BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **227,** 125–130 (1996)
ARTICLE NO. 1477

Naturally Occurring Protease Inhibitors Potent against the Human Immunodeficiency Virus

David Noever

Faculty of Pharmaceutical Sciences, Protease Inhibitor Laboratories, Inc., Box 4, Belle Mina, Alabama 35615

Received July 29, 1996

Inhibition of protease enzymes can render the human immunodeficiency virus (HIV) non-infectious in vitro. To enhance bioavailability and pharmacokinetic activity, 86 new blocking agents against HIV-1 protease are derived by screening genome sequences from naturally occuring enzymes. The new agents are rank-ordered according to their chemical distance from a known set of HIV-protease inhibitors; the scoring methods have previously demonstrated 92% success in classifying a given amino acid sequence prior to testing for antiviral potency. The work has: 1) generalized the empirical work on HIV-PR to more than double the number of published peptides for blocking PR activity; 2) rank-ordered the inhibitors according to their chemical distance from the consensus sequence; 3) identified at least 28 gut enzymes with known bioavailability (>10%) in vivo; 4) classified the family groupings of protease inhibitors in a hierarchical tree. Compared to the library of best known peptides, 19 of the natural sequences are closer to the consensus library than existing inhibitors.

Active HIV proteases are required to condense a viral protein core in subsequent generations (1-4) and therefoe protease inhibitors serve as principal targets for disease regulation (5). The present effort evaluates natural products to derive candidate protease inhibitors. Natural HIV protease inhibitors, in particular, are known to render the virus non-infectious (6), for example as derived from fragments of insulin, actin and autolysis proteins; similarly a number of anti-carcinogenic protease inhibitors have strong potency (nM concentrations) (7-8). Natural enzymes have known survivability in gut conditions (7), specific enzymatic activity and pharmological profiles with low toxicity, high selectivity and potential oral availability.

HIV-PR inhibitors were generalized from the chemical proximity to a known database (6, 12). The classification was able to find short peptide sequences in a matrix of selected natural compounds.

MATERIALS AND METHODS

Selection of chemicals. Short oligonucleotides were identified from the amino acid sequence of the HIV genomes and natural enzyme sequences. Representative compounds for proteases and their inhibitors were selected from several published databases (5-6, 9-10) and their amino acid sequences split into potency categories based on their inhibitory activity in previously reported PR inhibitory assays. The screened sequences included: HIV-PR, lysozyme, HIV1-GAG, HIV1-POL, HIV1-ENV, HIV2-GAG, HIV2-POL, HIV2-Q,HIV2X, HIV2R, HIV2-TAT, HIV2-ART, HIV2-F, papain, stem bromelain, fruit bromelain, ficin, chymopapain, streptopapain, pepsin, plasmin, phosphoglucosolipids, and ribonuclease. These compounds were selected based on their antiviral scores in the HIV-PR assay, with fragments of lysozyme (which does not inhibit HIV-PR) acting as negative control sequences.

Protease inhibitor assay. For performing the assay, homogeneous prepartions of HIV-1 protease were obtained and all peptides were prepared by solid phase peptide synthesis, as described elsewhere (9). All inhibitors were characterized by amino acid composition and reversed-phase HPLC analysis and were pure by these criteria. Kinetic analyses were performed in total volume 250 mL consisting of the enzyme HIV-PR (0.1547 mg), substrate HIV-1 (60 mM), buffer (NH 4HCO3 0.4 mg), DMSO (0.8%) and octapeptide inhibitor candidates (16, 32,48,64,80 mM). The solution (pH 4.7) was incubated at 37° C for 3 min. Kinetic constants (K_m and % inhibition) were determined using the ENZFITTER program (Elsevier-BIOSOFT, Cambridge. UK). The estimated error for all determinations was <20%.

Scoring the database. Substitution positions are shown in Fig. 1 and include the single-letter amino acid representa-

