Aryl ureas represent a new class of anti-trypanosomal agents

Xiaohui Du [1], Elizabeth Hansell [4], Juan C Engel [4], Conor R Caffrey [4], Fred E Cohen [1,2,3] and James H McKerrow [3,4]

Background: The trypanosomal diseases including Chagas' disease, African sleeping sickness and Nagana have a substantial impact on human and animal health worldwide. Classes of effective therapeutics are needed owing to the emergence of drug resistance as well as the toxicity of existing agents. The cysteine proteases of two trypanosomes, *Trypanosoma cruzi* (cruzain) and *Trypanosoma brucei* (rhodesain), have been targeted for a structure-based drug design program as mechanistic inhibitors that target these enzymes are effective in cell-based and animal models of trypanosomal infection.

Results: We have used computational methods to identify new lead scaffolds for non-covalent inhibitors of cruzain and rhodesain, have demonstrated the efficacy of these compounds in cell-based and animal assays, and have synthesized analogs to explore structure activity relationships. Nine compounds with varied scaffolds identified by DOCK4.0.1 were found to be active at concentrations below 10 μM against cruzain and rhodesain in enzymatic studies. All hits were calculated to have substantial hydrophobic interactions with cruzain. Two of the scaffolds, the urea scaffold and the aroyl thiourea scaffold, exhibited activity against *T. cruzi* in vivo and both enzymes in vitro. They also have predicted pharmacokinetic properties that meet Lipinski's 'rule of 5'. These scaffolds are synthetically tractable and lend themselves to combinatorial chemistry efforts. One of the compounds, 5′(1-methyl-3-trifluoromethylpyrazol-5-yl)-thiophene 3′-trifluoromethylphenyl urea (**D16**) showed a 3.1 μM IC_{50} against cruzain and a 3 μM IC_{50} against rhodesain. Infected cells treated with **D16** survived 22 days in culture compared with 6 days for their untreated counterparts. The mechanism of the inhibitors of these two scaffolds is confirmed to be competitive and reversible.

Conclusions: The urea scaffold and the thiourea scaffold are promising leads for the development of new effective chemotherapy for trypanosomal diseases. Libraries of compounds of both scaffolds need to be synthesized and screened against a series of homologous parasitic cysteine proteases to optimize the potency of the initial leads.

[1]Department of Cellular and Molecular Pharmacology and Medicine, University of California, San Francisco, CA 94143-0450, USA
[2]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143-0450, USA
[3]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0450, USA
[4]Department of Pathology, Veterans Affairs Medical Center, University of California, San Francisco, CA 94143-0450, USA

Correspondence: Fred E Cohen
E-mail: cohen@cmpharm.ucsf.edu

Keywords: Cysteine protease; DOCK; Structure-based drug design; Thiourea; Trypanosomal disease; Urea

Received: **16 February 2000**
Revisions requested: **31 March 2000**
Revisions received: **19 June 2000**
Accepted: **29 June 2000**

Published: **1 August 2000**

Chemistry & Biology 2000, **7**:733–742

1074-5521/00/$ – see front matter
© 2000 Published by Elsevier Science Ltd.
PII: S1074-5521(00)00018-1

Introduction

Trypanosomal parasites are the cause of global health problems [1]. Chagas' disease is the leading cause of heart disease in Latin America [2]. Over 16 million people are infected, with up to 80% prevalence in endemic areas, and over 90 million are at risk [3]. It is caused by *Trypanosoma cruzi*, a protozoan parasite that is transmitted to humans from the bite of a blood-sucking insect. The trypomastigote form of the insect enters the host bloodstream, ultimately invades a cardiac muscle cell and then transforms into the intracellular amastigote form. Amastigotes replicate within cells, transform back to trypomastigotes and rupture the cell, releasing the infectious form back into the bloodstream and other cells, amplifying the infection [4–6]. Currently, there is no satisfactory treatment for this parasitic infection. Because of significant toxicity, chemotherapy with nifurtimox or benznidazole must be carried out under close medical supervision [7]. In addition to dermatotoxicity and digestive disorders, benznidazole induces chromosomal damage in chagasic children [8].

The major cysteine protease of *T. cruzi*, cruzain, is pivotal for the development and survival of the parasite within the host cells. Tissue culture studies of *T. cruzi* viability in the presence of mechanism-based inhibitors suggest that cruzain is necessary for the parasite to progress through its life cycle. Treatment of infected mice with covalent inhibitors of cruzain has resulted in parasitological cure of the mice [9]. Taken together, these results argue for the pharmacologic utility of inhibitors of cruzain.

In central Africa, *Trypanosoma brucei gambiense* and *T. b. rhodesiense* are the causative agent of sleeping sickness in humans, and *T. b. brucei*, *Trypanosoma congolense* and *Trypa-*

nosoma vivax are responsible for 'Nagana' in cattle. African trypanosomes live extracellularly in blood and tissue fluids of their mammalian host and are transmitted by the bite of tsetse flies. This complex of diseases is characterized by waves of parasitemia in the blood with eventual infiltration of the brain and central nervous system. Without treatment, the infection is fatal. Over 50 million people in 36 countries are at risk of acquiring the infection and 25 000–50 000 new cases are reported from 200 foci annually [10,11]. Chemotherapy of African sleeping sickness is precarious [12]. Only four drugs are available and three of these were developed over 50 years ago. Treatment with these drugs can fail and is associated with serious toxicity. Rhodesain is the major cysteine protease of *T. b. rhodesiense* (Caffrey et al., manuscript in preparation) and shares 70.0% similarity in primary structure with cruzain. Targeting rhodesain, and its ortholog brucipain from *T. b. brucei*, with peptidyl diazomethane and phenyl vinyl sulfone inhibitors kills trypanosomes in culture and cures mice infected with the parasite [13,14]. Thus, rhodesain is a promising target for the development of new trypanosomicidal drugs.

Non-covalent protease inhibitors have demonstrated success in treatment of diseases including angiotensin-converting enzyme inhibitors for hypertension and HIV protease inhibitors for AIDS [15,16]. Encouraged by these results, we are pursuing non-covalent inhibitors of cruzain and rhodesain for treatment of trypanosomal diseases. Our goal is to identify compounds that are easy to synthesize from relatively inexpensive starting materials so that the eventual therapeutics are economically accessible to the people who live in the areas where these diseases are endemic.

