Vol. 181, No. 1, 1991
November 27, 1991

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
Pages 402-408

THE RAPID IDENTIFICATION OF HIV PROTEASE INHIBITORS THROUGH THE SYNTHESIS AND SCREENING OF DEFINED PEPTIDE MIXTURES

R.A. Owens*, P.D. Gesellchen, B.J. Houchins, and R.D. DiMarchi

Lilly Research Laboratories, Indianapolis, Indiana 46285

Received September 20, 1991

Small peptides with activity as enzyme inhibitors, hormone antagonists, or other peptide mimetics, can be identified by synthesizing and screening large numbers of peptides as defined mixtures. Several coupling reactions, each with a different amino acid, can be conducted simultaneously and then combined to generate a near equimolar mixture before coupling additional residues. The peptide mixtures are recovered free from the matrix and in quantities sufficient for screening in many assays. We describe the rapid identification of a potent peptide inhibitor of human immunodeficiency virus protease from twenty-two mixtures containing more than 240,000 tetrapeptides.
© 1991 Academic Press, Inc.

Small molecules that possess high binding affinity and specificity for enzymes or cellular receptors are attractive therapeutic agents. Suitable, orally bioavailable, small molecular weight surrogates for natural macromolecules have proven difficult to obtain through a chemical strategy directed at systematically altering the structure of the natural ligand (1). Where enzyme or receptor structure is known, *de novo* design of high affinity ligands represents an alternative and ever-improving approach (2). Random screening, of compounds with no known structural similarity to the natural ligand, is a low probability approach that has however proven successful (3). Historically, the latter compounds have come from either a large accumulation of previously prepared organic structures or a collection of natural products derived from fermentation broths. Recently, there has been renewed interest in the development of techniques which enhance this selection process, and thereby increase the probability of success (4-8).

Since the initial development of solid-phase peptide synthesis (9), several approaches have been developed which allow peptides to be synthesized

*To whom correspondence should be addressed.

Abbreviations: Bzl, Benzyl, Boc, tert-Butoxycarbonyl, Bom, Benzyloxymethyl, BSA, bovine serum albumin, DIC, 1,3-diisopropylcarbodiimide, DMF, N,N-dimethyformamide, HF, hydrofluoric acid, HIV, human immunodeficiency virus, HOBT, 1-hydroxybenzotriazole hydrate, MBHA, *p*-methylbenzyhydrylamine, MES, 2-[N-Morpholino]ethanesulfonic acid, Z, benzyloxycarbonyl, FITC, Fluoresceinisothiocyanate.

0006-291X/91 $1.50
Copyright © 1991 by Academic Press, Inc.
All rights of reproduction in any form reserved.

simultaneously (4,5,10-13). The most notable for screening purposes is the solid-phase peptide synthesis of individual sequences on polyethylene pins (4). This allows the integration of synthesis and screening to select a sequence which mimics the natural ligand in its binding to a specific protein. The evaluation of synthetic peptide ligands is performed in the presence of the support used in synthesis. While this can facilitate the physical aspects of binding analysis, it restricts the type of biological assays that can be employed. Of additional concern is the contribution that the matrix makes to the determination of macromolecular binding.

Recently, recombinant DNA technology has been applied to generate large peptide libraries of random amino acid sequences (6-8). The ability to rapidly amplify a selected peptide by growth of the host microorganism is a definite advantage of the biosynthetic methodologies. However, these methods have their limitations, which most notably includes the restriction to natural amino acids and a linear ordering of the amino acid sequence.

To increase the speed of synthesis and analysis and to generate a large library of free peptides, we elected to make peptides as defined mixtures. A series of small acetylated tetrapeptide amides were synthesized where only one position was unique and the remaining positions contained a near equimolar mixture of different amino acids. The total number of peptides in a mixture was dependent upon the length of the peptide and the total number of different amino acids selected for synthesis. We report the rapid identification of a potent peptide-based inhibitor for HIV protease,a potential target of antiviral therapy (14), from mixtures consisting of nearly a quarter of a million compounds.

