Potent, Low-Molecular-Weight Non-Peptide Inhibitors of Malarial Aspartyl Protease Plasmepsin II

Tasir S. Haque,[†] A. Geoffrey Skillman,[‡] Christina E. Lee,[†] Hiromu Habashita,[†] Ilya Y. Gluzman,[§] Todd J. A. Ewing,[‡] Daniel E. Goldberg,[§] Irwin D. Kuntz,*,[‡] and Jonathan A. Ellman*,[†]

Department of Chemistry, University of California, Berkeley, Berkeley, California 94720, Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California 94143, and Howard Hughes Medical Institute, Departments of Medicine and Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri 63110

A number of single-digit nanomolar, low-molecular-weight plasmepsin II aspartyl protease inhibitors have been identified using combinatorial chemistry and structure-based design. By identifying multiple, small-molecule inhibitors using the parallel synthesis of several focused libraries, it was possible to select for compounds with desirable characteristics including enzyme specificity and minimal binding to serum proteins. The best inhibitors identified have K_i's of 2−10 nM, molecular weights between 594 and 650 Da, between 3- and 15-fold selectivity toward plasmepsin II over cathepsin D, the most closely related human protease, good calculated log P values (2.86−4.56), and no apparent binding to human serum albumin at 1 mg/mL in an in vitro assay. These compounds represent the most potent non-peptide plasmepsin II inhibitors reported to date.

Introduction

Malaria is a parasitic disease that afflicts 300−500 million people worldwide, killing 1−2 million annually.[1] It is estimated that up to 40% of the world's population lives in regions where malaria is endemic. *Plasmodium falciparum* is the most dangerous form of the four malaria parasites that infect humans and is responsible for more than 95% of malaria-related deaths.[2] Increasing resistance of *P. falciparum* to existing therapies has heightened concerns about malaria in the international health community. However, there has been little economic incentive for the development of new drug-based antimalarial therapies. While quinine has been used for hundreds of years to treat malaria, its use as an antimalarial agent is hindered by parasite resistance. Chloroquine, an inexpensive quinine derivative first identified as an antimalarial agent in the 1930s, also faces widespread resistance throughout the world. Other commonly used antimalarial drugs such as mefloquine and Fansidar can provide effective treatment in cases of chloroquine resistance, but parasites resistant to these alternative therapies have also been reported.[3] Indeed, some *P. falciparum* strains have been identified that are resistant to all known antimalarial drugs.[4] The alarming spread of parasite resistance clearly indicates that new strategies are necessary for finding safe and effective means of treating malaria. For this reason, increased attention has focused on the plasmepsin aspartyl proteases of *P. falciparum*[5] as potential chemotherapeutic targets. Subnanomolar, peptidic inhibitors with modest activity in cell culture (IC$_{50}$ > 20 µM) have been reported,[6] and most recently, combinatorial methods have been used to identify lower molecular weight statine-based inhibitors.[7] The most potent inhibitor

(IC$_{50}$ = 50 nM) showed ∼6-fold selectivity over cathepsin D.[7] Herein, we report application of combinatorial chemistry and structure-based design methods to rapidly identify several potent (K_i = 2−10 nM) and selective (up to 15-fold over cathepsin D), low-molecular-weight inhibitors using several iterative focused libraries.

The *Plasmodium* parasite invades human erythrocytes and consumes up to 75% of the hemoglobin present.[8] Three enzymes have been identified that digest hemoglobin in an acidic parasite food vacuole: a cysteine protease, falcipain, and two aspartyl proteases, plasmepsins I and II (Plm I and II).[5] The two plasmepsins have a high degree of sequence homology to one another (73% identical) and are closely related in structure to human cathepsin D (Cat D). However, neither of the plasmepsins is targeted by any of the currently available antimalarial therapies. Of the two aspartyl proteases present in the malarial acidic food vacuole, Plm II was chosen as a primary target because the crystal structure was available[6] and because sufficient quantities of expressed enzyme could be obtained for high-throughput screening of compound libraries. At the initiation of the project, Plm I was only available via purification of parasite extract, whereas Plm II could be expressed and obtained in larger quantities.[9] The high degree of sequence homology of the two plasmepsins suggested that by targeting Plm II, both enzymes could ultimately be inhibited by the same compounds.

Our approach to obtaining potent Plm II inhibitors began by comparing the crystal structures of Plm II and Cat D both complexed to the peptide-based natural product pepstatin. Plm II and Cat D have a significant sequence homology (35%) and have an even higher sequence homology in the active site region. We had previously reported the successful identification of potent Cat D inhibitors by the design, synthesis, and screening of a library of 1039 mechanism-based inhibitors incorporating the hydroxyethylamine isostere (Fig-

[†] University of California, Berkeley.
[‡] University of California, San Francisco.
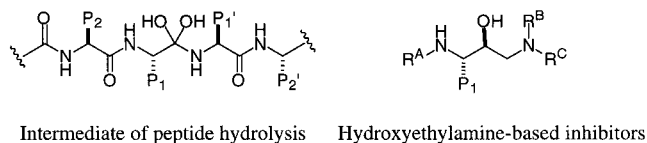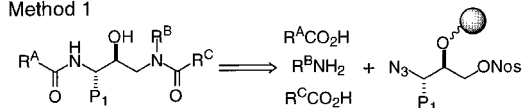[§] Washington University School of Medicine.

Intermediate of peptide hydrolysis   Hydroxyethylamine-based inhibitors

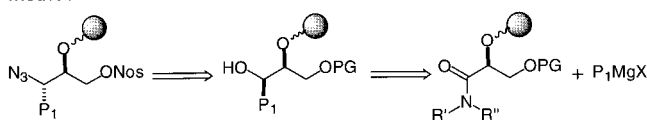**Figure 1.** Tetrahedral intermediate of hydrolysis and hydroxyethylamine isostere.

Figure 2. Synthesis approaches toward aspartyl protease inhibitors.[14,15]

ure 1).[10] Due to the close similarity of Plm II and Cat D, we screened this library against Plm II to identify lead compounds.

We next used six iterative libraries to optimize the Plm II inhibitors. First, each of the three sites of variation (R^A, R^B, R^C in Figure 1) was examined individually. By synthesizing these compounds in parallel, we were able to examine relatively large numbers of side chains at each position. Two methods were used to select subsets of side chains from the thousands of commercially available amines and acylating agents. One set of side chains was selected by modeling side chains in the Plm II active site using the program DOCK,[11-13] and the second set of side chains was selected by clustering to maximize the diverse display of functionality. We have previously shown that modeling can be more efficient at generating potent molecules;[10] however, by using both modeling and diversity clustering, we hoped to find compounds that would be anticipated to bind in the active site (modeled compounds), while simultaneously investigating whether we could identify side chains with unanticipated binding interactions (diverse compounds). The final two library iterations involved the simultaneous variation of two or more sites around the hydroxyethylamine scaffold to search for cooperative effects.

Results and Discussion

Chemistry. The hydroxyethylamine isostere has served as a particularly effective mechanism-based inhibitor of aspartyl proteases, where the tetrahedral intermediate of the peptide bond cleavage is approximated by the stable hydroxyethylamine compound (Figure 1). The isostere is also amenable to the introduction of a wide variety of side chains on both sides of the secondary alcohol whose position corresponds to the scissile bond of the peptide substrate.

We have previously reported two solid-phase synthesis sequences to prepare libraries of hydroxyethylamine-based inhibitors. In the first reported approach (Figure 2: method 1), diverse functionality is introduced at three sites, R^A, R^B, and R^C, using amines and acylating agents.[14] In the second reported approach, multiple side

chains can also be introduced at the P_1 position using Grignard reagents (Figure 2: method 2), thereby allowing diverse functionality to be introduced at all possible sites of the inhibitor structure.[15] Both synthesis sequences were used to prepare libraries targeting Plm II.
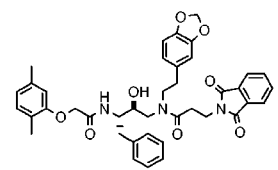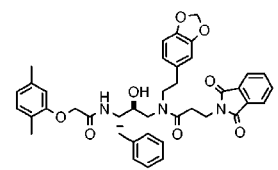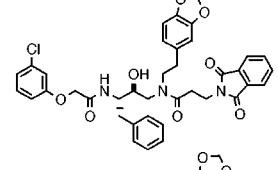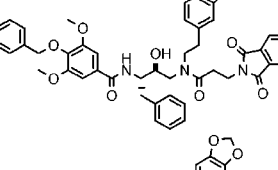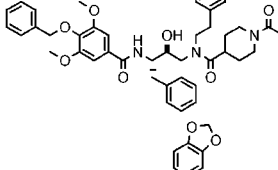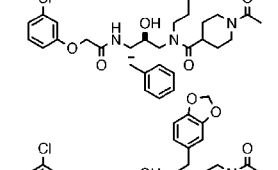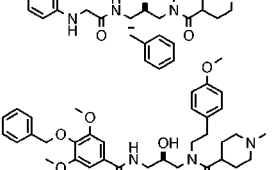
Lead Identification. A library of 1000 compounds ($10 \times 10 \times 10$ side chains at the three sites of variation) targeting Cat D and a second 39-compound library optimized toward Cat D were screened against Plm II using a fluorogenic peptide-substrate assay with an inhibitor concentration of 1 μM.[16] Of the 1039 compounds screened, 13 compounds showed greater than 50% inhibitory activity against Plm II. Compounds **1** and **2**, the most potent of the 13 crude compounds, were prepared on a larger scale, purified, and evaluated as representative lead compounds (Table 1). Both of these compounds were validated as submicromolar inhibitors of Plm II, as well as being potent Cat D inhibitors ($K_i =$ 15 and 220 nM for **2** against Cat D and Plm II, respectively).