Combinatorial chemistry and structure-based drug design strategies are both based on the hypothesis that increasingly potent agents can be found when large families of molecules are screened. To maximize screening efforts targeting diseases endemic to developing regions of the world, we are considering families of non-covalent small molecule inhibitors and related cysteine protease targets in parallel. Explicitly, members of a chemical series that are active against one parasitic protease are tested against a series of other homologous cysteine proteases. This concept proved useful in our previous work developing a series of acyl hydrazide and chalcone-based non-covalent inhibitors for cruzain and rhodesain. This series was originally developed to block falcipain, a cysteine protease from *Plasmodium falciparum* [17], that is 30% identical to cruzain. The acyl hydrazides were further optimized for cruzain. However, the optimization of the bisaryl acylhydrazides failed to produce inhibitors that were more potent than 300 nM [18,19] for cruzain and 420 nM for rhodesain [20]. To develop more potent inhibitors for cruzain and rhodesain and identify potential new inhibitor scaffolds, we initiated a computational screen of ACD (Available Chemicals Directory) database targeted against recently solved crystal structures of cruzain [21,22]. New lead structures have been identified, that will allow us to pursue more potent inhibitors for cruzain and rhodesain owing to the high degree of homology between the two enzymes.

Results and discussion

Altogether 21 compounds were selected for testing against cruzain and rhodesain following a computational screen of the ACD97 database by DOCK4.0.1. Nine out of these 21 molecules showed IC_{50} values of less than 10 μM against

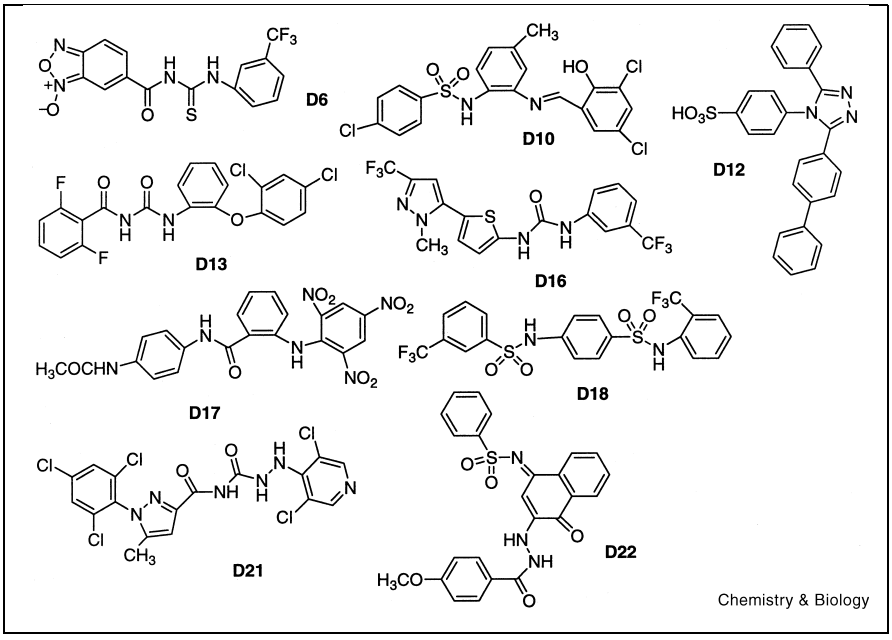


Figure 1. Structures of compounds having IC_{50} values below 10 μM for cruzain and rhodesain from a computational screening of the ACD database by DOCK4.0.1.

cruzain (Figure 1). The same nine compounds showed IC_{50} values of less than 10 μM against rhodesain. Two other compounds (**D1** and **D20**) have IC_{50} values below 10 μM for rhodesain but were one order of magnitude less active against cruzain. Altogether, 11 molecules have IC_{50} values below 10 μM against rhodesain. This demonstrates a concrete advantage of our strategy for increasing the number of effective inhibitors by testing against a set of related targets. We have found on more than one occasion that a given analog selected for its efficacy against one target protease is found to be more active against a related cysteine protease. Two of the nine compounds that were active against both proteases below 10 μM (**D13** and **D21**) were tested in a cell culture assay of *T. cruzi* infection of mammalian cells. No toxicity against the mammalian cells was observed. For both compounds, inhibition of parasite replication was observed. Untreated mammalian cells infected with parasites died at 6 days from parasitolysis of cells. Cells treated with **D13** or **D21** survived 10 days.

The nine compounds in Figure 1 provide a wide variety of new scaffolds for potential antiparasitic drugs, including ureas, thioureas, amides, sulfonamides, triazoles and semicarbozides. In medicinal chemistry, urea bonds have been used as critical structural elements in enzyme inhibitors [16]. Thioureas have been found in many biological interesting molecules [23–25] such as herbicides, insecticides and plant growth regulators. Sulfonamides can be found in numerous molecules that possess biological activity, including many antibiotics [26,27]. Triazoles have been found in drugs such as antifungal agent fluconazole and in aromatase inhibitors for estrogen-dependent diseases such as breast cancer [28,29]. A common feature of all these structures is that they have a barbell-like structure with two aromatic groups joined by a linker. The molecular variety apparent in Figure 1 suggests that cruzain and rhodesain accept a wide variety of inhibitors that contain two aryl groups connected with an appropriate linking scaffold.

The DOCK-generated enzyme–inhibitor complexes were re-examined to identify any common factors that could explain their activity. All nine compounds that were active at concentrations below 10 μM against cruzain appeared on the energy score list. Two of them also appeared on the contact score list. This indicates that unlike previous work with the malarial enzyme falcipain, electrostatic issues could not be neglected with cruzain. An examination of the molecular surface of cruzain calculated using the GRASP program (Figure 2) shows that while the S2 pocket is large and prefers hydrophobic residues, it has a certain degree of electronegativity. The S1 and S3 pockets are also electronegative. The S1′ pocket is neutral. (For notational convenience, the amino acid residues on the acyl side of the scissile bond are denoted P1, P2…P*n* and those on the leaving group side are labeled P1′, P2′…P*n*′. The corresponding binding sites on the enzyme are S1, S2…S*n* and S1′, S2′…S*n*′.)