MATERIALS AND METHOD

Peptide synthesis t-Boc-strategy with DIC/HOBT activated esters was employed using a RaMPS apparatus (Dupont) and a MBHA resin (Advanced Chem Tech) to generate C-terminally amidated peptides. Boc-amino acids used in syntheses were D-Ala, Asp(OBzl), D-Glu(OBzl), Phe, D-Phe, His(Bom), Ile, Lys(2Cl-Z) , D-Lys(2Cl-Z), D-Leu, Met, D-Asn, Pro, D-Pro, Gln, Arg(Tos), Ser(Bzl), D-Val, Trp, D-Trp, Tyr(2Br-Z), D-Tyr(2,6Cl-Z), and Statine (Advanced Chem Tech).

The polypropylene reaction vessels were filled with 0.2 gm of resin. A solution of 10% diisopropyl ethyl amine in DMF was used to neutralize the alpha-amine. Removal of the t-Boc protecting group was achieved in trifluroacetic acid/methylene chloride/anisole (50:45:5) for 30 min. Amino acid coupling was achieved by *in situ* activation of 4 equivalents of t-Boc-amino acid and HOBt with an equal amount of DIC in DMF. The reaction was rocked end-over-end for 60 min and then washed, neutralized, and subjected to recoupling with the same residue. Mixtures were generated by combining the resins from a round of synthesis, mixing, drying overnight, and redividing into 0.2 gm aliquots for addition of the next residue. A single 4 hour acetylation with10 equivalents of acetic acid, HOBt, and DIC was used as the terminal step of synthesis.

HF cleavage The peptide amides were removed from the resin and side-chain deprotected in HF/*p*-thiolcresol/*m*-cresol (90:5:5) at 0°C for 60 minutes using a multiple peptide cleavage apparatus (Multiple Peptide Systems). The HF was vacuum distilled and the peptides were precipitated with 50 mL of diethylether. After collection by filtration, the peptides were washed extensively with diethylether, solubilized in aqueous 50%

acetic acid containing 30% acetonitrile, concentrated by distillation and then subsequently lyophilized. The recovered peptides were weighed and redissolved in aqueous acetonitrile/acetic acid to prepare lyophilized aliquots.

Amino acid analysis Amino acid analysis was performed on a Beckman 330. The peptide resins were hydrolyzed with refluxing for 21 hours in conc HCl/acetic aced/phenol (2:1:1) and similarly peptides were hydrolyzed 21 hours in 6N HCl. Thioglycolic acid was added when Trp was a known amino acid.

HIV protease assay HIV-1 protease inhibitory activity was measured against a soluble substrate of Biotin-Gly-Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gly-Lys(FITC)-OH (15). In microwell titer plates, HIV-1 protease (16) and putative inhibitors are mixed for 60 minutes in MES-BSA buffer (0.05 M MES, 0.02 M NaCl, 0.002 M dithiothreitol, BSA 2.0 mg/ml, pH 5.5) prior to incubation with substrate for 12 hours. Proteolysis solution is transferred to a 96 well Fluoricon Assay Plate containing Fluoricon Avidin Assay Particles in TBSA (0.02 M Tris, 0.15 M NaCl, 1.0 mg/mL bovine serum albumin, pH 7.5). Unbound fluorescence was removed by filtration and sample detection was performed by excitation at 485 nm with reading of the resulting epifluorescence at 535 nm.

RESULTS AND DISCUSSION

Twenty-two amino acids, representing a variety of functional side chains and stereochemistry were selected as starting blocks of synthesis. In addition, Statine, a known "transition-state" analog for aspartic acid proteinases (17), was included in the protocol (at the second residue only) to increase the likelihood that an inhibitor of reasonable potency would be present in the mixtures. Tetrapeptides were chosen as the size for synthesis as it is highly desirable to restrict the total molecular size of a potential enzyme inhibitor to facilitate subsequent molecular modeling. Tetrapeptide mixtures, with only one defined residue per mixture and all combinations of the twenty-two amino acids in the remaining positions would represent a large synthetic diversity such that any single peptide would represent less than 0.01% of the total mixture.

