Identification of Substrate-Analog Trypsin Inhibitors through the Screening of Synthetic Peptide Combinatorial Libraries†

Jutta Eichler and Richard A. Houghten*

Torrey Pines Institute for Molecular Studies, 3550 General Atomics Court, San Diego, California 92121

Received February 16, 1993; Revised Manuscript Received July 27, 1993®

ABSTRACT: Synthetic peptide combinatorial libraries (SPCLs), which are made up in total of tens to hundreds of millions of peptides, enable the systematic screening for biologically active peptides in virtually all *in vitro* and even *in vivo* assay systems. In the current study, the applicability of this method to the identification of peptide enzyme inhibitors was investigated using trypsin as the model enzyme. A specifically designed library of hexapeptide mixtures was synthesized on cotton carriers and screened. The synthetic approach, using cotton as a solid support, was modified so that the deprotected peptides remained attached to the cotton carrier until they were released into solution directly prior to being assayed. Following an iterative process of synthesis and screening, in which all of the positions of the sequence were successively defined, a number of individual hexapeptides with trypsin inhibitory activity were identified. The most active, defined individual peptide sequence was then reincorporated into a new library, now made up of dodecapeptide mixtures. The iterative screening and synthesis of this library led to a dodecapeptide with improved inhibitory activity when compared to the hexapeptide from which it was derived.

Synthetic peptides are essential tools in various areas of biomedical research, as well as for the development of diagnostics and pharmaceuticals. Over the past decade, the need for large numbers of synthetic peptides has lead to the development of synthetic methods able to generate thousands of individual peptides or peptide analogs for large screening programs. A crucial step in accelerating the synthesis efficiency was the development of multiple synthesis techniques. Utilizing the basic concept of solid phase synthesis (Merrifield, 1963), numerous methods for the synthesis of large numbers of peptides using different technologies and carrier materials have been successfully introduced to the methodological pool of peptide chemistry (Geysen et al., 1984; Houghten, 1985; Frank & Döring, 1988; Daniels et al., 1989; Krchňák et al., 1989; Eichler et al., 1991; Fodor et al., 1991). However, the recently developed concepts of systematically synthesizing and screening all possible sequence combinations for peptides of a given length go far beyond the capabilities of even these powerful methods. This fact is well illustrated by the number of possible sequence combinations (64 000 000) for a hexapeptide containing only the 20 genetically coded amino acids. Since individual synthesis of such a large number of peptides currently remains a practical impossibility, several concepts of generating libraries of peptide mixtures, either by chemical synthesis (Furka et al., 1991; Houghten et al., 1991; Lam et al., 1991; Zuckerman et al., 1992) or by recombinant DNA approaches (Scott & Smith, 1990; Devlin et al., 1990; Cull et al., 1991; Felicia et al., 1991), have recently been shown to be feasible to identify a variety of peptide ligands [reviewed by Jung and Beck-Sickinger (1992), Birnbaum and Mosbach (1992), and Houghten et al. (1992)].

Enzyme-inhibiting peptides are among the peptide ligands of potential therapeutic interest. Multicomponent tetrapeptide mixtures have already been used in an earlier study for the identification of HIV protease inhibitors (Owens et al., 1991). Trypsin was chosen for the current study, since numerous polypeptide and protein inhibitors of trypsin from various plant and animal sources are known to range in length from 30 to several hundred amino acids. These inhibitors are highly specific, limited proteolysis substrates for their target enzymes (Laskowski & Kato, 1980). Their amino acid sequence contains at least one peptide bond called the reactive site ($-P_1-P_1'-$) (Ozawa & Laskowski, 1966), which specifically interacts with the active site of the cognate enzyme and is identical to the potential cleavage site. Thus, trypsin inhibitors have lysine or arginine at the P_1 position. For inhibitors, in contrast to normal substrates, the dissociation constant of the enzyme–inhibitor complex is extremely low, resulting in very slow hydrolysis. It is anticipated that short peptides with a potential reactive site in their sequence should be capable of inhibiting this enzyme.

Using peptide libraries already available in our laboratory, represented by the formulas Ac-O_1O_2XXXX-NH_2, consisting of 400 peptide mixtures, and Ac-O_1XXXXX-NH_2, Ac-XO_2-XXXX-NH_2, ... through Ac-XXXXXO_6-NH_2, consisting of 120 peptide mixtures, we have not been able to identify peptide mixtures with a trypsin inhibition of more than 20%.

The hexapeptide library initially screened in this study was designed to represent all possible reactive sites for trypsin inhibitors with the P_1 residue (lysine or arginine) in every position of the sequence except the last. This library was made up of 10 groups of peptide mixtures, represented by the formulas Ac-KOXXXX, Ac-XKOXXX, Ac-XXKOXX, Ac-XXXKOX, Ac-XXXXKO, Ac-ROXXXX, Ac-XROXXX, Ac-XXROXX, Ac-XXXROX, and Ac-XXXXRO. Each of these groups consists of 20 different peptide mixtures representing a total of 2 606 420 individual peptides. In each peptide mixture, four positions (X) represent a mixture of 19 of the 20 genetically coded amino acids (cysteine excluded), whereas O represents those positions occupied by individual amino acids (A through Y). This library was designed by combining the concept of synthetic peptide combinatorial

† This work was funded in part by Houghten Pharmaceuticals Inc., San Diego, CA.

® Abstract published in *Advance ACS Abstracts,* September 15, 1993.

0006-2960/93/0432-11035$04.00/0

libraries (SPCLs)[1] (Houghten et al., 1991), based on successively defining all positions of a peptide mixture by an iterative process of synthesis and screening, with the concept of positional scanning peptide libraries (Pinilla et al., 1992), in which six separate hexapeptide libraries are used to define one position each in a peptide sequence.