All the hits are calculated to make substantial hydrophobic interactions with cruzain. Table 1 shows a list of cruzain residues that each hit is thought to have hydrophobic interactions with as calculated by LIGPLOT [30]. These

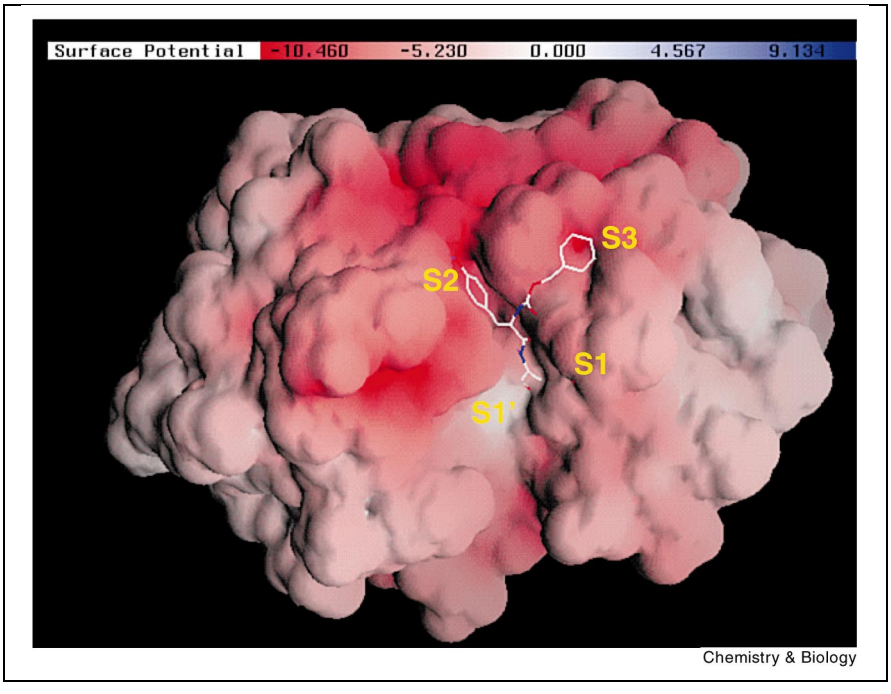


Figure 2. The molecular surface of cruzain with Z-YA-FMK bound calculated by GRASP. Red color indicates that the surface is electronegative and purple color indicates that the surface is electropositive. The subsites are labeled yellow in the figure.

Table 1
The calculated hydrophobic interactions of the cruzain–inhibitor complexes by DOCK.

Compound	Cruzain residues that have hydrophobic interactions with the compound
D6	Gly65†, Asp158
D10	Gly23, Ser64, Asp158
D12	Cys25, Gly65, Met68, Ala133, Asp158, His159, Trp177
D13	Met68, Ala133, Asp158
D16	Gly23, Cys25, Gly65
D17	Gly23, Cys25, Trp26, Gly65, Gly66, Leu67, Met68, Ala133, Asp158
D18	Cys25, Leu67, Met68, Ala133, Leu157, Asp158, His159, Trp177, Glu205
D21	Gly23, Cys25, Leu157
D22	Gln19, Cys25, Gly65, Leu67, Ala133, Leu157, Trp177, Glu205

†Residues are color-coded according to the subsite pockets they are located. S3 pocket residues: purple; S2 pocket residues: red; S1 pocket residues: green; S1′ pocket residues: blue. The black residues are not easily associated with one of the pockets.

residues are color-coded according to the subsite pocket that they are expected to occupy (see Figure 2 for the pockets location). Six of the nine hits have calculated hydrophobic interactions with Asp158. As Asp158 has a carboxylate group, we are exploring the introduction of an additional amino group on the ligand to increase the likelihood of forming hydrogen bonding interactions with Asp158. Two of the six hits (**D12** and **D18**) are calculated to have additional interactions with two other S1′ pocket residues: Trp177 and His159. Many of these compounds are calculated to interact with S2 pocket residues such as Leu67, Met68, Ala133, Leu157 and Glu205.

Seven of the hits (**D10**, **D12**, **D13**, **D17**, **D18**, **D21**, **D22**) exploited the S2 and S1′ pockets by orienting one aromatic end into each of these subsites in the putative docked geometry. None of the other tested compounds including the other two hits exploited the S1′ pocket in their calculated binding mode. A representative figure of the putative binding mode between **D13** and cruzain is shown in Figure 3.

Lipinski has described desired ranges for certain properties thought to be important for pharmacokinetics and drug development. They are CLogP <5, number of hydrogen bond donors ≤ 5, number of hydrogen bond acceptors ≤ 10 and molecular weight <500 [31]. A compound that fulfills at least three out of the four criteria is said to adhere to Lipinski's 'rule of 5'. His rule was used as a filter in helping us to choose the reasonable scaffolds as potential lead scaffolds. Table 2 lists the values of these properties for the new scaffolds and suggests that these compounds are reasonable starting points for a drug discovery effort. From a synthetic point of view, most of these compounds are easy to derivatize (e.g. **D6**, **D10**, **D16**, **D17**, **D18** and **D21**).

As the active scaffolds are synthetically tractable structures with the ability to inhibit the enzymes and block parasite replication, a mimetic search was conducted to look for chemically related molecules and test their general activities. GenX [32,33] is a program used to match compounds that have the same substructure or pharmacophore based on three dimensional similarity. It employs a sub-graph isomorphism algorithm. Here we employed this program for a computational miniscreen of the analogs. The analogs in ACD possessing atoms within 2 Å root mean square (rms) of the desired core scaffolds were identified and examined visually to yield an additional 38 compounds for investigation as cruzain/rhodesain inhibitors. These include sixteen thioureas (**D6**), three aroyl ureas (**D13**), nine ureas (**D16**), three imine sulfonamides (**D10**), one benzamide (**D17**) and six sulfonamides (**D18**). There are very few analogs of **D12**, **D21** and **D22** in the ACD.

Among these 38 compounds, five thioureas, three aroyl

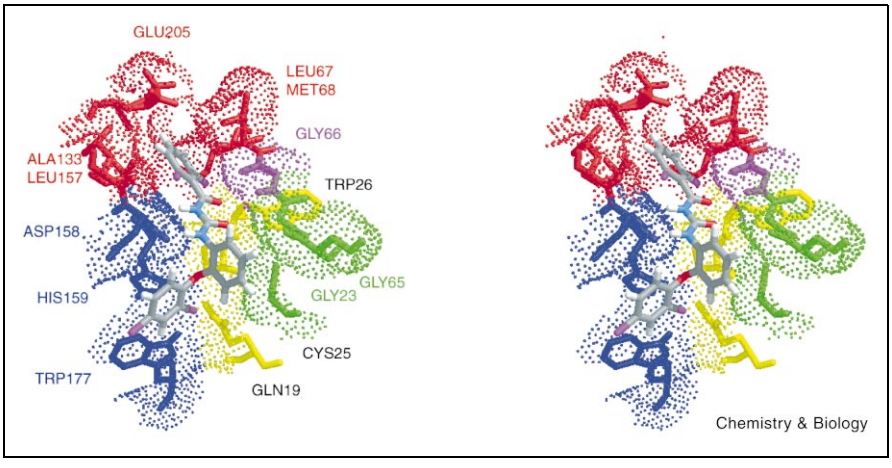


Figure 3. Stereoview of the calculated orientation of compound **D13** in the active site of cruzain by DOCK. The color coding of the cruzain residues is consistent with that of Table 1. The residues are color-coded according to the subsite pocket they belong. S3 pocket residues: purple; S2 pocket residues: red; S1 pocket residues: green; S1′ pocket residues: blue. The yellow residues labeled black are not easily associated with one of the pockets. **D13** is color-coded by atom types, C: gray; N: blue; O: red; F, Cl: purple.

