen with the naturally occurring L-amino acid mixture (Table I). These results are qualitatively consistent with other studies of the substrate specificity of the human collagenases. Sottrup-Jensen and Birkedal-Hansen (12) profiled the collagenase-mediated hydrolysis of a panel of serum protease inhibitors. These authors showed that the human a$_3$-macroglobulin bait region is readily cleaved by collagenase at a Gly-Leu scissile bond with Arg at the P; position (12). Also, Netzel-Arnett et al. (13) recently reported a study of alternative peptide substrates for collagenase. Again, the enzyme displayed latitude at P; with hydrophobic amino acids and arginine being well tolerated, while Glu substitution led to a sharp reduction in turnover.

Substrate mapping of P; with naturally occurring L-amino acids did not identify any peaks indicative of substantially enhanced substrates versus Leu (Table II). However, turnover of the miscellaneous amino acids substrate mixture at the P; position revealed several substrates with enhanced turnover. Inspection of Table II reveals a marked preference for medium sized nonbranched and uncharged hydrophobic amino acids as substrates. The peptide containing Cys(Me) at P; was found to be the optimal replacement for Leu P; while norvaline and norleucine showed modest increases in turnover.

To verify that Cys(Me) at P; and His at P; enhance turnover, substrates incorporating these individual substitu-

### Table I

| Substrate turnover for various P;'-substituted peptides |
|-----------------------------------------------|
| **Amino acid substitution at P;** | **Relative ($k_{cat}/K_m$)** |
| L-Amino acids |  |
| Lys | 1.4 |
| His | 3.2 |
| Arg | 2.7 |
| Gln | 1.3 |
| Asn | 1.4 |
| Ser | 1.4 |
| Thr | 1.5 |
| Ala | 1.1 |
| Pro | NT* |
| Tyr | 1.9 |
| Val | 1.9 |
| Met | 2.3 |
| Ile | 1.5 |
| Leu | 1.1 |
| Trp | 1.0* |
| Phe | 1.0 |
| Miscellaneous amino acids |  |
| Ornithine and (p-NH$_2$)Phe (co-elution) | 1.5 |
| Sar or 6-Ala | NT |
| Aba* isomer | NT |
| Sar or 6-Ala | NT |
| 8-Aminooctanoic acid | 0.8 |
| Aba* isomer | 0.5 |
| Aba* isomer | NT |
| Aba* isomer | 1.3 |
| Cys(Me) | 1.3 |
| Norvaline | 1.2 |
| (N-Me)Leu and (N-Me)Leu (co-elution) | NT |
| Norleucine | 1.2 |
| Tyr(Me) | 1.0 |
| (p-NH$_2$)-Phe | 1.1 |
| homo-Phe and (N-Me)Phe (co-elution) | 0.2 |
| Tyr(ET) | 0.8 |
| (p-Cl)Phe | 0.6 |
| Cyclohexylalanine | 1.1 |

---

* NT, no turnover; indicates that the rate was less than 10% of the internal standard.

---

| Amino acid substitution at P; | Relative ($k_{cat}/K_m$) |
|-------------------------------|--------------------------|
| L-Amino acids |  |
| Lys | NT* |
| His | NT* |
| Arg | NT* |
| Gln | 0.5 |
| Asn and Ser (co-elution) | NT* |
| Gly and Thr (co-elution) | 0.4 |
| Asp and Glu (co-elution) | NT* |
| Ala | 0.2 |
| Pro | 0.4 |
| Tyr | 0.2 |
| Val | NT* |
| Met | 0.8 |
| Ile | 0.7 |
| Leu | 1.0* |
| Phe | NT* |
| Trp | NT* |
| Miscellaneous amino acids |  |
| Ornithine | 0.2 |
| (p-NH$_2$)-Phe | NT* |
| Sar or 6-Ala and Aba* isomer (co-elution) | 0.2 |
| Sar or 6-Ala | 0.9 |
| Aba* isomer | 0.7 |
| Aba* isomer | NT* |
| Cys(Me) | 3.2 |
| Norvaline | 1.5 |
| 8-Aminooctanoic acid | 0.3 |
| Norleucine | 1.1 |
| Tyr(Me) | 0.8 |
| (N-Me)Leu | NT* |
| (N-Me)Leu | NT* |
| (p-NH$_2$)-Phe | 0.9 |
| homo-Phe or (N-Me)Phe | 0.4 |
| homo-Phe or (N-Me)Phe and Tyr(ET) (co-elution) | 0.4 |
| (p-Cl)Phe | 0.5 |
| Cyclohexylalanine | NT* |

---

* NT, no turnover; indicates that the rate was less than 10% of the internal standard.

