Pergamon

Bioorganic & Medicinal Chemistry Letters 8 (1998) 2315–2320

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

IDENTIFICATION OF POTENT INHIBITORS OF *PLASMODIUM FALCIPARUM* PLASMEPSIN II FROM AN ENCODED STATINE COMBINATORIAL LIBRARY

Carolyn DiIanni Carroll,*[a] Hitesh Patel,[b] Theodore O. Johnson,[b] Tao Guo,[b] Marc Orlowski,[a] Zhen-Min He,[b] Cullen L. Cavallaro,[b] Joan Guo,[c] Anna Oksman,[d] Ilya Y. Gluzman,[d] James Connelly,[c] Daniel Chelsky,[a] Daniel E. Goldberg,[d] and Roland E. Dolle[b]

From the Departments of [a]Biology, [b]Chemistry, and [c]Analytical Chemistry, Pharmacopeia, Inc., Princeton, NJ 08540, U.S.A. and the [d]Howard Hughes Medical Institute, Departments of Medicine and of Molecular Microbiology, Washington University, St. Louis, MO 63110, U.S.A.

Received 31 March 1998; accepted 28 May 1998

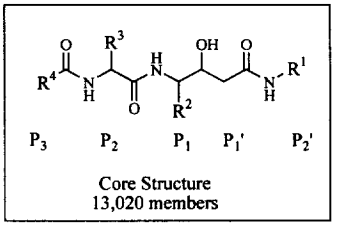
Abstract. An encoded 13,020-member combinatorial library was synthesized containing a statine core. Evaluation of this library with plasmepsin II, an aspartyl protease required for hemoglobin metabolism in the malaria parasite, led to the identification of potent and selective inhibitors as well as novel structure–activity relationships.

Malaria kills 2 million people per year and drug resistance to current therapies is a significant problem. *Plasmodium falciparum* encodes 2 aspartyl proteases, plasmepsin I and II, which are involved in the digestion of hemoglobin, a major nutrient source for the parasite.[1,2] Inhibition of hemoglobin degradation results in the death of the parasite.[3–5] Screening of encoded combinatorial libraries has identified inhibitors of enzymes[6] and receptors[7] and permitted the accumulation of data describing structure–activity relationships. After each synthetic step an encoding molecule or tag is attached to the bead indexing that unique step of the reaction in this library.[8–10] The tags are attached to the bead through an oxidatively labile linker while the compounds are attached to the bead through a photolabile linker, allowing selective removal of either the tag or compound. The tags are polyhalogenated aromatics with a variable alkyl chain, and are analyzed by electron capture gas chromatography after hydrolysis from the solid support. Thus, "reading" the tags uniquely identifies the synthetic path used to generate an individual compound.

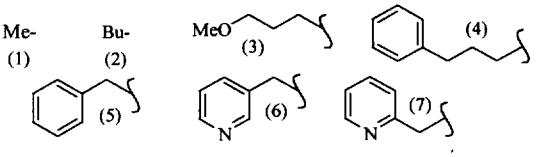
In an attempt to identify inhibitors and structure–activity relationships for this new antimalarial target, the plasmepsins, an encoded combinatorial library was synthesized containing a statine core (Table 1), a known transition state mimetic for aspartyl proteases.[11–13] The selection of synthons for the structure-based library, PL 792, was based on information available in the general literature. Because of the putative preferences for lipophilic amino acids at the scissile bond (P_1 residue), both leucine- and phenylalanine-statine (R^2 synthons) were incorporated into the library (Table 1).[14,15] The alanine-statine synthon, where the lipophilic side chain is absent, was considered a control synthon, as library compounds derived thereof would most likely be inactive. The inclusion of thirty-one amino acids (R^3 synthons, P_2 residue) and twenty N-capping elements (R^4 synthons, P_3 position) in library PL 792 provided diversity at both the P_2 and P_3 positions and resulted in a 13,020 member library (Table 1). It is the P_2 and P_3 regions of the inhibitors which impart the potency and selectivity to the statine-based inhibitors.[12,13]

0960-894X/98/$ - see front matter
PII: S0960-894X(98)00419-3

Table 1. Synthons for PL 792.
(Substituent number is shown in parenthesis.)



