Identification of Novel Parasitic Cysteine Protease Inhibitors by Use of Virtual Screening. 2. The Available Chemical Directory

Prashant V. Desai,[†] Akshay Patny,[†] Jiri Gut,[‡] Philip J. Rosenthal,[‡] Babu Tekwani,[§] Anuradha Srivastava,[§] and Mitchell Avery*[,†,§,||]

Department of Medicinal Chemistry, School of Pharmacy, P.O. Box 1848, University of Mississippi, University, Mississippi 38677-1848, Department of Chemistry and Biochemistry, P.O. Box 1848, University of Mississippi, University, Mississippi 38677-1848, Department of Medicine, San Francisco General Hospital, University of California, San Francisco, Box 0811, San Francisco, California 94143, and National Center for Natural Products Research, School of Pharmacy, P.O. Box 1848, University of Mississippi, University, Mississippi 38677-1848

Received June 17, 2005

The incidence of parasitic infections such as malaria, leishmaniasis, and trypanosomiasis has been steadily increasing. Since the existing chemotherapy of these diseases suffers from lack of safe and effective drugs and/or the presence of widespread drug resistance, there is an urgent need for development of potent, mechanism-based antiparasitic agents against these diseases. Cysteine proteases have been established as valid targets for this purpose. The Available Chemical Directory consisting of nearly 355 000 compounds was screened in silico against the homology models of plasmodial cysteine proteases, falcipain-2, and falcipain-3, to identify structurally diverse non-peptide inhibitors. The study led to identification of 22 inhibitors of parasitic cysteine proteases out of which 18 compounds were active against falcipain-2 and falcipain-3. Eight compounds exhibited dual activity against both enzymes. Additionally, four compounds were found to inhibit *L. donovani* cysteine protease. While one of the cysteine protease inhibitors also exhibited in vitro antiplasmodial activity with an IC_{50} value of 9.5 μM, others did not show noticeable antiplasmodial activity up to 20 μM. A model identifying important pharmacophoric features common to the structurally diverse falcipain-2 inhibitors has also been developed. Very few potent non-peptide inhibitors of the parasitic cysteine proteases have been reported so far, and identification of these novel and chemically diverse inhibitors should provide leads to be optimized into candidates to treat protozoal infections.

Introduction

According to reports by World Health Organization, nearly 3.5 billion people or more than 50% of the world's population suffer from some type of parasitic infection.[1] Many of these diseases cause chronic infections, leading to severe morbidity in numerous cases. Although efforts are being made to control these diseases, the incidence of parasitic infections has been increasing at an alarming rate.[2] Most of these diseases have received limited consideration from the pharmaceutical industry for development of new drugs[3] or vaccines.[4] Malaria, leishmaniasis, and trypanosomiasis are major parasitic diseases in developing countries caused by protozoa of the genus *Plasmodium*, *Leishmania*, and *Trypanosoma*, respectively. Worldwide 350−500 million clinical episodes of malaria are estimated to occur each year, resulting in more than 1 million deaths.[5] Of the four species of *Plasmodium* responsible for the disease, *Plasmodium falciparum* is the most lethal. The increasing resistance of malaria parasites to conventional drug therapy is a worldwide public health threat. Chagas' disease (American trypanosomiasis) is an endemic disease prevalent mostly in Latin America and is caused by *Trypanosoma cruzi*. It affects nearly 16−18 million people and is responsible for about 50 000 deaths each year.[6] Leishmaniasis continues to remain a severe public health problem in nearly 88 countries, with an estimated global prevalence of 12 million cases annually.[7] In addition, co-infection with *Leishmania* is emerging as an extremely serious complication of HIV infection.[8] Thus, the existing situation clearly suggests an urgent need for development of potent, mechanism-based antiparasitic agents against these diseases.

Cysteine proteases, with their unique nucleophilicity, constitute a pivotal class of enzymes with proposed key roles in multiple processes such as immunoevasion, enzyme activation, hydrolysis of host proteins, and cellular invasion in parasites. They are also characterized by adaptability to different substrates and stability in diverse biological environments.[9] Inhibition of parasitic cysteine proteases thus presents a potential strategy for combating parasitic infections. Three structurally different classes of cysteine proteases serve as targets for drug design: papain family, caspases, and Picornaviridae family.[10] Papain-like cysteine proteases from *L. donovani*, *P. falciparum* (falcipain-2 and falcipain-3), and *Trypanosoma cruzi* (cruzain) have been identified and functionally characterized.[9,11−13] Falcipain-2 plays a key role in hemoglobin hydrolysis in trophozoites. Knockout of this enzyme causes inhibition of hemoglobin hydrolysis in trophozoites and increased susceptibility to cysteine protease inhibitors, although the knockout is not lethal.[14] Falcipain-3, which is expressed later in the life cycle,[12] appears to be essential, as the enzyme could not be knocked out, but replacement of the gene with a functional copy was successful (Sijwali and Rosenthal, unpublished results). The other two *P. falciparum* cysteine proteases, namely, falcipain-1[15] and falcipain-2′ (Sijwali and Rosenthal, unpublished results), could be knocked out without deleterious effects. Thus, it appears that falcipain-2 and falcipain-3 are the key cysteine proteases essential for the parasite survival.

Numerous studies have established the effectiveness of peptidyl cysteine protease inhibitors against cultured *P. falci-*

*Corresponding author. Tel: +1-662-915-5879. Fax: +1-662-915-5638. E-mail: mavery@olemiss.edu.

[†] Department of Medicinal Chemistry, School of Pharmacy, University of Mississippi.

[‡] University of California.

[§] National Center for Natural Products Research, School of Pharmacy, University of Mississippi.

[||] Department of Chemistry and Biochemistry, University of Mississippi.

10.1021/jm0505765 CCC: $33.50
Published on Web 02/14/2006

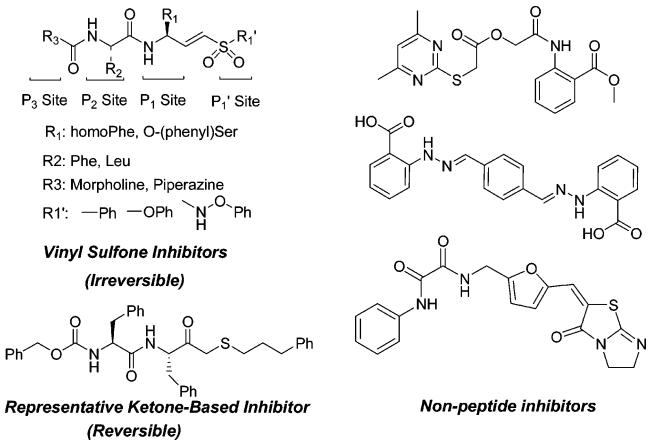

Figure 1. Representative structures of reported parasitic cysteine protease inhibitors.