Table 2
The pharmacokinetic properties of the nine active compounds identified by DOCK according to Lipinski's rules.

Compound	Molecular weight	CLogP	H-bond donors	H-bond acceptors	'Rule of 5' criteria met
D6	382	4.14	2	7	all
D10	470	5.62	2	5	3
D12	454	3.94	1	6	all
D13	437	6.61	2	5	3
D16	434	5.75	2	5	3
D17	480	3.84	3	5	all
D18	524	4.85	2	6	3
D21	509	6.44	3	3	2
D22	461	3.73	2	8	all

ureas, five ureas, two sulfonamides and one benzamide were observed to have IC_{50} values less than 10 μM against cruzain (see Figure 4, Tables 3 and 4 for structures). These compounds are also active against rhodesain at concentrations below 10 μM. None of the three imine sulfonamides tested has IC_{50} values below 10 μM. These results indicated that at least five of the scaffolds: thiourea, aroyl urea, urea, sulfonamide and benzamide, have some degree of generality and are possible new leads for cruzain inhibitors. These scaffolds were further compared with the information from database screening and assay testing.

It is interesting to note that **D61** (Figure 4) differs from **D17** by having a non-aryl central ring and a benzyl-substituted amine instead of an aryl-substituted amine without loss of activity. This indicates that the benzamide scaffold allows some flexibility. However, this change raises the CLogP, rendering it less attractive. **D17** remains as a potential lead but more information is needed about the general activity of this series before a meaningful synthetic strategy can be developed.

The high molecular weight and CLogP values associated with the sulfonamides (**D56**, **D58** in Figure 4) and the related compounds render this scaffold less desirable.

The aroyl thiourea and urea scaffolds are common to the largest number of active compounds identified computationally and documented experimentally for cruzain and rhodesain. Five compounds containing each scaffold were tested to detemine more accurate IC_{50} values against cruzain and rhodesain (Tables 3 and 4). The compounds containing the aroyl thiourea scaffold generally meet the requirements of Lipinski's rule with low molecular weight and appropriate CLogP. The analogs can be synthesized easily in a one step reaction through addition of an aryl amine to an aroyl isothiocyanate. When comparable substituents are considered, the compounds containing the

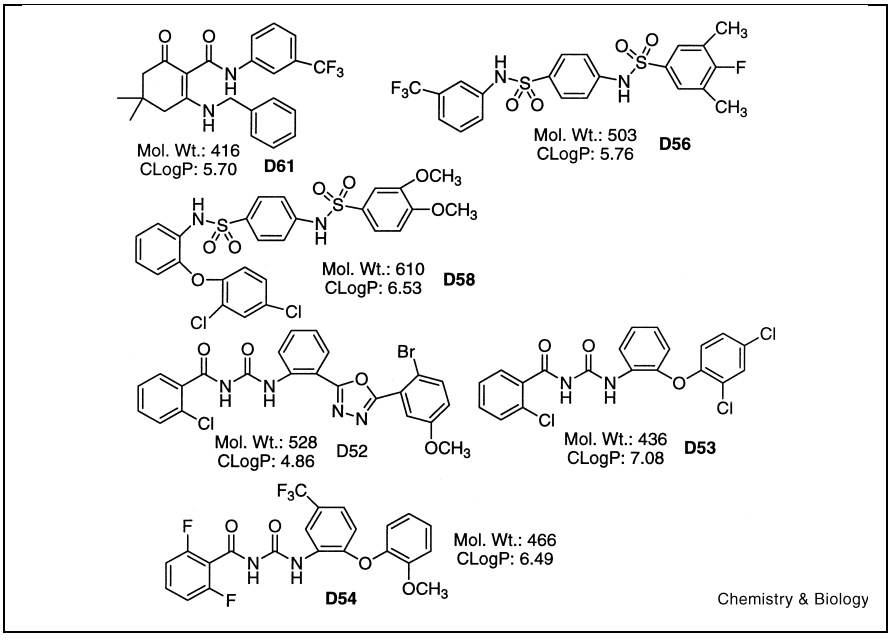


Figure 4. Structures of compounds having activity below 10 μM possessing benzamide, sulfonamide, and aroyl urea scaffolds at a second stage miniscreen by GenX. The molecular weights and calculated CLogP values of each compound are shown in the figure.

Table 3
Structures, physical properties and detailed IC_{50} values of the active compounds possessing the thiourea scaffold.

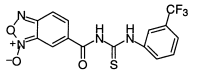
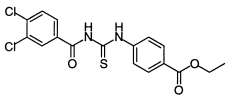
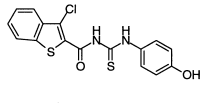
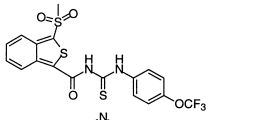
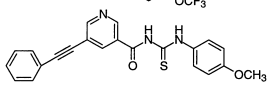
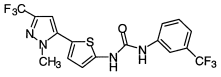
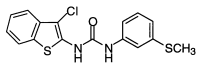
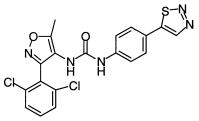
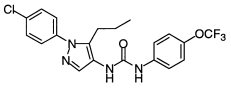
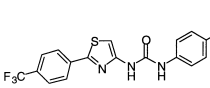
Compound	Structure	MW	CLogP	IC50 (μM) Cruzain	IC50 (μM) Rhodesain
D6		382	4.14	4.8	5
D23		397	4.99	1.9	2
D28		363	3.1	2.7	4
D34		475	3.55	6.9	4.2
D38		387	3.62	3.7	3.5
D37	Not tested further	446	5.66	<10	<10

Table 4
Structures, physical properties and detailed IC_{50} values of the active compounds possessing the urea scaffold.