R1 =

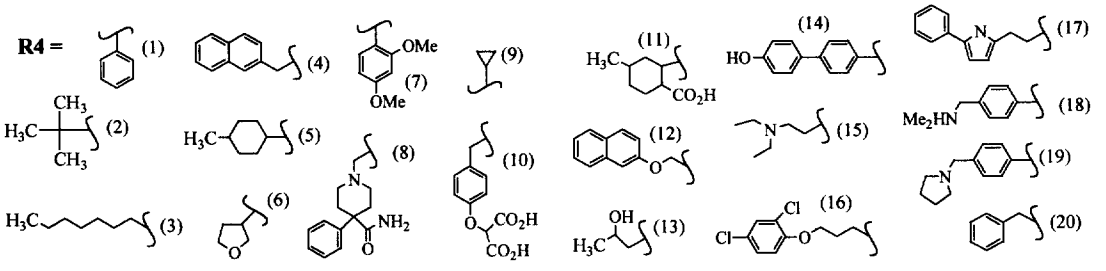


R2 =

R3 =

l-Alanine (1)	d-Histidine (9)	l-Glutamine (17)	l-Valine (25)
d-Alanine (2)	l-Isoleucine (10)	l-Hydroxyproline (18)	l- Lysine (26)
l-Arginine (3)	d-Leucine (11)	l-Phenylglycine (19)	l-Ornithine (27)
l-Asparagine (4)	l-Leucine (12)	d-Proline (20)	l-Pipecolic acid (28)
l-Pyridylalanine (5)	l-Serine (13)	l-2-Thienylalanine (21)	l-Homophenylalanine(29)
l-Aspartic acid (6)	l-Phenylalanine (14)	l-Threonine (22)	l-1-Naphthylalanine (30)
l-Cysteine (7)	l-*t*-Butylglycine (15)	l-Tryptophan (23)	l-2-Naphthylalanine (31)
d-Glutamine (8)	l-Cyclohexylalanine(16)	l-Citrulline (24)	

R4 =



Plasmodium falciparum plasmepsin II was expressed and purified from *E.coli*.[16] The plasmepsin II substrate, DABCYL-γ-aminobutyric acid-Glu-Arg-Met-Phe-Leu-Ser-Phe-Pro-EDANS, is based on the principles of fluorescence energy transfer and the sequence spans the known cleavage site in hemoglobin and hydrolysis occurs at the expected Phe/Leu site.[16] Since cathepsin D is 35% identical in amino acid sequence to plasmepsin II and is the most homologous human enzyme to plasmepsin II, it was chosen as the representative aspartyl protease for the selectivity screen. Splitting samples from the combinatorial library and screening these identical samples with both plasmepsin II and cathepsin D resulted in the identification of potent and selective inhibitors. A summary of the selectivity of the library is illustrated in Figure 1, which contains the data for 2,815 individual wells from multiple 96-well plates containing a single compound per well in which an identical portion of the bead eluate was assayed with plasmepsin II and cathepsin D. Region A contains compounds that are inhibitors of plasmepsin II but are not very inhibitory for cathepsin D. Region B illustrates compounds which are better inhibitors of cathepsin D than plasmepsin II. Compounds which fall near the dashed line have similar inhibitory activities for both enzymes. The tags from the beads containing compounds with inhibitory activity were decoded. Sixty-four inhibitory compounds were decoded for plasmepsin II and 32 inhibitory compounds were decoded for cathepsin D.

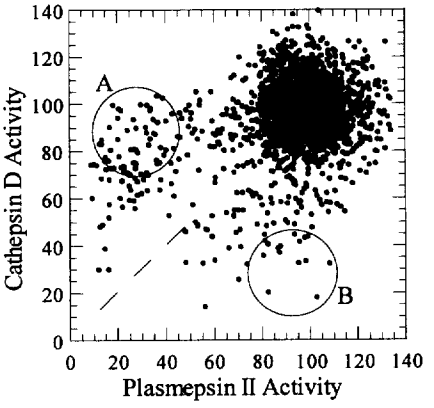


Figure 1. Selectivity plot of plasmepsin II activity versus cathepsin D activity. Each data point represents the result of a single compound tested with each protease; 2,815 data points are included. Activity is expressed as % activity remaining for each enzyme.