Each of the twenty-two amino acids were individually and exhaustively coupled to a (MBHA) resin. The resins were combined to yield a mixture consisting of an approximately equimolar concentrations of each amino acid. Mixture positions are denoted as $\mathbf{X_n}$, where **X** represents a mixture of twenty-two amino acids and **n** represents the residue position in the order of coupling to the resin.

A test tripeptide, Ac-Thr-Gly-$\mathbf{X_1}$-NH_2, was prepared to evaluate whether this approach would yield the expected amino acid content. Amino acid analysis of the tripeptide resin and the tripeptide mixture (Table 1) suggests that any particular peptide may be present at a relative concentration that varies no more than two-fold from theoretical. The differences reflects inaccuracies in preparation of the $\mathbf{X_1}$**-MBHA** resin and inefficiencies in recovery following cleavage and processing. There is an observable bias for the more hydrophilic peptides to be elevated in relative concentration.

As the first step in the process, twenty-two separate Boc-amino acid-MBHA resins were prepared. The resins were then combined to give Boc-$\mathbf{X_1}$**-MBHA**. Twenty-three unique Boc-amino acid-$\mathbf{X_1}$-MBHA dipeptide resins were then prepared utilizing the 22

Table 1. Amino acid analysis of the tripeptide mixture: Ac-Thr-Gly-X_1-NH_2

RESIDUE	THEORY	PEPTIDE	RESIN
Gly	22	22.0	22.0
Thr	22	19.8	21.4
Asp	2	2.2	2.4
Glu	2	2.3	2.4
Ser	1	1.4	1.6
Pro	2	1.7	1.8
Ala	1	1.6	1.5
Val	1	1.1	0.9
Met	1	1.0	1.2
Ile	1	0.8	0.7
Leu	1	1.0	0.6
Tyr	2	1.1	1.2
Phe	2	1.6	1.5
His	1	1.9*	1.2
Lys	2	1.9	2.2
Arg	1	0.4+	1.3
Trp	2	1.6	1.7

Levels have been normalized to Glycine at 2291 nmol/mg resin or 2291 nmol/mg peptide.
*Low level from incomplete removal of Tos . +Arg(Tos) coelutes with His.

Boc-amino acids from step one and Boc-statine. All twenty-three of these dipeptide resins were combined to generate a new mixed resin designated Boc-**Z_2-X_1-MBHA**, where **Z** is equal to **X** plus statine. In similar fashion the twenty-two Boc-amino acids were again individually coupled and combined to generate a new mixed tripeptide resin designated Boc-**X_3-Z_2-X_1-MBHA**. Finally, an additional sequence of coupling was performed to yield twenty-two separate tetrapeptide resins designated Boc-**D_4-X_3-Z_2-X_1-MBHA**, where **D_4** represents a defined single amino acid. Each of the peptide resin mixtures was acetylated and designated as Ac-**D_4-X_3-Z_2-X_1-MBHA**. The peptide-resin mixtures were then cleaved from the resin with liquid HF, and subsequently lyophilized to give 22 separate mixtures each of which contained 11,132 different tetrapeptides amides. A calculated total of greater than 244,904 acetylated tetrapeptide amides were represented in these 22 mixtures.

Each of the twenty-two peptide mixtures was evaluated for an ability to inhibit HIV protease. Where appreciable activity was identified, synthesis of a new set of tetrapeptides was performed wherein the **X_3** position was then defined. This constituted preparation of 22 new tetrapeptide mixtures each of which possessed 506 distinct peptides, designated Ac-**K_4-D_3-Z_2-X_1**-NH_2 (where **K_4** represents the known amino acid and **D_3** represents one of the twenty-two amino acids). This process of combining synthesis with screening was repeated until the sequence of a preferred tetrapeptide inhibitor from one or more of the mixtures was identified.