Compound	Structure	MW	CLogP	IC50 (μM) Cruzain	IC50 (μM) Rhodesain
D16		434	5.75	3.1	3
D45		349	5.13	3	4
D46		446	5.06	10	10
D49		439	5.55	2.9	3.5
D51		398	6.14	1.2	2
D47	Not tested further	405	6.43	<10	<10

urea scaffold have lower molecular weights than those with the aroyl thiourea scaffold, yet they share similar hydrogen bond donating and accepting capabilities. In addition, the urea moiety is more stable than a peptide bond. As a result, much attention has been paid to the syntheses and applications of substituted ureas in recent years. They are essential components of several drug candidates including potent HIV reverse transcriptase inhibitors [34], CCK-B receptor antagonists [35] and endothelin antagonists [36]. The synthesis of this type of scaffold is straightforward either by addition of an amine to an isocyanate or an amine to an activated carbamate [34–36]. The CLogP of the urea scaffold is 0.3 units higher than the aroyl thiourea scaffold, but it can be lowered conveniently by the addition of polar functional groups or nitrogen-containing groups such as pyridine or morpholine where the formation of HCl salt will increase aqueous solubility.

Encouraged by these results, some of the compounds from the aroyl thiourea scaffold and the urea scaffold were characterized in the cell culture assay of *T. cruzi* infection of mammalian cells. (At the time of cell culture assay, we were running out of **D28**. Due to the difficulty of setting up the cell culture assay and the long time needed to purchase additional **D28**, no further testing was performed on **D28**.) The results are shown in Table 5. Compounds containing the urea scaffold were efficacious in the cell culture screen. Infected mammalian cells treated with **D16** survived *22* days and cells treated with **D46** and **D51** survived 14 and 12 days, respectively. By comparison, the compounds with the best in vivo activity from the acyl hydrazide series resulted in cells surviving 17 days. Typically, cells only survive 6 days in the absence of effective treatment.

In the aroyl thiourea scaffold, the infected cells treated with **D34** survived 8 days. However, the activity of the other thioureas was disappointing. At this point, we initiated structure–activity studies. One representative compound, **D217**, was synthesized (Figure 5). It contained a six-membered aryl ring connected with a five-membered heteroaromatic ring at one side of the thiourea. This feature is present in some of the active compounds possessing the urea scaffold, such as **D16**, **D46**, **D47**, **D49** and **D51**. From our modeling work, we believe that the six–five combination might interact with the S2 pocket better than a single six-membered ring. **D217** also has a benzyl-substituted aryl ring on the other end of the molecule

Table 5
Results of cell culture assay of some of the active compounds possessing aroyl thiourea and urea scaffolds.

Compound	Results (days)	Compound	Results (days)
Control	6	**D16**	22
D23	crystals	**D45**	toxic
D34	8	**D46**	14
D38	6	**D49**	toxic
D217	18	**D51**	12

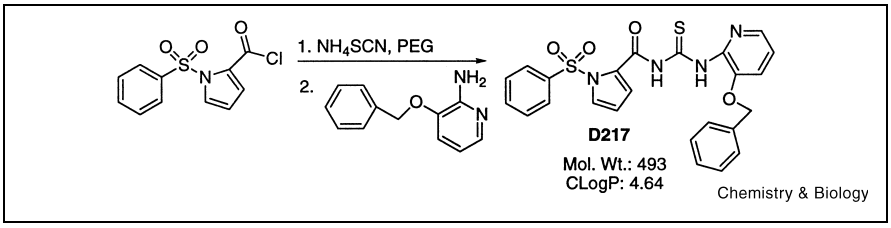


Figure 5. Synthesis of one of the compounds possessing the aroyl thiourea scaffold, **D217**.

which might result in better interaction in S1′ pocket. Before generating a library of compounds possessing features similar to **D217** for a detailed SAR study, it was tested to obtain some preliminary structure–activity information. **D217** was found to have an IC_{50} of 3 μM against cruzain and 3.8 μM against rhodesain. The infected cells treated with **D217** survived 18 days, more than twice as long as the parent compound.

A representative set of inhibitors (**D51**, **D16**, **D23**, **D34**, two inhibitors each in the urea and the aroyl thiourea scaffolds) were studied for their selectivity. To our delight, they show good selectivities. They are selective for cruzain and rhodesain over papain, and cathepsin B within the papain family. They do not inhibit the serine proteinase, trypsin. Table 6 shows the detailed IC_{50} values of the four inhibitors with various enzymes. A study of the time-dependent inhibition of the enzyme with five compounds including **D51**, **D16**, **D23**, **D34** demonstrated no decrease in enzyme activity (data not shown). This indicates that they are reversible inhibitors. In another study of inhibition efficacy as a function of increasing substrate concentration, the above four inhibitors were found to show similar behavior as leupeptin, a classical reversible competitive inhibitor. The % of inhibition of the enzyme was reduced with the increase of the concentration of substrate (data not shown).

At this point, both the thiourea and urea scaffolds looked promising based on their mechanism, specificity and their in vivo and in vitro enzyme assay. Efforts are underway to optimize these leads. Apart from their straightforward synthesis, they share synthetic building blocks. The same aryl acid and aryl amine building blocks can be used for the generation of compound libraries for both scaffolds and should augment the efficiency of the lead optimization process (Figure 6). Parallel synthesis can be applied for the generation of compound libraries as a result of these reaction conditions.

The aroyl urea scaffold is closely related to the aroyl thiourea scaffold (Figure 4). Thus, the aroyl thiourea synthetic strategy can be applied to the aroyl urea scaffold. While cell culture data for leads from this series (e.g. **D13**) look promising, the hydrophobicity of these compounds as captured by the CLogP value is problematic. This will limit analog options if we wish to remain with compounds that share desirable pharmacokinetic features.

Significance

Our goal remains to bridge standard medicinal chemistry strategies for compound optimization with ideas rooted in structural studies of the enzyme inhibitor interaction. Particular attention will be paid to adding groups that facilitate interactions with accessory subsite specificity pockets on cruzain. Based on our experiences with these new lead scaffolds and with the acyl hydrazides series, one or two pocket interactions with cruzain can result in inhibitors with activities at micromolar and even high nanomolar concentrations. We anticipate that building interactions with accessory subsites will enhance the potency of the inhibitors substantially. As specified in Table 1, these inhibitors are expected to interact with the S2 and S1′ pockets. Additional fragments can be added to the inhibitors so that they can interact with the S3 pocket. As the S2 and S3 pockets of cruzain are electronegative, substituents such as amino

Table 6
Selectivity of urea and aroyl thiourea inhibitors.

Inhibitor		IC_{50} (μM)			
		Cruzain	Papain	Cathepsin B	Trypsin
D51	Urea	0.8 (±0.01)	>50	10.8 (±0.5)	>50
D34	Thiourea	2.1 (±0.05)	>50	11 (±0.2)	>50
D23	Thiourea	1.5 (±0.01)	>50	10.7 (±0.2)	>50
D16	Urea	2.0 (±0.01)	>50	11 (±0.01)	>50
Leupeptin		<0.1	0.18 (±0.02)	0.25 (±0.05)	0.11 (±0.01)

The assays in this table were done in a different time from those of Tables 3 and 4, as a result, the IC_{50} values are slightly different among the same inhibitors.