parum and in murine malaria models.[9] Most of the earlier attempts to design cysteine protease inhibitors have focused on irreversible inhibitors such as vinyl sulfones (Figure 1) which are known to inhibit papain-family proteases, including cruzain[9] and falcipains.[16] These inhibitors tend to form covalent bonds with the thiolate of the catalytic cysteine and have nanomolar IC_{50} values. Although these compounds possess impressive potency, poor selectivity for parasitic cysteine proteases over the human cysteine proteases remains an issue of significant concern.[17] Also, it is advantageous to design reversible inhibitors to minimize potential toxicity.[17] Several highly potent and reversible inhibitors possessing an electrophilic isostere have been identified as inhibitors of human cysteine proteases.[18] Recently, aldehyde-based and ketone-based peptide compounds have been reported to be potent and reversible inhibitors of falcipains and cruzain.[19,20] The ketone-based compounds (Figure 1) are proposed to form transition-state hemithioacetal complexes with cysteine proteases. However, it is beneficial to identify non-peptide inhibitors, which are less subject to degradation by host proteases and thereby more likely to offer in vivo activity. To the best of our knowledge, only a few non-peptide inhibitors of falcipain-2 have been reported so far, and all of these have IC_{50} values in the micromolar range.[13,21−26]

The mature domains of falcipain-2, falcipain-3, cruzain, and the cysteine protease from *L. donovani* share more than 40% sequence homology, while the binding site residues exhibit about 90% homology.[9,11] Cocrystal structures of cruzain complexed with several peptide inhibitors have been reported.[27,28] While crystal structures for falcipain-2 and falcipain-3 are not yet reported, homology models for these enzymes have been developed by our research group.[26,29] On the basis of the superimposition of these structures, it has been suggested that the residues in the binding sites are highly conserved across the protozoal cysteine proteases.[13] Thus, it might be possible to identify common inhibitors of these protozoal enzymes to treat malaria, leishmaniasis, and trypanosomiasis.

Virtual screening is an emerging technology that is gaining an increased role in the drug discovery process.[30−32] The technique involves analyzing large collections of compounds, leading to smaller subsets for biological testing. It is now perceived as a complementary approach to experimental screening (high-throughput screening) and, when coupled with structural biology, promises to enhance the probability of success in the lead identification stage of the drug discovery process. Structure-based virtual screening requires computational fitting of compounds into an active site of a receptor by use of sophisticated algorithms, followed by scoring and ranking of these compounds to identify potential leads. In the first round of virtual screening against falcipain-2 and falcipain-3 by use of the ChemBridge database, a total of 24 diverse non-peptide inhibitors were identified.[13] Out of these, 12 compounds were found to be dual inhibitors of falcipain-2 and falcipain-3. Some of the representative inhibitors from this set are shown in Figure 1. In continuation of this initial effort, herein we report the results of our second round of virtual screening using the Available Chemical Directory (ACD),[33] one of the largest structure-searchable databases of commercially available compounds, leading us to identification of chemically diverse compounds as novel cysteine protease inhibitors. In addition, pharmacophore modeling has been carried out on the active set of compounds to identify common features relevant to their biological activity.

Experimental Section

Computational Tools. All calculations were performed on Silicon Graphics Octane 2 workstations, equipped with two parallel R12000 processors, V6 graphics board, and 512 MB memory. Sybyl 7.0 (Tripos Inc., St. Louis, MO) SPL scripts were used for the database filtration, while physicochemical parameters related to ADME were calculated using Cerius 2 (Accelrys Inc., San Diego, CA). Concord (Tripos Inc., St. Louis, MO) was used to generate three-dimensional (3D) coordinates of the compounds. GOLD[34] version 2.2 (CCDC, Cambridge, U.K.) was utilized for docking studies, whereas pharmacophore modeling was carried out using Catalyst 4.9 (Accelrys Inc., San Diego, CA).

Protein Preparation. Falcipain-2 and falcipain-3 possess significant structural homology in their binding site domains, and hence an overlapping binding site could be easily identified, making sure that all critical residues lining the active pocket were included in this definition. The protein preparation for the model structures was carried out as described earlier.[13] The protonation states of the charged residues were assigned by adding the hydrogen atoms at pH 5 in InsightII 2002 (Accelrys Inc., San Diego, CA). This was done to mimic their protonation state in the acidic food vacuole of *P. falciparum*. Asp, Glu, Arg, and Lys were treated as charged unless they were located in a hydrophobic environment. For both falcipain-2 and falcipain-3 models,[26,29] amino acid residues enclosed within a 12 Å radius from the catalytic sulfur atom (Cys42 in falcipain-2 and Cys51 in falcipain-3) were included in the binding site definition for docking studies.

Preliminary Docking Study. The crystal structure of cruzain with a vinyl sulfone inhibitor resolved at 1.60 Å (PDB code: 1F2A)[28] was used for preliminary docking studies and to establish the protocol, as previously described.[13] Interestingly, the poses that corresponded to the highest fitness score given by GOLD were also found to be closest to the experimental binding mode. The docked pose and experimental binding mode differed by only 1.44 Å root-mean-square deviation (rmsd) for the standard mode. The corresponding values for the 2-times speedup and the 7−8 times speedup GOLD settings were 1.8 and 2.1 Å, respectively. Orientation of the P1−P3 residues of the ligand in the docked poses was similar to that observed in the crystal structure. Hydrogen bonds with Gln19, Gly66, Asp158, and Trp177 of cruzain were also reproduced in the docked structures. The protocols thus provided a reasonable prediction of the experimental binding mode for the vinyl sulfone ligand. Additionally, 12 active compounds against falcipain-2 and falcipain-3, identified during the first round of virtual screening, were docked along with 988 randomly selected compounds in the homology model of falcipain-2 to study the enrichment provided by the docking protocol by use of the "GOLD score" to rank the compounds. Three different settings in GOLD were utilized, as mentioned above. More than 90% of the active compounds were retrieved by selecting the top 19% of the data set in the standard mode setting. The same percentage of active compounds could be retrieved by selecting the top 20% and top 28% of the data set for the 2-times speedup and 7−8 times speedup

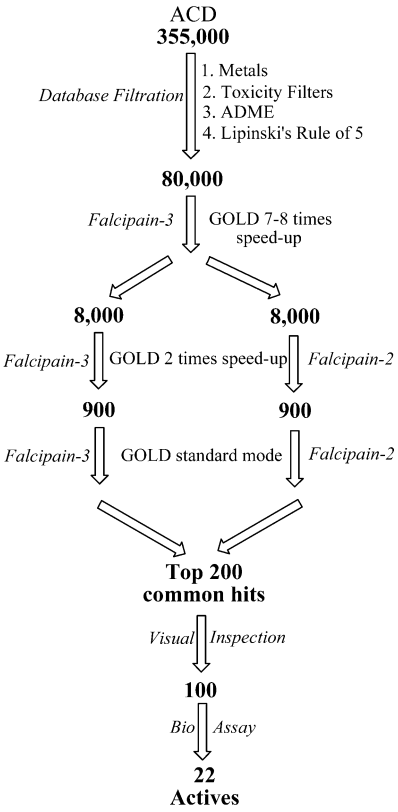

Figure 2. Protocol employed for virtual screening. The 7−8 times speedup, 2-times speedup, and standard mode represent three different sets of default parameters used for docking.

runs, respectively. These validation studies using parasitic cysteine protease structures clearly suggest that the docking protocol employed here is able to distinguish between active and inactive compounds.