MATERIALS AND METHODS

The general use of cotton as a carrier for solid phase peptide synthesis, described earlier (Eichler et al., 1991), was modified as described below for the current study.

Derivatization of the Cotton Carrier

Pretreatment. Twenty discs cut out of commercially available cotton fabric (diameter 4.7 cm) were shaken in 50 mL of 25% trifluoroacetic acid (TFA) in dichloromethane (DCM) for 20 min and then stacked one on top of the other in a flat bottom Büchner funnel of the same diameter as the discs. The funnel was placed on top of a vacuum flask; after removal of the excess 25% TFA/DCM by vacuum filtration, the cotton discs were washed with DCM (3 × 10 mL), 5% diisopropylethylamine (DIEA)/DCM (3 × 10 mL), and DCM (3 × 10 mL). The washings were carried out by soaking the cotton discs in the wash solution, followed by its removal with vacuum filtration. After the final wash, the cotton discs were removed from the Büchner funnel and air dried.

Acylation. Fmoc-Glycine (1.118 g, 4.0 mmol), hydroxybenzotriazole (HOBt, 540 mg, 4.0 mmol), *N*-methylimidazole (NMI, 637 μL, 8.0 mmol), and diisopropylcarbodiimide (DIC, 626 μL, 4.0 mmol) were dissolved in 6.7 mL of dimethylformamide (DMF). The cotton discs were soaked in this solution for 3 h. The cotton discs were then transferred to the Büchner funnel, and washed with DMF (3 × 10 mL) and DCM (2 × 10 mL) as described above. This procedure was repeated a second time, yielding a substitution of 22 μmol/disc.

Capping. The 20 cotton discs were soaked in 8 mL of a mixture of acetic anhydride/NMI/DMF 1:2:3 (v/v/v) for 60 min. The cotton discs were then washed with DMF (3 × 10 mL) and DCM (2 × 10 mL) as described above.

Synthesis of Peptide Mixtures

The general coupling cycle was carried out as follows: (i) Fmoc-deprotection (20% piperidine/DMF, 15 min); (ii) wash (3 × DMF, 2 × DCM); (iii) coupling (0.3 M Fmoc-aa/HOBt/DIC in DMF, 90 min or 2h); (iv) wash (3 × DMF; 2 × DCM). The amino acids were side chain protected as follows: Arg(Pmc); Asn(Trt); Asp(OtBu); Cys(Trt); Gln(Trt); Glu(OtBu); His(Trt); Lys(Boc); Ser(tBu); Thr(tBu); Tyr(tBu).

Fmoc Deprotection. The 20 cotton discs placed into the Büchner funnel were soaked in 10 mL of 20% piperidene/DMF for 15 min. The excess piperidine solution was then removed by vacuum filtration, and the cotton discs were washed with DMF (3 × 10 mL) and DCM (2 × 10 mL) as described above.

Amino Acid Coupling: Coupling of the Same Amino Acid to All Cotton Discs. The Fmoc-amino acid to be coupled (2.4 mmol), HOBt (324 mg, 2.4 mmol), and DIC (376 μL, 2.4 mmol) were dissolved in 7.6 mL of DMF. The deprotected cotton discs were soaked in this solution for 90 min. The excess activated amino acid solution was then removed by vacuum filtration, and the cotton discs were washed with DMF (3 × 10 mL) and DCM (2 × 10 mL) as described above.

Coupling of Different Amino Acids to Each Cotton Disc (O-Coupling). The 20 genetically coded amino acids (A through Y, 0.12 mmol each) were separately dissolved in 0.4 mL of a 0.3 M solution of HOBt and DIC in DMF (324 mg of HOBt and 376 μL of DIC dissolved in 7.6 mL of DMF). The deprotected cotton discs were labeled with the letters corresponding to the amino acids A through Y and soaked in their respective amino acid solutions for 90 min. All of the cotton discs were then placed in the Büchner funnel, the excess activated amino acid solution was removed by vacuum filtration, and the cotton discs were washed with DMF (3 × 10 mL) and DCM (2 × 10 mL) as described above.

Coupling of Amino Acid Mixtures (X-Coupling). A 0.3 M solution of 19 of the 20 genetically coded amino acids (cysteine excluded) was prepared using a previously established molar ratio (see Table I) along with HOBt in DMF. The aliquots (7.6 mL each) were stored at −20 °C. Before coupling, a mixture aliquot was warmed to room temperature and DIC (376 μL, 2.4 mmol) was added. Following a 20-min activation period, the deprotected cotton discs were soaked in this solution for 2 h. The excess activated amino acid solution was then removed by vacuum filtration, and the cotton discs were washed with DMF (3 × 10 mL) and DCM (2 × 10 mL) as described above.

Acetylation of the N-Terminus. After coupling of the last amino acid or amino acid mixture, the cotton discs were Fmoc-deprotected and washed. The deprotected cotton discs were soaked in 8 mL of a mixture of acetic anhydride/pyridine/DMF 1:2:3 (v/v/v) for 60 min. The excess acetylation solution was then removed by vacuum filtration, and the cotton discs were washed with DMF (3 × 10 mL) and DCM (2 × 10 mL) as described above.

Deprotection of Side Chains. The acetylated cotton discs were added to a bottle containing 30 mL of 50% TFA and 5% triisobutylsilane in DCM. After 2 h, the TFA solution was poured off, 100 mL of DCM was added, and the bottle was shaken for 2 min. This wash procedure was repeated twice more with DCM, three times with 5% DIEA/DCM, and finally three times with DCM. The cotton discs were blotted between layers of filter paper and air dried. With a standard office hole puncher, small discs (diameter 7 mm) were cut out, bottled, and kept at 4 °C.