Iterative Optimization of Individual Sites. The two compounds described above were used to begin optimization against Plm II. Initially, each of three possible sites of variation (R^A, R^B, and R^C) was individually optimize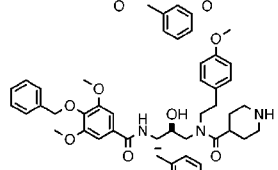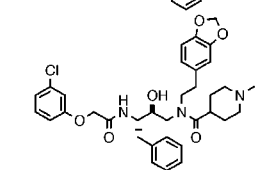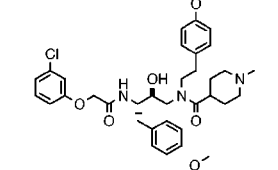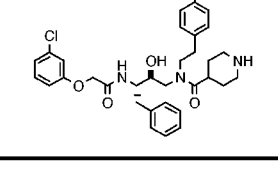d. This iterative approach allowed for the exploration of relatively large numbers of side chains at each site without having to resort to the extremely large numbers of compounds that would be necessary if all sites were examined simultaneously. Subsequent libraries probed for cooperativity between two or more sites around the hydroxyethylamine core.

1. Library R^A. The first site examined was the R^A side chain. The two other variable sites were fixed based on the structures of the lead inhibitors **1** and **2** (R^B = 3,4-(methylenedioxy)phenethylamine; R^C = 3-phthalimidopropionic acid). A list of commercially available acylating agents was generated by screening the Available Chemical Directory[17] and selecting for characteristics such as chemical compatibility, molecular weight, and cost. From this list, two sets of side chains were established for R^A (Figures 3 and 4). The diverse set was generated by clustering the acylating agents to explore a broad range of functionality at the site (side chains A1-A44, Figure 3). More specifically, the selection technique involved a complete linkage hierarchical agglomerative clustering of side chains based on the atomic content and connectivity of each molecule. The modeled set of side chains was designed to fit specifically into the enzyme active site. DOCK[11-13] was used to model each acylating agent in the appropriate part of the crystallographic active site of Plm II (side chains A45-A74, Figure 4).[6] Nearly all of the 50 highest-scoring modeled side chains had molecular weights between 250 and 300 Da. Since low-molecular-weight (MW) inhibitors were desired, we selected the 18 highest-scoring side chains with a MW between 250 and 300 Da and the 12 highest-scoring side chains with a MW < 250 Da.

Upon screening the R^A library, compound **3**, one of the modeled compounds, had significantly improved inhibitory activity compared to the lead compounds (Table 1). At the R^A site, this inhibitor incorporated side chain A52 (4-(benzyloxy)-3,5-dimethoxybenzoic acid). A

Table 1

Entry No.		mol. wt.	ClogP[*]	OH+NH	N+O	"Rule of 5" fulfillments	K_i (nM)	
							Cat D	Plm II
1		692	5.86	2	11	1	–	300 ± 22
2		698	5.71	2	11	1	15 ± 2[**]	220 ± 6
3		800	5.50	2	13	1	–	100 ± 3
4		752	2.78	2	12	2	4.3 ± 0.5	4.8 ± 0.6
5		650	2.99	2	10	3	1.9 ± 0.6	4.0 ± 0.4
6		649	2.86	3	10	3	1.3 ± 0.2	3.0 ± 0.3
7		710	4.35	2	10	3	58 ± 10	4.1 ± 0.2
8		696	3.50	3	10	3	71 ± 16	9.0 ± 0.2
9		622	4.00	2	9	3	5.8 ± 0.6	2.0 ± 0.1
10		608	4.56	2	8	3	9.8 ± 1.6	2.0 ± 0.1
11		594	3.71	3	8	3	63 ± 12	4.3 ± 0.1

[*]ClogP values calculated for the neutral compounds [**]The K_i for compound **2** was determined previously.[9]
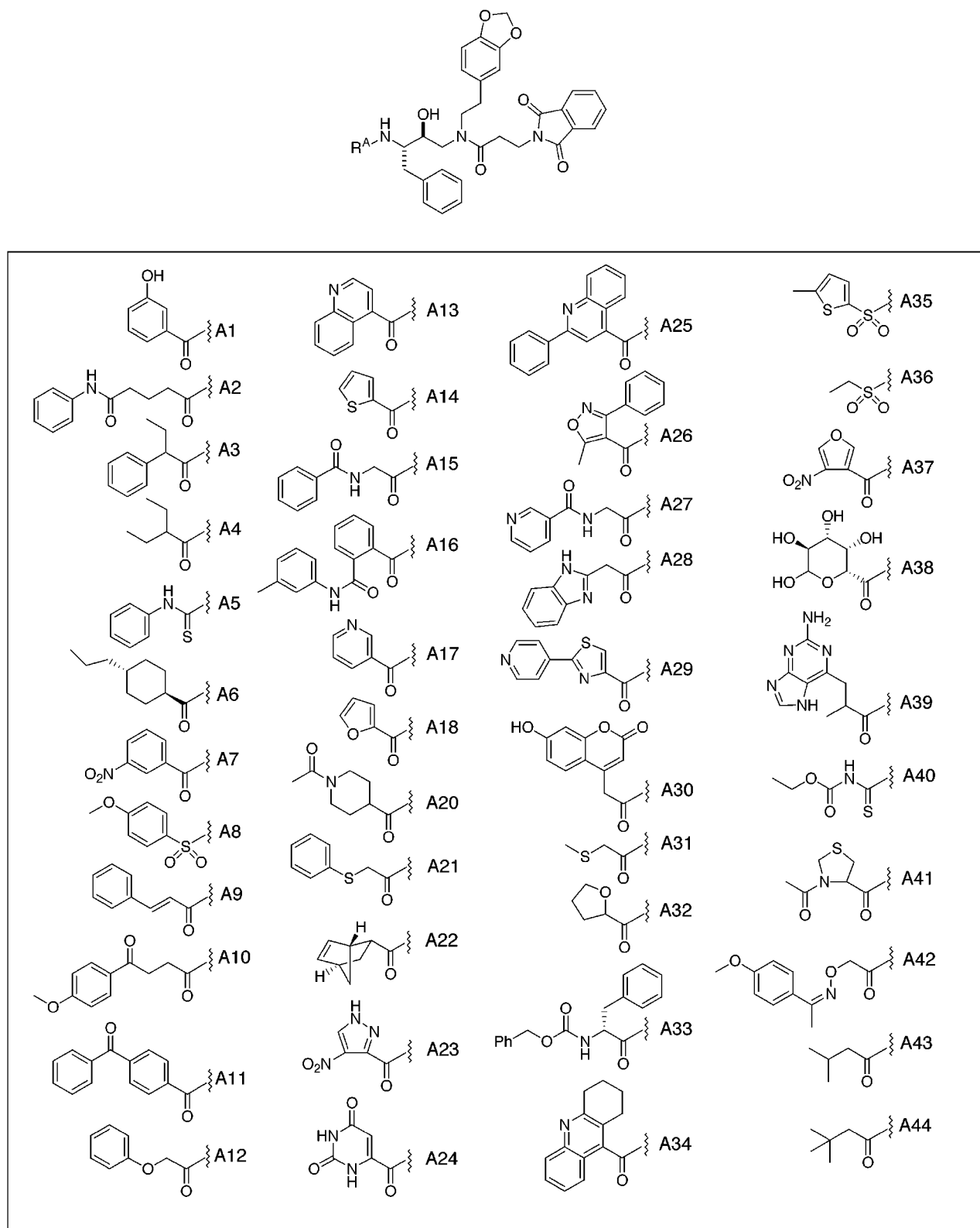
Figure 3. Diverse side chains for library R^A.

K_i of 100 nM was determined for analytically pure inhibitor **3** against Plm II, a 2−3-fold improvement over the initially identified lead compounds. A 12-member library was prepared examining variants of side chain A52 in an attempt to determine the nature of the increased affinity of **3** to the Plm II active site (Figure 5). The IC_{50} values of the entire 12-member library are at least 10-fold higher than that of inhibitor **3** toward Plm II. For example, the compound containing the R^A = A52e (4-(benzyloxy)-3-methoxybenzoic acid) has a K_i of 3.7 μM against Plm II, demonstrating that removal of just one of the methoxy groups from the R^A side chain results in significantly decreased affinity to the Plm II active site. This is consistent with the modeled structure in which both methoxy groups make significant interactions with the protein.

Figure 4. Modeled side chains for library R^A.

2. Library R^B. In a fashion similar to the R^A Library, we next examined the functionality at the R^B site. The chemically compatible amine side chains were clustered to maximize diversity and modeled in the Plm II active site to generate two lists of side chains (Supporting Information Figure 3). It should be noted that some