Database Filtration. The Available Chemical Directory (version 2003.4)[33] containing approximately 355 000 compounds was utilized for virtual screening. When we set the Tanimoto similarity index of 0.7 as the redundancy level, 70% of the compounds in this database can be considered structurally diverse with respect to the ChemBridge database utilized in our previous virtual screening project. Database filtration was performed to collect only druglike compounds for the docking studies (Figure 2). Molecules with metals were removed, and in the case of salts, the counterions were stripped, followed by neutralization of the primary ionic species by use of a Sybyl 7.0 SPL script. One of the key issues in modeling ligand−protein complexes is setting the appropriate ionization state of the protein residues and/or the ligand functional groups.[35] The solution to this problem is quite complicated considering that it is not possible to have a *single* global model of ionization for each protein−ligand system since protons are not static and the ionization state of residues is a group function.[35,36] In an interesting report, Fornabaio et al. have attempted to address this issue by utilizing a novel protocol called "computational titration" involving the generation of parallel models of multiple ionization ensembles for each distinct protonation level that might exist for a given set of ligand−protein complexes.[37] However, such techniques are not feasible for virtual screening involving thousands of ligands. In the present case, during the preliminary docking studies to assess enrichment (vide supra), it was observed that the best results were obtained when the protein was treated as charged while the ligands were treated as neutral species. Hence, it was decided to dock the ligands in an un-ionized state for the current virtual screening.

It is beneficial to remove compounds with potentially toxic functional groups from databases, since the presence of such groups in hits generated by virtual screening might not appear to be toxicologically relevant until a later stage of development. The database was thus screened for toxic functional groups, and compounds containing metallic atoms, aromatic nitro groups, *N*-oxides, aldehydes, nitrogen and sulfur mustards, chloramines, and isocyanides were removed. Estimation of ADME properties and their optimization during a drug design project is a critical activity. The database was filtered on the basis of ADME related parameters to eliminate nondruglike molecules by use of a similar approach as described before.[13] The filtration was performed by calculating ADME_absorption parameters in Cerius 2. Molecules with an absorption level score of 0 (zero: good absorption) were selected. Average values for the polar surface area (PSA) and ALogP98 for the selected compounds were 65 and 2.1, respectively. From these, 35% of compounds categorized as "good to very soluble" were retained. The database was then subjected to the Lipinski's rule of five.[38] Compounds with molecular weight less than 500, number of H-bond donors less than 5, and acceptors less than 10 were retained. One recent study has suggested lower limits for the number of donors in druglike molecules;[39] however, the present study did not utilize this new information. Finally, the filtered database, comprising nearly 80 000 compounds, was submitted to Concord to generate 3D coordinates and subsequently used for docking studies.

Docking. Virtual screening involving docking of large databases can be computationally very expensive, so an approach that optimizes the balance between the precision of docking and the time required for the process is necessary. The initial stages of receptor-based virtual screening are generally executed to discard many compounds quickly, retaining only those which fit the receptor in an overall favorable topology. Exhaustive docking for the retained compounds can consequently be carried out to estimate their binding mode and interactions with the target receptor. Thus, in the present study, docking processes were carried out in three consecutive stages, employing three different settings in GOLD as described previously.[13] Falcipain-2 and falcipain-3 are highly homologous, and model structures of these two proteases have suggested a similar 3D structure for the binding sites.[26,29] Hence, the first stage of the docking studies (7−8 times speedup) was conducted for only one of these two enzymes. The distal end of the S2 pocket in falcipain-3 has been proposed to be narrower than that in falcipain-2.[29] Thus, the compounds which bind to falcipain-3 are also likely to fit falcipain-2, while the reverse may not always be true. This was, in fact, found to be true in our first round of virtual screening, where all 12 compounds that were active against falcipain-3 were also active against falcipain-2, but 10 compounds were only active against falcipain-2.[13] The database was thus docked only against the falcipain-3 model in the first stage, significantly reducing the computational time.

As depicted in Figure 2, an initial set of 355 000 compounds in the ACD was reduced to about 80 000 after the database filtration described above. These 80 000 compounds were submitted to the first stage of docking against the falcipain-3 model. From this run, the top 8000 (10%) compounds were selected for the second phase of docking, which was carried out against both falcipain models. Lastly, the top 900 compounds (11%) from each run were submitted to the third and final stage of docking. At each stage only the highest ranking binding poses were stored for subsequent analysis. The top 200 common hits were visually inspected based on the following criteria: (1) reasonable internal geometry of the ligand in the binding pose; (2) proximity of the electrophilic center of the ligand (if any) to the catalytic cysteine in the proposed binding mode; and (3) complementarity between the ligand and protein surfaces in terms of spatial occupancy and hydrophilic/hydrophobic regions. On the basis of these observations, a total of 100 compounds were selected and purchased from the respective vendors. The identity and purity (more than 95%) of the compounds were confirmed based on LC−MS and/or ^{1}H NMR data provided by the suppliers. Finally, the selected compounds were examined for in vitro enzyme inhibition in assays against falcipain-2 and falcipain-3. Since the *L. donovani* cysteine protease is highly homologous to the malarial cysteine proteases, the selected compounds were also tested for their ability to inhibit this enzyme. Moreover, in vitro parasite

growth inhibition assays were conducted against cultured *P. falciparum* erythrocytic stages and *L. donovani* promastigotes.

Pharmacophore Modeling. The falcipain-2 inhibitors identified in the present study were utilized to generate a pharmacophore model to identify common pharmacophoric features present in this set of structurally diverse compounds. The Hip-Hop algorithm[40] was utilized to identify and overlay common chemical features of the compounds. The algorithm uses active compounds to derive the pharmacophore without taking their biological data into account. This approach was relevant, since the range of inhibition exhibited by the active compounds was narrow, with less than 2 log units of variation. This range was not large enough to generate any meaningful activity-based (predictive) pharmacophore model by use of the Catalyst/Hypogen algorithm.[41] Hence, on the basis of the assumption that most active compounds bind in a manner similar to the receptor binding site, a qualitative 3D model of the common features required for binding of ligands to the receptor was developed. The docked conformations of the active compounds were included for generation of the pharmacophore model. The following features were included: hydrogen bond acceptor, hydrogen bond donor, hydrophobic, hydrophobic aliphatic, hydrophobic aromatic, and ring aromatic. Ten different hypotheses were generated using the "Common Features only" option. The parameters associated with hypothesis generation (see Supporting Information) were varied iteratively until reasonable pharmacophore hypotheses were generated. The 10 hypotheses were visually inspected, and the selection of pharmacophore hypothesis was based on its rank, ability to fit the active compounds, and ability to explain proposed interactions with the target as suggested by the docking study. The top five hypotheses appeared to have at least three features consistent with the docked poses. Among these, all four features of the fourth hypothesis were observed to be consistent with the docked poses, and hence it was selected for further analysis.

Conformational Search and Similarity Index. A conformational search by use of Catalyst was carried out for the falcipain-2 inhibitors identified in the present study. A maximum of 250 conformations were generated for each compound by use of the "Best Quality" option, and all conformations within 20 kcal/mol of the minimum energy conformation were collected. The receptor bound conformation for each compound was then overlapped with all the conformations generated, and a best fit was obtained for each pair.