The starting twenty-two tetrapeptide mixtures (Ac-**D_4-X_3-Z_2-X_1-NH_2**) were initially screened for HIV inhibitory activity at a total peptide concentration of 15 mM. Table 2 (column 2) shows the degree of inhibition observed for each mixture. A structure-

Table 2. Percent inhibition of HIV protease by acetylated-tetrapeptide amides

DEFINED(D)	D_4-X_3-Z_2-X_1	Phe-D_3-Z_2-X_1	Phe-Ile-D_2-X_1	Phe-Ile-Sta-D_1
Phe	103	7	16	0
D-Phe	57	23	15	43
Tyr	98	8	4	0
D-Tyr	68	5	4	0
Trp	81	9	1	0
D-Trp	65	6	7	19
Met	96	5	19	0
Ile	102	68	21	0
D-Val	30	28	15	1
D-Leu	51	0	18	68
Pro	9	0	21	0
Gln	58	0	18	0
D-Asn	27	0	14	0
Ser	17	0	18	0
D-Ala	11	0	6	5
Asp	117	0	15	0
D-Glu	78	0	15	9
Arg	17	0	8	8
His	8	0	21	0
Lys	29	0	16	0
D-Lys	5	0	0	0
Sta	---	---	83	---

Inhibition is calculated from the retention of fluorescence with inhibitor compared to the incubation mixture without inhibitors.

activity relationship is evident. The preference in the defined (**D_4**) residue is for an aromatic or hydrophobic residue of L stereochemistry. There was also a strong response to an acidic residue in this position. A dose response analysis for the Ac-Phe-**X_3-Z_2-X_1**-NH_2 tetrapeptide mixture yielded an IC_{50} of 4.4 mM. Assuming one unique bioactive peptide among the 11,132 peptides present in this mixture one can calculate a predicted IC_{50} of 400 nM for that peptide. If there are multiple potent sequences then the final IC_{50} will be somewhere between 400 nM and 4.4 mM.

Based upon this first set of results, two sets of twenty-two tetrapeptide mixtures were synthesized with the third position (**D_3**) defined and the fourth position set as either aspartic acid (Ac-Asp-**D_3-Z_2-X_1**-NH_2) or phenylalanine (Ac-Phe-**D_3-Z_2-X_1**-NH_2). The resulting 44 mixtures were assayed at a total peptide concentration of 506 uM. The aspartic acid series displayed no specific preference for position 3 (data not shown) and was not evaluated any further. However, with the phenylalanine series, isoleucine was found to be preferred in the third position (Table 2, column 3). A dose response analysis of this mixture yielded an IC_{50} of 112 uM. Again one can predict an IC_{50} for one unique tetrapeptide within this series(221 nM).

Next, the second position of the tetrapeptide was defined (**D_2**) while phenylalanine and isoleucine were held constant at the first two residues. The 23 tetrapeptide mixtures (Ac-Phe-Ile-**D_2-X_1**-NH2) were assayed at 220 uM total peptide concentration for their ability to inhibit the HIV protease. As predicted, statine was found to be the residue