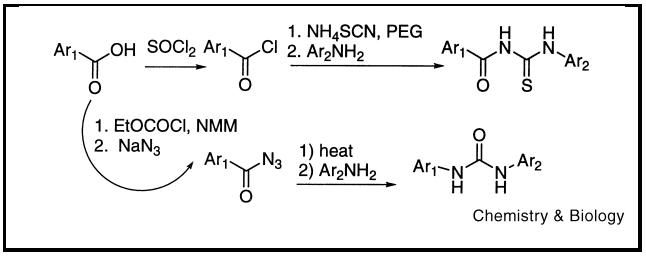


Figure 6. The synthesis of the compounds possessing the aroyl thiourea and urea scaffolds can apply common building blocks.

groups in these pockets might increase the potency of the inhibitors (Figure 2).

When compared to the acyl hydrazide scaffold, the aroyl thiourea, the urea and the aroyl urea scaffolds all have better hydrogen bonding capabilities. Though all the scaffolds have aryl groups at both ends, the angle and distance between the two ends vary from one linker to the other. The urea scaffold is more stable than the acyl hydrazide scaffold and has a lower molecular weight. What is more encouraging is that a compound containing the urea scaffold has already demonstrated better activity in a cell culture model of trypanosomal infection than the compound having the best activity in the acyl hydrazide series.

Materials and methods

Computational screening

Several crystal structures of cruzain bound to covalent inhibitors are available including cruzain bound with benzoyl tyrosine alanine fluoromethyl ketone (Z-YA-FMK), cruzain bound with benzoyl arginine alanine fluoromethyl ketone (Z-RA-FMK) and cruzain bound with WRR256 [37]. These were compared to identify any significant differences. The rms deviation between cruzain bound to Z-YA-FMK and cruzain bound to Z-RA-FMK is 0.42 Å and for cruzain bound to Z-YA-FMK and cruzain bound to WRR256 is 0.31 Å. These differences are compatible with the structural distinctions between identical proteins studied in distinct crystal forms. For convenience, the crystal structure of cruzain bound to Z-YA-FMK was used in our study. The movement of the Cys25 thiol before and after the covalent linkage was also examined [38]. As the change is relatively small (1.5 Å), no efforts were made to remodel the cruzain structure. Histidines are considered positively charged (hydrogen on both nitrogens). The united atom Kollman charges were applied to each residue in cruzain. These values were loaded from the BIOPOLYMER menu in SYBYL program.

DOCK4.0.1 is an automated method to screen small-molecule databases for ligands that could bind to a target protein [39]. The algorithm first generates the molecular surface of the protein with the dms function within the UCSF MIDASPLUS program [40], which is a feature derived from the original Connoly's surface program [41,42]. The surface is used to generate a negative image of the active site of the protein with sets of overlapping spheres. A grid was constructed that enclosed the entire active site. The grid was used by DOCK to evaluate the steric boundary of the protein and electrostatic and van der Waals interactions between the ligand and the protein. The ligand is then fitted to the spheres of the negative image of the protein and the interaction energy is calculated. The quality of fit of a ligand to the active site is decided either by shape complementarity (contact score) or by molecular mechanics force-field energy (energy score) using the AMBER potential [43]. In our case, the residues within 7 Å of Z-YA-FMK were selected for constructing the molecular surface. Forty-one spheres were selected to represent the active site of cruzain. A grid with an additional 2 Å margin for the box enclosing the spheres was constructed.

There are 175 413 small molecules in the ACD97 database (Molecular Design Limited, San Leandro, CA, USA), separated into 21 sections. The separation is for calculation convenience so that each section did not have more than 10 000 compounds. The molecules that have heavy atom range between 10 and 40 within the 175 413 compounds were screened. The structures of the molecules in the database were calculated by CONCORD, a heuristic algorithm developed by R. Pearlman at the University of Texas. Orientations of the ligands were sampled with 5000 as a limit during the DOCK run. Standard rigid body minization in the DOCK program was applied. We did not calculate multiple conformations for the ligands. The search took about 4 days on a silicon graphics server with four parallel R10 000 processors. During the DOCK search, the top 50 molecules by energy score and the top 50 molecules by contact score in each section were saved. The resulting 2200 compounds were visually examined several times for complementarity of the van der Waals surfaces and hydrogen bond interactions using the molecular display software MIDASPLUS. Twenty-one commercially available compounds were selected for testing against cruzain and rhodesain. When we chose these 21 compounds from the 2200 list, the DOCK scores are not critical in making a decision here, although compounds that fit well with higher score are preferred and compounds appearing both on contact score list and energy score list are preferred. Peptides are not considered as we are interested in non-hydrolyzable small molecules only. Compounds need to be able to occupy the S2 pocket as this is a deep and large pocket specific for each cysteine protease. Compounds with complicated structures are less desirable owing to synthetic considerations. If repetitive appearance of compounds with common scaffolds occurred, one or two compounds that appear to have the best fitting are chosen for testing. The visual inspection of the 2200 initial list was done in triplicate to achieve consistency and to avoid viewer bias owing to fatigue.

The CLogP values were calculated for each compound using the software CLogP 4.51 (Daylight Chemical Information Systems, Santa Fe, NM, USA) which is an implementation of the work of Hansch [44].

Synthesis and screening of inhibitor analogs against recombinant cruzain

Synthesis of **D217**: 1-phenylsulfonyl-2-pyrrole carbonyl chloride (50 mg, 0.18 mmol), was dissolved in anhydrous dichloromethane. PEG600 (5 mg, 0.0081 mmol) and ammonium thiocyanate (21 mg, 0.27 mmol) were added to the reaction. After stirring for 1 h under nitrogen at room temperature, 2-amino-3-benzoloxypyridine (35 mg, 0.17 mmol) was added to the mixture and the reaction was stirred for 2 h. Inorganic salts were filtered off from the reaction and the resulting solution was concentrated in vacuo. Crystallization from dichloromethane/methanol mixed solvent afforded 30 mg of **D217** (33% yield). ^{1}H NMR (400 MHz, DMSO-*d6*) 8.04 (m, 3H), 7.87 (s, 1H), 7.73 (t, 1H, J=7.5 Hz), 7.61 (t, 3H, J=7.6 Hz), 7.48 (d, 2H, J=7.2 Hz), 7.31 (m, 4H), 7.20 (m, 1H), 6.49 (t, 1H, J=3.2 Hz), 5.24 (s, 2H). MS(FAB) (M+H) calculated for $C_{24}H_{20}N_4O_4S_2$ 493.6, found 493.0.