The Tanimoto similarity indices[42] for the compounds were calculated using standard Unity 2D fingerprints as implemented in the Selector module of Sybyl 7.0. The Tanimoto coefficient between compounds A and B is calculated using the formula below, where N_A and N_B are the number of bit sets in fingerprints of compounds A and B, respectively, and N_{AB} is the number of bit sets common to both compounds.

$$\text{Tanimoto(A, B)} = N_{AB}/(N_A + N_B - N_{AB})$$

Biological Evaluation. IC_{50} values against recombinant falcipain-2 and falcipain-3 were determined as follows. Equal amounts of recombinant enzymes were incubated for 30 min at room temperature in 100 mM sodium acetate, pH 5.5, and 10 mM dithiothreitol with different concentrations of tested inhibitors. Inhibitor solutions were prepared from stocks in DMSO (maximum concentration of DMSO in the assay was 1%). After 30 min incubations, the substrate Z-Leu-Arg-AMC (benzoxycarbonyl-Leu-Arg-7-amino-4-methyl-coumarin) in the same buffer was added to a final concentration of 25 μM. Fluorescence was monitored for 15 min at room temperature in a Labsystems Fluoroskan Ascent spectrofluorometer. IC_{50} values were determined from plots of percent activity over compound concentration using GraphPad Prism software.

Effects of compounds on parasite development were determined by flow cytometry.[14] Chloroquine resistant W2 strain *P. falciparum* parasites were synchronized with 5% sorbitol and cultured at 1% parasitemia and 2% hematocrit under the atmosphere of 3% O_2, 6% CO_2, and 91% N_2 in RPMI-1640 medium supplemented with 10% human serum. Inhibitors were added from DMSO stocks (maximum concentration of DMSO equal to 0.1%). After 48 h, the culture medium was removed and replaced with 1% formaldehyde in PBS pH 7.4, and cells were fixed for 48 h at room temperature. Fixed parasites were transferred into 0.1% Triton-X-100 in PBS containing 1 nM YOYO-1 dye (Molecular Probes). Parasitemia was determined from dot plots (forward scatter vs fluorescence acquired on a FACSort flow cytometer) by use of CellQuest software (Beckton Dickinson). IC_{50} values for growth inhibition were determined from plots of percent control parasitemia over inhibitor concentration by use of GraphPad Prism software.

For the in vitro testing of the compounds against *Leishmania* cysteine protease (LCP), the major cysteine protease purified from *Leishmania donovani* promastigotes by affinity chromatography on Con A Sepharose was used.[43] The activity was assayed using a synthetic peptide substrate, Z-Phe-Arg-AMC (benzoxycarbonyl-Phe-Arg-7-amino-4-methyl-coumarin). The assay was carried out in 96 well microtiter plates. The assay mixture contained sodium acetate buffer (100 mM, pH 5.5), dithiothreitol (10 mM), 0.1% Triton-X-100, EDTA (1 mM), 10 μL of purified enzyme, and compounds in varying concentrations diluted from the stock prepared in DMSO. The plate was preincubated at 37 °C for 10 min. The reaction was initiated by addition of the substrate (10 μM final concentration), and the plates were further incubated for 30 min at 37 °C. The proteolytic activity was monitored by fluorometric quantification of 7-amino-4-methylcoumarin liberated by proteolytic cleavage of the synthetic substrate. The plate was read on a microplate reader at an emission wavelength of 355 nm and excitation wavelength of 460 nm for the fluorescence.

Antileishmanial activity of the compounds was tested in vitro on a culture of *Leishmania donovani* promastigotes. In a 96 well microplate assay, compounds with appropriate dilution were added to the *Leishmania* promastigotes culture (2×10^6 cell/mL). The plates were incubated at 26 °C for 72 h, and growth of *Leishmania* promastigotes was determined by the alamar blue assay.[44] Pentamidine and amphotericin B were used as the standard antileishmanial agents. All the compounds were simultaneously tested for cytotoxicity on VERO (monkey kidney fibroblast) cells by the neutral red assay.[45] IC_{50} values for each compound were computed from the growth inhibition curve.

Results and Discussion

Out of 100 compounds submitted for biological testing, 18 were found to be low micromolar inhibitors of falcipain-2 (Figure 3 and Table 1). As predicted, falcipain-3 was less strongly inhibited, but seven compounds had IC_{50} values below 50 μM against this enzyme; all falcipain-3 inhibitors also inhibited falcipain-2. This result validates our decision to use only the falcipain-3 model in the first stage of docking. Keeping in mind the lack of experimental structures for the target enzymes and the scarcity of non-peptide inhibitors reported for the malarial cysteine proteases, we find that the results are encouraging.

While one of the falcipain-2 inhibitors, compound **4**, also exhibited in vitro antiplasmodial activity with an IC_{50} value of 9.5 μM, others did not show noticeable antiplasmodial activity up to 20 μM. It is surprising that compound **4** exhibited antimalarial activity at a lower concentration than that required for enzyme inhibition. This might be due to selective uptake of the compound by the parasite food vacuole, the apparent principal site of falcipain activity, or involvement of an unknown mechanism. It is worth mentioning that all 18 active compounds were noncytotoxic at least up to 50 μM. At the same time, it is rather disconcerting to note the poor in vitro growth inhibition profile of these compounds. As indicated by Lipinski et al.,[38] most compounds failing the Lipinski's rule of five may prove to be troublesome; however, the converse is not true. That is, compounds passing this filter can still prove troublesome in