0006-291X/96 $18.00

ANLAEEA	0.00
AEELAEIF	0.00
PFIFEEEP	0.00
ARVLAEAM	0.00
ARVLFEAL	0.00
ARVLFEAL	0.00
ARVLPEAL	0.00
DTVLEEMS	0.00
GQVNYEEF	0.00
SLNLRETN	0.00
DQILIEIC	0.00
TQIMFETF	0.00
DDLFFEAD	0.00
PIVGAETF	0.00
PTLLTEAP	0.00
ARNYPEAL	0.00
ELEFPEGG	0.00
GDALLERN	0.00
DAINTEFK	0.00
SFNFPQJT	0.00
SFNFPQJT	0.00
ARVLFQAL	0.00
VEVAEEEE	0.00
SFIGMESA	0.00
SQNYPIV	0.00
LPVNGEFS	0.00
GSHLVEAL	0.00
SCNFPQJT	0.00
AECFRIRD	0.00
SYNFPQJT	0.00
SFTFPQJT	0.00
YFNFPQJT	0.00
AETFYVDK	0.00
ARVLFIAL	0.00
SQNFPIVQ	0.00
ARVLFVAL	0.00
SQNFPIVQ	0.00
REAFRVFD	0.00
TLNFPISP	0.00
TFNFPQJT	0.00
YEEFVQMM	0.00
AQTFYVNL	0.00
SFYFPQJT	0.00
ARVLFAAL	0.00
ARVLFTAL	0.00
ARVLFNAL	0.00
ARVLFDAL	0.00
PGNFLQSR	0.00
FRSGVETT	0.00
KLVFFAE	0.00
ATIMMQRG	0.00
QITLWQRP	0.00
SFNYPQJT	0.00
SQNYPIEQ	0.00
SQNYPIVP	0.00
SFNYPQDT	0.00
SFNYPQVT	0.00
SQNYPIIQ	0.00
RKILFLDG	0.00
SRSLYASS	0.00
AQNYPIVL	0.00
SFNYPQJI	0.00
SQNYPIVL	0.00
SQNYPIVQ	0.00
SQNYPVVQ	0.00
SFNYPLIT	0.00
KELYPLTS	0.00
SQNYPVVQ	0.00
SFNYLQJT	0.00
SQNYPIVQ	0.00
SQNYPLVQ	0.00
SQNYAIVQ	1.00
SFNSPQJT	0.00
SQNYAIVQ	0.00
AEAMSQVT	0.00
SQNYPIVE	0.00
RQNYPIVQ	1.00
SFNYGQJT	0.00
GGVYATRS	0.00
SQNYPFVQ	0.00
SQNYPNVQ	0.00
SFNYPPIT	0.00
SQNYPAVQ	0.00
SQNYPDVQ	1.00
TQNYPIVQ	0.00
SQNYPKVQ	1.00
SQNPPIVQ	1.00
SQNYPKVQ	1.00
SQNYTIVQ	0.00
SQNYPGVQ	1.00
SQNSPIVQ	1.00
SQNYLIVQ	0.00
FESNFNTN	1.00
RQANFLGQ	0.00
SQNYDIVQ	1.00
SQQYPIVQ	1.00
SQNYKIVQ	1.00
SKNYPIVQ	0.00
SNNYPIVQ	0.00
AAMLRHGL	1.00
SQKYPIVQ	1.00
HLVEALYL	0.00
AYVAYRNR	1.00
NATNRNTD	1.00
ATNRNTDG	1.00
ETTALVCD	0.00
ASVNCAKI	1.00
LDNYRGYS	1.00
SALLSSDI	1.00
NLCNIPCS	1.00
AAAMLRHG	1.00
LGNYVCAA	1.00
QJNSRYYC	1.00
ALLSSDIT	1.00
GSTDYGIL	1.00
STDYGILQ	1.00
INSRYYCD	1.00
NAYVAYRN	1.00
RNLCNIPC	1.00
DNGRTPGS	1.00
VFGRCQLA	1.00
DNYRGYSL	1.00
LFESNFNT	1.00
NGRTPGSR	1.00
RCKGTDVQ	1.00
TASVNCAK	1.00
KVFGRCQL	1.00
GNYVCAAL	1.00
VSDGDGMN	1.00
CKGTDVQA	1.00
IVSDGDGM	1.00

FIG. 1. Database of experimentally derived HIV-1- protease inhibitors rank-order according to their respective chemical distance from the aligned consensus sequence. Rank ordering shows the chemical distance which splits the sequence list according to experimentally determined inhibitors (denoted with 0.0) and non-inhibitors (denoted with 1.0). Non-inhibitors are lysozyme fragments and the chemical distance is a related metric derived by previous workers (refn.(5,6)). The remaining letters in each row represent the single-letter code for each site substitution on the peptide chain.