IC_{50}, inhibitor mechanism and specificity determinations. All other inhibitors were purchased from Aldrich and Ryan Scientific. Inhibitors were screened for effectiveness against the *T. cruzi* cathepsin L-like protease (cruzain) using purified recombinant protein [45]. Cruzain (2 nM) was incubated with 500–10 000 nM inhibitor in 100 mM sodium acetate buffer (pH 5.50) and 5 mM dithiothreitol (buffer A), for 5 min at room temperature. Buffer A containing Z-Phe-Arg-AMC (Bachem, K_m = 1 μM) was added to enzyme inhibitor to give 20 μM substrate in 200 μl, and the

increase in fluorescence (excitation at 355 nM and emission at 460 nM) was followed with an automated microtiter plate spectrofluorimeter (Molecular Devices, spectraMAX Gemini). Inhibitor stock solutions were prepared at 20 mM in DMSO and serial dilutions were made in DMSO (0.7% DMSO in assay). Controls were performed using enzyme alone, and enzyme with DMSO. IC_{50} values were determined graphically using inhibitor concentrations in the linear portion of a plot of inhibition versus log [I] (seven concentrations tested with at least two in the linear range). IC_{50} values for other cysteine proteases were determined similarly: rhodesain (recombinant, cysteine protease from *T. b. rhodesiense*, EC 3.4.22.X, gift of Conor Caffrey) at 3 nM and 10 μM Z-Phe-Arg-AMC ($K_m = 1$ μM); papain (EC 3.4.22.2, Sigma) at 6 nM enzyme and 100 μM Z-Phe-Arg-AMC ($K_m = 50$ μM); cathepsin B (bovine spleen, EC 3.4.22.1, Sigma) at 10 nM enzyme and 100 μM Z-Phe-Arg-AMC ($K_m = 110$ μM); and cathepsin S (recombinant human, EC 3.4.22.27, gift of Dieter Bromme) at ~10 nM enzyme and 20 μM Z-Phe-Arg-AMC ($K_m = 20$ μM). The IC_{50} determinations for the serine proteinase trypsin (bovine pancreatic, EC 3.4.21.4, Cal Biochem) were done in 100 μM Tris pH 8 and 2 mM $CaCl_2$ with 20 μM Z-Phe-Arg-AMC and enzyme at 1.25 ng per 200 μl assay. Time dependence of inhibition was determined by incubating cruzain with inhibitors, DMSO or enzyme alone at room temperature for time points between 30 s and 2 h and the activity determined as above. Substrate inhibitor competition with cruzain and selected inhibitors was done with 0.5–90 μM Ac-Lys-Glu-Lys-Leu-Arg-AMC (gift of Smith Kline Beecham) ($K_m = 1$ μM). Leupeptin (Roche), a classical competitive inhibitor, was used as a control. Similar experiments were done using cathepsin S and 2–100 μM Z-Phe-Arg-AMC. Enzyme (0.1 × volume) was added to a mixture of inhibitor and substrate (0.9 × volume), with the % of DMSO kept constant at 1.5%. Controls of enzyme alone and enzyme with DMSO were done. Inhibition mechanism was determined by plots of % inhibition against substrate concentration (K.F. Tipton, Chapter 4, 'Patterns of Enzyme Inhibition', in Enzymology Labfax, ed. P.C. Engel).

Tissue culture models of T. cruzi infection: growth inhibition of T. cruzi amastigotes by cysteine protease inhibitors

J774 macrophages were cultured in RPMI 1640 medium and 5% heat-inactivated FCS (RPMI medium). For growth inhibition assays, J774 macrophages were irradiated (3000 rad) to arrest cell growth and cultured on coverglasses within six-well plates for 24 h at 37°C. After infection with *T. cruzi* trypomastigotes of the Y strain for 3 h, monolayers were washed with RPMI medium and treated with inhibitors at 7.5 μM in RPMI medium. Inhibitor stocks were made at 7.5 mM in DMSO and all assays included DMSO (0.01–0.02% v/v) controls. The inhibitors were evaluated in *T. cruzi*-infected macrophage cultures for 24 days. Trypomastigote output, indicative of the completion of the intracellular cycle, was then assayed in treated and untreated cultures to determine growth inhibition of intracellular *T. cruzi* amastigotes. In untreated cultures, parasites destroy mammalian cells at the end of their 6 day cycle of replication.

Acknowledgements

This work was supported by NIH Program Project Grant AI35707 and a NIH postdoctoral fellowship to X.D. (1 F32 AI10293-01).