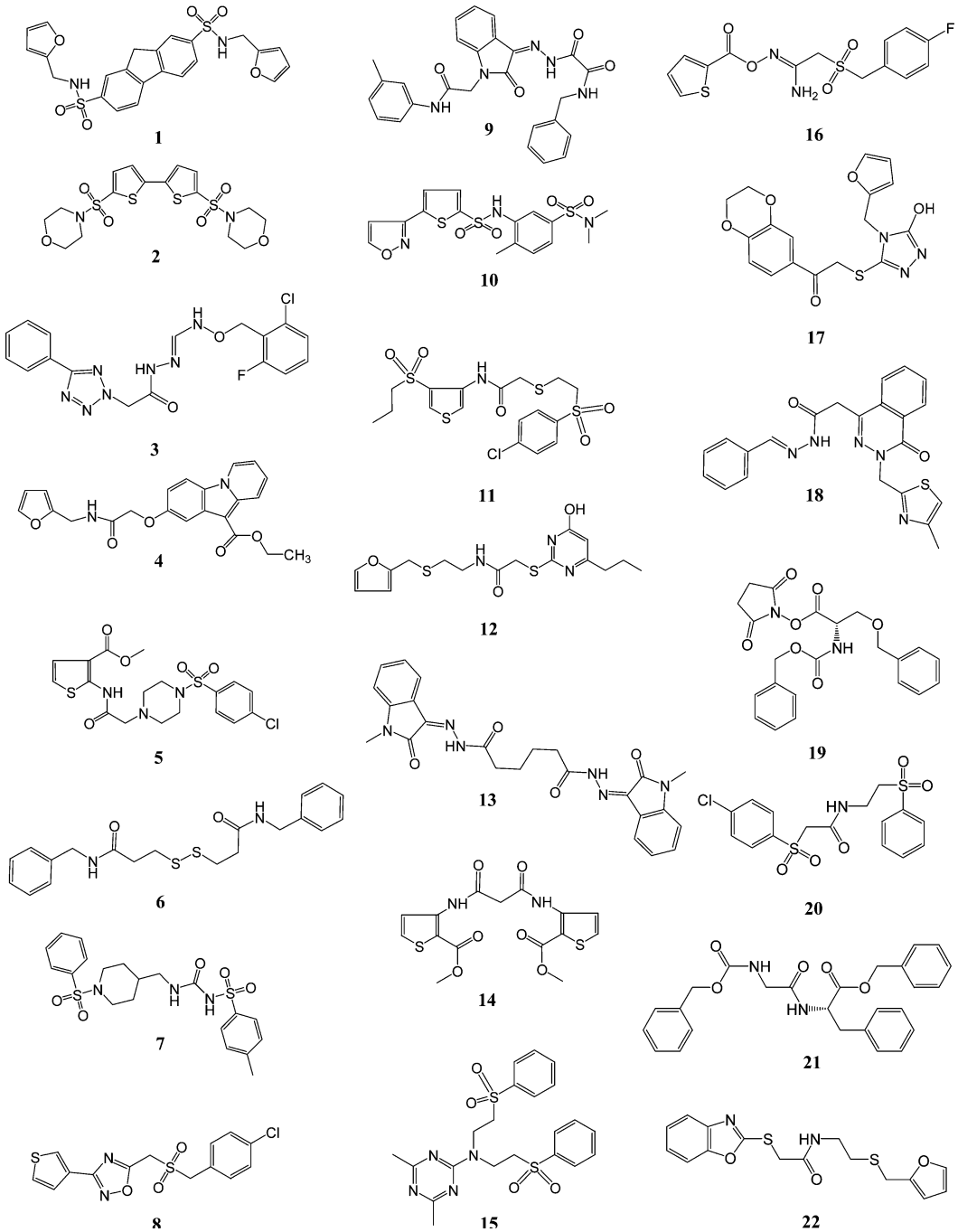

Figure 3. Structures of parasitic cysteine protease inhibitors identified using virtual screening.

experimental studies. Thus, although the falcipain-2 inhibitors identified in the present study satisfy the Lipinski's rule of five, insufficient cell permeability might still be one of the possible reasons for their poor antiplasmodial activity. It should also be noted here that in most cases cysteine protease inhibitors exhibiting potent antiplasmodial activity have been reported to have low nanomolar or better IC_{50} values against falcipain-2 and falcipain-3.[16,20] Considering this observation, further improvement in the enzyme inhibition profile of the present set of compounds might be required to enhance the antiplasmodial activity.

Superimposition of the docked poses of inhibitors (Table 1, compounds **1**−**18**) in the binding site of the falcipain-2 model provided by the last stage of docking is shown in Figure 4. All the ligands seem to have reasonable interactions with the S2 pocket, one of the major determinants of specificity for this

Table 1. Compounds Active against Falcipain-2 and Falcipain-3[a]

compd	FP-2[b] IC_{50} (μM)	FP-3[c] IC_{50} (μM)	compd	FP-2[b] IC_{50} (μM)	FP-3[c] IC_{50} (μM)
1	1.4	11.4	**11**	3.4	>50.0
2	5.4	37.7	**12**	4.6	>50.0
3	10.8	38.5	**13**	6.5	>50.0
4	13.8	31.2[d]	**14**	10.4	>50.0
5	14.3	49.2	**15**	13.2	>50.0
6	15.3	18.8	**16**	50.0	>50.0
7	19.0	34.0	**17**	52.3	>50.0
8	47.8	56.6	**18**	54.3	>50.0
9	1.6	>50.0	**E-64**	0.015	0.075
10	2.9	>50.0			

[a] All compounds were found to be noncytotoxic at least up to 50 μM in VERO cells. [b] Falcipain-2. [c] Falcipain-3. [d] The IC_{50} value for antiplasmodial activity is 9.5 μM.

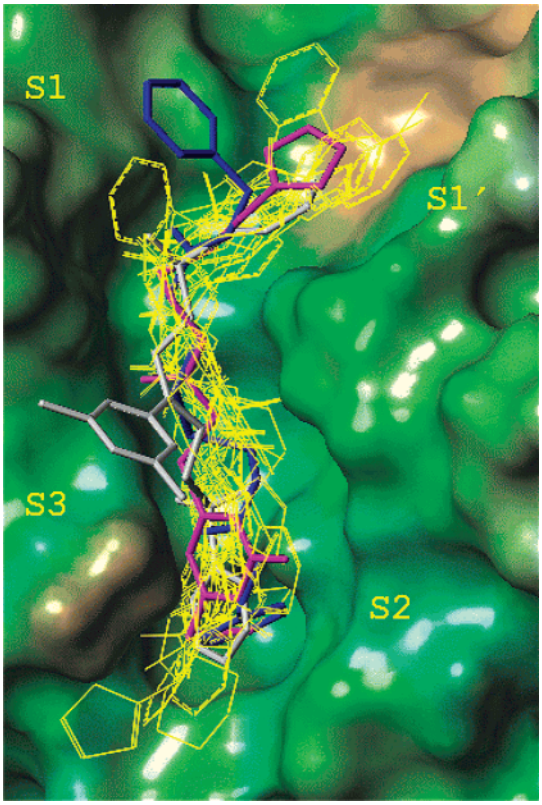

Figure 4. Superimposition of the docked conformations of inhibitors in the homology model of falcipain-2. Inhibitors are colored yellow except compounds **6** (blue), **12** (magenta), and **15** (white). The enzyme binding site is shown as a Connolly surface colored by lipophilic potentials, and the subsites are labeled as S1, S1′, S2, and S3.

class of cysteine proteases.[46] While the figure clearly reveals that collectively the compounds appear to span four major subsites of the binding pocket, S1, S1′, S2, and S3, there are subtle differences in specific interactions by individual ligands as proposed by the docking studies. For example, compound **15** (white in Figure 4) shows additional interactions with the S3 pocket but no significant interactions with the S1 pocket. On the other hand, compound **6** (blue in Figure 4) appears to have fewer interactions with the S3 pocket than those with the S1 pocket. Finally, compound **12** (magenta in Figure 4) along with several other compounds seems to effectively occupy the S1′ pocket.

A closer view of the proposed binding mode obtained by the final docking stage for compound **12** in the homology model of falcipain-2 is shown in Figure 5A. The inhibitor appears to fit well in the active pocket, showing several hydrogen-bonding and van der Waals interactions with the binding site residues. The oxygen atom of the terminal furan ring makes a hydrogen bond with a side chain amide proton of Gln36. In addition, the furan ring is also involved in $\pi-\pi$ stacking interactions with the aromatic ring of Trp206 in the S1′ pocket. The sulfur atom near the furan ring can accept a hydrogen bond from the neighboring imidazole proton of His174 with a distance of 2.8 Å. The amide linker of the inhibitor appears to form a hydrogen bond with the backbone carbonyl of Asn173 and the backbone amide proton of Gly183. Additionally, the pyrimidine nitrogen atoms appear to accept hydrogen bonds from the backbone amide proton of Ile85 and the side chain hydroxyl of Ser149 in the S2 pocket. The *n*-propyl group attached to the pyrimidine ring is involved in hydrophobic interactions with Ile85, Ala235, and Phe236. Most of the falcipain-2 inhibitors identified in the present study appear to have similar interactions with the residues in the active pocket.