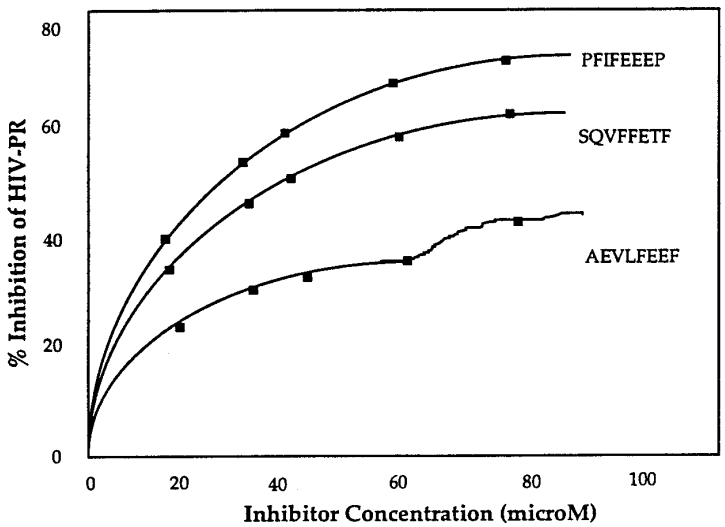

FIG. 2. Experimentally determined activity (K_i) for HIV-protease inhibitors for database (Fig. 1. PFIFEEP) and two de novo inhibitors (SQVFFETF) and (AEVLFEEF). Activity is percent inhibition and (not shown) linear on log rate scale.

tion (i.e A=alanine..,etc). These 20 substitutions were arranged in any of 8 enzyme blocking sites. For predicting inhibition, the classification has partitioned the database of PR inhibitors with 92% success (6,12). As collected in Fig. 2, the data on HIV-PR inhibition has been ranked according to its antiviral potency and chemical distance from an average vector corresponding to the library of known sequences. The corresponding negative control sequences (which do not show inhibitory activity against HIV-PR) are selected from lysozyme and are shown in Fig. 2.

RESULTS

Three test sequences (one control and two de novo octapeptides) were found to inhibit HIV-PR activity with IC_{50} in the μM range.

On the HIV-1 genome (Fig. 3), two short-chain peptides (HIV-PR 33-42; DTVLEEMNL and HIV-PR 63-71; DQIPVEICG)) show probabilities for self-inhibition of protease. The remainder of the HIV genome which shows inhibitory potential for the protease appears elsewhere in the sequence: HIV-GAG; HIV-POL; HIV-ENV. Similarly, short segments of the HIV-2 genome also show significant inhibition for HIV-1PR, but appear in GAG, POL, Q, R, F, and TAT sections. In particular, the relatively shorter portion (229 residues) of the HIV-2-F genome provides 10 peptides with potential inhibition of the HIV-1 infectious spread in vitro. Although this F-segment represents less than 5% of the 4000 or so genomic length for HIV-2, it contains more than 30% of the inhibitory peptides. The availability of these peptides for blocking protease, can be supplied either synthetically or otherwise as part of the viral lifecycle.

Digestive enzymes showing HIV-PR inhibition included segments of papain, stem bromelain, fruit bromelain, ficin, chymopapain, streptopapain, pepsin, plasmin, phosphoglucosolipids, and ribonuclease. The best candidates were found in plasmin and hirudin, with pepsin, papain, stem bromelain and ficin showing similarly strong inhibitory potential. Papain, an inhibitor of cysteine proteases, shows a comparable influence for inhibiting the aspartyl HIV protease.

DISCUSSION

The most widely used AIDS therapeutic currently prescribed is 3′-azidothymidine (AZT, Zidovudine) which shows a 100 times lower potency (mM concentrations for inhibition) against