Figure 5. Modified A52 side chains for library R^A (A52−A52l).
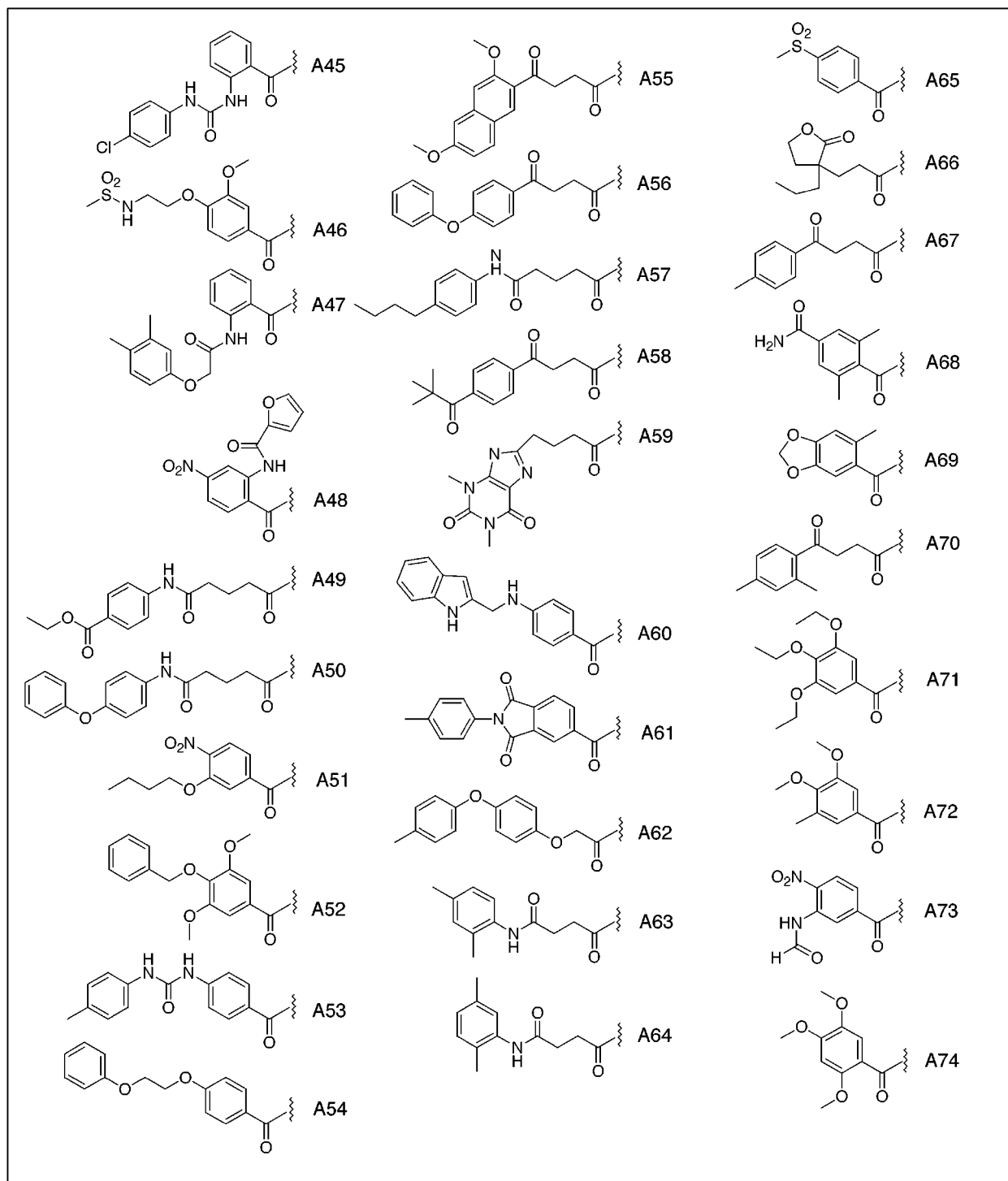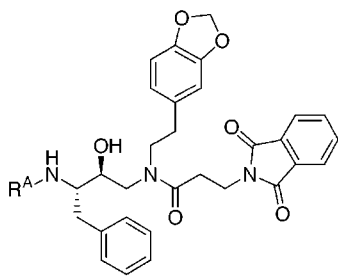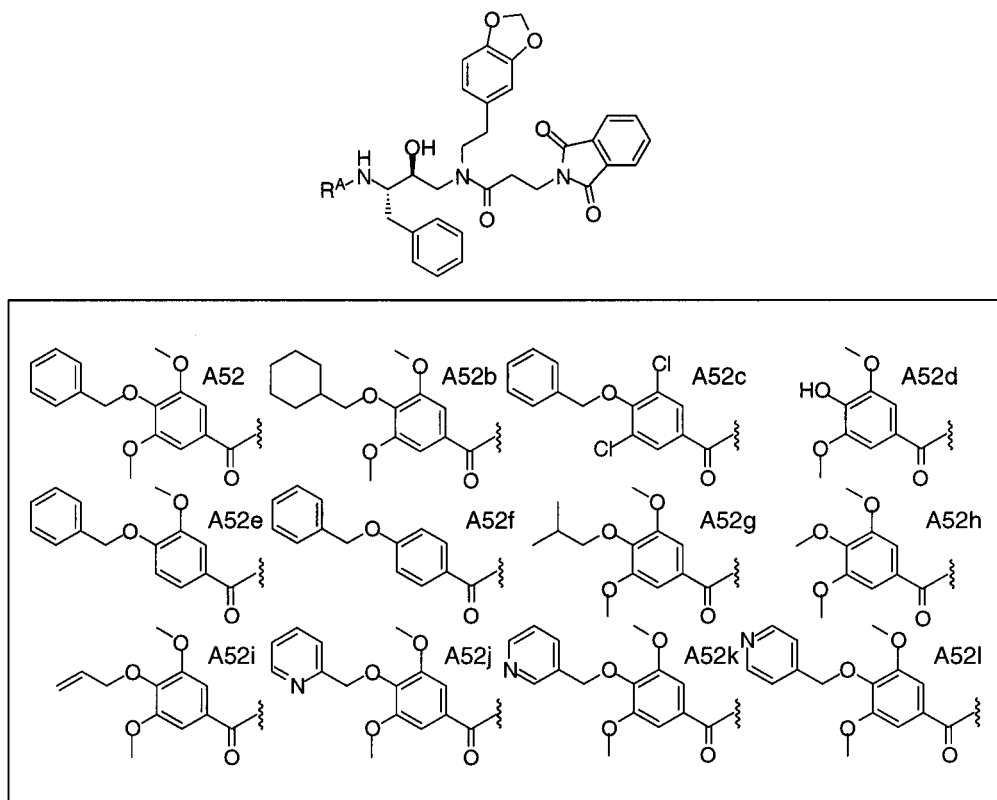
difficulty was experienced when trying to model side chains into the S_1' and S_2' pocket of Plm II in the pepstatin A/Plm II crystal structure (the pocket for the R^B and R^C side chains). Because pepstatin A does not fill the S_1' subsite, the site is somewhat collapsed in the crystal structure. When the hydroxyethylamine scaffold was modeled in the Plm II active site, there was close contact between the enzyme and the scaffold near the attachment point of the R^B and R^C side chains. Although the R^B and R^C side chains could be modeled in, they were highly constrained. We hypothesized that the enzyme might relax significantly to accommodate inhibitors, by opening the "flap" loops that cover the aspartyl protease active site. Indeed, this hypothesis was substantiated by the crystal structure of compound **3** bound to Plm II.[18] In that structure, the closest contact across the S_1' subsite (γ-carbon of V78 to ζ-carbon of F294) was 8.11 Å compared to only 5.4 Å in the pepstatin/Plm II complex.[6]

No novel R^B side chains were identified that showed significant improvement over the lead compounds. Therefore, we carried the 3,4-(methylenedioxy)phenethylamine (R^B = B25) into subsequent libraries.

3. Library R^C. The R^C site was the next site to be examined. Due to the difficulty of modeling the R^B and R^C side chains, only a diverse set of side chains was selected for inclusion into the library (Figure 6). Due to chemical compatibility considerations with the reaction conditions used in the library synthesis, the list of diverse side chains for R^C differs from that used for the R^A library. For example, compounds containing reducible functionality were eliminated from the R^C set because of the subsequent $SnCl_2$ reduction required to expose the R^A amine coupling site.

After screening the R^C library, several new side chains were identified as having both good inhibitory potency against the enzyme *and* reduced molecular weight compared to the R^C side chain from the previously identified inhibitors. Compounds containing these side chains were prepared on a larger scale and purified for evaluation with Plm II. Of the several new R^C side chains examined in the crude assay, compound **4** containing the *N*-acetylnipecotic acid side chain (C33) was the most potent, as well as the most promising in terms of potential for future modification via the piperidine amine. Analytically pure **4** with a K_i of 4.8 nM is 20-fold more potent than compound **3** and marked our first single-digit nanomolar inhibitor to Plm II (Table 1).

Inhibitor Optimization by Combinatorial Evaluation at Multiple Sites. 1. Library R^AR^C. One of the initial goals of this project was to identify potent, *low-molecular-weight* inhibitors of Plm II. The compounds identified above, while being potent Plm II inhibitors, do not lie within our desired molecular weight upper limit of 650 Da based on known aspartyl protease drugs. To gain insight into the structure−activity relationship of the Plm II inhibitors with the active site and also to identify low-molecular-weight compounds, a library of 80 compounds was synthesized where both the R^A and R^C positions were simultaneously varied (Figure 7). This library allowed us to explore any cooperative effects between the R^A and R^C sites that might enhance the affinity of the inhibitors for the Plm II active site. While the scaffold weighs only 163 Da, the inclusion of the 4-(benzyloxy)-3,5-dimethoxybenzoic acid side chain (A52; MW = 288 Da as the free acid) in a potent inhibitor would make it difficult to achieve an overall molecular
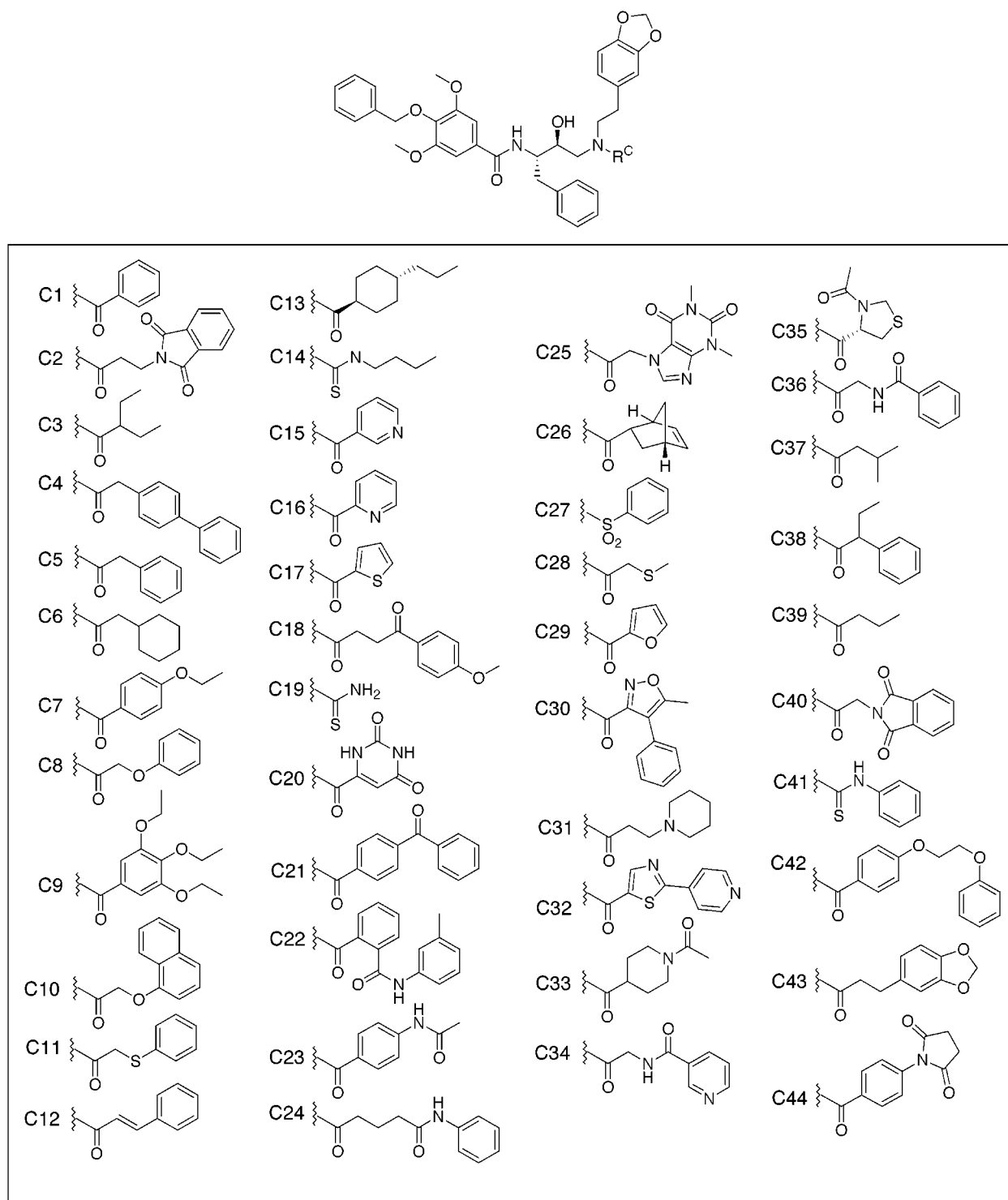
Figure 6. Diverse side chains for library R^C.