To identify common features responsible for activity of this diverse set of compounds, 10 different pharmacophore hypotheses were generated with scores ranging from 109.6 to 136.6. Out of these, the hypothesis that correlated well with the ligand−receptor interactions in the docked complex is shown in Figure

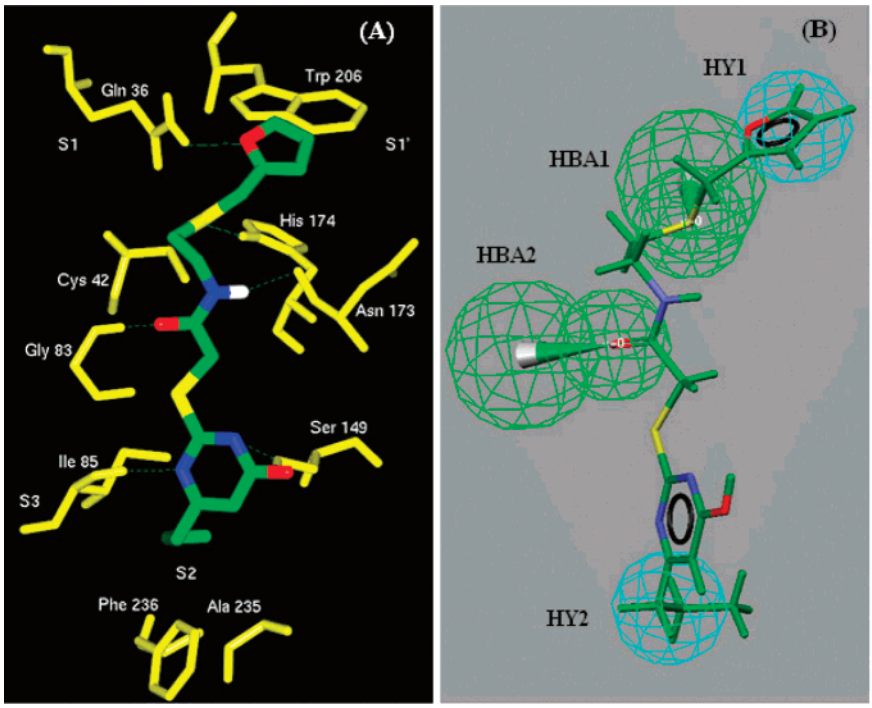

Figure 5. Binding mode and pharmacophore fitting for compound **12**. (A) Proposed binding mode in homology model of falcipain-2. The compound is shown as sticks colored by atom types. Critical residues of the binding pocket are shown as capped sticks in yellow. The hydrogen bonds are shown as dotted green lines. The subsites are labeled as S1, S1′, S2, and S3. (B) Pharmacophore model superimposed on compound **12**. The inhibitor is shown as sticks colored by atom types. HY1 and HY2 represent the hydrophobic regions, whereas HBA1 and HBA2 represent the hydrogen bond acceptor regions.

Table 2. Compounds Active against *Leishmania donovani* Cysteine Protease[a]

compd	LCP[b] IC_{50} (μM)
19	23.5
20	24.9
21	31.4
22	43.0
leupeptin	0.01

[a] All compounds were found to be noncytotoxic at least up to 50 μM in VERO cells. [b] *Leishmania donovani* cysteine protease.

5B, superimposed with one of the hits, compound **12**. The model is composed of four chemical features: two hydrogen-bond acceptors HBA1 and HBA2 (green) and two hydrophobic regions HY1 and HY2 (cyan). Compound **12** maps well onto the four features of the pharmacophore model. The hydrophobic furan ring and the *n*-propyl group appear to be aligned with HY1 and HY2, respectively, while the polar sulfur atom near the furan ring and carbonyl in the amide linker occupy the HBA1 and HBA2 regions, respectively. The model also shows remarkable complementarity to the observed ligand-binding site interactions in the proposed binding mode (Figure 5A). The HY1 feature surrounding the furan ring of the inhibitor indicates a region corresponding to hydrophobic space in proximity of the S1′ pocket and correlates to the $\pi-\pi$ interaction with Trp206. Similarly, HY2 near the *n*-propyl chain of the inhibitor corresponds to the hydrophobic region formed by Ile85, Phe236, and Ala235 in the S2 pocket. The presence of HBA1 near the sulfur atom might indicate its ability to form a hydrogen bond with His174. Finally, HBA2 near the amide carbonyl of the inhibitor suggests that a donor atom from the receptor in this region can form a hydrogen-bonding interaction. This is clearly seen as the case with the Gly83 backbone amide proton forming a critical hydrogen bond with the carbonyl oxygen of the inhibitor. These pharmacophoric features are common to several falcipain-2 inhibitors identified in the present study, and thus their presence and a few key individual interactions might mediate their activity. Also, more than 60% of the falcipain-2 inhibitors identified from the ChemBridge database[13] appear to fit the present pharmacophore model. Interestingly, in the case of compounds with electrophilic groups such as α-heterosubstituted carbonyl (compound **17**) or hydrazone (compounds **3**, **9**, **13**, **16**, and **18**), the carbonyl oxygen appears to be aligned with the HBA1 region (except for compounds **3** and **16**). Considering the corresponding location of the HBA1 region in the binding pocket (Figure 5), the electrophilic groups would be expected to be placed in the proximity of the catalytic cysteine (Cys42 in falcipain-2), resulting in strong chemical interactions. Similar observations can also be made for the hydrazone and the imine containing falcipain-2 inhibitors identified in the previous virtual screening project.[13]

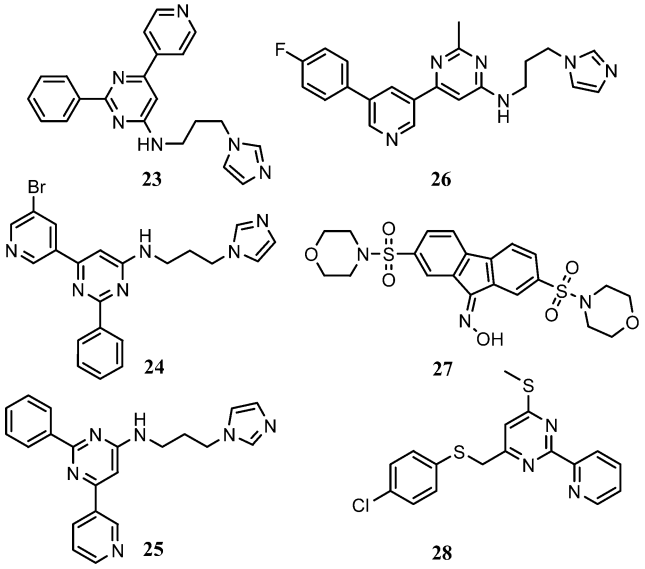

Figure 6. Structures of in vitro growth inhibitors of *P. falciparum* and *L. donovani*.

It should be noted that the present pharmacophore model is unable to distinguish between the compounds inhibiting only falcipain-2 and the compounds exhibiting dual inhibition of falcipain-2 and falcipain-3. As discussed earlier, the model structures of falcipain-2 and falcipain-3 reveal that the binding site residues are highly conserved but that falcipain-3 has a smaller S2 pocket than falcipain-2.[29] It is not surprising that the present pharmacophore model is not able to identify differences between these similar enzymes, since it is challenging to explain steric issues in a pharmacophore model. Several recent studies have shown that inclusion of pharmacophore-based filtration in addition to receptor-based virtual screening can significantly improve the probability of identifying potential hits.[47,48] Thus, the current pharmacophore model would be useful as an additional filter during future virtual screening.

