Pergamon

Bioorganic & Medicinal Chemistry Letters 8 (1998) 3203–3206

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

EVALUATION OF A STRUCTURE-BASED STATINE CYCLIC DIAMINO AMIDE ENCODED COMBINATORIAL LIBRARY AGAINST PLASMEPSIN II AND CATHEPSIN D

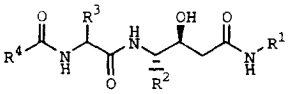
Carolyn DiIanni Carroll,*[a] Theodore O. Johnson,[b] Shewei Tao,[b] Giorgio Lauri,[b] Marc Orlowski,[a] Ilya Y. Gluzman,[c] Daniel E. Goldberg,[c] and Roland E. Dolle[b]

From the Departments of [a]Biology and [b]Chemistry, Pharmacopeia, Inc., Princeton, NJ 08543, U.S.A. and the [c]Howard Hughes Medical Institute, Departments of Medicine and of Molecular Microbiology, Washington University, St. Louis, MO 63110, U.S.A.

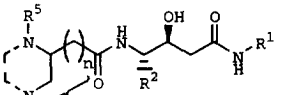
Received 18 August 1998; accepted 2 October 1998

Abstract. A structure-based 18,900-member combinatorial library was synthesized containing a statine template and three cyclic diamino acids as potential P_1', P_2-P_4 surrogates. Evaluation of this encoded library against two aspartyl proteases, plasmepsin II and cathepsin D, led to the identification of selective inhibitors for each enzyme.

Aspartyl proteases are involved in many biological pathways in humans, parasites, plants, fungi, and retroviruses.[1] New members of the protease superfamily continue to be identified, a recent example is the malarial protease plasmepsin II.[2] We recently described the utility of an encoded 13,020 member statine dipeptide library (PL 792) as a general screening tool for rapidly establishing P_2 specificity preferences for this class of protease.[3] This library was screened against human liver cathepsin D and the malarial protease, plasmepsin II. The assay results revealed a broad P_2 specificity for cathepsin D and a rather strict requirement for β-branched amino acids at the P_2 position for plasmepsin II. Here we describe the synthesis and evaluation of another encoded 18,900 member statine library (PL 799) containing cyclic diamino acids. In contrast to our first library, the design of PL 799 scaffold was based on a docking exercise[4] using the X-ray crystal structure of a cathepsin D-pepstatin complex,[5] which indicated that the functionalized piperazine acetic acid scaffold is sterically compatible with the active site. One substituent, either the R^4 or R^5 may bind along the S_3-S_4 subsites and the other substituent may be oriented towards the S_1' subsite, which was not explored by our first statine library.[3] Thus piperazine acetic acid and the two structurally related cyclic amino acids (R^3, Table 1) were incorporated into the library as well as commercially available statines (R^2). The inclusion of 15 carboxylic acid or carboxaldehydes (R^4 synthons) and twenty acids (R^5 synthons) in library PL 799 provided diversity at the putative P_1', P_3, and P_4 positions. This resulted in a 18,900 member library (Scheme 1). While our initial comparison of the cyclic diamino scaffolds 1 and 2 found that the (S)-piperazine acid could attain the closest overlap with bound pepstatin, in subsequent docking experiments either isomer could be preferred depending on the choice of R^4 and R^5 sidechains.



PL 792
13,020 members
(reference 3)