Since combinatorial library PL792 was synthesized with various substituents in each of the 4 positions around the statine core structure (Table 1), the frequency that the substituents are observed in inhibitory compounds in each of the variable positions can be analyzed. Figure 2 illustrates the frequency that each substituent is observed in inhibitory compounds for plasmepsin II and for cathepsin D (comparison of compounds in regions A and B in Figure 1). For example, in the R^2 position, three substituents are possible but only substituents corresponding to large hydrophobic substituents, phenylalanine- and leucine-statine are observed in inhibitory compounds for both aspartyl proteases (Figure 2). The small hydrophobic alanine-statine was not observed, indicating that a large hydrophobic residue is required to inhibit each protease as predicted from studies of substrate specificity. Comparison of the crystal structures of plasmepsin II and cathepsin D indicates that the phenyl rings forming the S_1 subsite generate the same binding pocket for these 2 enzymes.[17] Similarly, in the R^1 position, 7 substituents are present in the library; however, only the lipophilic groups are observed in inhibitory compounds for both enzymes while the pyridyl substituents are not observed (Figure 2) in either enzyme. Therefore, as anticipated, the R^1 and R^2 positions are necessary for inhibitory activity but do not impart significant selectivity between plasmepsin II and cathepsin D in this library.

As illustrated in Table 1, there are 31 substituents in the R^3 position. For compounds that show preference for plasmepsin II, only Ile (substituent 10) and Val (substituent 25) were observed in the R^3 position (Figure 2). Leu (substituent 12), a structural isomer of Ile, is present in the library but is not observed in the plasmepsin II-inhibitory compounds (Figure 2). These data strongly suggest that a β-branched carbon is preferred in this position. In contrast, cathepsin D has broader specificity in the P_2 position since 9 of the 31 substituents were observed in inhibitors (Figure 2). In the R^4 position, 10 of the 20 possible substituents are observed in the plasmepsin II-inhibitory compounds (Figure 2). These substituents are either basic, polar or hydrophobic indicating that a relatively broad range of substituents are tolerated at this position. As anticipated with the design of the library, it appears that the R^3 and R^4 substituents impart the selectivity to the statine-based inhibitors which is in agreement with the literature.[12,13] This type of frequency analysis is useful for deriving a general description of the structure–activity relationships but examination of resynthesized and purified compounds is essential (see below).

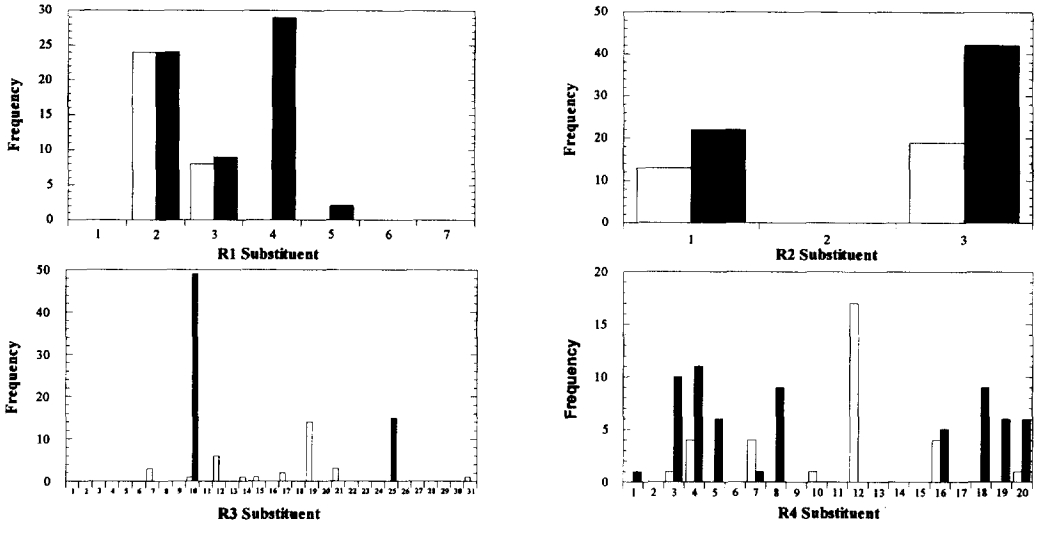


Figure 2. Substituent frequency. Frequency that each substituent was observed in the R1, R2, R3, and R4 positions of inhibitory compounds for plasmepsin II (filled in dark bars) and cathepsin D (open bars). This analysis compares structures that had a preference for either plasmepsin II or cathepsin D (region A and region B, respectively, from Figure 1). The structure for each substituent number is listed in Table 1.