A conformational search was carried out for the falcipain-2 inhibitors identified in the present study, and conformations were then superimposed on the proposed receptor bound conformations. For every docked conformation, a matching conformation was found in the set of generated conformers. Further, for 13 out of the 18 active compounds, the relative energy for the best fit conformation was within 10 kcal/mol of the lowest energy conformation, and 4 out of the top 5 active compounds met this criterion. Thus, the proposed binding modes of the active compounds do not appear to have significant intramolecular strain.

The average pairwise Tanimoto similarity index[42] was 0.30 for the set of 100 tested compounds, suggesting that the compounds were chemically diverse in nature (lower Tanimoto

Table 3. Biological Data for Compounds Exhibiting Antimalarial as Well as Antileishmanial Activity

compd	antimalarial W2[a] IC_{50} (μM)	antileishmanial[b] IC_{50} (μM)	antilesihmanial[b] IC_{90} (μM)	cytotoxicity[c] IC_{50} (μM)
23	0.13	50.5	98.2	NC
24	0.44	7.8	15.4	32.2
25	0.94	12.6	56.1	40.7
26	2.6	41.2	85.0	41.2
27	6	52.7	81.0	NC
28	20	50.0	105.6	NC
artemisinin	0.008			
amphotericin B		0.04	0.3	
pentamidine		2.5	8.3	

[a] Antimalarial activity against chloroquine resistant (W2) strain of *P. falciparum*. [b] Antileishmanial activity against *Leishmania donovani* promastigotes. [c] Cytotoxicity in VERO (monkey kidney fibroblast) cells; NC: non cytotoxic at least up to 50 μM.

values are indicative of more chemical diversity). The average pairwise Tanimoto similarity index for the 18 falcipain-2 inhibitors (Table 1) was 0.28, also suggesting a chemically diverse profile. Moreover, the 18 active compounds were compared with the 22 active compounds previously identified in our virtual screening of the ChemBridge database.[13] This comparison showed that if we set the Tanimoto similarity index of 0.7 as the redundancy level, 89% of the compounds in the current set are structurally novel with respect to the former set. This implies that most of the inhibitors identified in the second virtual screening run by use of the ACD are chemically diverse from those discovered in the previous run.

Four additional compounds were identified as inhibitors of *L. donovani* cysteine protease in micromolar concentrations (Table 2, Figure 3) with no apparent cytotoxicity at least up to 50 μM. Moreover, six compounds (Figure 6) exhibited in vitro inhibition of both *P. falciparum* and *L. donovani* in micromolar concentrations as shown in Table 3. Since none of these compounds were active against the respective cysteine proteases, a different mechanism of action appears to be involved. Apart from these compounds, an additional 13 compounds showed activity against either *P. falciparum* or *L. donovani* with IC_{50} values in the micromolar range. The biological data and structures of these compounds are provided in the Supporting Information (Table S1, Figure S1).

Conclusions

The Available Chemical Directory comprising nearly 355 000 compounds was screened in silico against the homology models of falcipain-2 and falcipain-3 to identify structurally diverse parasitic cysteine protease inhibitors. The report provides an example of optimum utilization of computational resources implementing stepwise docking strategy for virtual screening. Out of 100 compounds selected for biological screening, 22 cysteine protease inhibitors were identified. Eighteen compounds were active against falcipain-2 and eight were also active against falcipain-3. Additionally, four compounds inhibited the *L. donovani* cysteine protease at micromolar concentrations. However, the relatively poor in vitro growth inhibition profiles of these compounds need to be addressed during the subsequent lead optimization stage. The pharmacophore model developed in this study would be useful as an additional filter to query different databases and retrieve compounds with features known to be critical for inhibitory activity. Since few potent non-peptide inhibitors of the parasite cysteine proteases have been reported so far, identification of these novel and chemically diverse inhibitors provides initial leads for optimization into more potent and efficacious drug candidates to treat protozoal infections.

Acknowledgment. This work was supported by CDC Cooperative Agreement UR3/CCU4186520-03. We thank Prasenjit Mukherjee for useful discussion.