Synthesis of Individual Peptides

The individual defined peptides (termed "no X" peptides) were synthesized on a prototype of a multiple peptide synthesizer using cotton as carrier (Lebl et al., 1992). These syntheses will be reported in detail elsewhere. Briefly, peptides were assembled via standard Fmoc/tBu strategy on an acid-labile linker (*N*-Fmoc-2,4-dimethoxy-4′-(carboxymethyloxy)-benzhydrylamine). After coupling of the last amino acid, the N-terminus was acetylated. Upon treatment with TFA, the peptides were cleaved and isolated as their C-terminal amides.

Trypsin Inhibition Assay

Peptide–Cotton Discs. The 7-mm diameter peptide–cotton discs (0.5 μmol/disc) were placed into a 96-well assay plate, and 50 μL of 0.1 N NaOH was added to each well to cleave the peptides from the cotton carrier. After 15 min, 100 μL

[1] Abbreviations: DCM, dichloromethane; DIC, diisopropylcarbodiimide; DIEA, diisopropylethylamine; DMF, dimethylformamide; Fmoc, 9-fluorenylmethoxycarbonyl; HOBt, hydroxybenzotriazole; NMI, *N*-methylimidazole; SPCL, synthetic peptide combinatorial library; TFA, trifluoroacetic acid; OtBu, *tert*-butyl ester; tBu, *tert*-butyl ether; Trt, trityl; Pmc, pentamethylchromansulphonyl; Boc, *tert*-butyloxycarbonyl; RP-HPLC, reversed-phase high-performance liquid chromatography.

Table I: Molar Ratio for the Components of the Amino Acid Mixture Used for X Couplings

amino acid	molar ratio	relative incorporation[a]		
		average	highest	lowest
A	0.22	1.13	1.35	0.66
D	0.47	1.89	2.59	1.41[b]
E	0.62	1.83	2.26	1.39[c]
F	0.35	0.91	1.39	0.61
G	0.20	1.15	1.38	0.70
H	0.72	1.17	1.42	0.88
I	2.51	0.89	1.22	0.60
K	0.59	0.86	1.28	0.61
L	0.48	1.04	1.22	0.62
M	0.34	1.09	1.30	0.61
N	1.65		see data for D	
P	0.20	1.08	1.30	0.74
Q	2.03		see data for E	
R	1.98	0.97	1.20	0.62
S	0.80	1.00	1.32	0.60
T	2.18	0.84	1.28	0.61
V	1.85	0.94	1.29	0.60
W	0.99		not determined	
Y	0.81	1.02	1.37	0.63

[a] Determined by amino acid analysis of XO dipeptides. [b] Together with N. [c] Together with Q.

of 0.1 M Tris buffer containing 0.025 M $CaCl_2$ (pH 7.8) and 15 μL of trypsin (from bovine pancreas, 8600 units/mg of solid) solution (0.05–0.5 mg/mL 0.02 M HCl) were added. Thorough mixing was ensured by vigorous stirring of the cotton discs in the solutions with the pipette tip. After incubation for 30 min, the chromogenic substrate, N^{α}-benzoyl-DL-arginine-*p*-nitroanilide (100 μL, 2.3 mM in water) was added. An aliquot (150 μL) was taken out of each well and transferred to another well. Absorbances of the transferred solutions were read at 405 nm after 30 min using a Titertek multichannel photometer and expressed as a percentage of the control (Ac-Gly-cotton disc).

Cleaved Individual Peptides. Fifteen microliters of trypsin solution (0.05 mg/mL 0.02 M HCl), 50 μL of peptide solution (1 mM in water), and 100 μL of buffer were mixed in the wells of a 96-well assay plate and incubated for 30 min. Then 100 μL substrate solution was added. The buffer and substrate solutions were the same as described above. After 30 min, absorbances were read at 405 nm and expressed as a percentage of the control (well without peptide). The values obtained with serial dilutions of the peptide solutions were used to determine the IC_{50} values. Both inhibition assays were carried out at room temperature.

RESULTS

The X positions of the peptide mixtures were incorporated by coupling of a mixture of 19 amino acids, instead of the normally coupled single amino acid. The molar ratio of amino acids in the coupling mixture is shown in Table I. This ratio was derived from the representation of each amino acid in dipeptide mixtures of the formula XO (determined by amino acid analysis), in which the X position was incorporated by coupling an equimolar ratio of the 19 amino acids to one of the 20 natural amino acids. Table I also shows the amino acid analysis data of the dipeptide mixtures XO when the recalculated molar ratio was used for the X coupling. The ratio between D and N, as well as E and Q, in the coupling mixture was set according to their coupling rate ratio. Tryptophan, which could not be determined by amino acid analysis, was added to the coupling mixture according to its coupling rate relative to other amino acids. Coupling rates