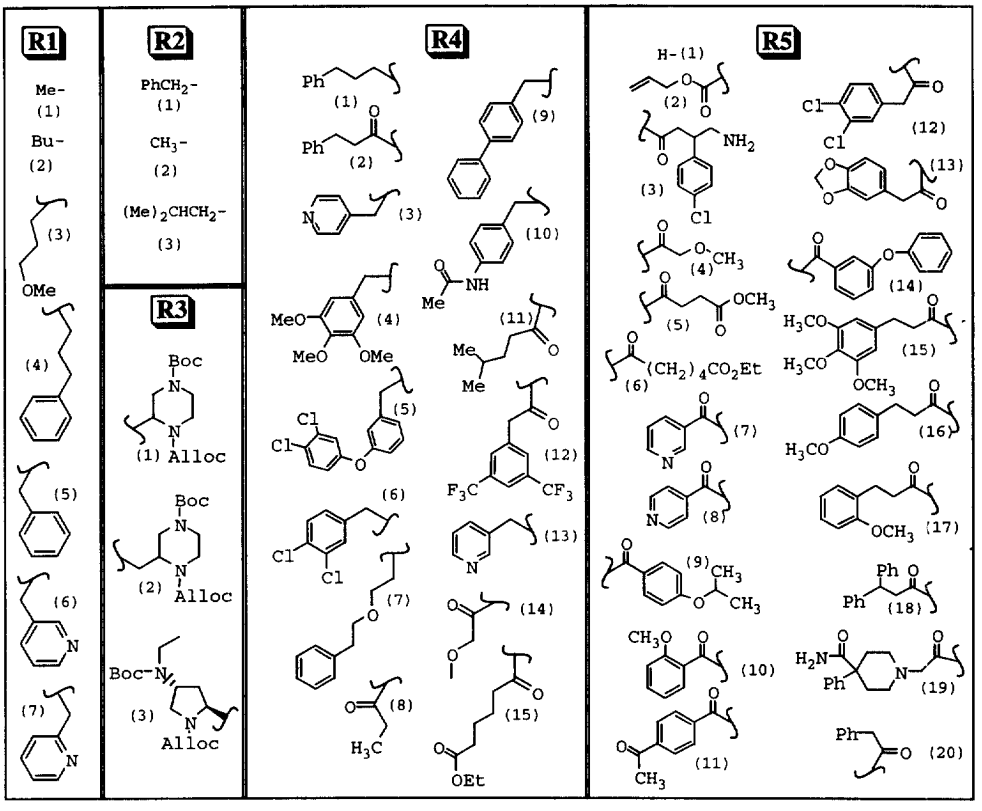


PL 799
18,900 members
(this work)

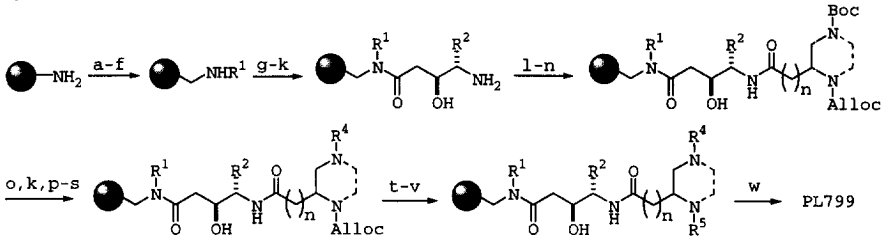
Library PL 799 was evaluated against cathepsin D and plasmepsin II.[6] Structures for decoding were selected on the basis of < 50% activity remaining in the assay with a screening concentration of ~5 μM. A total of 69 structures were decoded for cathepsin D while 37 structures were decoded for plasmepsin II. The

PII: S0960-894X(98)00554-X

Table 1. Synthons for PL 799. (Substituent number shown in parenthesis.)



Scheme 1. Synthesis of the Library PL799.[a]



[a]*Reagents and conditions:* (a) TentaGel™ resin (S-NH_2, 0.3 mmol/g) apportioned into 15 reaction vessels; (b) 3 equiv each Bis-Fmoc-Lys, HOBT, 5 equiv DIC, CH_2Cl_2; (c) encoded using three tags as per ref 7; (d) 30% piperidine-DMF, 1 h; (e) 5 equiv each 4-bromomethyl-3-nitrobenzoic acid, HOBT, 8 equiv DIC, CH_2Cl_2, 3 h; (f) one of seven R^1NH_2 as per Table 1: 10 equiv amine, THF, 8 h; (g) pool and split into three reaction vessels; (h) one of three R^2-Boc-protected statines as per Table 1: 4 equiv each statine , HATU, 8 equiv iPr_2EtN, DMF, 3 h; (i) encoded using two tags as per ref 7; (j) pool; (k) 50% TFA-CH_2Cl_2, 1 h; (l) split into three reaction vessels; (m) one of three R^3-Boc-Alloc-protected cyclic diamino acids as per Table 1: 4 equiv each diamino acid, HATU, 8 equiv iPr_2EtN, 6 h; (n) encoded using three tags as per ref 7; (o) pool; (p) split into fifteen reaction vessels; (q) one of fifteen R^4 aldehydes or carboxylic acids (R^4 replaces the Boc-group on R^3) as per Table 1: 30 equiv R^4CHO, 20 equiv $NaBH_3CN$, 2% HOAc-DMF 8 h, or 4 equiv each R^4CO_2H, HATU, 8 equiv iPr_2EtN, 6 h; (r) encoded using five tags as per ref 7; (s) pool; (t) 0.5 equiv $(Ph_3)_4Pd$, 10 equiv Bu_3SnH, CH_2Cl_2, 1 h; (u) split into twenty reaction vessels; (v) one of twenty R^5 carboxylic acids (R^5 replaces the Alloc-group on R^3) as per Table 1: 4 equiv each R^5CO_2H, HATU, 8 equiv iPr_2EtN, 6 h; (w) *hv* (365 nm), MeOH, 3 h.