weight goal of less than 650 Da. Therefore, a wide range of smaller R^A side chains was included in an attempt to identify new $R^A R^C$ side chain combinations resulting in smaller inhibitors. Several phenoxyacetic acids that were present in or similar to the initially identified lead inhibitors were also included at R^A (A75−A77).

Upon evaluation against Plm II, several trends became clear. First, compounds containing R^C side chains related to cyclohexyl carboxylic or substituted acetic acids tended to inhibit Plm II in the fluorogenic assay (C6, C13, C33, and C46, Figure 7), with the *N*-acetylnipecotic acid side chain (C33) being present in the most

potent compounds. Second, at the R^A position, the 4-(benzyloxy)-3,5-dimethoxybenzoic acid side chain (A52) as well as all three of the phenoxyacetic acid side chains (A75−A77) provided potent inhibitors. On the basis of the crude assay data, accurate K_i determinations were performed on analytically pure **5** in Table 1 (K_i = 4.0 nM). Inhibitor **5** is slightly more potent than **4**, but with a significantly reduced molecular weight (650 Da) relative to **4** (752 Da). Compound **6**, the aniline derivative of the 3-chlorophenoxyacetic R^A side chain in **5**, was individually synthesized and evaluated to determine whether a hydrogen bond-accepting group in

Figure 7. Side chains for library $R^A R^C$.

the R^A side chain would significantly affect the binding of the inhibitors. As demonstrated by the K_i value of 3.0 nM for analytically pure **6**, there is a slight improvement in inhibitory potency upon changing the oxygen to nitrogen in the R^A side chain.
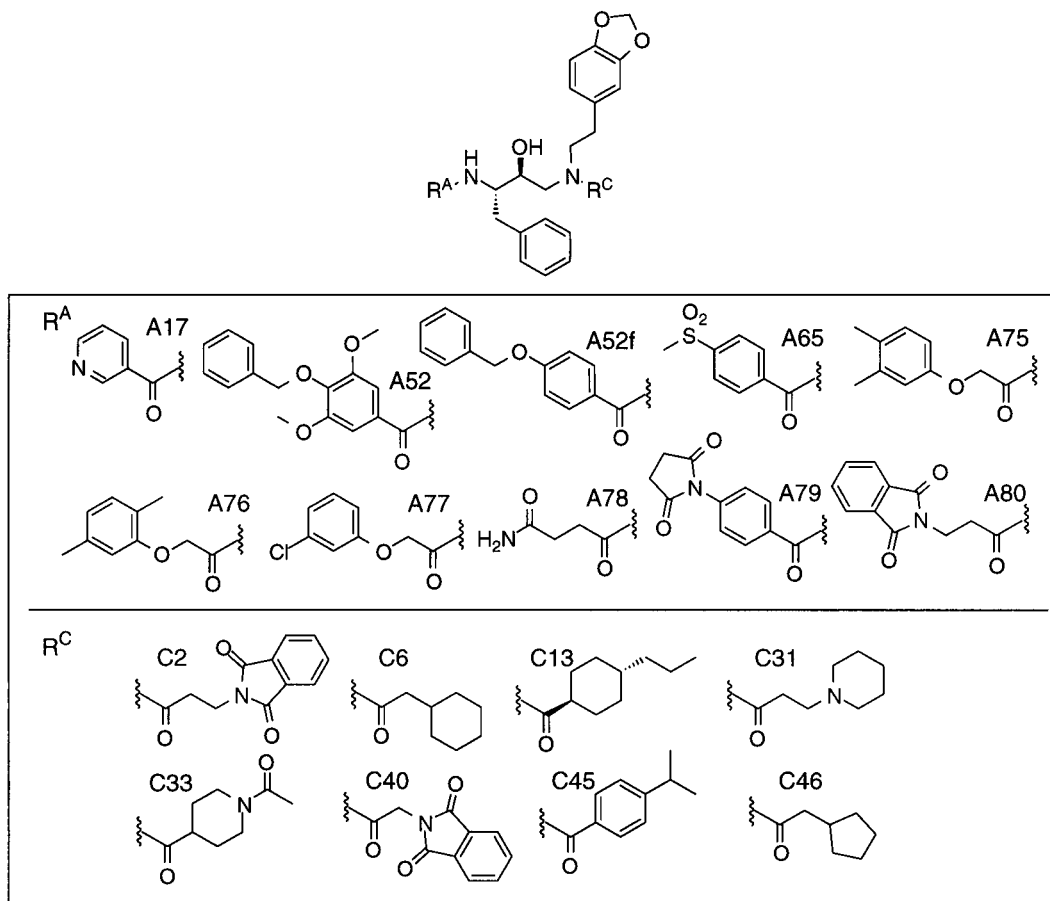
2. Library $R^A R^B R^C P_1$. The final library simultaneously examined all four sites around the hydroxyethylamine scaffold (Figure 8). The isobutyl moiety (leucine equivalent) was included at the P_1 position to attempt to identify smaller Plm II inhibitors. The R^A position was designed to explore several phenoxyacetic acids and cinnamic acids, where the cinnamic acids provide a rigidified phenoxyacetic acid equivalent. At the R^B position, both the 3,4-(methylenedioxy)phenethylamine (B25) and the smaller 4-methoxyphenethylamine (B55) were included. The R^C collection is primarily focused on exploring isonipecotic acid derivatives (C47, C48), proline derivatives (C51, C52), cyclohexanecarboxylic acid (C49), and isobutyric acid side chains (C50). Nipecotic acid and isonipecotic acid were selected for this position based on good activity of *N*-acetylnipecotic acid (C33) as well as by modeling the side chain into the crystal structure of **3** bound to Plm II.[18]

The compounds containing the isobutyl P_1 side chain were uniformly less potent in the crude assay than the compounds containing the benzyl P_1 side chain. To investigate this preference, we compared the isobutyl P_1 side chain in the pepstatin–Plm II complex[6] with the benzyl P_1 side chain in the compound **3**–Plm II complex.[18] The S_1 subsite is part of a large hydrophobic groove which extends into the S_3 subsite and can easily accommodate either the benzyl or the isobutyl side

chain. While the isobutyl side chain makes hydrophobic contacts (atom center distance < 4.0 Å) with three residues (I32, Y77, and I123), the benzyl side chain makes hydrophobic contacts with six residues (I32, Y77, I123, F111, F120, and the β-C of S79), including two phenylalanine contacts which are not present with the isobutyl side chain. This structural analysis is consistent with the preferred cleavage site for Plm II (-Phe-Leu-), where the phenylalanine side chain would occupy the S_1 pocket of the protease.[19]

Both R^B side chains in the library resulted in potent inhibitors. The effectiveness of the 4-methoxyphenethylamine (B55) was a moderate step toward our goal of reducing the size of the Plm II inhibitors with no cost in binding affinity, as demonstrated by compounds **9** and **10** (Table 1). There was a preference for isonipecotic acids at R^C and either A52 or phenoxyacetic acids at R^A, especially the 3-chlorophenoxyacetic acid (A77). Of the numerous inhibitors identified in the final library, several of the more potent inhibitors were selected for scaleup and purification for exact K_i determination (compounds **7**–**11**, Table 1). Analysis of the purified compounds revealed several small, highly potent Plm II inhibitors (compounds **9**–**11**, Table 1).