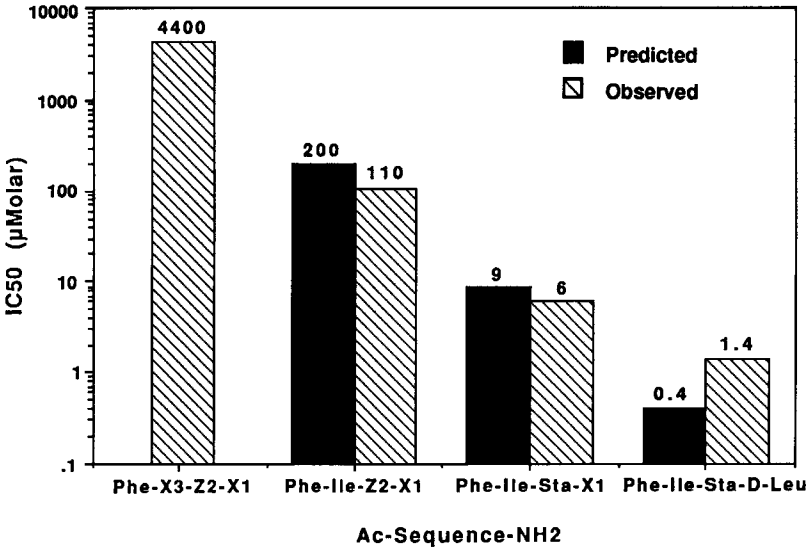

Figure 1. Comparison of the predicted and observed potencies as each position of Ac-Phe-$\mathbf{X_3}$-$\mathbf{Z_2}$-$\mathbf{X_1}$-NH_2 was defined. Predicted potencies were calculated by dividing the observed IC_{50} of Ac-Phe-$\mathbf{X_3}$-$\mathbf{Z_2}$-$\mathbf{X_1}$-NH_2 by the number of residues in each of the mixed positions ($\mathbf{X_3}$=22; $\mathbf{Z_2}$=23; $\mathbf{X_1}$=22). Observed potencies were measured in the assay on the most active mixture from each round of synthesis.

contributing most significantly to the observed inhibition. The IC_{50} (6 uM) measured for the mixture of twenty-two tetrapeptides generates a calculated final IC_{50} of 273 nM.

The last series of syntheses yielded 22 defined tetrapeptides, Ac-Phe-Ile-Sta-$\mathbf{D_1}$-NH_2, which were tested at 10 uM each. Only two residues (D-leucine and D-phenylalanine) were observed to possess significant activity (Table 2, column 5). The IC_{50} for the D-leucine tetrapeptide was determined to be 1.4 uM. The preference of a D-leucine at the C-terminal position of the peptide was entirely unpredicted by any published structure-activity relationship of HIV protease inhibitors. In Figure 1, a comparison of the projected and the observed inhibition using the IC_{50} for the initial phenylalanine based tetrapeptide mixture is shown. A single inhibitor possessing an IC_{50} only 3.5-fold greater than originally predicted was rapidly identified starting from a mixture containing more than 11,000 peptides.

The most potent tetrapeptide identified (Ac-Phe-Ile-Sta-D-Leu-NH_2) served as a base for further structural analysis. Amino acid residues not present in the original mixture and others that displayed appreciable activity at intermediate points of analysis were examined as individual changes. Conversion of the fourth residue from D-Leu to D-Ile, Leu, or Val increased the IC_{50} of the parent tetrapeptide in each instance more than tenfold. The identical substitutions for Ile at the second residue yielded decreases in potency for D-Ile and Leu, and a twofold increase in potency for Val. Further analysis of the N-terminal preference in the series, where the second residue was Val, identified Ac-Trp-Val-Sta-D-Leu-NH_2 as an inhibitor with an IC_{50} of 200 nM. This represents a 7-fold increase in potency from the original compound.

The importance of peptide length was also examined. By increasing the length of Ac-Trp-Val-Sta-X_1-NH_2 to Ac-Trp-Val-Sta-Val-X_1-NH_2 the IC50 decreased from 810 nM to 50 nM. Within this series a defined pentapeptide was observed to possess an IC50 of 5 nM. This represents a 40-fold increase in potency from the starting tetrapeptide (Ac-Trp-Val-Sta-D-Leu-NH_2), or a 280-fold increase from the originally identified inhibitor (Ac-Phe-Ile-Sta-D-Leu-NH_2). Further synthetic modifications of these peptide inhibitors are currently in progress.