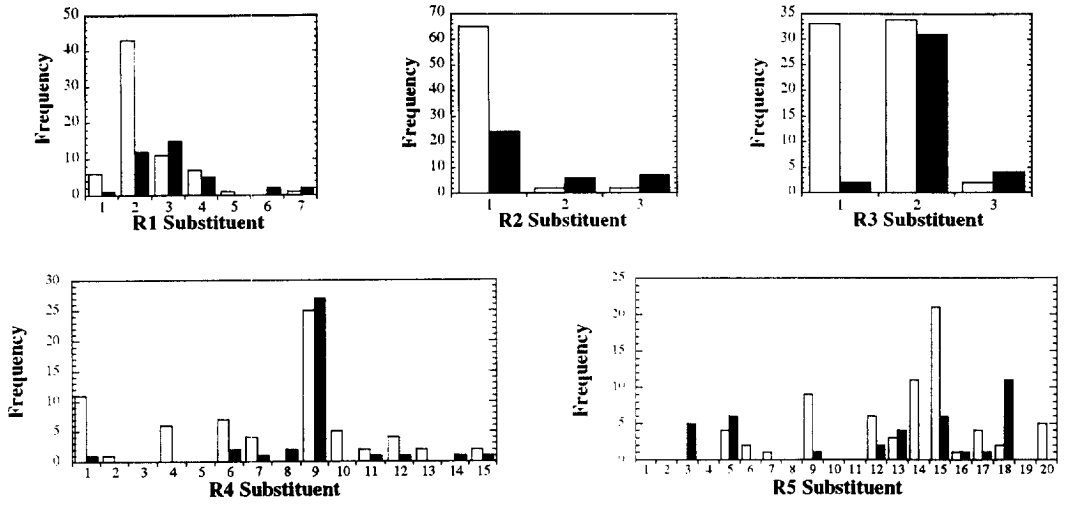


Figure 1. Substituent frequency. Frequency that each substituent was observed in the R^1, R^2, R^3, R^4 and R^5 positions of inhibitory compounds for plasmepsin II (dark bars) and cathepsin D (open bars). The structure for each substituent number is listed in Table 1.

frequency at which each of the five variable positions was observed in the inhibiting compounds was analyzed (Figure 1). Both enzymes favored alkyl amides at the R^1 position and phenylalanine-statines at the R^2 position. This later observation is consistent with the known precedent for large hydrophobic P_1 residues for the two enzymes.[8] These identical substituents in the R^2 position were also included in the library PL 792.[3] There, both phenylalanine- and leucine-statine were observed in compounds inhibitory for plasmepsin II and cathepsin D. Therefore, the frequency that substituents were observed in the R^2 position between the 2 libraries, PL 799 and PL 792, is different indicating that the substituents in other positions can influence or affect the substituent under evaluation.

There were 3 cyclic diamino acids synthons in the R^3 position. Cathepsin D demonstrated an equal preference for the piperazine containing substituents (substituents 1 and 2) versus the proline substituent (substituent 3). In contrast, plasmepsin II preferred only substituent 2, where substituents 1 and 2 only differ by a single methylene group. A variety of substituents were observed for each enzyme in the R^4 and R^5 positions (Figure 1), with the enzymes favoring large hydrophobic arylalkyl substituents.

In order to confirm and evaluate the inhibitory compounds observed in this encoded library, compounds must be resynthesized and purified. Table 2 contains the 4 compounds synthesized from PL 799 and their corresponding K_i values for plasmepsin II and cathepsin D. Compound PS 172564 was synthesized because it was the most potent and selective inhibitor observed during screening for plasmepsin II. This compound is a competitive inhibitor of plasmepsin II with a K_i value of 490 nM and exhibits ~90-fold selectivity for plasmepsin II versus cathepsin D.[9] PS 349374 and PS 990762 were synthesized because they were the most potent inhibitors of cathepsin D observed during screening and contained the selective piperazine substituent in