Screening Potent Inhibitors for Desired Characteristics. By employing an iterative library design, synthesis, and evaluation sequence, multiple single-digit nanomolar inhibitors were identified that displayed a significant level of structural variety. This diverse set of inhibitors enabled selection of compounds based on desirable characteristics distinct from potency.
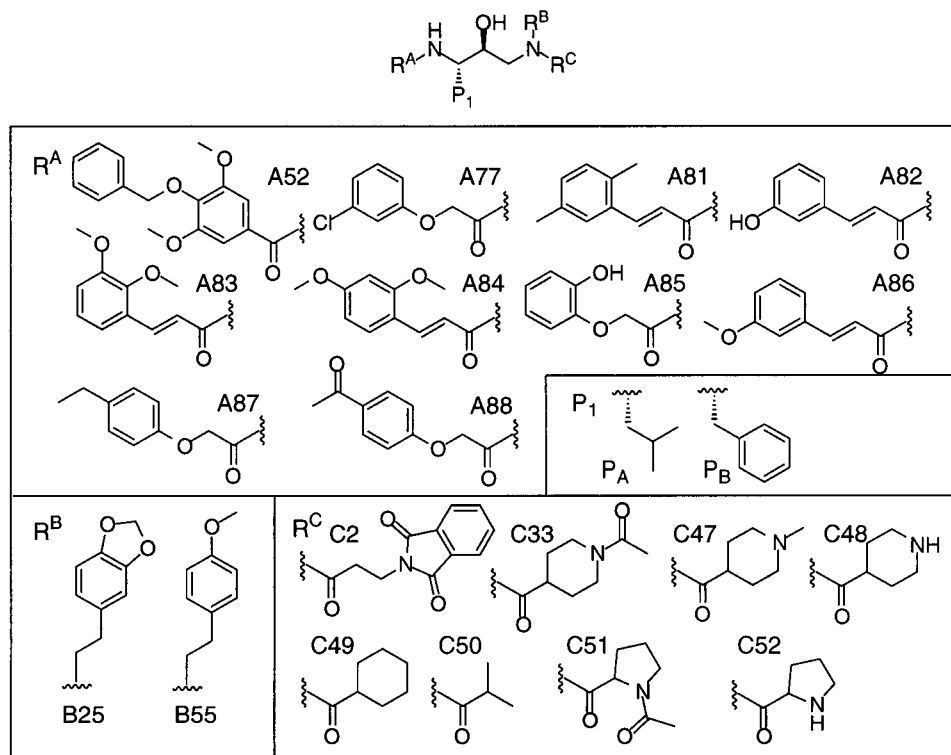
Figure 8. Side chains for library $R^A R^B R^C P_1$.

First, we evaluated the most potent Plm II inhibitors for activity against the most highly homologous human protease, cathepsin D. Plm II selectivity would be desired from a therapeutic standpoint, since it may be undesirable to inhibit human cathepsin D. While some of the most potent Plm II inhibitors, **4**−**6**, showed no selectivity for Plm II over Cat D, several of the single-digit nanomolar, low-molecular-weight inhibitors, **7**, **8**, and **11**, showed up to 15-fold selectivity (Table 1). Comparison of the selective and nonselective inhibitors indicates that the piperidine-based side chains at the R^C position are likely the most important contributor to the observed selectivity.

A representative set of the most potent Plm II inhibitors (compounds **4**, **10**, and **11**) were also assayed in a buffer containing human serum albumin (HSA). Binding to serum proteins could significantly reduce the effectiveness of the inhibitors in vivo as has been observed for some HIV protease inhibitors.[20] At 1 mg/mL of HSA, the potencies of inhibitors **4**, **10**, and **11** were not affected, indicating that little if any binding to HSA occurs (the fluorescence-based enzyme assay could not be performed at higher concentrations due to interactions between HSA and the enzyme or substrate). These results suggest that HSA binding will likely not diminish inhibitor potency significantly, even at the high concentrations of HSA found in serum (30−45 mg/mL).

The potent Plm II inhibitors were also screened for desired calculated properties thought to be important for pharmacokinetics, including calculated log P (Clog P) values, number of hydrogen bond acceptors, and number of hydrogen bond donors (Table 1). Desired ranges for these properties described by Lipinski based upon the evaluation of compound databases include Clog P < 5, number of hydrogen bond donors ≤ 5, number of hydrogen bond acceptors ≤ 10, and molecular weight <

500 Da.[21] A compound that fulfills at least three out of the four criteria is in agreement with Lipinski's "Rule of 5." The most potent and selective inhibitors, **7**, **8**, and **11**, satisfy three out of the four criteria (Table 1). Although the molecular weights of the most potent inhibitors (594−650 Da) are above Lipinski's upper limit of 500 Da, it is significant that they are within the molecular weight range of the currently approved HIV protease inhibitors.[22]

Finally, the top Plm II inhibitors, compounds **9**−**11**, were determined to be moderately *more* potent toward Plm I than Plm II enabling simultaneous inhibition of both aspartyl proteases in the parasite digestive vacuole with a *single* inhibitor. These inhibitors were also determined to have 1−2 μM IC$_{50}$ values for inhibition of parasite growth in cultured parasite-infected human erythrocytes.

Conclusion

By applying a rational and directed approach to the iterative design of a small number of libraries, a diverse collection of potent, low-molecular-weight inhibitors of Plm II were rapidly identified. The use of focused libraries made it possible to identify multiple potent inhibitors with some structural variety, allowing us to select a subset of the potent inhibitors that have additional desirable characteristics. For example, inhibitors **5**, **6**, and **9**−**11** stand out as being of low molecular weight (594−650 Da), having desirable Clog P characteristics (2.86−4.56 calculated values), having ≤ 5 hydrogen bond donors, having ≤ 10 hydrogen bond acceptors, containing *no* chiral centers outside of the hydroxyethylamine core, and being made up of commercially available side chains. Inhibitors **7** and **11** also show approximately 15-fold selectivity over the most closely related human aspartyl protease cathepsin D. In addition, there is no observable decrease in inhibitory
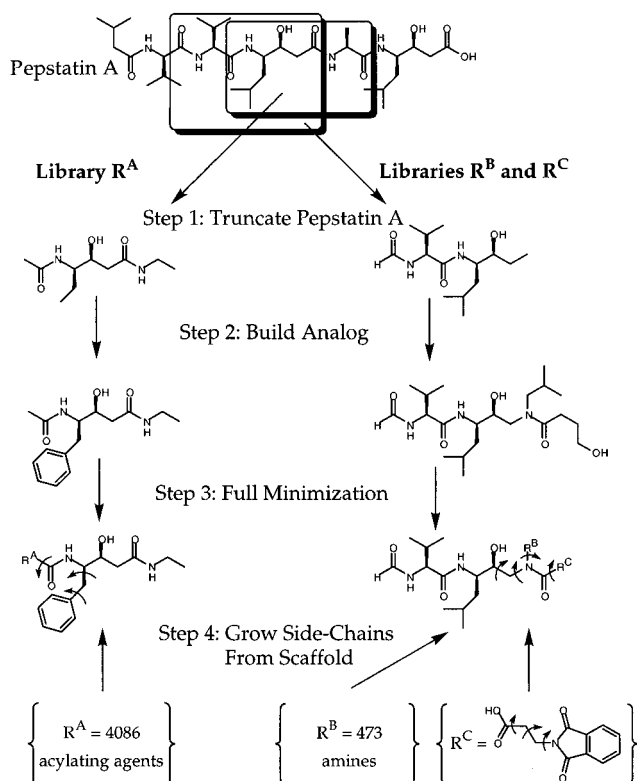
Figure 9. Structure-based design strategies.

potency for compounds **4**, **10**, and **11** when assayed in buffer containing HSA at 1 mg/mL (the upper limit of the assay), suggesting that Plm II inhibition should not be dramatically affected even at the significantly greater in vivo HSA concentrations. Notably, compound **11** has all of these desired characteristics with a K_i of 4.3 nM. The compounds discussed in this paper represent the most potent and lowest-molecular-weight Plm II inhibitors to be reported. We are continuing to examine related compounds in in vitro and cell-based assays. Finally, we believe that the lead identification and optimization protocols described in this report could successfully be applied to rapidly identify potent, low-molecular-weight, and specific inhibitors to any aspartyl protease.

Experimental Section

Computational Methods. 1. Structure Examination. To use a structure-based drug design approach to the optimization libraries, we first examined the crystal structure of a Plm II/pepstatin A complex from *P. falciparum*. This crystal structure has a resolution of 2.7 Å and an *R* value of 0.195. There are two copies ("a" and "b") of the protease in the unit cell, each with a copy of pepstatin A bound in the active site. The unit cell also contained 122 crystallographic waters. The two copies of the protease are quite similar (0.92 Å αC rmsd). We chose to use copy "a" for library screening based on the slightly lower *B* factors and slightly better bond angles in its copy of pepstatin.

2. R^A Scaffold Generation. On the basis of previous modeling and crystallography of Cat D,[6,23] we assumed that our hydroxyethylamine inhibitors would bind in a manner similar to the crystallographic pepstatin. To model the R^A side chain, we truncated pepstatin A and modified the statine side chain to a benzyl side chain (Figure 9: library R^A, steps 1 and 2). This scaffold was minimized for 50 steps within a rigid Plm II active site under the AMBER[24] force field in Sybyl 6.3[25] (Figure 9: library R^A, step 3). This scaffold orientation and conformation were used as the basis for all R^A library modeling.

3. Reagent Selection and Preparation (R^A and R^C Libraries). The web-based program UC_Select,[26] which incorporates Daylight toolkits,[27] was used to select potential reagents for each library from the Available Chemicals Directory (ACD).[17] A variety of acylating agents were selected to generate the R^A library, including carboxylic acids, acid halides, isocyanates, isothiocyanates, and sulfonyl halides. Compounds were limited to a 100–300 Da molecular weight range. Compounds with acid halides, anhydrides, azos, sulfonic acids, >1 nitro, epoxides, azides, >4 halides, peroxides, macrocycles, metals, or boron were eliminated by default. Furthermore, compounds with aldehyde, alkyl halide, amine, hydroxylamine, thiol, or multiple acylating functionalities were eliminated. Suppliers were limited to Aldrich, Fluka, Sigma, Calbiochem, ICN, Pfaltz & Bauer, TCI America, Lancaster, Acros Organics, Maybridge International, and Trans World. Final lists included 3458 carboxylic acids, 224 acid halides, 103 isocyanates, 150 isothiocyanates, and 158 sulfonyl halides. Each of these side chains was converted to the appropriate amide derivative using a Daylight toolkit program;[26,27] then Cartesian coordinates were generated for each side chain using distance geometry. Each side chain was then transformed into the orientation of the scaffold and attached by overlaying the appropriate amide of the scaffold with the amide-like bond of the side chain. This resulted in the generation of 4086 unique scaffold side chain compounds.

Similarly for the R^C library, a molecular weight range of 30–275 Da was used. Acylating agents were identified using the screening methods listed above for the R^A site, with the additional elimination of side chains containing alcohol, nitro, or phenol groups.