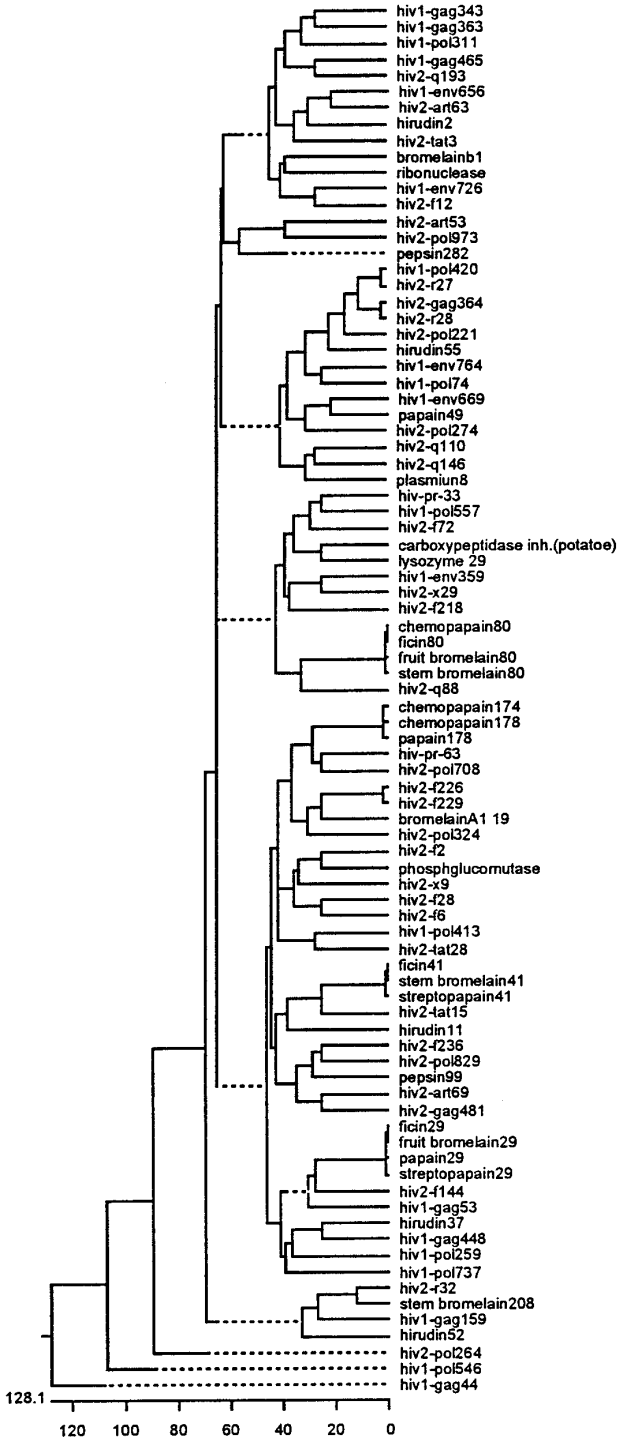

FIG. 3. Phylogenetic tree using clustal method with PAM250 residue weight table for the HIV-PR library sequences and screened natural enzyme sources. Standard nomentclature for source of the enzyme and small numbers label the sequence site of cleavage for initiating the octapeptide reading frame (1–8 is left to right, with HIV-PR inhibition between sites 4–5).

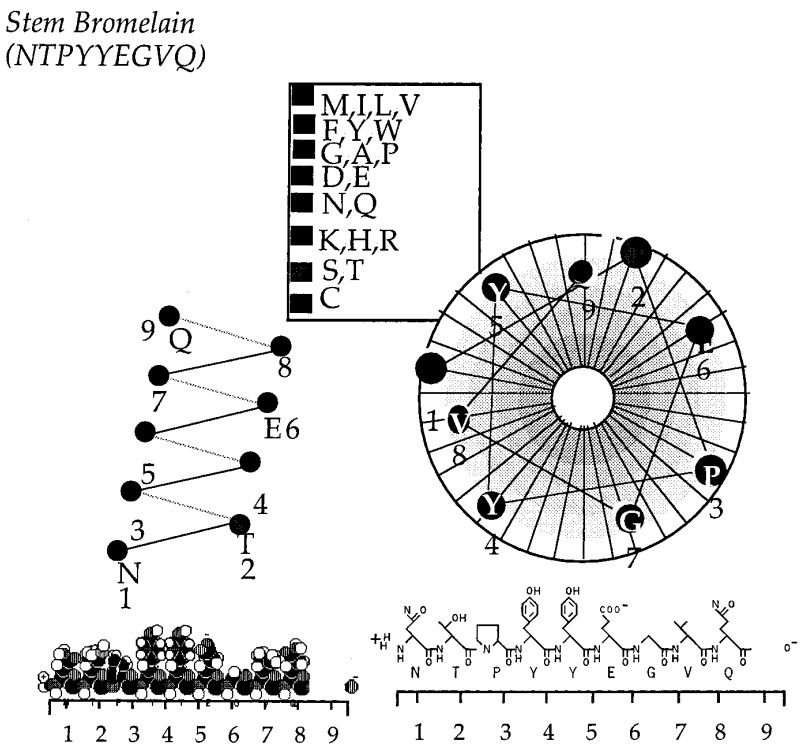

FIG. 4. Example analysis for stem bromelain sequence as a naturally occuring and predicted HIV-protease inhibitor. Shown are the helical wheel, beta net and both its linear and chemical formula (calculated with DNA Star).