Supporting Information Available: Biological data and structures of compounds exhibiting either antimalarial or antileishmanial activity without inhibition of corresponding cysteine proteases, list of charged residues used for the docking studies, and Catalyst parameters employed for hypothesis generation. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Lecaille, F.; Kaleta, J.; Bromme, D. Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design. *Chem. Rev.* **2002**, *102*, 4459−4488.
(2) Brooks, D. R.; Hoberg, E. P. Parasite systematics in the 21st century: opportunities and obstacles. *Trends Parasitol.* **2001**, *17*, 273−275.
(3) McKerrow, J. H. Development of cysteine protease inhibitors as chemotherapy for parasitic diseases: insights on safety, target validation, and mechanism of action. *Int. J. Parasitol.* **1999**, *29*, 833−837.
(4) Abath, F. G. Development of vaccines against human parasitic diseases: tools, current status and perspectives. *Expert Opin. Invest. Drugs* **2000**, *9*, 301−310.
(5) *WHO World Malaria Report*; World Health Organization: Geneva, Switzerland, 2005.
(6) Godal, T.; Nagera, J. *Tropical Diseases;* WHO Division of Control in Tropical Disease, World Health Organization: Geneva, Switzerland, 1990; pp 12−13.
(7) *WHO, Weekly Epidemiol. Rec.* **2002,** *77*, 365−372.
(8) *The Leishmaniasis and Leishmania/HIV Co-Infections. Fact Sheet Number 116;* World Health Organization: Geneva, Switzerland, 2000.
(9) Sajid, M.; McKerrow, J. H. Cysteine proteases of parasitic organisms. *Mol. Biochem. Parasitol.* **2002**, *120*, 1−21.
(10) Hernandez, A. A.; Roush, W. R. Recent advances in the synthesis, design and selection of cysteine protease inhibitors. *Curr. Opin. Chem. Biol.* **2002**, *6*, 459−465.
(11) Shenai, B. R.; Sijwali, P. S.; Singh, A.; Rosenthal, P. J. Characterization of native and recombinant falcipain-2, a principal trophozoite cysteine protease and essential hemoglobinase of Plasmodium falciparum. *J. Biol. Chem.* **2000**, *275*, 29000−29010.
(12) Sijwali, P. S.; Shenai, B. R.; Gut, J.; Singh, A.; Rosenthal, P. J. Expression and characterization of the Plasmodium falciparum haemoglobinase falcipain-3. *Biochem. J.* **2001**, *360*, 481−489.
(13) Desai, P. V.; Patny, A.; Sabnis, Y.; Tekwani, B.; Gut, J.; Rosenthal, P.; Srivastava, A.; Avery, M. Identification of novel parasitic cysteine protease inhibitors using virtual screening. 1. The ChemBridge database. *J. Med. Chem.* **2004**, *47*, 6609−6615.
(14) Sijwali, P. S.; Rosenthal, P. J. Gene disruption confirms a critical role for the cysteine protease falcipain-2 in hemoglobin hydrolysis by Plasmodium falciparum. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4384−4389.
(15) Sijwali, P. S.; Kato, K.; Seydel, K. B.; Gut, J.; Lehman, J.; Klemba, M.; Goldberg, D. E.; Miller, L. H.; Rosenthal, P. J. Plasmodium falciparum cysteine protease falcipain-1 is not essential in erythrocytic stage malaria parasites. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8721−8726.
(16) Shenai, B. R.; Lee, B. J.; Alvarez-Hernandez, A.; Chong, P. Y.; Emal, C. D.; Neitz, R. J.; Roush, W. R.; Rosenthal, P. J. Structure−activity relationships for inhibition of cysteine protease activity and development of Plasmodium falciparum by peptidyl vinyl sulfones. *Antimicrob. Agents Chemother.* **2003**, *47*, 154−160.
(17) Engel, J. C.; Doyle, P. S.; Hsieh, I.; McKerrow, J. H. Cysteine protease inhibitors cure an experimental Trypanosoma cruzi infection. *J. Exp. Med.* **1998**, *188*, 725−734.
(18) Leung, D.; Abbenante, G.; Fairlie, D. P. Protease inhibitors: current status and future prospects. *J. Med. Chem.* **2000**, *43*, 305−341.
(19) Huang, L.; Lee, A.; Ellman, J. A. Identification of potent and selective mechanism-based inhibitors of the cysteine protease cruzain using solid-phase parallel synthesis. *J. Med. Chem.* **2002**, *45*, 676−684.
(20) Lee, B. J.; Singh, A.; Chiang, P.; Kemp, S. J.; Goldman, E. A.; Weinhouse, M. I.; Vlasuk, G. P.; Rosenthal, P. J. Antimalarial activities of novel synthetic cysteine protease inhibitors. *Antimicrob. Agents Chemother.* **2003**, *47*, 3810−3814.
(21) Batra, S.; Sabnis, Y. A.; Rosenthal, P. J.; Avery, M. A. Structure-based approach to falcipain-2 inhibitors: synthesis and biological evaluation of 1,6,7-trisubstituted dihydroisoquinolines and isoquinolines. *Bioorg. Med. Chem.* **2003**, *11*, 2293−2299.
(22) Chiyanzu, I.; Clarkson, C.; Smith, P. J.; Lehman, J.; Gut, J.; Rosenthal, P. J.; Chibale, K. Design, synthesis and anti-plasmodial evaluation in vitro of new 4-aminoquinoline isatin derivatives. *Bioorg. Med. Chem.* **2005**, *13*, 3249−3261.
(23) Chiyanzu, I.; Hansell, E.; Gut, J.; Rosenthal, P. J.; McKerrow, J. H.; Chibale, K. Synthesis and evaluation of isatins and thiosemicarbazone derivatives against cruzain, falcipain-2 and rhodesain. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3527−3530.
(24) Dominguez, J. N.; Leon, C.; Rodrigues, J.; Gamboa de Dominguez, N.; Gut, J.; Rosenthal, P. J. Synthesis and evaluation of new antimalarial phenylurenyl chalcone derivatives. *J. Med. Chem.* **2005**, *48*, 3654−3658.
(25) Greenbaum, D. C.; Mackey, Z.; Hansell, E.; Doyle, P.; Gut, J.; Caffrey, C. R.; Lehrman, J.; Rosenthal, P. J.; McKerrow, J. H.; Chibale, K. Synthesis and structure−activity relationships of parasiticidal thiosemicarbazone cysteine protease inhibitors against Plasmodium falciparum, Trypanosoma brucei, and Trypanosoma cruzi. *J. Med. Chem.* **2004**, *47*, 3212−3219.

(26) Sabnis, Y.; Rosenthal, P. J.; Desai, P.; Avery, M. A. Homology modeling of falcipain-2: validation, de novo ligand design and synthesis of novel inhibitors. *J. Biomol. Struct. Dyn.* **2002**, *19*, 765−774.
(27) Cazzulo, J. J. Proteinases of Trypanosoma cruzi: potential targets for the chemotherapy of Changas desease. *Curr. Top. Med. Chem.* **2002**, *2*, 1261−1271.
(28) Brinen, L. S.; Hansell, E.; Cheng, J.; Roush, W. R.; McKerrow, J. H.; Fletterick, R. J. A target within the target: probing cruzain's P1′ site to define structural determinants for the Chagas' disease protease. *Struct. Folding Des.* **2000**, *8*, 831−840.
(29) Sabnis, Y. A.; Desai, P. V.; Rosenthal, P. J.; Avery, M. A. Probing the structure of falcipain-3, a cysteine protease from Plasmodium falciparum: comparative protein modeling and docking studies. *Protein Sci.* **2003**, *12*, 501−509.
(30) Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375−382.
(31) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.
(32) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening—an overview. *Drug Discovery Today* **1998**, *3*, 160−178.
(33) Elsevier MDL. http://www.mdl.com.
(34) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(35) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. The determinants of pK_as in proteins. *Biochemistry* **1996**, *35*, 7819−7833.
(36) Dixon, H. B.; Clarke, S. D.; Smith, G. A.; Carne, T. K. The origin of multiply sigmoid curves of pH-dependence. The partitioning of groups among titration p*K* values. *Biochem. J.* **1991**, *278* (Part 1), 279−284.
(37) Fornabaio, M.; Cozzini, P.; Mozzarelli, A.; Abraham, D. J.; Kellogg, G. E. Simple, intuitive calculations of free energy of binding for protein−ligand complexes. 2. Computational titration and pH effects in molecular models of neuraminidase-inhibitor complexes. *J. Med. Chem.* **2003**, *46*, 4487−4500.
(38) 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 Delivery Rev.* **1997**, *23*, 3−25.
(39) Proudfoot, J. R. The evolution of synthetic oral drug properties. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1087−1090.
(40) Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. Identification of common functional configurations among molecules. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563−571.
(41) *Catalyst*, version 4.9; Accelrys: San Diego, CA.
(42) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood behavior: a useful concept for validation of "molecular diversity" descriptors. *J. Med. Chem.* **1996**, *39*, 3049−3059.
(43) Labriola, C.; Sousa, M.; Cazzulo, J. J. Purification of the major cysteine proteinase (cruzipain) from Trypanosoma cruzi by affinity chromatography. *Biol. Res.* **1993**, *26*, 101−107.
(44) Mikus, J.; Steverding, D. A simple colorimetric method to screen drug cytotoxicity against Leishmania using the dye alamar blue. *Parasitol. Int.* **2000**, *48*, 265−269.
(45) Babich, H.; Borenfreund, E. Cytotoxicity of T-2 toxin and its metabolites determined with the neutral red cell viability assay. *Appl. Environ. Microbiol.* **1991**, *57*, 2101−2103.
(46) Gillmor, S. A.; Craik, C. S.; Fletterick, R. J. Structural determinants of specificity in the cysteine protease cruzain. *Protein Sci.* **1997**, *6*, 1603−1611.
(47) Polgar, T.; Keseru, G. M. Virtual screening for beta-secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J. Med. Chem.* **2005**, *48*, 3749−3755.
(48) Evers, A.; Klabunde, T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.* **2005**, *48*, 1088−1097.

JM0505765