---

* All $k_{cat}/K_m$ relative to Dnp-Pro-Leu-Gly-Leu-Trp-Ala-D-Arg-NH$_2$.

* Sar, sarcosine.

* Aba isomers used are aminobutyric acid, aminoisobutyric acid/((N-Me)Ala, and γ-aminobutyric acid.
tions. Dnp-Pro-Leu-Gly-Cys(Me)-Trp-Ala-d-Arg-NH₂ and Dnp-Pro-Leu-Gly-Leu-His-Ala-d-Arg-NH₂ were synthesized and tested in a competition assay with the standard substrate as above. The $k_{cat}/K_m$ values of these individual substrates were 2.9 and 4.8 relative to Dnp-Pro-Leu-Gly-Leu-Trp-Ala-d-Arg-NH₂, in reasonable agreement with the values derived from the initial digests. Cleavage of these modified compounds at the Gly-Leu bond was verified by co-elution of the product, Dnp-Pro-Leu-Gly-OH, with a synthetic standard on reverse phase HPLC analysis. In order to assess the effect of simultaneous modification at $P_1$ and $P_2$, we prepared Dnp-Pro-Leu-Gly-Cys(Me)-His-Ala-d-Arg-NH₂. The double substituted substrate had a $k_{cat}/K_m$ value of 9.7 relative to Dnp-Pro-Leu-Gly-Leu-Trp-Ala-d-Arg-NH₂, representing an order of magnitude increase in turnover.

Genetic approaches have also been described that have the capacity to generate larger numbers of peptide/protein combinations for screening libraries. Reidhaar-Olson and Sauer (14) utilized the technique of combinatorial cassette mutagenesis to map substitutions that allowed retention of functional repressor activity. Recently there have been reports describing the utility of epitope libraries constructed from phage 1USE5 (15) and random peptide libraries from coliphage M13 (16). Unfortunately, the genetic approaches are only able to identify proteins containing the naturally occurring L-amino acids.

The concept of screening peptide mixtures for a specific activity was originally described in the mimotope mapping technique of Geysen et al. (17). Standard solid-phase peptide synthetic techniques have been used to assemble mixtures of peptides (18). These chemical approaches permit the incorporation of unnatural amino acids allowing the preparation of unique synthetic mixtures of protease substrates. The power of this optimization method lies in the combinatorial capacity for individual site mapping. The only perceived limitation of this technique is the inability to assign distinct $k_{cat}$ and $K_m$ values to individual peptides. However, the method offers the significant advantage that relative rates for a large number of substrates can be determined with a single measurement. Incorporation of optimal substitutions can quickly lead to order of magnitude enhancements of substrate turnover as seen above.

A detailed knowledge of the subsite specificity of collagenase is important for the development of synthetic substrates and inhibitors. Modifications which increase specificity (i.e. $k_{cat}/K_m$) may translate into corresponding increases in inhibitor potency (19-22). The results above represent a systematic approach to studying this specificity. In the example described, more than 60 amino acid substitutions were quickly surveyed at the $P_1$ and $P_2$ positions, respectively. The results suggest a significant altitude exists at $P_2$ with $>20$ natural and unnatural amino acids giving rate enhancement over tryptophan (Table I). However, the $P_1$ position has a rather narrow specificity with only 3 amino acids proving better than leucine (Table II).

In conclusion, the method described here is generally applicable for the rapid optimization of any protease substrate $k_{cat}/K_m$ and offers a unique and efficient approach for analysis of substrate specificity. High performance liquid chromatography/continuous flow fast atom bombardment mass spectrometry offers a rapid and reliable way of assigning structure and analyzing substrate hydrolysis in complex mixtures without having to first isolate the individual peptides. The results obtained in terms of the substrate specificity ratio $k_{cat}/K_m$ indicate which peptides in the mixture merit further study. To assure that a given substitution is truly rate-enhancing it is important to verify the structure of enhanced substrates by direct chemical synthesis and enzymatic evaluation to confirm the identity of the cleavage site.