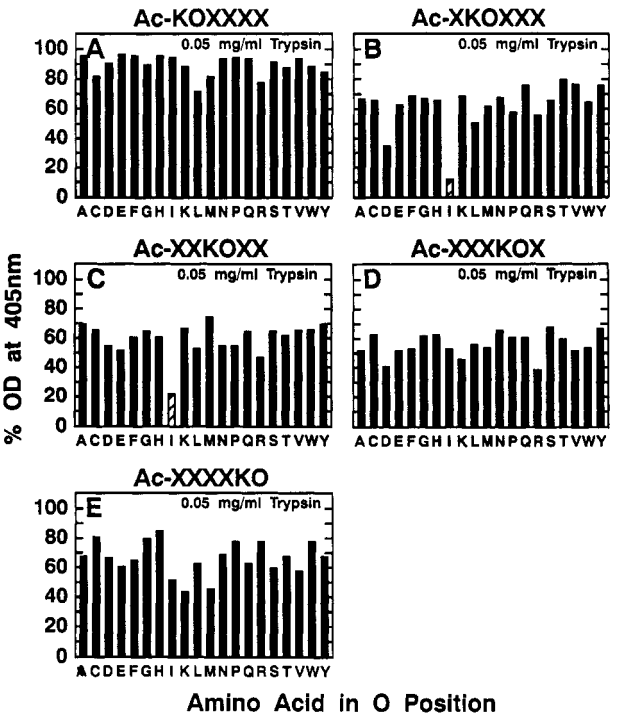

FIGURE 1: Initial screening of the hexapeptide library for trypsin inhibiting peptide mixtures. Each panel represents one of the 10 groups of peptide mixtures making up the library (data for mixtures having RO as defined positions not shown). Each bar represents trypsin inhibition by a peptide mixture defined in the O position with one of the 20 amino acids. Hatched bars indicate peptide mixtures that were chosen to be carried through the iterative process of successively defining of all of the positions of the sequence.

of amino acids (data not shown) were determined by measuring their percent coupling after a 5-min reaction time.

The results of the initial screening of the peptide library are shown in Figure 1. Two peptide mixtures, Ac-XKIXXX and Ac-XXKIXX, were found to strongly inhibit the tryptic hydrolysis of N^{α}-benzoyl-DL-arginine-*p*-nitroanilide. Interestingly, inhibition of greater than 20% (less than 80% OD) was not found for any of the mixtures in which RO made up the defined positions (data not shown). By defining the X positions adjacent to the lysine in Ac-XKIXXX and Ac-XXKIXX, 40 new mixtures with three X positions (Ac-OKIXXX and Ac-XOKIXX) were synthesized and screened. Ac-AKIXXX and Ac-XTKIXX were found to be the strongest trypsin inhibitors (Figure 2A,B). Upon defining the position adjacent to isoleucine, it was found that Ac-AKIYXX and Ac-XTKIFX had the strongest inhibitory activity at the 2X stage (Figure 2C,D), whereas phenylalanine, lysine, arginine and tyrosine were similarily effective at this position for both peptide mixtures. Ac-AKIYRX, Ac-AKIYDX, and Ac-XTKIFT were the mixtures with the strongest inhibitory potency at the 1X stage (Figure 2E,F). Upon defining the final X position, the individual peptide sets Ac-AKIYRO, Ac-AKIYDO, and Ac-OTKIFT were synthesized and screened, both as cotton–peptide discs (Figure 3) as well as isolated peptide amides. Table II shows the IC_{50} values (concentration necessary for 50% inhibition) of these peptides as compared to basic pancreatic trypsin inhibitor (BPTI). The IC_{50} value of BPTI (82 nM), at a trypsin concentration of 193 nM, was found to be in general accordance (18% deviation) with the 1:1 enzyme–inhibitor stoichiometry determined by Kunitz (1947).

In addition, overlapping hexapeptides bearing the reactive sites of the following inhibitors were synthesized and tested:

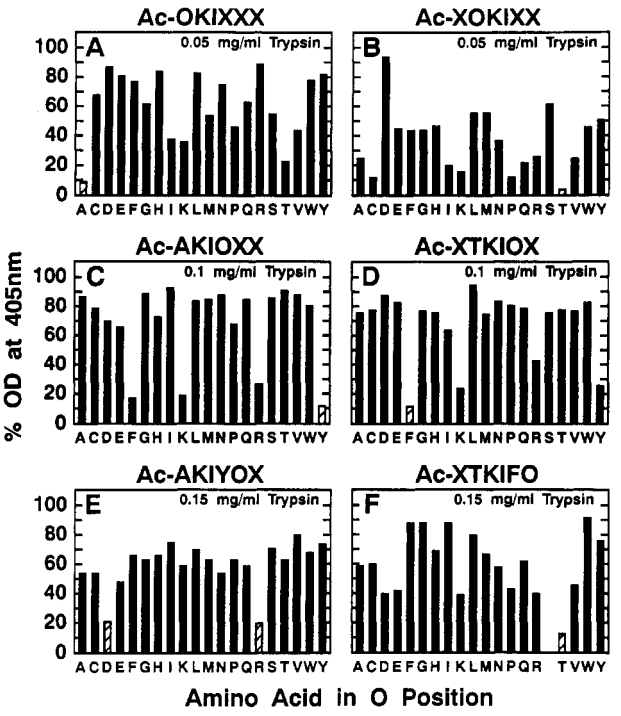

FIGURE 2: Iterative process of defining the positions in the hexapeptide mixtures identified in the initial screening [A, C, and E for Ac-XKIXXX; B, D, and F (Ac-XTKIFS is missing) for Ac-XXKIXX]. Hatched bars indicate peptide mixtures that were chosen to be further defined.

Bowman–Birk inhibitor (Ac-ACTKSN-NH_2, Ac-CTKSNP-NH_2, Ac-TKSNPP-NH_2, reactive site: -KS-); basic pancreatic trypsin inhibitor (Ac-GPCKAR-NH_2, Ac-PCKARI-NH_2, Ac-CKARII-NH_2, reactive site: -KA-); human and porcine pancreatic secretory trypsin inhibitor (Ac-GCTKIY-NH_2, Ac-CTKIYN-NH_2, Ac-TKIYNP-NH_2, Ac-GCPKIY-NH_2, Ac-CPKIYN-NH_2, Ac-PKIYNP-NH_2, reactive site: -KI-); Kazal inhibitor from bovine pancreas (Ac-GCPRIY-NH_2, Ac-CPRIYN-NH_2, Ac-PRIYNP-NH_2, reactive site: -RI-); trypsin inhibitor from corn (Ac-IPGRLP-NH_2, Ac-PGRLPP-NH_2, Ac-GRLPPL-NH_2, reactive site: -RL-). The IC_{50} values of these peptides are shown in Table III.