In order to confirm and evaluate the inhibitory compounds observed in this library, compounds must be synthesized and purified. Compounds were chosen to be synthesized to quantitate the structure–activity relationships observed during screening, and therefore, compounds which were predicted to prefer plasmepsin II or cathepsin D as well as compounds which would inhibit both enzymes were synthesized. Table 2 contains the compounds synthesized from PL 792 and their corresponding K_i values for plasmepsin II and cathepsin D. The K_i values of the plasmepsin II-inhibitory compounds range from 50 to 600 nM and all compounds exhibited competitive inhibition. Compound PS 273800 is the most potent inhibitor of plasmepsin II synthesized from PL 792 with a K_i value of 50 nM and is sixfold selective for plasmepsin II over cathepsin D (Table 2). More potent inhibitors of plasmepsin II have been identified,[17] however, these compounds have higher molecular weights. Some additional structure–activity relationships may be developed by examining the compounds listed in Table 2. Comparison of the R^1 positions in compounds PS 429694 and PS 749213 indicate that the lipophilic substituents (butyl amine and 3-phenyl-1-propylamine) have similar potency for plasmepsin II which is consistent with the frequency data. The frequency data (Figure 2) also indicated that only phenylalanine- and leucine-statine were observed in the R^2 position. Comparison of the K_i values for compounds containing phenylalanine-statine (PS 189863, PS 707194, and PS 361691-1) to the corresponding compounds containing leucine-statine (PS 777621, PS 273800, and PS 699506, respectively) indicates that only slight changes in potency are observed for these hydrophobic substituents in both proteases. Stereoisomers of the statine hydroxyl group in this position (see compounds PS 154636-1 and -2) result in over a 50-fold difference in potency for plasmepsin II indicating that the S stereochemistry of the hydroxyl group is preferred and this is identical to the stereochemistry of the hydroxyl group in statine found in pepstatin A.

Table 2. Analysis of Resynthesized Compounds

Compound	Substituent[a] R1	R2	R3	R4	Plas II K_i (nM)	Cat D K_i (nM)
PS 273800	2	3	10	4	50	320
PS 707194	2	1	10	4	90	50
PS 189863	2	1	10	12	110	500
PS 777621	2	3	10	12	180	560
PS 444035	2	1	10	19	220	1200
PS 662477	2	1	10	7	220	2400
PS 429694	2	1	10	8	410	5500
PS 222036	4	1	10	18	440	1300
PS 725074	2	1	25	7	500	1600
PS 361691-1	2	1	19	12	560	200
PS 154636-1	2	1	19	7	590	2100
PS 749213	4	1	10	8	600	3900
PS 679304	2	3	12	12	5600	920
PS 699506	2	3	19	12	5800	110
PS 290351	2	3	14	12	12000	1300
PS 361691-2	2	1	<u>19</u>	12	15000	700
PS 731167	2	1	17	12	>30000[b]	200
PS 154636-2	2	<u>1</u>	19	7	>30000[b]	6200

[a]refers to substituents listed in Table 1; underline indicates opposite stereochemistry; [b]not inhibitory at 30 μM.

In general, comparison of compounds with single substituent changes indicates the potency rank of residues in the R^3 position is substituent 10 (Ile) [most potent] > substituent 25 (Val) > substituent 19 (phenylglycine) > substituent 12 (Leu) > substituent 14 (Phe) [least potent] > substituent 17 (Gln) [not inhibitory] for plasmepsin II. More specifically, changing the R^3 substituent from Gln (substituent 17) to the β-branched Ile (substituent 10), comparison of PS 731167 and PS 189863, results in over a 270-fold gain in potency for plasmepsin II. Compound PS 731167 is > 150-fold more selective for cathepsin D than plasmepsin II possibly due to the glutamine in the R^3 position. Furthermore, replacing the β-branched Ile (PS 777621) with the structural isomer, Leu (PS 679304) results in a 31-fold decrease in potency. This agrees with the frequency data which indicated that the R^3 position prefers β-branched substituents.

In general, the order of potency of inhibitors containing different substituents in the R^4 position is substituent 4 (2-naphthyl) [most potent] = substituent 12 (2-naphthoxy) > substituent 7 (2,4-dimethoxyphenyl) = substituent 19 (4-[1-pyrollidino]-methylphenyl) > substituent 8 (1-phenylcyclohexanecarboxamide) = substituent 18 (N-methylbenzylamine) [least potent]. Compound PS 429694 is 13-fold selective for plasmepsin II versus cathepsin D which may be due to the basic substituent in the R^4 position (comparison of compounds PS 429694 and PS 707194). The IC_{50} values for compounds PS 189863, PS 444035, and PS 429694 were ~10 μM, ~10 μM, and ~20 μM against *Plasmodium falciparum* in cell culture, respectively; these values are similar to other compounds tested previously.[3,17] Additional cell culture potency is required prior to further drug development.