References

1. McKerrow, J.H., Sun, E. & Rosenthal, P.J. (1993). The proteases and pathogenicity of parasitic protozoa. *Annu. Rev. Microbiol.* **47**, 821–853.
2. Libow, L.F., Beltranni, V.P., Silvers, D.N. & Grossman, M.E. (1991). Post-cardiac transplant reactivation of Chagas' disease diagnosed by skin biopsy. *Cutis* **48**, 37–40.
3. Godal, T. & Nagera, J. (1990). Tropical diseases. In *WHO Division of Control in Tropical Diseases*, pp. 12–13, World Health Organization, Geneva.
4. Bonaldo, M.C., D'Escoffier, L.N., Salles, J.M. & Goldenberg, S. (1981). Characterization and expression of proteases during *Trypanosoma cruzi* metacyclogenesis. *Exp. Parasitol.* **73**, 44–51.
5. Harth, G., Andrews, N., Mills, A.A., Engel, J., Smith, R. & McKerrow, J.H. (1993). Peptide-fluoromethyl ketones arrest intracellular replication and intercellular transmission of *T. cruzi*. *Mol. Biochem. Parasitol.* **58**, 17–24.
6. Meirelles, M.N.L., Juliano, L., Carmona, E., Silva, S.G., Costa, E.M., Murta, A.C.M. & Scharfstein, J. (1992). Inhibitors of the major cysteinyl proteinase (GP57/51) impair host cell invasion and arrest the intracellular development of *Trypanosoma cruzi* in vitro. *Mol. Biochem. Parasitol.* **52**, 175–184.
7. Kirchhoff, L.V. (1993). American trypanosomiasis (Chagas' disease) – a tropical disease now in the United States. *New Engl. J. Med.* **329**, 639–644.
8. Gorla, N.B., Ledesma, O.S., Barbieri, G.P. & Larripa, I.B. (1988). Assessment of cytogenetic damage in chagasic children treated with benznidazole. *Mutat. Res.* **206**, 217–220.
9. Engel, J.C., Doyle, P.S., Hsieh, E. & McKerrow, J.H. (1998). Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection. *J. Exp. Med.* **188**, 725–734.
10. Molyneux, D.H. (1997). In *Trypanosomiasis and Leishmaniasis: Biology and Control* (Hide, G., Mottram, J.C., Coombs, G.H. & Holmes, P.H., eds.), pp. 39–50, CAB Int., Oxford.
11. WHO (1998). Control and surveillance of African trypanosomiasis. *World Health Org. Tech. Rep. Ser.* **881**.
12. Croft, S.L., Urbina, J.A. & Brun, R. (1997). In *Trypanosomiasis and Leishmaniasis: Biology and Control* (Hide, G., Mottram, J.C., Coombs, G.H. & Holmes, P.H., eds.), pp. 245–257, CAB Int., Oxford.
13. Scory, S., Caffrey, C.R., Stierhof, Y.-D., Ruppel, A. & Steverding, D. (1999). *Trypanosoma brucei*: killing of bloodstream forms in vitro and in vivo by the cysteine proteinase inhibitor Z-phe-ala-CHN2. *Exp. Parasitol.* **91**, 327–333.
14. Caffrey and Steverding, unpublished results.
15. Williams, G.H. (1988). Converting-enzyme inhibitors in the treatment of hypertension. *New Engl. J. Med.* **319**, 1517–1525.
16. Lam, P.Y.S., et al., & Erickson-viitanen, S. (1994). Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **263**, 380–384.
17. Rings, C.S., Sun, E., McKerrow, J.H., Lee, G.K., Rosenthal, P.J., Kuntz, I.D. & Cohen, F.E. (1993). Structure-based inhibitor design by using protein models for the development of antiparasitic agents. *Proc. Natl. Acad. Sci. USA* **90**, 3583–3587.
18. Li, R., et al., & Kenyon, G.L. (1996). Structure-based design of parasitic protease inhibitors. *Biol. Med. Chem.* **4**, 1421–1427.
19. Flaherty, T., unpublished work.
20. Troeberg, L., et al., & Cohen, F.E., Chalcone and acyl hydrazide cysteine proteinase inhibitors are trypanocidal for *Trypanosoma brucei brucei* (submitted).
21. Mcgrath, M.E., Eakin, A.E., Engel, J.C., McKerrow, J.H., Craik, C.S. & Fletterick, R.J. (1995). The crystal structure of cruzain: a therapeutic target for Chagas' disease. *J. Mol. Biol.* **247**, 251–259.
22. Gillmor, S.A., Craik, C.S. & Fletterick, R.J. (1997). Structural determinants of specificity in the cysteine protease cruzain. *Protein Sci.* **6**, 1603–1611.
23. Rasmussen, C.R., et al., & Molinari, A.J. (1988). Improved procedures for the preparation of cycloalkyl-, arylalkyl-, and arylthioureas. *Synthesis*, 456–459.
24. Zhang, Y. & Wei, T. (1997). Phase-transfer-catalyzed synthesis of *N*-aryl-*N'*-(2-chlorobenzoyl)-thiourea derivatives. *Synth. Commun.* **27**, 751–756.
25. Wei, T. & Zhang, Y. (1998). Synthesis of *N*-aroyl-*N'*-hydroxyethyl-(hydroxyphenyl)thiourea derivatives under the condition of phase transfer catalysis. *Synth. Commun.* **28**, 2851–2859.
26. Spink, W.W. (1942). *Sulfanilamide and Related Compounds in General Practice*. The year book publishers, Inc.
27. Lilly, E. (1953). *Eli Lilly and Company Antibiotic Therapy: Sulfonamide Therapy*.
28. Tsuruoka, A., Kaku, Y., Kakinuma, H., Tsukada, I., Yanagisawa, M. & Naito, T. (1997). Synthesis and antifungal activity of novel thiazole-containing triazole antifungals. *Chem. Pharm. Bull.* **45**, 1169–1176.
29. Okada, M., et al., & Fujikura, T. (1996). Studies on aromatase inhibitors. I. Synthesis and biological evaluation of 4-amino-4*H*-1,2,4-triazole derivatives. *Chem. Pharm. Bull.* **44**, 1871–1879.
30. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. (1995). LIGPLOT:

a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134.
31. Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3–25.
32. Perris, V., Wallace, A.C., Prusiner, S.B. & Cohen, F.E. (2000). Structure-based identification of therapeutic agents that inhibit prion diseases. *Proc. Natl. Acad. Sci. USA* **97**, 6073–6078.
33. Ullman, J.R. (1976). An algorithm for subgraph isomorphism. *J. Assoc. Comput. Mach.* **23**, 31–42.
34. Hogberg, M., et al., & Backbro, K. (1999). Urea-PETT compounds as a new class of HIV-1 reverse transcriptase inhibitors. 3. Synthesis and further structure–activity relationship studies of PETT analogues. *J. Med. Chem.* **42**, 4150–4160.
35. Castro, J.L., et al., & Matassa, V.G. (1996). Controlled modification of acidity in cholecystokinin B receptor antagonists: *N*-(1,4-benzodiazepin-3-yl)-*N*′-[3-tetrazol-5-ylamino)phenyl]ureas. *J. Med. Chem.* **39**, 842–849.
36. Von Geldern, T.W., et al., & Opgenorth, T.J. (1996). Azole endothelin antagonists. 2. Structure–activity studies. *J. Med. Chem.* **39**, 968–981.
37. Brinen, L., unpublished results. University of California, San Francisco, CA.
38. Wallace, A.C., Borkakoti, N. & Thornton, J.M. (1997). TESS: A geometric hashing algorithm for deriving 3D coordinate templates for searching structural databases. Application to enzyme active sites. *Protein Sci.* **6**, 2308–2323.
39. Ewing, T.J.A. & Kuntz, I.D. (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comp. Chem.* **18**, 1175–1189.
40. Ferrin, T.E., Huang, C.C., Jarvis, L.E. & Langridge, R. (1988). The MIDAS display system. *J. Mol. Graph.* **6**, 13–27.
41. Connolly, M.L. (1983). Analytical molecular surface calculation. *J. Appl. Cryst.* **16**, 548–558.
42. Connolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709–713.
43. Pearlman, D.A., et al., & Kollman, P.A. (1995). Amber4.1. University of California at San Francisco, San Francisco, CA.
44. Hansch, C. & Leo, A.J. (1979). Substituent Constants for Correlation Analysis in Chemistry and Biology. Wiley (Interscience), New York.
45. Eakin, A.E., McGrath, M.E., McKerrow, J., Fletterick, R.J. & Craik, C. (1995). Production of crystallizable cruzain, the major cysteine proteinase from *Trypanosoma cruzi*. *J. Mol. Biol.* **268**, 6115–6118.