4. Docking of R^A Side Chains (R^A Library). The "anchor and grow" algorithm implemented in DOCK[28] was used to model each scaffold and side chain from the virtual library in the Plm II active site. For each side chain, the scaffold, in its minimized pose (scaffold orientation and conformation), was used as the "anchor". The benzyl side chain and the R^A side chain were "grown" from this scaffold in the following manner (Figure 9: library R^A, step 4). First, the molecule's flexibility was preanalyzed. Each side chain was broken into rigid fragments separated by flexible bonds, and a list of preferred torsion angles for each flexible bond was determined. The molecule was divided into layers starting with the anchor fragment. The first layer contains the rigid fragments which attach to the anchor fragment at each "growth point". The second layer contains subsequent rigid fragments which attach to fragments from the first layer. Additional layers were generated until the entire molecule had been analyzed. Second, the molecule was "grown" into place. The rigid anchor was placed in its minimized pose (vide supra). The fragments from the first layer were placed in each of their acceptable torsion positions. These partial molecules were scored (vide infra) and minimized using the six degrees of freedom associated with the pose as well as torsional degrees of freedom from the current layer and the two previous layers. The list of partial molecules was ranked according to score and rmsd from the best-scoring partial molecule, and the top 25 partial molecules were used as starting positions for subsequent layers of growth. For each additional layer, all torsions of that layer were again generated, minimized, and ranked. At each layer, the top 25 fragments were saved for further growth. This greedy method resulted in 25 final conformations. The best-scoring molecule was saved for comparison with other molecules in the database. For this study, scores consisted of the intermolecular van der Waals and electrostatic terms as well as intramolecular van der Waals and electrostatic terms excluding 1,2 and 1,3 terms from the AMBER force field.[24]

The best scoring compounds from the DOCK runs were divided into two categories based on molecular weight ("large" = 250–300 Da and "small" = 100–249 Da). The 100 best-scoring compounds were examined on graphics and evaluated based on conformation and hydrogen bond formation. The best 74 compounds were clustered using Daylights connectivity structure measure,[27] a Tanimoto similarity metric, and a

complete linkage hierarchical clustering algorithm. This minimal clustering used an intracluster similarity of 0.85 which generated 64 total clusters with 55 singletons. The best-scoring compound in each cluster was considered the cluster representative, and clusters with members predicted to bind in different binding modes were separated into individual clusters. Each remaining molecule was assessed graphically. Compounds were eliminated if they had large hydrophobic groups extending into solvent, extended aliphatic chains, no heteroatoms, no hydrogen bonds, amino acid subunits, or hydrogen bond clashes. There were 30 "large" compounds and 5 "small" compounds remaining. The top scoring 25 "large" compounds and the 5 "small" compounds were selected for synthesis. We sought 20 additional "small" compounds. The next 250 best scoring compounds were then analyzed. Sixty-nine of them had R^A side chains with a MW < 250 Da. These compounds were assessed graphically with the same criteria used above, and 31 additional compounds were selected. The top 20 of these were selected for synthesis.

5. Clustering of R^A Side Chains (R^A and R^C Libraries). All unique R^A side chains were clustered to identify a representative set of 50 compounds. The structure of each reagent was encoded in a binary fingerprint using the Daylight connectivity method.[27] An all-by-all similarity distance matrix was calculated using a Tanimoto similarity metric.[27] Complete linkage hierarchical clustering was carried out until only 50 clusters remained. This corresponded to a minimum internal cluster similarity of 0.1659 and no singletons. One compound for each cluster was chosen as a representative of that cluster. A similar clustering method was applied to side chains for the R^C site in the R^C library, using a modified list of acylating agents (as discussed above).

6. R^B and R^C Scaffold Generation (R^B and R^C Libraries). Similarly to the R^A library, the scaffold for R^B and R^C libraries was generated by starting with the crystallographic structure of pepstatin A bound to Plm II. Pepstatin A was truncated as shown in Figure 9 (libraries R^B and R^C, step 1). A hydroxyethylamine scaffold was then built, and leucine and serine side chains were attached at the R^B and R^C side chain positions, respectively (Figure 9: libraries R^B and R^C, step 2). These side chains were minimized in a rigid receptor similarly to the R^A scaffold (Figure 9: libraries R^B and R^C, step 3). The side chains were then truncated, and the resulting fragment orientation and conformation were used as the scaffold for all R^B and R^C libraries.

7. R^B Reagent Selection (R^B Library). Similarly to R^A and R^C reagents, R^B reagents were selected using the UC_Select program. Primary amines were selected from the ACD 95.1. In addition to the defaults used in R^A and R^C selection, compounds with alcohol, aldehyde, alkyl halide, amino acid, carboxylic acid, hydroxylamine, nitro, phenol, thio, phosphoric acid or ester, and phosphonic acid or ester functional groups were eliminated. Suppliers were limited as with the R^A and R^C reagents. Furthermore, reagents were limited to a molecular weight range of 30–275 Da. This selection process generated a list of 473 potential amine side chains for the R^B library.

8. Docking of R^B Side Chains (R^B Library) (Supporting Information Figure 3). For each R^B reagent, R^C was fixed with the 3-phthalimidopropionic acid side chain (C2, Supporting Information Figure 3). Molecules were generated by attaching the appropriate R^B side chain and the R^C = C2 chain to the support. Docking was started with the rigid scaffold anchor described above. The R^B and R^C side chains were grown in layer by layer as described for the R^A library (Figure 9: libraries R^B and R^C, step 4). The 100 top-scoring compounds were examined visually and clustered as described above. Sixteen of these top-scoring compounds had already been used in the initial screening libraries. After removing compounds which did not pass the criteria described for the R^A library, the top 47 compounds were selected for synthesis.

9. Clustering R^B Side Chains (R^B Library) (Supporting Information Figure 3). The 473 R^B reagents were clustered in a manner similar to the R^A and R^C reagents. The 50 R^B clusters, however, contained 13 singletons, with only 71 reagents in the largest cluster. The minimum internal similarity among the 50 clusters was 0.36.

10. Calculation of Clog P Values. A SMILES string for each compound in its neutral form was generated. Structures were made from each SMILES with the PRADO utility and checked visually for accuracy.[27] Finally, we used the Clog P program to calculate an octanol/water partition coefficient for each compound based on its SMILES string.[27]

Synthesis Methods. 1. Library Synthesis (Method 1). Libraries R^A and R^B were synthesized on solid support as reported previously (Figure 2).[14] Side chains with questionable stability were tested by exposure to trifluoroacetic acid for 1 h at room temperature (equivalent to extreme cleavage conditions). Side chains with questionable acylation/amine displacement characteristics were checked via incorporation into minimal hydroxyethylamine test compounds. Compounds were then cleaved from support and validated by TLC and MALDI-MS. Libraries were synthesized in a spatially separate array in a 96-well format, using commercially available Merrifield resin (Novabiochem). Note that libraries derived from docked and/or diverse side chains often contained several components less than the number listed in the computational methods, since some of the selected side chains were either back-ordered or no longer available. Efforts were made to replace such side chains with other side chains from the same diverse cluster whenever possible. Acylating side chains containing alcohols or phenols were incorporated into libraries R^A and R^B using 0.3 M side chain, 0.3 M dicyclohexylcarbodiimide (DCC), 0.3 M *N*-hydroxysuccinimide (HOSu) and 0.9 M *i*-Pr$_2$EtN in *N*-methylpyrrolidinone (NMP) for a minimum of 4 h. Several side chains (particularly cinnamic acid derivatives and *N*-phthaloyl-β-alanine) were coupled using 0.3 M *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) instead of PyBOP due to observed precipitation in the PyBOP coupling solution. Random compounds from each library (10–15% of total library) were checked by MALDI-MS for presence of desired molecular ion peak. The desired molecular ion peak was observed in 90–100% of the compounds checked.

2. Library Synthesis (Method 2). The R^C, R^AR^C, and $R^AR^BR^CP_1$ libraries were synthesized on solid support as recently reported.[15] The P_1 side chain is incorporated by Grignard addition to a support-bound pyrrolidine amide (Figure 2). After diastereoselective reduction of the resulting ketone, nosylation and azide displacement of the secondary alcohol, and removal of the primary alcohol protecting group and subsequent nosylation, the support-bound scaffold segment was provided. Modifications of coupling conditions similar to those used in method 1 (vide supra) were also applied to libraries synthesized using method 2.

3. High-Throughput Plm II Assay. Compounds were tested for inhibitory activity against Plm II in 96-well microtiter plates using a fluorometric high-throughput assay.[29] The fluorogenic substrate used in the assay was DABCYL-GABA-Glu-Arg-Nle-Phe-Leu-Ser-Phe-Pro-EDANS.[30] The assay was performed in Microfluor "W" (DYNEX Technologies, Inc., Chantilly, VA) fluorescence microtiter plates, and readings were obtained on a Perkin-Elmer LS-50B fluorometer with an attached 96-well plate reader. An excitation wavelength of 336 nm and emission wavelength of 490 nm were used, with a 430-nm emission cutoff filter in place. Typical substrate concentrations of 1 μM and Plm II concentrations of 1.25–2.5 nM were used. Assays were performed in a 0.1 M sodium acetate buffer (pH = 5.0) containing 10% glycerin, and 0.01% Tween-20. All libraries were initially assayed using crude products at 1 μM inhibitor concentration and assuming a 50% overall yield. Active compounds were further diluted and assayed. Inhibitors were dissolved in a DMSO stock solution prior to addition to the buffer. Assays were performed in 5% DMSO to ensure dissolution of the inhibitors.

K_i Determinations. The Plm II assays for the fully characterized compounds were performed in a quartz cuvette (Starna Cells, Inc., Atascadero, CA) with a Perkin-Elmer LS-

50B spectrometer (excitation = 336 nm, emission = 490 nm). The substrate (Bachem California, Inc., Torrence, CA) used was as reported for the high-throughput assay. A 0.1 M sodium acetate buffer (pH = 5.0) with 10% glycerin and 0.01% Tween-20 was used with a final concentration of 0.6−0.7 nM Plm II. The Plm II proenzyme was stored in H_2O and activated in the sodium acetate buffer (pH = 5.0) described above. In a typical assay with a final volume of 600 μL, to 545 μL of buffer was added 30 μL of inhibitor in DMSO, followed by 15 μL of Plm II stock (buffer). The mixture was incubated at 25 °C for 4.5 min followed by the addition of 10 μL of substrate stock (buffer). The change of fluorescence intensity was recorded as a function of time. Assays were performed in duplicate or triplicate at five to six inhibitor concentrations for each K_i determination. Data were fitted by nonlinear regression analysis to the equation derived by Williams and Morrison.[31]

$$\nu = \frac{\nu_0}{2E_t}*\left\{\sqrt{\left[\left(K_i\left(1+\frac{S}{K_m}\right)+I_t-E_t\right)^2+4K_i\left(1+\frac{S}{K_m}\right)E_t\right]}-\left[K_i\left(1+\frac{S}{K_m}\right)+I_t-E_t\right]\right\}$$

The K_m for the substrate was determined to be 0.75 μM by using a Lineweaver−Burke plot (the K_m was previously determined by Goldberg and co-workers to be 0.96 μM).[8] The variables S, E_t, and I_t are the concentrations of substrate, active enzyme, and inhibitor, respectively.