The hexapeptide with the strongest inhibitory activity (Ac-AKIYRP-NH_2) was chosen to serve as a motif in a library of 20 dodecapeptide mixtures with five X positions: Ac-XXXAKIYRPOXX represents 49 521 980 individual peptides. Following the above described iterative process of synthesis and screening with this new library, it was found that Ac-XXXAKIYRPDXX was the strongest inhibitor at the 5X stage (Figure 4A), Ac-XXGAKIYRPDXX at the 4X stage (Figure 4B), Ac-XXGAKIYRPDKX at the 3X stage (Figure 4C), Ac-XYGAKIYRPDKX at the 2X stage (Figure 4D), Ac-XYGAKIYRPDKM at the 1X stage (Figure 4E), and Ac-YYGAKIYRPDKM at the stage of individual peptides (Figure 4F). Ac-YYGAKIYRPDKM-NH_2 was also the most active among the individually defined isolated peptides. The IC_{50} values of these are shown in Table IV. Table V shows the IC_{50} values for the replacement sets of Ac-YYGAKIYRPDKM-NH_2, which were synthesized by successively substituting each position of the sequence with alanine, leucine, and the corresponding D-amino acids.

DISCUSSION

To date, libraries of synthetic peptide mixtures have primarily been synthesized on polystyrene and polyamide polymeric resins, which are the most commonly used carriers in solid-phase peptide synthesis. In the current study, we synthesized peptide libraries on cotton, which has recently been shown to be a simple and inexpensive alternative carrier for multiple peptide synthesis (Eichler et al., 1991). Cotton is a hydrophilic carrier that is compatible with both organic solvents and aqueous solutions. This allows for the synthesis, as well as the ready release, of synthesized peptide mixtures (or individual peptides) into an aqueous bioassay medium. The synthetic process is thus facilitated by avoiding separate cleavage and extraction steps. This is beneficial because of the possibility of losing components of the peptide mixture during those steps. An important precondition for using this method is the even distribution of modified hydroxyl groups on the cotton. Determination of the substitution of Fmoc-Gly–cotton discs found that there were differences of less than 5% between individual discs as well as on different spots of the same cotton disc.

Since glycine, as used here, is always the first amino acid to be coupled to the cotton, diketopiperazine formation upon Fmoc-deprotection of the second coupled amino acid is expected to occur and may lead to underrepresentation of some amino acids. To determine the extent of this reaction, each Fmoc-amino acid was coupled to a separate Gly–cotton disc of known substitution. After deprotection, Fmoc-isoleucine was coupled to all of the discs, and the substitution was determined again. The difference from the original substitution serves as a measure of the loss due to diketopiperazine formation. These losses were found to be relatively small, with the highest loss found for Gly-Gly (7%).

A central point in synthesizing peptide mixtures is to have as close to an equal representation as possible of all amino acids in the X positions. The DCR method (divide, couple, recombine) (Houghten et al., 1991) or split synthesis procedure (Lam et al., 1991; Furka et al., 1991) ensures equimolar representation of amino acids at each position by the process of coupling each amino acid to separate portions of the carrier resin, followed by subsequent mixing of all resin portions before dividing them again prior to the next coupling step. This method, however, is not as readily carried out for the synthesis of peptide mixtures with defined positions within the sequence, since in this case the resin has to be divided into 400 different portions for every X coupling which follows the coupling of the defined position.

An alternative means to generate peptide mixtures is through the coupling of amino acid mixtures (Geysen et al., 1986; Houghten et al., 1988; Flynn et al., 1991). Blake and Litzi-Davis (1992) generated multicomponent peptide mixtures by separately coupling three different groups of amino acid mixtures in an approximately equimolar ratio. Recently, a positional scanning peptide library composed of six hexapeptide mixtures with one defined and five X positions represented as Ac-OXXXXX-NH_2, Ac-XOXXXX-NH_2, ... through Ac-XXXXXO-NH_2 has been synthesized by coupling mixtures consisting of 18 amino acids for the X positions (Pinilla et al., 1992).

Due to differences in size, steric hindrance of side chains, and protecting groups, the various amino acid derivatives couple at different rates. These differences must be reflected in the ratio of amino acids used for the coupling mixture. The molar ratio of amino acids used in this study was deduced from the incorporation ratio obtained by coupling of an equimolar mixture of amino acids. As shown in Table I, the average incorporation of amino acids was approximately equimolar. The deviations from the average values reflect variations in incorporation of the individual amino acids when

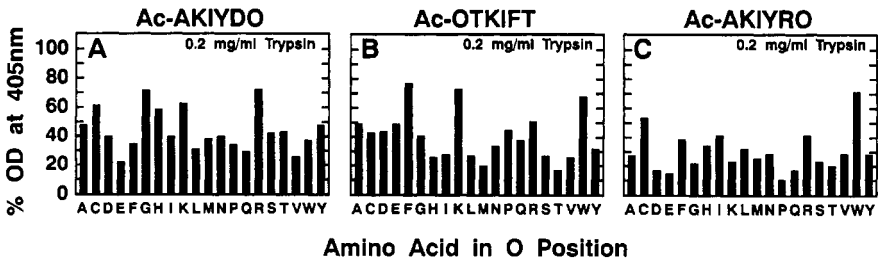

FIGURE 3: Final stage of the iterative process: trypsin inhibition by individual hexapeptides.