The method is also amenable to the identification of small peptides which mimic natural peptide receptor agonists. We have recently identified the peptide CCK-4 (Trp-Met-Asp-Phe-NH_2) through use of this methodology (18). One additional modification of this synthetic process which we are investigating involves the use of a diamine-based building block to prepare branched chain mixtures. This approach may prove of particular importance in the identification of novel small peptides which mimic natural peptide receptor agonists.

ACKNOWLEDGMENTS

We thank Dave Carrender, Joe Manetta, and Theresa Gygi for their assistance in evaluating the peptides in the HIV Protease assay, and Bob Rothenberger and Charlie Campbell for their technical expertise in peptide synthesis.

REFERENCES

1. Freidinger, R.M. Anderson, P.S., Bock, B.G., Chang, R.S.L., DiPardo, R.M., Evans, B.E., Garsky, V.M., Lotti, V.J., Rittle, K.E., Veber, D.F., Whitter, W.L. (1988) in *Peptides, Chemistry and Biology: Proceedings of the Tenth American Peptide Symposium*, G. Marshall, Ed. (ESCOM, Leiden,) 97-100.
2. Perchonock, C.D.,Uzinskas, I., McCarthy, M.E., Erhard, K.F., Gleason, J.G., Wasserman, M.A., Muccitelli, R.M., DeVan, J.F., Tucker, S.S., Vickery, L.M., Kirchner, T., Weichman, B.M., Mong, S. Scott, M.O., Chi-Rosso, G., Wu, H-L., Crooke, S.T., Newton, F.F. (1986) *J. Med. Chem.* **29**, 1442 -1452.
3. Evans, B.E., Bock, M.G., Rittle, K.E., DiPardo, R.M., Whitter, W.L., Veber, D.F., Anderson, P.S., Freidinger, R.M. (1986) *PNAS* **83**, 4918.
4. Geysen, H. M., Meloen, R. H., and Bareling, S. J. (1984) *PNAS* **81**, 3998-4002.
5. Houghten, R. A. (1985) *PNAS* **82**, 5131-5135.
6. Scott ,J. K. and Smith, G. P. (1990) *Science* **249**, 386-390;
7. Devlin, J. J., Panganiban, L. C., Devlin, P. E. (1990) *Science* **249**, 404-406;
8. Cwirla, S.E., Peters, E. A., Barrett, R. W., Dower, W. J. (1990) *PNAS* **87**, 6378-6382.
9. Merrifield, R.B. (1963) *J. Am. Chem. Soc.* **85**, 2149-2154.
10. Tjoeng, F. S., Towery, D.S., Bulock, J.W., Wipple, D.E., Fok, K.F., Williams, M.H., Zupec, M.E., Adams, S.P. (1990) *Int. J. Peptide Protein Res.* **35**, 141-146;
11. Hudson, D. (1988) *J. Org. Chem.* **53**, 617-624;
12. Frank, R. and Doring ,R. (1988) *Tetrahedron,* **44**, 6031-6040,.
13. Furka, A., Sebestyen, J., Asgedom, M., Dibo, D. (1991) *Int. J. Peptide Protein Res.* **37**, 487-493.
14. Mitsuya, H., Yarchoan, R., Broder, S. (1990) *Science* **249**,1533-1544
15. Manetta, J.V., Lai, M-H. T., Osborne, H.E., Dee, A., Margolin, N., Sportsman, J.R., Vlahos, C.J., Yan, S.B., Heath, W.F., in press.
16. Margolin, N., Heath, W. F., Osborne, E., Lai, M., Vlahos C. (1991) *Bich. Biph. Res. Comm.*, **167**, 554-560;
17. Marciniszyn, J.Jr., Hartsuck, J. A., Tang, J. A. (1976) *JBC* **25**, 7088.
18. Owens, R.A., DiMarchi, R.D., Houchins, B.J., Gesellchen, P.D., manuscript in preparation.