The Plm II assays containing HSA were performed using the same conditions described above except the to the sodium acetate buffer was added 1 mg of HSA (nondenatured from Calbiochem, San Diego, CA) per 1 mL of buffer. Control experiments performed with only enzyme and substrate excluding inhibitor, in buffer with HSA concentrations above 1 mg/mL, resulted in a constant or even decrease in fluorescence intensity over a 20-min time period. This is presumably due to HSA interacting with either the enzyme or the substrate.

The human liver cathepsin D (Calbiochem, San Diego, CA) assays were performed in a fashion similar to the Plm II assays. The same substrate that was used in the Plm II experiments was employed for the cathepsin D assays. A 0.1 M formic acid buffer (pH = 3.7) was used with a final concentration of 0.7 nM cathepsin D. In a typical assay with a final volume of 600 μL, to 560 μL of buffer was added 20 μL of inhibitor in DMSO, followed by 10 μL of cathepsin D stock (0.05% Triton X-100 in H_2O). The mixture was incubated at 25 °C for 4.5 min followed by the addition of 10 μL of substrate stock (DMSO). The K_m was determined to be 0.87 μM using a Lineweaver−Burke plot.

***P. falciparum*-Infected Erythrocyte Assay.** Assays of inhibitory activity in cultured parasite-infected human erythrocytes were performed as previously described.[19]

Scaled-Up Synthesis of Inhibitors. Selected inhibitors were synthesized on a larger scale for purification and exact K_i determinations. Compounds were typically synthesized on a 20−30-mg scale on solid support, using the previously described library synthesis method 1 (Figure 2). The compounds were then cleaved from support and purified by silica gel flash chromatography to yield the desired pure compounds. Special care was taken to promptly remove residual acid and immediately purify compounds containing the 4-(benzyloxy)-3,5-dimethoxybenzoic acid side chain, as moderate loss of the terminal benzyl group was observed under the 1:1 trifluoroacetic acid:1,2-dichloroethane cleavage conditions.

Compound 1: ^{1}H NMR (400 MHz, CDCl₃) δ 2.10 (m, 1 H), 2.25 (s, 3 H), 2.29 (s, 3 H), 2.40−2.81 (m, 3 H), 2.95 (m, 3 H), 3.38 (m, 2 H), 3.64 (m, 1 H), 3.82 (m, 3 H), 4.30 (m, 1 H), 4.45 (m, 2 H), 5.91 (m, 2 H), 6.41−6.57 (m, 2 H), 6.63 (d, J = 7.8 Hz, 1 H), 6.72 (bd, J = 7.1 Hz, 1 H), 7.02 (d, J = 7.5 Hz, 1 H), 7.13−7.30 (m, 8 H), 7.72 (m, 2 H), 7.83 (m, 2 H); FAB-HRMS

calcd for $C_{40}H_{42}N_3O_8$ (M + H⁺) = 692.29719, obsd = 692.29892. Anal. ($C_{40}H_{41}N_3O_8\cdot0.5H_2O$) C, H, N.

Compound 2: ^{1}H NMR (300 MHz, CDCl₃/CD₃OD) δ 2.40 (m, 2 H), 2.58 (t, J = 7.0 Hz, 2 H), 2.82 (m, 2 H), 2.98 (dd, J = 3.4, 14.1 Hz, 1 H), 3.31−3.46 (m, 3 H), 3.74−3.88 (m, 3 H), 4.13 (m, 1 H), 4.32 (d, J = 14.8 Hz, 1 H), 4.41 (d, J = 14.8 Hz, 1 H), 5.84 (s, 2 H), 6.41 (dd, J = 1.6, 7.9 Hz, 1 H), 6.49 (d, J = 1.6 Hz, 1 H), 6.60 (d, J = 7.9 Hz, 1 H), 6.74 (dd, J = 1.9, 8.3 Hz, 1 H), 6.89 (m, 1 H), 6.94 (d, J = 8.0 Hz, 1 H), 7.07−7.22 (m, 6 H), 7.67 (m, 2 H), 7.78 (m, 2 H); FAB-HRMS calcd for $C_{38}H_{37}N_3O_8Cl$ (M + H⁺) = 698.22692, obsd = 698.22560. Anal. ($C_{38}H_{36}N_3O_8Cl$) C, H, N.

Compound 3: ^{1}H NMR (400 MHz, CDCl₃) δ 2.47 (m, 2 H), 2.57 (t, J = 7.2 Hz, 2 H), 2.85 (d, J = 7.9 Hz, 2 H), 2.93 (m, 1 H), 3.21 (m, 1 H), 3.51 (m, 1 H), 3.64 (s, 3 H), 3.75 (m, 1 H), 3.83 (s, 3 H), 4.12 (m, 2 H), 5.75 (bs, 1 H), 5.91 (dd, J = 1.5, 3.1 Hz, 2 H), 6.43 (d, J = 7.7 Hz, 1 H), 6.53 (m, 3 H), 6.65 (d, J = 7.8 Hz, 1 H), 7.07−7.39 (m, 12 H), 7.72 (m, 2 H), 7.84 (m, 2 H); FAB-HRMS calcd for $C_{46}H_{46}N_3O_{10}$ (M + H⁺) = 800.31832, obsd = 800.31861. Anal. ($C_{46}H_{45}N_3O_{10}$) C, H, N.

Compound 4: ^{1}H NMR (400 MHz, CDCl₃) δ 1.04 (bs, 1 H), 1.27 (m, 1 H), 1.41−1.70 (m, 1 H), 2.03 (d, J = 2.9 Hz, 3 H), 2.25 (m, 1 H), 2.36 (m, 1 H), 2.67 (m, 2 H), 2.83 (m, 2 H), 3.00−3.14 (m, 2 H), 3.70−3.95 (m, 3 H), 3.85 (s, 6 H), 4.28 (m, 1 H), 4.55 (bd, J = 14.0 Hz, 1 H), 5.04 (s, 2 H), 5.91 (s, 2 H), 6.45 (d, J = 7.7 Hz, 1 H), 6.55 (d, J = 1.7 Hz, 1 H), 6.61 (d, J = 9.4 Hz, 1 H), 6.70 (d, J = 7.9 Hz, 1 H), 6.91 (d, J = 3.1 Hz, 1 H), 7.24−7.34 (m, 8 H), 7.46 (d, J = 6.6 Hz, 2 H). Anal. ($C_{43}H_{49}N_3O_9\cdot0.8H_2O$) C, H, N.

Compound 5: ^{1}H NMR (400 MHz, CDCl₃) δ 1.35−1.80 (m, 5 H), 2.03 (d, J = 3.9 Hz, 3 H), 2.20 (m, 1 H), 2.36 (m, 1 H), 2.65 (m, 2 H), 2.80−2.95 (m, 4 H), 3.35−3.50 (m, 2 H), 3.77 (m, 1 H), 4.20 (dd, J = 8.0, 8.9 Hz, 1 H), 4.39 (d, J = 14.8 Hz, 1 H), 4.47 (d, J = 14.8 Hz, 1 H), 4.48 (m, 1 H), 4.96 (bd, 1 H), 5.91 (s, 2 H), 6.44 (m, 1 H), 6.54 (m, 1 H), 6.71 (d, J = 7.8 Hz, 1 H), 6.79 (d, J = 7.0 Hz, 1 H), 6.92 (s, 1 H), 7.00 (m, 1 H), 7.15−7.27 (m, 5 H). Anal. ($C_{35}H_{40}N_3O_7Cl$) C, H, N.

Compound 6: ^{1}H NMR (400 MHz, CDCl₃) δ 1.26 (m, 1 H), 1.40−1.62 (m, 3 H), 1.82 (bs, 1 H), 2.05 (s, 3 H), 2.20 (m, 1 H), 2.36 (m, 1 H), 2.67 (m, 2 H), 2.80−2.99 (m, 4 H), 3.45 (m, 2 H), 4.17 (dd, J = 7.6, 16 Hz, 1 H), 4.45 (m, 1 H), 4.53 (bd, J = 13 Hz, 1 H), 4.86 (bd, J = 29 Hz, 1 H), 5.93 (s, 2 H), 6.44 (m, 2 H), 6.56 (m, 2 H), 6.73 (m, 2 H), 6.98 (dd, J = 4.4, 9.6 Hz, 1 H), 7.08 (t, J = 8.0 Hz, 1 H), 7.17−7.32 (m, 5 H). Anal. ($C_{35}H_{41}N_4O_6Cl\cdot1.3H_2O$) C, H, N.

Compound 7: ^{1}H NMR (400 MHz, CDCl₃) δ 1.18−1.45 (bm, 1 H), 1.64−1.87 (bm, 4 H), 2.02 (m, 1 H), 2.24 (bs, 3 H), 2.66 (m, 2 H), 2.83 (m, 2 H), 3.03 (d, J = 7.2 Hz, 2 H), 3.09 (d, J = 13.4 Hz, 1 H), 3.45 (m, 1 H), 3.47 (m, 3 H), 3.75 (s, 6 H), 3.77 (m, 1 H), 4.27 (m, 1 H), 5.04 (s, 2 H), 6.66 (d, J = 9.1 Hz, 1 H), 6.81 (d, J = 9.4 Hz, 2 H), 6.92 (s, 2 H), 6.94 (d, J = 8.6 Hz, 2 H), 7.18−7.35 (m, 8 H), 7.46 (d, J = 6.6 Hz, 2 H). Anal. ($C_{42}H_{51}N_3O_7$) C, H, N.