Table II: Trypsin Inhibitory Activity of Defined Hexapeptides (IC_{50}, μM)

O position	Ac-AKIYRO-NH_2	Ac-AKIYDO-NH_2	Ac-OTKIFT-NH_2
A	200	270	278
C	205	282	223
D	122	216	263
E	108	144	197
F	220	ND[a]	ND
G	155	321	190
H	242	270	ND
I	200	186	177
K	135	246	355
L	198	228	190
M	232	240	133
N	195	230	205
P	46	202	235
Q	144	180	185
R	242	324	220
S	179	246	210
T	157	228	164
V	230	222	220
W	272	246	290
Y	237	ND	220
BPTI	0.082		

[a] ND, not determined.

Table III: Trypsin Inhibitory Activity of Synthetic Fragments of Naturally Occurring Trypsin Inhibitors

peptide sequence	IC_{50} (μM)
Ac-ACTKSN-NH_2	>600
Ac-CTKSNP-NH_2	>600
Ac-TKSNPP-NH_2	>600
Ac-GPCKAR-NH_2	>600
Ac-PCKARI-NH_2	320
Ac-CKARII-NH_2	290
Ac-GCTKIY-NH_2	148
Ac-CTKIYN-NH_2	140
Ac-TKIYNP-NH_2	102
Ac-GCPKIY-NH_2	335
Ac-CPKIYN-NH_2	170
Ac-PKIYNP-NH_2	102
Ac-GCPRIY-NH_2	>600
Ac-CPRIYN-NH_2	>600
Ac-PRIYNP-NH_2	>600
Ac-IPGRLP-NH_2	>600
Ac-PGRLPP-NH_2	>600
Ac-GRLPPL-NH_2	>600

the mixture was coupled to different individual amino acids. Taking into account these variations, as well as the fact that the ratio of coupling rates between amino acids may vary depending on its position in the sequence, it becomes clear that the ideal of equimolar representation of all components of the peptide mixture cannot readily be obtained by coupling of amino acid mixtures. Recent studies with receptor binding (Dooley & Houghten, 1993) and ELISA (Pinilla et al., 1992) assays have shown, however, that the same peptide ligands were identified screening a peptide library synthesized by the DCR method and another library prepared by coupling mixtures of amino acids.

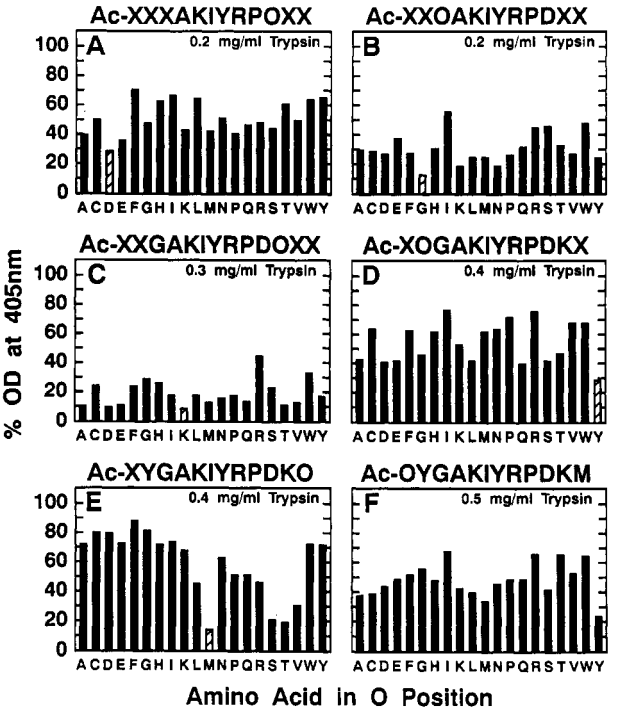

FIGURE 4: Iterative process of defining the positions in the dodecapeptide mixture. Each panel represents one more defined position. Hatched bars indicate peptide mixtures chosen to be further defined.

Table IV: Trypsin Inhibitory Activity of Defined Dodecapeptides

peptide sequence	IC_{50} (μM)
Ac-AYGAKIYRPDKM-NH_2	20
Ac-CYGAKIYRPDKM-NH_2	21
Ac-DYGAKIYRPDKM-NH_2	22
Ac-EYGAKIYRPDKM-NH_2	29
Ac-FYGAKIYRPDKM-NH_2	26
Ac-GYGAKIYRPDKM-NH_2	ND[a]
Ac-HYGAKIYRPDKM-NH_2	24
Ac-IYGAKIYRPDKM-NH_2	27
Ac-KYGAKIYRPDKM-NH_2	26
Ac-LYGAKIYRPDKM-NH_2	24
Ac-MYGAKIYRPDKM-NH_2	22
Ac-NYGAKIYRPDKM-NH_2	24
Ac-PYGAKIYRPDKM-NH_2	28
Ac-QYGAKIYRPDKM-NH_2	27
Ac-RYGAKIYRPDKM-NH_2	24
Ac-SYGAKIYRPDKM-NH_2	20
Ac-TYGAKIYRPDKM-NH_2	28
Ac-VYGAKIYRPDKM-NH_2	22
Ac-WYGAKIYRPDKM-NH_2	20
Ac-YYGAKIYRPDKM-NH_2	10

[a] ND, not determined.