HIV-PR than the naturally occurring and test peptides (12). Because of side effects for AZT including bone marrow toxicity and severe peripheral neuropathy (15), the use of natural enzymes with equal or better potency, but less toxicity even at large doses, encourages further work on database development and testing. In this regard, vegetable-derived protease inhibitors (7-8) have demonstrated epidemiological significance in reducing the incidence of a number of other degenerative diseases such as cancer (colon, breast, prostate, oral and pharyngeal).

None of the 86 proposed peptides have previously appeared in HIV-PR inhibiting trials. A phylogenetic tree (Fig. 3) shows the relationship between the families of derived octapeptides. The minimum number of nodes tracing back to the origin is 13, so an average cluster can be defined with 6 to 7 members each. Among the 86 octapeptides, the two best scoring natural sequences both appear in the same general cluster. In contrast, the majority of natural compounds cluster in subclasses that remain distinct from the HIV genome segments. The sequence proximity between hirudin (2-10), HIV-1-ENV (656-664), HIV-2-ART (63-71) suggests possible overlap in function. The presence of blood enzymes like hirudin on the de novo list provides a starting structure for development of highly bioactive peptoids (16) and peptidomimetics. Hirudin, a polypeptide isolated from the leech (Hirudo medicinalis) contains 65 amino acid residues and serves as the starting structure for a number of analogues capable of inhibiting thrombin (17). With an inhibition constant K=20 fM, hirudin is the strongest known thrombin inhibitor, thus controlling various disease states like thrombosis, artherosclerosis, myocardial infarction and blood vessel contraction or dilation. Similarly ficin (41-49), stem bromelain (41-49) and streptopapain (41-49), all appear to overlap phylogenetically with HIV-1-GAG (448-456). Fig. 4 details the predicted secondary structure for stem bromelain as mainly turn

and coil regions with a highly hydrophilic nature. Both its antigenicity and surface probability are concentrated between enzymatic binding sites 4 and 1.

Compared to known peptide inhibitors, 19 of the natural sequences are predicted to behave better for antiviral activity, based on chemical proximity to the consensus library. In conclusion, promisingly high plasma levels for such natural compounds may make them an appealing addition to investigate the site requirements of the HIV-PR binding site, to target protease inhibition and to render non-infectious viral blockage.

REFERENCES

1. Sathyanarayana, B. K., and Wlodawer, A. (1993) *Curr. Sci.* **65,** 835–847.
2. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayan, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B. H. (1989) *Science* **245,** 616–621.
3. Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) *Nature* **337,** 617–620.
4. Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Scolnick, E. M., and Sigal, I. S. (1988) *Proc. Natl. Acad. Sci.* **85,** 4686–4690.
5. Poorman, R. A., Tomasselli, A. G., Heinrikson, R. L., and Kezdy, F. J. (1991) *J. Biol. Chem.* **266,** 14554–14561.
6. Chou, J. J. (1993) *Biopolymers* **33,** 1405–1414.
7. Rackis, J. J., Wolf, W. J., and Baker, E. C. (1986) Protease inhibitors in plant foods:Content and inactivation, *Adv. Exper. Med. Biol.* **199,** 299–347.
8. Kennedy, A. (1993) Protease Inhibitors as Cancer Chemopreventive Agents (Troll, W., and Kennedy, A., Eds.), p. 9, Plenum Press, New York.
9. Tommasselli, A. G., Hui, J. O., Adams, L., Lowery, D., Greenberg, B., Yem, A., Deibel, M. R., Zurcher, Neely, H., and Heinrickson, R. L. (1991) *J. Biol. Chem.* **266,** 14548–14553.
10. Henderson, L. E., Benveniste, R. E., Sowder, R. C., Copeland, T. D., Schultz, A. M., and Oroszlan, S. (1988) *J. Virol.* **62,** 2587–2595.
11. Noever, D. A., and Baskaran, S. Proc. Intl. Conf. Artificial Neural Nets and Genetic Algorithms, April 26, 1995, Ales France, 223.
12. King, E. (1994) *Scientific Computing and Automation* **5,** 35–41.
13. Holland, J. H. (1975) Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI.
14. Forrester, S. (1993) Genetic Algorithms: Principles of Natural Selection Applied to Computation, *Science* **261,** 872–878.
15. Miller, M., Schneider, J., Jaskolski, M., Sathyanarayan, Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., and Wlodawer, A. (1989) *Science* **246,** 1149–1152.
16. Gante, J. (1994) *Angew. Chem. Int. Ed. Engl.* **33,** 1699–1720.
17. Maraganore, J. M., Bourdon, P., Jablonski, J., Ramachandran, K. L., and Fenton, J. W. (1990) *Biochemistry* **29,** 7095.