Conclusions. Screening of a statine-based encoded combinatorial library led to the identification of potent and selective inhibitors for plasmepsin II as well as cathepsin D. In addition, structure–activity characterization generated new information for inhibitors of plasmepsin II; specifically, a preference for β-branched amino acids at the P_2 position. These compounds inhibit the parasite in cell culture demonstrating that the plasmepsin enzymes are good candidates for antimalarial therapy.

Acknowledgments. We would like to thank Dr. Sergei V. Gulnik and Dr. John W. Erickson (National Cancer Institute) for providing an initial supply of a cathepsin D substrate. We would also like to thank Dr. Nolan H. Sigal and Dr. John J. Baldwin for their support of this project. The work at Washington University was supported by UNDP/ World Bank/ WHO Special Programme for Research and Training in Tropical Diseases (TDR).

References

1. Scheilbel, L. W.; Sherman, I. W. In *Malaria*; Wernsdorfer, W. H.; McGregor, I., Eds; Churchhill: Livingstone, 1988; pp 219–252.
2. Goldberg, D. E.; Slater, A. F. G.; Cerami, A.; Henderson, G. B. *J. Exp. Med.* **1991**, *173*, 961.
3. 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.
4. Rosenthal, P. J.; Meshnick, S. R. *Mol. Biochem. Parasitol.* **1996**, *83*, 131.
5. Francis, S. E.; Sullivan, D. J. Jr.; Goldberg, D. E. *Ann. Rev. Microbiol.* **1997**, *51*, 97.
6. Burbaum, J. J.; Ohlmeyer, M. J. H.; Reader, J. C.; Henderson, I.; Dillard, L. W.; Li, G.; Randle, T. L.; Sigal, N. H.; Chelsky, D.; Baldwin, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6027.
7. Apell, K. C.; Chung, T. D. Y.; Ohlmeyer, M. J. H.; Sigal, N. H.; Baldwin, J. J.; Chelsky, D. *J. Biomolecular Screening* **1996**, *1*, 27.
8. Ohlmeyer, M. H. J.; Swanson, R. N.; Dillard, L. W.; Reader, J. C.; Asouline, G.; Kobayashi, R.; Wigler, M.; Still, W. C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10922.
9. Baldwin, J. J.; Burbaum, J. J.; Henderson, I.; Ohlmeyer, M. H. J. *J. Am. Chem. Soc.* **1995**, *117*, 5588.
10. Nestler, P. H.; Bartlett, P.; Still, W. C. *J. Org. Chem.* **1994**, *59*, 4723.
11. Blundell, T. L.; Cooper, J.; Foundling, S. I.; Jones, D. M.; Atrash, B.; Szelke, M. *Biochemistry* **1987**, *26*, 5585 and Abdel-Meguid, S. S. *Med. Res. Rev.* **1993**, *13*, 731.
12. Shewale, J. G.; Takahashi, R.; Tang, J. In *Aspartic Proteinases and Their Inhibitors*; Kostka, V., Ed.; Wlater de Gruyter, 1986; pp 101–116.
13. Rich, D. H. In *Proteinase Inhibitors*; Barrett, A. J.; Salvensen, G., Eds.; Elsevier Science Publishers B. V.: Amsterdam, 1986; Chapter 5 and references therein.
14. Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Smith, C. E.; Duffin, K. L.; Goldberg, D. E. *J. Clin. Invest.* **1994**, *93*, 1602.
15. Rich, D. H.; Agarwal, N. S. *J. Med. Chem.* **1986**, *22*, 2519.
16. Luker, K. E.; Francis, S. E.; Gluzman, I. Y.; Goldberg, D. E. *Mol. Biochem. Parasitol.* **1996**, *79*, 71.
17. Silva, A. M.; Lee, A. Y.; Gulnik, S. V.; Majer, P.; Collins, J.; Bhat, T. N.; Collins, P. J.; Cachau, R. E.; Luker, K. E.; Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Goldberg, D. E.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034.