The trypsin inhibition assay described originally by Fritz et al. (1966) was modified for use in a 96-well assay plate. Because all peptides making up the library screened in this study are potential substrates for trypsin, with some of them potentially acting as inhibitors, the incubation time prior to

Table V: Trypsin Inhibitory Activity of the Alanine, Leucine, and D-Amino Acid Replacement Sets of Ac-YYGAKIYRPDKM-NH_2

sequence	IC_{50} (μM)	sequence	IC_{50} (μM)
Ac-YYGAKIYRPDKA-NH_2	30	Ac-YYGAKLYRPDKM-NH_2	111
Ac-YYGAKIYRPDAM-NH_2	17	Ac-YYGALIYRPDKM-NH_2	>600
Ac-YYGAKIYRPAKM-NH_2	18	Ac-YYGLKIYRPDKM-NH_2	27
Ac-YYGAKIYRADKM-NH_2	33	Ac-YYLAKIYRPDKM-NH_2	38
Ac-YYGAKIYAPDKM-NH_2	58	Ac-YLGAKIYRPDKM-NH_2	18
Ac-YYGAKIARPDKM-NH_2	132	Ac-LYGAKIYRPDKM-NH_2	24
Ac-YYGAKAYRPDKM-NH_2	188		
Ac-YYGAAIYRPDKM-NH_2	>600	AcYYGAKIYRPDK**m**-NH_2[a]	33
Ac-YYAAKIYRPDKM-NH_2	94	Ac-YYGAKIYRPD**k**M-NH_2	24
Ac-YAGAKIYRPDKM-NH_2	12	Ac-YYGAKIYRP**d**KM-NH_2	27
Ac-AYGAKIYRPDKM-NH_2	20	Ac-YYGAKIYR**p**DKM-NH_2	12
		Ac-YYGAKIY**r**PDKM-NH_2	48
Ac-YYGAKIYRPDKL-NH_2	31	Ac-YYGAK**I**yRPDKM-NH_2	76
Ac-YYGAKIYRPDLM-NH_2	40	Ac-YYGAK**i**YRPDKM-NH_2	294
Ac-YYGAKIYRPLKM-NH_2	25	Ac-YYGA**k**IYRPDKM-NH_2	>600
Ac-YYGAKIYRLDKM-NH_2	26	Ac-YYG**a**KIYRPDKM-NH_2	204
Ac-YYGAKIYLPDKM-NH_2	47	Ac-Y**y**GAKIYRPDKM-NH_2	35
Ac-YYGAKILRPDKM-NH_2	102	Ac-**y**YGAKIYRPDKM-NH_2	72

[a] Lowercase letters indicate D-amino acids.

the addition of the chromogenic substrate was extended to 30 min in order to avoid mistaking simple substrate competition for inhibition. After this period of time, normal substrates would be expected to be largely degraded, and only slowly hydrolyzed peptides (potential inhibitors), which are still interacting with the enzyme, are detected by decreased hydrolysis rate of the chromogenic substrate.

The synthetic method used here, which avoids separate cleavage and isolation processes, in combination with the multiple assay, greatly facilitates the process of generating and screening peptide libraries. It should be noted that this approach is not limited solely to the described trypsin inhibition assay but can be applied to any bioassay performed in a microtiter plate format.

The peptide mixtures initially found to inhibit trypsin (Ac-XKIXXX and Ac-XXKIXX), both have KI as defined positions within the sequence. These mixtures were carried through the iterative process for the successive defining of all of the X positions. This process revealed that both mixtures represent frame-shifted analogs of a group of related sequence motifs. Upon defining the positions adjacent to lysine, the same amino acids were found to be the most effective in both mixtures, but with different activities. This becomes more obvious upon defining the position adjacent to isoleucine. In both mixtures, the aromatic amino acids phenylalanine and tyrosine, as well as the basic amino acids lysine and arginine, were found to be effective in this position.

During the iterative process, the concentration of the trypsin solution used was increased 5-fold (from 0.05 mg/mL at the 4X stage to 0.2 mg/mL at the stage of defined peptides). Analogous to a decreasing peptide concentration, the increasing trypsin concentration indicates enhancing inhibitory activity as more positions of the sequence are defined. Serial dilutions of the peptides, even though they are released into solution prior to the inhibition assay, are not readily feasable using peptide–cotton discs. Vigorous mechanical agitation (e.g., sonication, vortexing) is usually necessary to not only chemically cleave, but also physically extract a peptide from the synthesis carrier. No inhibitory activity was detected when the cotton discs were removed from the wells after the the NaOH treatment (hydrolysis of the peptide-cotton ester), indicating that the peptides, although chemically cleaved, were still physically attached to the cotton discs. On the other hand, omitting the NaOH step also greatly reduces the inhibitory activity, which indicates that peptides in solution are more able to interact with the enzyme than are immobilized peptides.

Table II and Figure 3 illustrate the trypsin inhibitory activity of defined hexapeptides. It should be noted that in all but one case the peptides having the strongest inhibitory activity in the peptide–cotton disc assay also had the lowest IC_{50} values when tested in solution as isolated peptides. The order was reversed for Ac-MTKIFT and Ac-TTKIFT, but the difference between these two peptides was not significant. Ac-AKIYRP-NH_2 was identified as the hexapeptide with the strongest trypsin inhibitory activity. It is possible, however, that if alternative iterative routes were followed, such as choosing lysine or arginine at the 2X stage, other peptides with similar or even better inhibitory activity could be found.

Interestingly, there is a clear sequential resemblance between the peptide with the strongest inhibitory activity, which was identified by screening the hexapeptide library (Ac-AKIYRP-NH_2), and the peptides with the strongest inhibitory activity among the overlapping synthetic hexamer fragments of naturally occurring trypsin inhibitors (Ac-TKIYNP-NH_2 and Ac-PKIYNP-NH_2). It should be noted, however, that Ac-AKIYRP-NH_2 is more effective than Ac-TKIYNP-NH_2 or Ac-PKIYNP-NH_2. The IC_{50} of Ac-AKIYRP-NH_2 is approximately 500 times higher than that of BPTI, a 58-amino acid polypeptide, which indicates that Ac-AKIYRP-NH_2 is still a much better substrate for trypsin than is BPTI and therefore only acts as a partial, temporary inhibitor. This presumption was experimentally confirmed by RP-HPLC analysis of Ac-AKIYRP-NH_2, which showed that after 4 h the peptide was completely hydrolyzed.