Notes Added in Proof—Recently, several reports have appeared describing uses of synthetic peptide libraries. Birkett et al. (Birkett, A. J., Soler, D. F., Woiz, R. L., Bond, J. S., Wiseman, J., Berman, J., and Harris, R. B. (1991) Anal. Biochem. 196, 137-143) detailed the use of small synthetic peptide protease substrates and sequencing to elucidate protease substrate specificity. Flynn et al. (Flynn, G. C., Pohl, J., Flocco, M. T., and Rothman, J. E. (1991) Nature 353, 726-730) synthesized a known length, undefined sequence library and studied the peptide binding specificity of the heat shock protein family. Lao et al. (Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., and Knapp, R. J. (1991) Nature 354, 82-84) produced a "one-bead, one-peptide" library and reported screening of a β-endorphin monoclonal antibody. Finally, Houghten et al. (Houghten, R. A., Pinilla, C., Blonder, S. E., Appel, J. R., Dooley, C. T., and Coerwo, J. H. (1991) Nature 354, 84-86) described an iterative approach utilizing synthetic peptide combinatorial libraries to screen for monoclonal antibody binding or antimicrobial activity.

REFERENCES

1. Henderson, L. M., Copeland, T. D., Swonder, R. C., Schultz, A. M., and Oroslanz, S. (1988) in Human Retrovirus, Cancer, and AIDS: Approaches to Prevention and Therapy (Bolognesi, D., ed) pp. 135-147, Alan R. Liss, Inc., New York
2. Caprioli, R. M. (1990) Anal. Chem. 62, 477A-485A
3. Caprioli, R. M., Fan, T., and Cottrell, J. S. (1986) Anal. Chem. 58, 2949-2954
4. It, Y., Takeuchi, T., Ichii, D., and Goto, M. (1985) J. Chromatogr. 346, 161-166
5. Stack, M. S., and Gray, R. D. (1989) J. Biol. Chem. 264, 4277-4281
6. Fields, G. B., Van Wart, H. E., and Birkeled-Hansen, H. (1987) J. Biol. Chem. 262, 6221-6226
7. Masui, Y., Takechhi, T., Sakakihara, S., Hori, N., and Nagai, Y. (1977) Biochem. Med. 17, 215-221
8. Rutter, W. J., and Santi, D. V. (November 16, 1989) U. S. Patent WO 89/10931
9. Abeles, R. H., Frisell, W. R., and MacKenzie, C. G. (1960) J. Biol. Chem. 235, 855-856; Correction (1960) J. Biol. Chem. 235, no. 6
10. Fersht, A. (1985) in Enzyme Structure and Mechanism, (Fersht, A., ed) p. 111-112, W. H. Freeman and Co., New York
11. Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T., and Watson, J. T. (1987) Anal. Chem. 59, 2621-2625
12. Netzel-Jensen, L., and Birkeled-Hansen, H. (1989) J. Biol. Chem. 264, 393-401
13. Netzel-Arnett, S., Fields, G. B., Birkeled-Hansen, H., and Van Wart, H. E. (1991) J. Biol. Chem. 266, 6747-6755
14. Reidhaar-Olson, J. F., and Sauer, R. T. (1988) Science 241, 53-57
15. Scott, J. M., and Smith, G. P. (1990) Science 249, 386-390
16. Devlin, J. L., Panganiban, L. C., and Devlin, P. E. (1990) Science 249, 404-406
17. Geysen, H. M., Rodda, S. J., and Mason, T. J. (1986) Mol. Immunol. 23, 709-715
18. Tjeng, F. S., Towery, D. S., Bullock, J. W., Whipple, D. E., Fok, K. F., Williams, M. H., Zupec, M. E., and Adams, S. P. (1990) Int. J. Pept. Protein Res. 35, 141-146
19. Westerik, J. O., and Wolfenden, R. (1972) J. Biol. Chem. 247, 8195-8197
20. Thompson, R. C. (1973) Biochemistry 12, 47-51
21. Thompson, R. C., and Bauer, C. A. (1979) Biochemistry 18, 1552-1558
22. Birkett, P. A., and Marlowe, C. K. (1983) Biochemistry 22, 4618-4624