Table 2. Analysis of Resynthesized Compounds

Compound[b]	Substituent[a] R^1	R^2	R^3	R^4	R^5	Plas II K_i[c] (nM)	Cat D K_i[c] (nM)
PS 172564	2	1	2	9	3	490	~45,000[d]
PS 349374	2	1	1	7	15	~65,000[d]	1800
PS 681550	2	1	2	4	9	~40,000[d]	5300
PS 990762	2	1	1	15	15	~100,000[d]	1100

[a]refers to substituents listed in Table 1; [b]mixture of diastereomers; [c]K_i values are the average of at least 2 determinations; [d]IC_{50} value determined and are typically within 2-fold of K_i values.

the R^3 position. These compounds inhibited cathepsin D competitively with K_i values from 1 to 2 μM while much higher concentrations are necessary to inhibit plasmepsin II demonstrating that these compounds are ~35- to 90-fold selective for cathepsin D. Compound PS 681550 was synthesized containing the nonselective piperazine substituent in the R^3 position and cathepsin D-selective R^4 and R^5 substituents. This compound has a K_i value of 5 μM for cathepsin D and was only 8-fold selective for cathepsin D suggesting that piperazine substituent is an important selectivity determinant which is consistent with the frequency data in Figure 1.

In summary, an 18,900 member statine encoded combinatorial library was synthesized incorporating cyclic diamino acids potentially accessing the S_1', S_3, and S_4 enzyme binding sites. Screening of this library with plasmepsin II and cathepsin D led to the identification of selective inhibitors for both aspartyl proteases. Selectivity was not only dependent on the R^4 and R^5 substituent but also on the cyclic diamino acid in the R^3 position. With the identification of these inhibitors from this library, X-ray crystallographic structural studies can now be performed to confirm if these inhibitors are binding as the modeling exercise has predicted.

Acknowledgments. The work at Washington University was supported by UNDP/ World Bank/ WHO Special Programme for Research and Training in Tropical Diseases (TDR). DEG is a recipient of a Burroughs Wellcome Fund Scholar Award in Molecular Parasitology.

References & Notes.

1. Abdel-Meguid, S. S. *Med. Res. Rev.* **1993**, *13*, 731.
2. Francis, S. E.; Gluzman, I. Y.; Oksman, A.; Knickerbocker, A.; Mueller, B. M. L.; Sherman, D. R.; Russell, D. G.; Goldberg, D. E. *EMBO J.* **1994**, *13*, 306.
3. Carroll, C. D.; Patel, H.; Johnson, T. O.; Guo, T.; Orlowski, M.; He, Z.-M.; Cavallaro, C. L.; Guo, J.; Oksman, A.; Gluzman, I. Y.; Connelly, J.; Chelsky, D.; Goldberg, D. E.; Dolle, R. E. *Bioorg. & Med. Chem. Lett.* **1998**, *8*, 2315.
4. Computational docking studies were performed using MacroModel BatchMin 5.0 and the cathepsin D X-ray structure, 1LYB ref 5.
5. Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Hosur, M. V.; Sowder II, R. C.; Cachau, R. E.; Collins, J.; Silvia, A. M.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6796.
6. The plasmepsin assays were performed in 50 mM sodium acetate (pH 5.0), 1 mg/ml BSA, 0.01% Tween 20, 12.5% glycerol, 9% DMSO and 6 μM substrate (DABCYL-γ-aminobutyric acid-Glu-Arg-Met-Phe-Leu-Ser-Phe-Pro-EDANS). The cathepsin D assays were performed in 25 mM sodium formate (pH 3.5), 1 mg/ml BSA, 6% DMSO and 6 μM substrate (Ac-Glu-Glu(EDANS)-Lys-Pro-Ile-Met-Phe-Phe-Arg-Leu-Gly-Lys(DABCYL)-Glu-NH_2.
7. Baldwin, J. J.; Burbaum, J.J.; Henderson, I.; Ohlmeyer, M. H. J. *J. Am. Chem. Soc.* **1995**, *117*, 5588.
8. Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Smith, C. E.; Duffin, K. L.; Goldberg, D. E. *J. Clin. Invest.* **1994**, *93*, 1602 and Rich, D. H.; Agarwal, N. S. *J. Med. Chem.* **1986**, *22*, 2519.
9. Compound PS 172564 had an IC_{50} value of 3 μM against *Plasmodium falciparum* in cell culture. (Desjardin, R. E.; Canfield, R. J.; Haynes, J. D.; Chulay, J. D.; *Antimicrob. Agents Chemother.* **1979**, *16*, 710.)