A library of dodecapeptides, Ac-XXXAKIYRPOXX, was synthesized and screened in order to determine if the inhibitory activity of Ac-AKIYRP-NH_2 could be enhanced if it was part of a longer sequence. When the iterative process was carried out with this library (Figure 4), Ac-YYGAKIYR-PDKM-NH_2 was found to have the strongest inhibitory activity (Table IV). This peptide is approximately 5 times more effective than Ac-AKIYRP-NH_2. The individual alanine, leucine, and D-amino acid replacement sets of Ac-YYGA-KIYRPDKM-NH_2 were synthesized and screened in order to identify the importance and specificity of each position in the sequence for trypsin inhibition. The IC_{50} values of these peptides (Table V) indicate that -AKIYR- is the essential motif in the sequence, since the residues of this motif could not be substituted by any of the three amino acids tested

(alanine, leucine, and the corresponding D-amino acid) without dramatic loss in the inhibitory activity (the sole exception being the substitution of alanine by leucine). The residues on either side of the motif were found to be more replaceable and therefore are considered less specific. In ongoing further studies, the identified motif -AKIYR- is being structurally modified in order to improve its trypsin inhibitory activity.

ACKNOWLEDGMENT

We wish to thank Eileen Silva for her assistance in preparing this paper.

REFERENCES

Birnbaum, S., & Mosbach, K. (1992) *Curr. Opin. Biotechnol. 3,* 49–54.

Blake, J., & Litzi-Davis, L. (1992) *Bioconjugate Chem. 3,* 510–513.

Cull, M. G., Miller, J. F., & Schatz, P. J. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89,* 1865–1869.

Daniels, S. B., Bernatowicz, M. S., Cuoll, J. M., & Köster, H. (1989) *Tetrahedron Lett. 30,* 4345–4348.

Devlin, J. J., Panganiban, L. C., & Devlin, P. E. (1990) *Science 249,* 404–406.

Dooley, C. T., & Houghten, R. A. (1993) *Life Sci. 52,* 1509–1517.

Eichler, J., Bienert, M., Stierandová, A., & Lebl, M. (1991) *Peptide Res. 4,* 296–307.

Felicia, F., Castagnoli, L., Musacchio, A., Jappelli, R., & Cesareni, G. (1991) *J. Mol. Biol. 22,* 301–310.

Flynn, G. C., Pohl, J., Flocco, M. T., & Rothman, J. E. (1991) *Nature 353,* 726–730.

Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., & Solas, D. (1991) *Science 251,* 767–773.

Frank, R., & Döring, R. (1988) *Tetrahedron 44,* 6031–6040.

Fritz, H., Hartwich, G., & Werle, E. (1966) *Hoppe-Seyler's Z. Physiol. Chem. 345,* 150.

Furka, A., Sebestyen, F., Asgedom, M., & Dibo, G. (1991) *Int. J. Pept. Protein Res. 37,* 487–493.

Geysen, H. M., Meloen, R. H., & Barteling, S. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81,* 3998–4002.

Geysen, H. M., Rodda, S. J., & Mason, T. J. (1986) *Mol. Immunol. 23,* 709–715.

Houghten, R. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82,* 5131–5135.

Houghten, R. A., Appel, J. R., & Pinilla, C. (1988) in *Peptide Chemistry 1987* (Shiba, T., & Sakakibara, S., Eds.) pp 769–774, Protein Research Foundation, Osaka, Japan.

Houghten, R. A., Appel, J. R., Blondelle, S. E., Cuervo, J. H., Dooley, C. T., & Pinilla, C. (1992) *BioTechniques 13,* 412–421.

Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T., & Cuervo, J. H. (1991) *Nature 354,* 84–86.

Jung, G., & Beck-Sickinger, A. G. (1992) *Angew. Chem. 31,* 367–486.

Krchňák, V., Vágner, J., & Mach, O. (1989) *Int. J. Pept. Protein Res. 33,* 209–213.

Kunitz, M. (1947) *J. Gen. Physiol. 30,* 311–320.

Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., & Knapp, R. J. (1991) *Nature 354,* 82–84.

Laskowski, M., Jr., & Kato, I. (1980) *Annu. Rev. Biochem. 49,* 593–626.

Lebl, M., Stierandová, A., Eichler, J., Pátek, M., Pokorný, V., Jehnička, J., Mudra, P., Ženišek, K., & Kalousek, J. (1992) in *Innovation and Perspectives in Solid Phase Synthesis* (Epton, R., Ed.) pp 251–257, Intercept Ltd., Andover, MD.

Merrifield, R. B. (1963) *J. Am. Chem. Soc. 85,* 2149–2154.

Owens, R. A., Gesellchen, P. D., Houchins, B. J., & DiMarchi, R. D. (1991) *Biochem. Biophys. Res. Commun. 181,* 402–408.

Ozawa, K., & Laskowski, M., Jr. (1966) *J. Biol. Chem. 241,* 3955–3961.

Pinilla, C., Appel, J. R., Blanc, P., & Houghten, R. A. (1992) *BioTechniques 13,* 901–905.

Scott, J. K., & Smith, G. P. (1990) *Science 249,* 386–390.

Zuckerman, R. N., Kerr, J. M., Siani, M. A., Banville, S. C., & Santi, D. V. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89,* 4505–4509.