Compound 8: ^{1}H NMR (400 MHz, CDCl₃) δ 1.19 (bd, 1 H), 1.34 (bd, 1 H), 1.50 (bq, 1H), 1.90−2.20 (m, 4 H), 2.43 (m, 2 H), 2.67 (m, 2 H), 3.00−3.11 (m, 4 H), 3.43 (m, 3 H), 3.74 (s, 3 H), 3.79 (m, 1 H), 3.85 (s, 6 H), 4.30 (m, 1 H), 5.04 (s, 2 H), 6.88 (m, 4 H), 6.94 (s, 2 H), 6.95 (d, J = 8.5 Hz, 2 H), 7.15−7.34 (m, 7 H), 7.46 (d, J = 7.2 Hz, 2 H). Anal. ($C_{41}H_{49}N_3O_7\cdot0.9F_3CCO_2H$) C, H, N.

Compound 9: ^{1}H NMR (400 MHz, CDCl₃) δ 1.40 (m, 1 H), 1.67−1.95 (m, 4 H), 2.02 (m, 1 H), 2.23 (s, 3 H), 2.62 (t, J = 7.0 Hz, 2 H), 2.83−3.02 (m, 6 H), 3.34−3.46 (m, 2 H), 3.76 (m, 2 H), 4.19 (dd, J = 8.9, 8.9 Hz, 1 H), 4.38 (d, J = 14.8 Hz, 1 H), 4.46 (d, J = 14.8 Hz, 1 H), 5.12 (bs, 1 H), 5.90 (s, 2 H), 6.44 (dd, J = 1.6, 8.0 Hz, 1 H), 6.53 (d, J = 1.6 Hz, 1 H), 6.70 (d, J = 7.8 Hz, 1 H), 6.78 (dd, J = 2.0, 8.0 Hz, 1 H), 6.91 (t, J = 2.2 Hz, 1 H), 6.99 (d, J = 6.0 Hz, 1 H), 7.16−7.27 (m, 6 H). Anal. ($C_{34}H_{40}N_3O_6Cl$) C, H, N.

Compound 10: ^{1}H NMR (400 MHz, CDCl₃) δ 1.35 (m, 3 H), 1.52 (m, 1 H), 1.68 (bq, 2 H), 1.84 (m, 2 H), 2.27 (bs, 1 H), 2.64 (m, 2 H), 2.82−2.95 (m, 5 H), 3.04 (q, J = 7.4 Hz, 1 H), 3.39−3.47 (m, 2 H), 3.75 (s, 3 H), 3.77 (m, 1 H), 4.20 (dd, J = 8.7, 8.7 Hz, 1 H), 4.38 (d, J = 14.8 Hz, 1 H), 4.46 (d, J = 14.8

Hz, 1 H), 5.11 (bs, 1 H) 6.80 (m, 3 H), 6.93 (m, 3 H), 6.99 (d, $J = 8.8$ Hz, 2 H), 7.18–7.28 (m, 5 H). Anal. ($C_{34}H_{42}N_3O_5Cl\cdot 2F_3CCO_2H$) C, H, N.

Compound 11: ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.18 (m, 1 H), 1.38 (m, 1 H), 1.50–1.85 (m, 2 H), 2.17 (m, 1 H), 2.54–2.70 (m, 4 H), 2.87–2.98 (m, 4 H), 3.19 (bd, $J = 13.0$ Hz, 1 H), 3.42–3.60 (m, 2 H), 3.72 (s, 3 H), 3.73–3.80 (m, 1 H), 4.19 (dd, $J = 8.6, 8.6$ Hz, 1 H), 4.41 (d, $J = 15.0$ Hz, 1 H), 4.48 (d, $J = 15.0$ Hz, 1 H), 5.18 (bs, 1 H), 6.79 (m, 4 H), 6.96 (m, 5 H), 7.15–7.25 (m, 5 H). Anal. ($C_{33}H_{41}N_3O_5Cl_2\cdot0.7F_3CCO_2H$) C, H, N.

Supporting Information Available: Figures containing all side chains employed to prepare the initial cathepsin D library and the side chains used in the preparation of library R^B. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Wyler, D. J. Malaria-Overview and Update. *Clin. Infect. Dis.* **1993**, *16*, 449–458.
(2) *Malaria: Obstacles and Opportunities: a Report of the Committee for the Study on Malaria Prevention and Control: Status Review and Alternative Strategies, Division of International Health, Institute of Medicine.* Oaks, S. C., Jr.; Mitchell, V. S., Pearson, G. W., Carpenter, C. C. J., Eds.; National Academy Press: Washington, DC, 1991.
(3) Foote, S. J.; Cowman, A. F. The Mode of Action and the Mechanism of Resistance to Antimalarial Drugs. *Acta Trop.* **1994**, *56*, 157–171.
(4) Murray, M. C.; Perkins, M. E. Chemotherapy of Malaria. *Annu. Rep. Med. Chem.* **1996**, *31*, 141–150.
(5) Francis, S. E.; Sullivan, D. J., Jr.; Goldberg, D. E. Hemoglobin Metabolism in the Malaria Parasite *Plasmodium falciparum*. *Annu. Rev. Microbiol.* **1997**, *51*, 97–123.
(6) 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. Structure and Inhibition of plasmepsin II, a Hemoglobin-Degrading Enzyme from *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034–10039.
(7) Carroll, C. D.; Patel, H.; Johnson, T. O.; Guo, T.; Orlowski, M.; He, M.; Cavallaro, C. L.; Guo, J.; Oksman, A.; Gluzman, I. Y.; Connelly, J.; Chelsky, D.; Goldberg, D. E.; Dolle, R. E. Identification of Potent Inhibitors of *Plasmodium falciparum* Plasmepsin II from an Encoded Statine Combinatorial Library. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2315–2320.
(8) Goldberg, D. E. Hemoglobin Degradation in Plasmodium-Infected Red Blood Cells. *Semin. Cell Biol.* **1993**, *4*, 355–361.
(9) Luker, K. E.; Francis, S. E.; Gluzman, I. Y.; Goldberg, D. E. Kinetic Analysis of Plasmepsins I and II, Aspartic Proteases of the *Plasmodium falciparum* Digestive Vacuole. *Mol. Biochem. Parasitol.* **1996**, *79*, 71–78.
(10) Kick, E. K.; Roe, D. C.; Skillman, A. G.; Liu, G.; Ewing, T. J. A.; Sun, Y.; Kuntz, I. D.; Ellman, J. A. Structure-Based Design and Combinatorial Chemistry Yield Low Nanomolar Inhibitors of Cathepsin D. *Chem. Biol.* **1997**, *4*, 297–307.
(11) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
(12) Meng, E. C.; Gschwend, D. A.; Blaney, J. M.; Kuntz, I. D. Orientational Sampling and Rigid-Body Minimization in Molecular Docking. *Proteins: Struct. Funct. Genet.* **1993**, *17*, 266–278.
(13) Ewing, T. J. A.; Kuntz, I. D. Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening. *J. Comput. Chem.* **1992**, *18*, 1175–1189.
(14) Kick, E. K.; Ellman, J. A. Expedient Method For The Solid-Phase Synthesis Of Aspartic Acid Protease Inhibitors Directed Toward The Generation Of Libraries. *J. Med. Chem.* **1995**, *38*, 1427–1430.
(15) Lee, C. E.; Kick, E. K.; Ellman, J. A. General Solid-phase Synthesis Approach To Prepare Mechanism-Based Aspartyl Protease Inhibitor Libraries. Identification of Potent Cathepsin D Inhibitors. *J. Am. Chem. Soc.* **1998**, *120*, 9735–9747.
(16) Side chains for both the 1000-compound Cat D library and the 39-compound Cat D library are provided as Supporting Information (Supporting Information Figures 1 and 2).[10]
(17) Available Chemicals Database, version 95.1; Molecular Design Ltd., San Leandro, CA.
(18) Silva, A. Personal communication.
(19) Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Smith, C. E.; Duffin, K. L.; Goldberg, D. E. Order and Specificity of the *Plasmodium falciparum* Hemoglobin Degradation Pathway. *J. Clin. Invest.* **1994**, *93*, 1602–1608.
(20) Olson, R. E.; Christ, D. D. Plasma Protein Binding of Drugs. *Annu. Rep. Med. Chem.* **1996**, *31*, 327–336.
(21) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25.
(22) Kaldor, S. W.; Kalish, V. J.; Davies, J. F.; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.; Dressman, B. A.; Hatch, S. D.; Khalil, D. A.; Kosa, M. B.; Lubbehusen, P. P.; Muesing, M. A.; Patick, A. K.; Reich, S. H.; Su, K. S.; Tatlock, J. H. Viracept (Nelfinavir Mesylate, AG1343): A Potent, Orally Bioavailable, Inhibitor of HIV-1 Protease. *J. Med. Chem.* **1997**, *40*, 3979–3985 and references therein.
(23) Erickson, J. W. Personal communication.
(24) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
(25) Sybyl, version 6.3; TRIPOS Associates, St. Louis, MO.
(26) Skillman, A. G.; Kuntz, I. D. Unpublished results.
(27) ClogP, Fingerprint, Nearneighbors, Merlin, Prado, Rubicon, and Toolkit Software, version 4.51; Daylight Chemical Information Systems, Inc., Santa Fe, NM.
(28) Ewing, T. J. A.; Kuntz, I. D. Unpublished results.
(29) Matayoshi, E. D.; Wang, G. T.; Krafft, G. A.; Erickson, J. Novel Fluorogenic Substrates for Assaying Retroviral Proteases by Resonance Energy Transfer. *Science* **1990**, *247*, 954–958.
(30) Substrate was initally synthesized as described for similar fluorogenic substrates in Krafft, G. A.; Wang, G. T. Synthetic Approaches to Continuous Assays of Retroviral Proteases. *Methods Enzymol.* **1994**, *241*, 70–86, and later purchased (Bachem California Inc., Torrance, CA).
(31) Williams, J. W.; Morrison, J. F. The Kinetics of Reversible Tight-Binding Inhibition. *Methods Enzymol.* **1979**, *63*, 437–467.

JM980641T